

Benchmark report: FinanceBench

This benchmark report was generated automatically by the Ask This Guy platform. It measures the answers of an Ask This Guy agent on the FinanceBench public benchmark, under the conditions described below, and shows every question with the delivered answer and its verdict.

Report generated on 2026-10-02 20:52:32 UTC.

For more information, contact bonjour@askthisguy.com.

About FinanceBench

FinanceBench tests open-book question answering on the public filings of listed companies. Each question asks for a fact, a figure or a judgment that an analyst would find in, or compute from, a company's annual, quarterly or current reports: answering requires finding the right passage or table in long financial documents, then often a calculation or a short reasoning.

- **Developed by:** Patronus AI (Pranab Islam, Anand Kannappan, Douwe Kiela, Rebecca Qian, Nino Scherrer, Bertie Vidgen).
- **Published:** November 2023, with the paper "FinanceBench: A New Benchmark for Financial Question Answering".
- **Task:** answer a question about a public company from its filings: information extraction, numerical reasoning and logical reasoning.
- **Corpus:** 368 public filings of U.S. listed companies (10-K annual reports, 10-Q quarterly reports, 8-K current reports and earnings releases), in PDF.
- **Questions:** the open-source sample of 150 questions, drawn from the 10,231 questions of the full benchmark and spread over three types (metrics-generated, domain-relevant and novel-generated).
- **Expected answers:** each question comes with a reference answer and the evidence passage of the filing it relies on.
- **Official scoring:** none; the authors graded the answers by hand. In this report, each answer is graded by the judge described in the methodology.
- **License:** dataset under Creative Commons Attribution-NonCommercial 4.0 (CC BY-NC 4.0).
- **Reference:** [arXiv:2311.11944](https://arxiv.org/abs/2311.11944); github.com/patronus-ai/financebench; huggingface.co/datasets/PatronusAI/financebench.

Methodology

- **Questions:** All 150 evaluation questions of the agent that have an expected answer. Each question has an expected answer written beforehand by a domain expert.
- **Execution:** each question is asked in a new conversation with the agent, one question at a time (no parallel load), through the same answer pipeline as production conversations.
- **Latency:** time taken by the platform to produce the complete answer, from the question to the end of the answer.
- **Judging:** one call per question to an independent judge model (mistral/mistral-medium-3-5), which receives the question, the expected answer and the delivered answer, and returns OK or NOK with a one-sentence rationale. The prompt given to the judge model is reproduced in Appendix A.
- **Criterion:** OK when the delivered answer contains the essential information of the expected answer, without contradiction or factual error on the point asked. Additional correct information is accepted; wording, length, language and style are not judged. When the question sets no precision, a figure given with a different precision is the same figure if it rounds to the other one (for example 12.2% for 12%).
- **Errors:** a question without delivered answer counts as NOK.
- **Exclusions:** questions without expected answer are not part of the benchmark.
- **Disambiguation:** the original questions are each tied to one source filing, and some do not name its period (for example "this 10-K report"), while the agent searches a knowledge base holding several filings of the same company. When a question names no year, fiscal year or quarter, or no company, ATG appended one fixed sentence giving the period or the company of its source filing: "Added note by ATG for disambiguation: the question is referring to ...", shown with the question below. The question text is otherwise unchanged (11 question(s) in this run).
- **Human review:** a human reviewer can overrule a judge verdict. The question then takes the reviewer's verdict, and the report shows both verdicts with the reviewer's justification (4 question(s) in this run).

Summary

Metric	Value
Questions asked	150
Answers delivered	150
Errors (no answer, counted NOK)	0
OK	138 (92.0%)
NOK	12 (8.0%)
Verdicts changed by human review	4
Tool calls	229 (1.5 per question)
Latency p50 / p90	8.4 s / 18.0 s
Latency mean	10.5 s
Tools	Internal knowledge only

Test conditions

Condition	Value
Start (first question)	2026-10-02 16:32:05 UTC
End (last answer)	2026-10-02 17:03:58 UTC
Duration	31 min 52 s
AI policy	Unrestricted: all AI providers
Response mode	Fast
Question selection	All 150 evaluation questions of the agent that have an expected answer
Judge model	mistral/mistral-medium-3-5

Tools

Internal knowledge only. The agent answered from its internal knowledge only: it had no access to the Internet or to external services.

Tool	Category	Calls during the run
explore_document	Internal knowledge only	56
inspect_images	Internal knowledge only	0
knowledge_base_search	Internal knowledge only	165
query_documents	Internal knowledge only	8
read_conversation_dataset	Internal knowledge only	0

Knowledge base

To answer questions, the agent has access to 368 files.

By format

Format	Documents
PDF	368

By main language

Language	Documents
en	368

Results

Latency

Statistic (answered questions)	Seconds
p50 (median)	8.4 s
p90	18.0 s
Mean	10.5 s
Minimum	5.3 s
Maximum	40.5 s

Questions and answers

Q1. What is the FY2018 capital expenditure amount (in USD millions) for 3M? Give a response to the question by relying on the details shown in the cash flow statement.

Result: OK · Latency: 7.2 s · Tool calls: 1

Expected answer

\$1577.00

Delivered answer

\$1,577 million.

Per 3M's Consolidated Statement of Cash Flows (fiscal year ended December 31, 2018), the line "Purchases of property, plant and equipment (PP&E)" under Cash Flows from Investing Activities is \$(1,577) million, which is the capital expenditure figure.

For context, this compares to \$1,373 million in 2017 and \$1,420 million in 2016.

Cited documents: 3M_2018_10K.pdf

Judge rationale: The agent's answer matches the expected answer in value and units, with no contradictions or omissions.

Q2. Assume that you are a public equities analyst. Answer the following question by primarily using information that is shown in the balance sheet: what is the year end FY2018 net PPNE for 3M? Answer in USD billions.

Result: OK · Latency: 8.5 s · Tool calls: 1

Expected answer

\$8.70

Delivered answer

\$8.7 billion (exactly \$8,738 million).

From 3M's consolidated balance sheet at December 31, 2018: gross property, plant and equipment of \$24,873 million less accumulated depreciation of \$16,135 million equals net PP&E of \$8,738 million, i.e. \$8.738 billion.

Cited documents: 3M_2018_10K.pdf

Judge rationale: The agent's answer matches the expected answer when rounded to the same precision (\$8.7 billion).

Q3. Is 3M a capital-intensive business based on FY2022 data?

Result: **OK** · Latency: 7.1 s · Tool calls: 1

Expected answer

No, the company is managing its CAPEX and Fixed Assets pretty efficiently, which is evident from below key metrics:
CAPEX/Revenue Ratio: 5.1%
Fixed assets/Total Assets: 20%
Return on Assets= 12.4%

Delivered answer

No — 3M is not a capital-intensive business based on FY2022 data. It earns a high return on a large asset base while spending modestly on fixed assets relative to revenue.

The three indicators (FY2022, fiscal year ended December 31, 2022):

- **Capital expenditure / revenue:** $\$1,749\text{M} / \$34,229\text{M} = 5.11\%$ (modest)
- **Net fixed assets / total assets:** $\$9,178\text{M} / \$46,455\text{M} = 19.76\%$ (modest — goodwill of \$12,790M is the largest single asset line)
- **Return on assets:** net income attributable to 3M of \$5,777M → **12.44%** on the year-end balance (\$46,455M) and **12.35%** on the two-year average balance $((\$46,455\text{M} + \$47,072\text{M})/2 = \$46,763.5\text{M})$

A high ROA (~12.4%) combined with capex of only ~5% of revenue and net PP&E of under one-fifth of total assets is the profile of a non-capital-intensive company, even though goodwill (not fixed assets) makes up most of the asset base.

Source: 3M 2022 Form 10-K (consolidated income statement, balance sheet, and cash flow statement).

Cited documents: 3M_2022_10K.pdf

Judge rationale: The agent's answer includes all essential metrics (CAPEX/Revenue ~5.1%, Fixed Assets/Total Assets ~20%, ROA ~12.4%) and aligns with the expected answer's conclusion that 3M is not capital-intensive.

Q4. What drove operating margin change as of FY2022 for 3M? If operating margin is not a useful metric for a company like this, then please state that and explain why.

Result: OK · Latency: 8.5 s · Tool calls: 1

Expected answer

Operating Margin for 3M in FY2022 has decreased by 1.7% primarily due to:
-Decrease in gross Margin
-mostly one-off charges including Combat Arms Earplugs litigation, impairment related to exiting PFAS manufacturing, costs related to exiting Russia and divestiture-related restructuring charges

Delivered answer

Operating margin **declined 1.7 percentage points, from 20.8% (FY2021) to 19.1% (FY2022)** — a drop from \$7,369M operating income on \$35,355M net sales to \$6,539M on \$34,229M.

Drivers (per 3M's own margin bridge):

- **Underlying (ex-special-items) margin fell 1.5 pp** (22.3% → 20.8%), driven by:
 - Raw material and logistics cost inflation: **−2.4 pp** (the main recurring headwind)
 - Organic growth, productivity, pricing and spending discipline: **+1.0 pp** (partially offsetting)
- **Special items added a net −1.7 pp drag in 2022** (vs. −1.4 pp in 2021), dominated by one-off items:
 - Net costs for significant litigation (PFAS, Combat Arms Earplugs, respirator/asbestos): **−6.7 pp**
 - Gain on the Food Safety split-off with Neogen: **+8.0 pp**
 - PFAS manufacturing exit costs: **−2.4 pp**
 - Russia exit charges: **−0.3 pp**; divestiture costs: **−0.2 pp**; divestiture-related restructuring: **−0.1 pp**

On relevance: Operating margin *is* a useful metric for 3M. It is a diversified industrial and consumer-products manufacturer, not a financial institution, so operating income relative to sales is a standard and meaningful measure of its core operating performance — the one caveat is that 2022's GAAP figure is heavily distorted by litigation and divestiture special items, which is why 3M also reports the adjusted (ex-special-items) margin of 20.8%.

Cited documents: 3M_2022_10K.pdf

Judge rationale: The agent's answer includes all essential drivers of the operating margin change (decrease in gross margin, one-off charges like litigation, PFAS exit, Russia exit, and restructuring) and does not contradict the expected answer.

Q5. If we exclude the impact of M&A, which segment has dragged down 3M's overall growth in 2022?

Result: OK · Latency: 6.3 s · Tool calls: 1

Expected answer

The consumer segment shrunk by 0.9% organically.

Delivered answer

Consumer — it was the only 3M segment with negative organic sales growth in 2022 (−0.9%), while the other three all grew organically.

Organic sales change by segment in 2022 (excluding M&A, i.e. acquisitions net of divestitures, and currency):

- Health Care: +3.2%
- Transportation and Electronics: +1.2%
- Safety and Industrial: +1.0%
- **Consumer: −0.9%**

Management attributed the Consumer decline to softening consumer-retail trends — consumers pulling back on discretionary spending and retailers cutting inventories — with home improvement falling and consumer health and safety flat, partially offset by demand for Scotch Blue painter's tape, Scotch-Brite, and Post-it solutions.

Source: 3M 2022 Form 10-K, "Performance by Business Segment."

Cited documents: 3M_2022_10K.pdf

Judge rationale: The agent's answer correctly identifies the consumer segment as the one with negative organic growth (−0.9%) in 2022, matching the essential information in the expected answer.

Q6. Does 3M have a reasonably healthy liquidity profile based on its quick ratio for Q2 of FY2023? If the quick ratio is not relevant to measure liquidity, please state that and explain why.

Result: OK · Latency: 6.1 s · Tool calls: 1

Expected answer

No. The quick ratio for 3M was 0.96 by Jun'23 close, which needs a bit of an improvement to touch the 1x mark

Delivered answer

No — 3M's quick ratio for Q2 FY2023 (quarter ended June 30, 2023) sits below 1.0 on both usual bases, so the quick ratio does not indicate a healthy liquidity profile.

- **Strict basis** = (cash & equivalents + short-term investments + receivables) / total current liabilities = $(\$4,258 + \$56 + \$4,947) / \$10,936 = \$9,261 / \$10,936 = \mathbf{0.85}$
- **Broad basis** = (total current assets – inventories) / total current liabilities = $(\$15,754 - \$5,280) / \$10,936 = \$10,474 / \$10,936 = \mathbf{0.96}$

Both figures are under 1.0, and the strict basis (the deciding one) is clearly below the 1.0 threshold, so the quick ratio alone would not be read as a healthy liquidity position. (Note: 3M itself describes its balance sheet as strong, supported by \$4.3B of cash/securities, an undrawn \$4.25B revolver, and a \$5B commercial paper program — but on the quick-ratio measure specifically, the answer is no.)

Cited documents: 3M_2023Q2_10Q.pdf

Judge rationale: The agent's answer correctly states that 3M's quick ratio is below 1.0 (0.96 on broad basis) and thus does not indicate a healthy liquidity profile, matching the essential information in the expected answer.

Q7. Which debt securities are registered to trade on a national securities exchange under 3M's name as of Q2 of 2023?

Result: OK · Latency: 6.5 s · Tool calls: 1

Expected answer

Following debt securities registered under 3M's name are listed to trade on the New York Stock Exchange:

- 1.500% Notes due 2026 (Trading Symbol: MMM26)
- 1.750% Notes due 2030 (Trading Symbol: MMM30)
- 1.500% Notes due 2031 (Trading Symbol: MMM31)

Delivered answer

Three debt securities are registered to trade on a national securities exchange under 3M's name as of Q2 2023, all on the New York Stock Exchange (per the cover page of the Form 10-Q for the quarter ended June 30, 2023):

Title of each class	Trading Symbol	Exchange
1.500% Notes due 2026	MMM26	New York Stock Exchange
1.750% Notes due 2030	MMM30	New York Stock Exchange
1.500% Notes due 2031	MMM31	New York Stock Exchange

Note that the 0.950% Notes due 2023 (MMM23), which appeared in the prior year's 10-K, is no longer listed here — consistent with that series having matured in 2023.

Cited documents: 3M_2023Q2_10Q.pdf

Judge rationale: The agent's answer includes all essential debt securities listed in the expected answer with matching details (interest rates, due years, and trading symbols) and does not contradict the reference.

Q8. Does 3M maintain a stable trend of dividend distribution? Added note by ATG for disambiguation: the question is referring to Q2 of FY2023.

Result: OK · Latency: 5.6 s · Tool calls: 1

Expected answer

Yes, not only they distribute the dividends on a routine basis, 3M has also been increasing the per share dividend for consecutive 65 years

Delivered answer

Yes — 3M maintains a stable and steadily increasing dividend trend.

- Q2 2023 dividend: **\$1.50 per share** (declared May 2023), up from **\$1.49** in Q2 2022.
- Q1 2023 was also \$1.50 (a 1% increase over the prior year's \$1.49), so the first six months of 2023 totaled \$3.00 per share vs. \$2.98 in 2022.
- This marked the **65th consecutive year of dividend increases**, and 3M has paid dividends continuously since 1916.

So the per-share payout is both stable quarter-to-quarter and growing year-over-year, with no cuts or pauses — a textbook stable dividend track record.

Cited documents: 3M_2023Q2_10Q.pdf

Judge rationale: The agent's answer confirms the stable and increasing dividend trend, including the 65-year consecutive increase, which matches the essential information in the expected answer.

Q9. What is the FY2019 fixed asset turnover ratio for Activision Blizzard? Fixed asset turnover ratio is defined as: FY2019 revenue / (average PP&E between FY2018 and FY2019). Round your answer to two decimal places. Base your judgments on the information provided primarily in the statement of income and the statement of financial position.

Result: OK · Latency: 6.5 s · Tool calls: 1

Expected answer

24.26

Delivered answer

24.26

- FY2019 net revenues: \$6,489 million
- PP&E, net at Dec 31, 2019: \$253 million; at Dec 31, 2018: \$282 million
- Average PP&E = $(253 + 282) / 2 = \$267.5$ million
- Fixed asset turnover = $6,489 / 267.5 = \mathbf{24.26}$

Cited documents: ACTIVISIONBLIZZARD_2019_10K.pdf

Judge rationale: The agent's answer matches the expected answer exactly and includes the correct calculation steps.

Q10. What is the FY2017 - FY2019 3 year average of capex as a % of revenue for Activision Blizzard?
Answer in units of percents and round to one decimal place. Calculate (or extract) the answer from the statement of income and the cash flow statement.

Result: OK · Latency: 7.1 s · Tool calls: 1

Expected answer

1.9%

Delivered answer

1.9%

Capex as a % of revenue (capex from the cash flow statement, net revenues from the income statement):

- FY2017: \$155M / \$7,017M = 2.21%
- FY2018: \$131M / \$7,500M = 1.75%
- FY2019: \$116M / \$6,489M = 1.79%

3-year average = (2.21% + 1.75% + 1.79%) / 3 = 1.91% ≈ **1.9%**

(For reference, the aggregate method — total capex \$402M / total revenue \$20,006M — gives 2.0%.)

Source: Activision Blizzard 2019 Form 10-K.

Cited documents: ACTIVISIONBLIZZARD_2019_10K.pdf

Judge rationale: The agent's answer matches the expected answer (1.9%) and includes correct calculations and rounding.

Q11. You are an investment banker and your only resource(s) to answer the following question is (are): the statement of financial position and the cash flow statement. Here's the question: what is the FY2015 operating cash flow ratio for Adobe? Operating cash flow ratio is defined as: cash from operations / total current liabilities. Round your answer to two decimal places.

Result: OK · Latency: 10.3 s · Tool calls: 2

Expected answer

0.66

Delivered answer

0.66

- Cash from operations (FY2015, year ended Nov 27, 2015): \$1,469,502 thousand
- Total current liabilities (Nov 27, 2015): \$2,213,556 thousand
- Ratio = $1,469,502 / 2,213,556 = 0.6639 \rightarrow \mathbf{0.66}$

(Adobe's fiscal 2015 ended November 27, 2015.)

Cited documents: ADOBE_2015_10K.pdf

Judge rationale: The agent's answer matches the expected answer (0.66) and includes the correct calculation and rounding.

Q12. What is Adobe's year-over-year change in unadjusted operating income from FY2015 to FY2016 (in units of percents and round to one decimal place)? Give a solution to the question by using the income statement.

Result: OK · Latency: 7.6 s · Tool calls: 1

Expected answer

65.4%

Delivered answer

65.4%

From Adobe's consolidated statements of income (in thousands):

- FY2015 operating income (year ended Nov 27, 2015): \$903,095
- FY2016 operating income (year ended Dec 2, 2016): \$1,493,602

Change = $(1,493,602 - 903,095) / 903,095 = 590,507 / 903,095 = 0.6539 \rightarrow \mathbf{65.4\%}$

(Adobe's fiscal year ends in late November/early December, so FY2015 ended Nov 27, 2015 and FY2016 ended Dec 2, 2016.)

Cited documents: ADOBE_2016_10K.pdf

Judge rationale: The agent's answer matches the expected answer exactly in value and precision, with no contradictions or omissions.

Q13. What is the FY2017 operating cash flow ratio for Adobe? Operating cash flow ratio is defined as: cash from operations / total current liabilities. Round your answer to two decimal places. Please utilize information provided primarily within the balance sheet and the cash flow statement.

Result: OK · Latency: 6.2 s · Tool calls: 1

Expected answer

0.83

Delivered answer

0.83

Adobe's FY2017 (fiscal year ended December 1, 2017) operating cash flow ratio:

- Cash from operations: \$2,912.9M (net cash provided by operating activities)
- Total current liabilities: \$3,527.5M (as of December 1, 2017)
- Ratio: $2,912,853 / 3,527,457 = \mathbf{0.83}$

Both figures are taken directly from the Consolidated Statements of Cash Flows and the Consolidated Balance Sheets in Adobe's FY2017 10-K.

Cited documents: ADOBE_2017_10K.pdf

Judge rationale: The agent's answer matches the expected answer exactly (0.83) and includes the correct calculation and source references.

Q14. Does Adobe have an improving operating margin profile as of FY2022? If operating margin is not a useful metric for a company like this, then state that and explain why.

Result: OK · Latency: 5.7 s · Tool calls: 1

Expected answer

No the operating margins of Adobe have recently declined from 36.8% in FY 2021 to 34.6% in FY2022. A drop by 2.2% in a year.

Delivered answer

No — Adobe's operating margin declined as of FY2022 (fiscal year ended December 2, 2022).

- FY2022: operating income \$6,098M ÷ revenue \$17,606M = **34.64%**
- FY2021: operating income \$5,802M ÷ revenue \$15,785M = **36.76%**
- Change: **−2.12 percentage points** (a decline)

Operating margin is a perfectly relevant metric for Adobe — it's a software company, not a financial institution, so operating profitability is a core measure of performance. The margin fell because operating expenses grew 15% (to \$9,343M), outpacing the 12% revenue growth, driven by higher compensation and increased marketing spend.

Cited documents: ADOBE_2022_10K.pdf

Judge rationale: The agent's answer correctly states that Adobe's operating margin declined from FY2021 to FY2022, provides the exact figures (34.64% vs. 36.76%), and confirms the drop, matching the essential information in the expected answer.

Q15. Does Adobe have an improving Free cashflow conversion as of FY2022?

Result: OK · Latency: 6.1 s · Tool calls: 1

Expected answer

Yes, the FCF conversion (using net income as the denominator) for Adobe has improved by ~13% from 143% in 2021 to 156% in 2022

Delivered answer

Yes — Adobe's free cash flow conversion improved in FY2022 (fiscal year ended December 2, 2022), rising from 1.43 in FY2021 to 1.56 in FY2022.

- FY2022: FCF = \$7,838M (operating cash flow) – \$442M (capex) = \$7,396M; ÷ net income \$4,756M = **1.56**
- FY2021: FCF = \$7,230M – \$348M = \$6,882M; ÷ net income \$4,822M = **1.43**

Change: +0.13 (1.43 → 1.56), driven by higher operating cash flow (+\$608M) and only a modest capex increase, while net income fell slightly.

Cited documents: ADOBE_2022_10K.pdf

Judge rationale: The agent's answer correctly states that Adobe's FCF conversion improved from 1.43 in FY2021 to 1.56 in FY2022, matching the essential information in the expected answer.

Q16. What is the quantity of restructuring costs directly outlined in AES Corporation's income statements for FY2022? If restructuring costs are not explicitly outlined then state 0.

Result: OK · Latency: 6.2 s · Tool calls: 1

Expected answer

0

Delivered answer

0.

Restructuring costs do not appear as a line item in AES Corporation's Consolidated Statements of Operations for FY2022 (year ended December 31, 2022). The statement reports revenue, cost of sales, operating margin, general and administrative expenses, interest expense/income, loss on extinguishment of debt, other expense/income, loss on disposal and sale of business interests, goodwill impairment expense, asset impairment expense, foreign currency transaction gains/losses, and other non-operating expense — but no separate restructuring line.

(Consistent with the MD&A note that there were no major-restructuring-program costs in 2020, 2021, or 2022.)

Cited documents: AES_2022_10K.pdf

Judge rationale: The agent's answer correctly states 0 and confirms no restructuring costs were explicitly outlined in the income statement.

Q17. Roughly how many times has AES Corporation sold its inventory in FY2022? Calculate inventory turnover ratio for the FY2022; if conventional inventory management is not meaningful for the company then state that and explain why.

Result: **NOK** · Latency: 9.3 s · Tool calls: 2

Expected answer

AES has converted inventory 9.5 times in FY 2022.

Delivered answer

About 12 times — AES sold through roughly 12× its average inventory in FY2022 (calendar year ended December 31, 2022).

Calculation (FY2022):

- Total cost of sales (COGS): \$10,069M
- Inventory: \$1,055M (Dec 31, 2022) and \$604M (Dec 31, 2021)
- Average inventory = $(1,055 + 604) / 2 = \$829.5\text{M}$
- **Average-basis turnover = $10,069 / 829.5 = 12.14\times$**
- Year-end-basis turnover = $10,069 / 1,055 = 9.55\times$

On the relevance of the metric: Conventional inventory management is not very meaningful for AES. AES is a power generator, not a merchandiser — it sells electricity, not its inventory. Its inventory is a raw input (fuel and other raw materials, \$733M, plus spare parts and supplies, \$322M) that is consumed to produce power, not a finished good held for resale. So a high turnover here mainly reflects how quickly fuel is burned and restocked, not how efficiently the company turns over a product line the way a retailer or manufacturer would. The ratio is still computable and it does show fuel is turned over quickly, but it should not be read as a measure of inventory-management performance in the usual retail/manufacturing sense.

Cited documents: AES_2022_10K.pdf

Judge rationale: The agent's answer provides a different figure (12.14×) for the inventory turnover ratio than the expected answer (9.5×), and the discrepancy is not due to rounding or precision differences.

Q18. Based on the information provided primarily in the statement of financial position and the statement of income, what is AES's FY2022 return on assets (ROA)? ROA is defined as: FY2022 net income / (average total assets between FY2021 and FY2022). Round your answer to two decimal places.

Result: OK · Latency: 8.6 s · Tool calls: 1

Expected answer

-0.02

Delivered answer

ROA = -0.02

Inputs (from the FY2022 Form 10-K):

- FY2022 net income (loss) attributable to The AES Corporation: **\$(546) million**
- Total assets FY2022: \$38,363 million; FY2021: \$32,963 million
- Average total assets = (38,363 + 32,963) / 2 = **\$35,663 million**

Calculation: -546 / 35,663 = **-0.0153** → **-0.02** (rounded to two decimals)

The negative result reflects AES's net loss for FY2022, driven largely by ~\$1.5 billion of pre-tax asset impairment expense and losses on sales of business interests.

Cited documents: AES_2022_10K.pdf

Judge rationale: The agent's answer matches the expected ROA value of -0.02, correctly calculated and rounded to two decimal places.

Q19. What is Amazon's FY2017 days payable outstanding (DPO)? DPO is defined as: $365 \times (\text{average accounts payable between FY2016 and FY2017}) / (\text{FY2017 COGS} + \text{change in inventory between FY2016 and FY2017})$. Round your answer to two decimal places. Address the question by using the line items and information shown within the balance sheet and the P&L statement.

Result: OK · Latency: 5.9 s · Tool calls: 1

Expected answer

93.86

Delivered answer

93.86 days

Using the FY2017 10-K balance sheet and income statement (in millions):

- **Average accounts payable (FY2016–FY2017):** $(25,309 + 34,616) / 2 = 29,962.5$
- **FY2017 cost of sales (COGS):** 111,934
- **Change in inventory (FY2016→FY2017):** $16,047 - 11,461 = 4,586$
- **Denominator:** $111,934 + 4,586 = 116,520$

DPO = $365 \times 29,962.5 / 116,520 = 93.86$ days

Cited documents: AMAZON_2017_10K.pdf

Judge rationale: The agent's answer matches the expected answer exactly, including the required precision of two decimal places.

Q20. What is Amazon's year-over-year change in revenue from FY2016 to FY2017 (in units of percents and round to one decimal place)? Calculate what was asked by utilizing the line items clearly shown in the statement of income.

Result: OK · Latency: 7.4 s · Tool calls: 1

Expected answer

30.8%

Delivered answer

30.8%

From the Consolidated Statements of Operations, total net sales (the revenue line) were:

- FY2016: \$135,987 million
- FY2017: \$177,866 million

Change = $(177,866 - 135,987) / 135,987 = 41,879 / 135,987 = 30.8\%$

(Note: Amazon's own MD&A rounds this to 31%, but the exact figure from the statement line items is 30.8%.)

Cited documents: AMAZON_2017_10K.pdf

Judge rationale: The agent's answer matches the expected answer exactly in value and precision, with no contradictions or omissions.

Q21. By drawing conclusions from the information stated only in the income statement, what is Amazon's FY2019 net income attributable to shareholders (in USD millions)?

Result: OK · Latency: 5.9 s · Tool calls: 1

Expected answer

\$11588.00

Delivered answer

\$11,588 million.

The FY2019 Consolidated Statements of Operations reports a single "Net income" line of \$11,588 million, with no separate deduction for non-controlling interests. Since the income statement presents no such allocation, the entire net income is attributable to shareholders.

Cited documents: AMAZON_2019_10K.pdf

Judge rationale: The agent's answer matches the expected net income figure of \$11,588 million in USD.

Q22. What is Amcor's year end FY2020 net AR (in USD millions)? Address the question by adopting the perspective of a financial analyst who can only use the details shown within the balance sheet.

Result: OK (human review) · Latency: 8.0 s · Tool calls: 1

Expected answer

\$1616.00

Delivered answer

\$1,615.9 million.

Reading straight off Amcor's consolidated balance sheet as of June 30, 2020 (the end of fiscal year 2020), the line item "Trade receivables, net" is reported at **\$1,615.9 million** (in millions, USD). This is the net figure — i.e., gross trade receivables already reduced by the allowance for doubtful accounts — as it appears on the face of the balance sheet.

For context, the prior year (June 30, 2019) stood at \$1,864.3 million, so net AR declined by \$248.4 million year over year.

Cited documents: AMCOR_2020_10K.pdf

Judge verdict: NOK · The agent's answer (\$1,615.9 million) does not match the expected answer (\$1616.00 million) due to a precision discrepancy that exceeds the allowed rounding.

Human review: OK · The figure is more specific in the agent's response

Q23. What was the key agenda of the AMCOR's 8k filing dated 1st July 2022?

Result: **OK** · Latency: 6.2 s · Tool calls: 1

Expected answer

Amcor Finance (USA), Inc. and Amcor Flexibles North America, Inc., entered into supplemental indentures relating to Guaranteed Senior Notes due 2026 and 2028. This involved the substitution of the Substitute Issuer (Amcor Flexibles North America) for the Former Issuer (Amcor Finance) and the assumption of covenants under the indentures. (In essence a novation agreement)

Delivered answer

The key agenda of Amcor's Form 8-K dated July 1, 2022 (earliest event reported June 30, 2022) was the **substitution of the issuer for its outstanding senior notes** — reported under Item 8.01 (Other Events).

Specifically, on June 30, 2022, Amcor Finance (USA), Inc. (the "Former Issuer") and Amcor Flexibles North America, Inc. (the "Substitute Issuer") — each a wholly-owned subsidiary of Amcor plc — entered into two supplemental indentures with Deutsche Bank Trust Company Americas (the Trustee):

- A **Second Supplemental Indenture** to the 2016 Indenture (governing the 3.625% Guaranteed Senior Notes due 2026 and the 4.500% Guaranteed Senior Notes due 2028), and
- A **First Supplemental Indenture** to the 2019 Indenture (governing the "New" 2026 and 2028 Notes issued in the June 2019 exchange offer).

In each case, the purpose was the **substitution of the Substitute Issuer for the Former Issuer** and the **assumption by the Substitute Issuer of the Former Issuer's covenants** under the indentures.

In short: it was a routine debt-structure housekeeping filing moving the issuer of Amcor's 2026 and 2028 guaranteed senior notes from one subsidiary (Amcor Finance USA) to another (Amcor Flexibles North America), with no change to the notes' terms.

Cited documents: AMCOR_2022_8K_2022-07-01.pdf

Judge rationale: The agent's answer contains all essential information from the expected answer (substitution of issuer, assumption of covenants, and the specific notes involved) and does not contradict it.

Q24. Has AMCOR's quick ratio improved or declined between FY2023 and FY2022? If the quick ratio is not something that a financial analyst would ask about a company like this, then state that and explain why.

Result: OK · Latency: 11.0 s · Tool calls: 3

Expected answer

The quick ratio has slightly improved from 0.67 times to 0.69 times between FY 2023 and FY 2022.(3.4% jump)

Delivered answer

Improved. Amcor's quick ratio rose from FY2022 to FY2023 (fiscal years ended June 30, 2022 and June 30, 2023).

Using the broad basis (current assets minus inventories, over current liabilities):

- FY2022: $(5,853 - 2,439) / 5,103 = 3,414 / 5,103 = \mathbf{0.67}$
- FY2023: $(5,308 - 2,213) / 4,476 = 3,095 / 4,476 = \mathbf{0.69}$

On the strict basis (cash + receivables over current liabilities):

- FY2022: $(775 + 1,935) / 5,103 = 2,710 / 5,103 = \mathbf{0.53}$
- FY2023: $(689 + 1,875) / 4,476 = 2,564 / 4,476 = \mathbf{0.57}$

The improvement is driven mainly by the drop in current liabilities (5,103 → 4,476, largely lower trade payables and the removal of the Russian-business held-for-sale items) outpacing the decline in quick assets. The quick ratio is a standard, relevant liquidity metric for Amcor — a manufacturer with meaningful inventories and working-capital cycles — so it is appropriate to ask about here.

Cited documents: AMCOR_2023_10K.pdf

Judge rationale: The agent's answer correctly states the quick ratio improved and provides the exact figures (0.67 to 0.69) matching the expected answer, with no contradictions.

Q25. What are major acquisitions that AMCOR has done in FY2023, FY2022 and FY2021?

Result: **OK** · Latency: 20.1 s · Tool calls: 6

Expected answer

Amcor completed these acquisitions during FY2023:

-100% equity interest of a flexibles manufacturing company in the Czech Republic

- 100% equity interest in a medical device packaging manufacturing site in Shanghai, China.
- acquisition of a New Zealand-based leading manufacturer of state-of-the-art, automated protein packaging machines.

Delivered answer

Amcor's major acquisitions in this window are all in **FY2023** — three bolt-on deals totaling roughly \$164 million. **FY2022 and FY2021 had no material acquisitions** (the cash-flow "Business acquisitions" line is \$0 in both years).

FY2023 (year ended June 30, 2023) — three bolt-on acquisitions:

- **Czech Republic flexible-packaging plant** — 100% equity interest in a world-class flexible packaging manufacturing plant, completed August 1, 2022; purchase consideration **\$59 million** (including a deferred \$5 million paid in Q1 FY2024).
- **Shanghai, China medical-device packaging site** — 100% equity interest in a medical device packaging manufacturing site, completed March 17, 2023; purchase consideration **\$60 million** (including \$20 million contingent consideration payable over three years).
- **New Zealand protein-packaging machinery** — a leading manufacturer of automated protein packaging machines, completed May 31, 2023; purchase consideration **\$45 million** (including \$13 million contingent consideration payable over two years).

FY2022 (year ended June 30, 2022): No material acquisitions — the cash-flow statement reports \$0 for business acquisitions.

FY2021 (year ended June 30, 2021): No material acquisitions — the cash-flow statement reports \$0 for business acquisitions.

Note: Amcor's fiscal year ends June 30, so "FY2023" = the year ended June 30, 2023. The FY2023 earnings release separately confirms the company "completed three bolt-on acquisitions" during the year. (A subsequent, non-material deal — a small Indian flexible-packaging manufacturer — was signed August 10, 2023, i.e., after FY2023 year-end.)

Sources: (FY2023 10-K, acquisitions note), (FY2022 10-K, cash-flow statement), (FY2023 Q4 earnings release).

Cited documents: AMCOR_2022_10K.pdf, AMCOR_2023_10K.pdf, AMCOR_2023Q4_EARNINGS.pdf

Judge rationale: The agent's answer includes all essential acquisitions listed in the expected answer for FY2023 and correctly states no material acquisitions for FY2022 and FY2021, with no contradictions.

Q26. What industry does AMCOR primarily operate in? Added note by ATG for disambiguation: the question is referring to FY2023.

Result: OK · Latency: 6.3 s · Tool calls: 1

Expected answer

Amcor is a global leader in packaging production for various use cases.

Delivered answer

The packaging industry.

Amcor is a global leader in developing and producing responsible packaging, with its business organized into two reportable segments — Flexibles (≈76% of FY2023 consolidated net sales) and Rigid Packaging (≈24%) — serving food, beverage, pharmaceutical, medical, home and personal-care, and other consumer goods end markets.

In short, Amcor's core business is packaging (flexible and rigid), making packaging its primary industry.

Cited documents: AMCOR_2023_10K.pdf

Judge rationale: The agent's answer correctly identifies the packaging industry as Amcor's primary industry, matching the essential information in the expected answer.

Q27. Does AMCOR have an improving gross margin profile as of FY2023? If gross margin is not a useful metric for a company like this, then state that and explain why.

Result: OK · Latency: 6.3 s · Tool calls: 1

Expected answer

No. For AMCOR there has been a slight decline in gross margins by 0.8%.

Delivered answer

No — Amcor's gross margin declined as of FY2023 (fiscal year ended June 30, 2023).

- FY2023: gross profit \$2,725M ÷ net sales \$14,694M = **18.5%**
- FY2022: gross profit \$2,820M ÷ net sales \$14,544M = **19.4%**
- Change: **−0.9 percentage points** (a 3% drop in gross profit dollars, from \$2,820M to \$2,725M)

Gross margin is a fully relevant metric for Amcor — it is a packaging manufacturer, not a financial institution, so gross margin is a standard and meaningful measure of its performance. The decline was driven mainly by the pass-through of higher raw material costs (which inflate both revenue and cost of sales but dilute the margin) and the impact of disposed operations (the sale of the Russian business), partly offset by favorable operating cost performance.

Cited documents: AMCOR_2023_10K.pdf

Judge rationale: The agent's answer correctly states that gross margin declined and provides supporting figures that align with the expected answer's essential information (a decline of ~0.8-0.9%).

Q28. What is the nature & purpose of AMCOR's restructuring liability as of Q2 of FY2023 close?

Result: OK · Latency: 11.0 s · Tool calls: 3

Expected answer

87% of the total restructuring liability is related Employee liabilities.

Delivered answer

As of Q2 FY2023 close (December 31, 2022), Amcor's restructuring liability was an accrual of **\$93 million** for the estimated costs of its ongoing restructuring plans, recorded under other current and other non-current liabilities on the balance sheet.

It is made up of three components:

- **Employee costs – \$81 million** (the dominant share, ~87%): severance and other employee-related exit costs
- **Fixed asset related costs – \$2 million**: facility/asset exit costs
- **Other costs – \$10 million**

In substance, the liability represents the estimated cash outlays Amcor expects to pay to complete its restructuring programs. The restructuring activity in the three and six months ended December 31, 2022 was **primarily driven by the Russia-Ukraine conflict** (the 2019 Bemis integration plan, which had fed prior-year charges, was substantially complete by the end of FY2022). Amcor expected the **majority of the \$93 million to be paid within the next twelve months**, so it is largely a current obligation.

For context, the balance moved from \$118 million at June 30, 2022 to \$93 million at December 31, 2022, after \$2 million of new charges to earnings, \$25 million of cash paid, and a \$2 million reversal of unused amounts.

Cited documents: AMCOR_2023Q2_10Q.pdf

Judge rationale: The agent's answer includes the essential information that 87% of the restructuring liability is related to employee liabilities, matching the expected answer.

Q29. What Was AMCOR's Adjusted Non GAAP EBITDA for FY 2023

Result: **OK** · Latency: 7.4 s · Tool calls: 1

Expected answer

AMCOR's Adj. EBITDA was \$2,018mn in FY 2023

Delivered answer

\$2,018 million was Amcor's Adjusted non-GAAP EBITDA for fiscal year 2023 (the twelve months ended June 30, 2023).

Per the company's fiscal 2023 earnings release, the adjusted non-GAAP results table reports EBITDA of \$2,018 million for the twelve months ended June 30, 2023, down 5% on a reported basis from \$2,117 million in fiscal 2022 but up 1% on a comparable constant currency basis.

For reference, the reconciliation builds from GAAP net income attributable to Amcor of \$1,048 million, adding back net income attributable to non-controlling interests (\$10M), tax expense (\$193M), net interest expense (\$259M), and depreciation & amortization (\$569M) to reach GAAP EBITDA of \$2,080 million, then adjusting for items such as hyperinflation impact (\$24M), Russia-Ukraine conflict impacts (~\$90M, net of the gain on the Russian business sale), amortization of acquired intangibles (\$160M), and other items to arrive at the \$2,018 million adjusted figure.

Note that Amcor's fiscal year ends June 30, so "FY 2023" here refers to the year ended June 30, 2023.

Cited documents: AMCOR_2023Q4_EARNINGS.pdf

Judge rationale: The agent's answer includes the essential figure (\$2,018 million) and does not contradict the expected answer.

Q30. How much was the Real change in Sales for AMCOR in FY 2023 vs FY 2022, if we exclude the impact of FX movement, passthrough costs and one-off items?

Result: OK · Latency: 6.8 s · Tool calls: 1

Expected answer

The Real Growth was flat in FY 2023 vs FY 2022.

Delivered answer

0% (in line / flat).

Amcors net sales rose 1% on a reported basis in FY2023 (\$14,694M vs \$14,544M, +\$150M). Stripping out the three items you named:

- FX movement: -3%
- Raw-material pass-through: +5%
- Items affecting comparability (one-offs, disposed/ceased ops): -1%

$1\% + 3\% - 5\% + 1\% = 0\%$ — i.e., sales were "in line" with the prior year on a comparable constant-currency basis, reflecting price/mix of +3% offset by volumes of -3%.

The 10-K states it directly: excluding the \$776M pass-through, \$426M negative currency impact and \$207M from disposed/ceased operations, the remaining variation was an increase of just \$7M, or 0%.

Cited documents: AMCOR_2023Q4_EARNINGS.pdf, AMCOR_2023_10K.pdf

Judge rationale: The agent's answer correctly states that the real change in sales was 0% (flat), matching the essential information in the expected answer.

Q31. Answer the following question as if you are an equity research analyst and have lost internet connection so you do not have access to financial metric providers. According to the details clearly outlined within the P&L statement and the statement of cash flows, what is the FY2015 depreciation and amortization (D&A from cash flow statement) % margin for AMD?

Result: OK · Latency: 5.9 s · Tool calls: 1

Expected answer

4.2%

Delivered answer

4.18%

- D&A (from cash flow statement, FY2015): **\$167M**
- Net revenue (from P&L, FY2015): **\$3,991M**
- D&A margin = $167 / 3,991 = 4.18\%$

(AMD's fiscal 2015 ended December 26, 2015.)

Cited documents: AMD_2015_10K.pdf

Judge rationale: The agent's answer (4.18%) matches the expected answer (4.2%) when rounded to one decimal place.

Q32. Does AMD have a reasonably healthy liquidity profile based on its quick ratio for FY22? If the quick ratio is not relevant to measure liquidity, please state that and explain why.

Result: OK · Latency: 6.3 s · Tool calls: 1

Expected answer

Yes. The quick ratio is 1.57, calculated as (cash and cash equivalents+Short term investments+Accounts receivable, net +receivables from related parties)/ (current liabilities).

Delivered answer

Yes — AMD had a reasonably healthy liquidity profile for FY2022 (fiscal year ended December 31, 2022), with the quick ratio comfortably above 1.0 on both usual bases.

- **Strict quick ratio** = (cash + short-term investments + receivables) / total current liabilities
= (\$4,835M + \$1,020M + \$4,126M) / \$6,369M = \$9,981M / \$6,369M = **1.57**
- **Broad quick ratio** = (total current assets – inventories) / total current liabilities
= (\$15,019M – \$3,771M) / \$6,369M = \$11,248M / \$6,369M = **1.77**

Both figures clear the 1.0 threshold, so the liquidity profile is healthy — AMD held roughly \$1.57–\$1.77 of quick assets for every dollar of current liabilities, and its cash plus short-term investments alone (\$5.86B) nearly covered all current liabilities.

The quick ratio is relevant here: AMD is a standard operating (semiconductor) company with inventories and a normal working-capital cycle, so a quick-ratio liquidity check applies.

Source: AMD 2022 Form 10-K, Consolidated Balance Sheets as of December 31, 2022.

Cited documents: AMD_2022_10K.pdf

Judge rationale: The agent's answer confirms the quick ratio is 1.57 and states AMD has a healthy liquidity profile, matching the essential information in the expected answer.

Q33. What are the major products and services that AMD sells as of FY22?

Result: **OK** · Latency: 6.5 s · Tool calls: 1

Expected answer

AMD sells server microprocessors (CPUs) and graphics processing units (GPUs), data processing units (DPUs), Field Programmable Gate Arrays (FPGAs), and Adaptive System-on-Chip (SoC) products for data centers; CPUs, accelerated processing units (APUs) that integrate CPUs and GPUs, and chipsets for desktop and notebook personal computers; discrete GPUs, and semi-custom SoC products and development services; and embedded CPUs, GPUs, APUs, FPGAs, and Adaptive SoC products.

Delivered answer

As of FY22 (fiscal year ended December 31, 2022), AMD is a global semiconductor company that sells its products across four reportable segments, plus occasional IP licensing. Here are the major products and services:

Data Center segment — server CPUs (AMD EPYC), data center GPUs (AMD Instinct), data processing units (DPUs, from Pensando), FPGAs (Virtex, Kintex), and Adaptive SoCs (Zynq, Versal), plus Alveo accelerator cards.

Client segment — CPUs, APUs (accelerated processing units that integrate CPU + GPU), and chipsets for desktop and notebook PCs (AMD Ryzen, Ryzen PRO, Threadripper, Athlon families).

Gaming segment — discrete GPUs (AMD Radeon), semi-custom SoC products (e.g., game console chips for PlayStation, Xbox, Steam Deck), and development services.

Embedded segment — embedded CPUs, GPUs, APUs, FPGAs, and Adaptive SoC products for automotive, aerospace, industrial, healthcare, and 5G markets.

Services / other — AMD also develops and sells software platforms (development tools, compilers, drivers for CPUs, GPUs, and FPGAs), provides design and development (custom/semi-custom silicon) services, and from time to time sells or licenses portions of its intellectual property (IP) portfolio.

The portfolio was notably broadened in 2022 by two acquisitions: **Xilinx** (February 2022), which added FPGAs and Adaptive SoCs, and **Pensando** (May 2022), which added DPUs. AMD's FY22 net revenue was \$23.6 billion, up 44% year-over-year.

Would you like a breakdown of revenue by segment, or a deeper look at any one product line?

Cited documents: AMD_2022_annualreport.pdf, AMD_2022_10K.pdf

Judge rationale: The agent's answer includes all essential product categories from the expected answer (server CPUs/GPUs, DPUs, FPGAs, Adaptive SoCs, client CPUs/APUs/chipsets, discrete GPUs, semi-custom SoCs, and embedded products) with no contradictions.

Q34. What drove revenue change as of the FY22 for AMD?

Result: **OK** · Latency: 8.5 s · Tool calls: 1

Expected answer

In 2022, AMD reported Higher sales of their EPYC server processors, higher semi-custom product sales, and the inclusion of Xilinx embedded product sales

Delivered answer

Revenue grew 44% in FY2022 (fiscal year ended December 31, 2022), from \$16.4 billion to a record \$23.6 billion — an increase of roughly \$7.2 billion.

The drivers, by segment:

- **Data Center +64%** (\$6.0B vs \$3.7B) — the biggest contributor, driven by higher sales of EPYC server processors.
- **Gaming +21%** (\$6.8B vs \$5.6B) — higher semi-custom (console SoC) sales, partly offset by lower discrete graphics sales on soft consumer demand.
- **Embedded +~1,750%** (\$4.6B vs \$0.2B) — a one-off step-up from the inclusion of Xilinx embedded product revenue after the February 2022 acquisition.
- **Client -10%** (\$6.2B vs \$6.9B) — the offset, from a 24% drop in unit shipments on a weak PC market and supply-chain inventory correction, only partly cushioned by a 19% rise in average selling price.

So the growth was led by Data Center (EPYC) and Gaming (semi-custom), amplified by the Xilinx acquisition, and partially offset by the Client/PC downturn.

Source: AMD 2022 Form 10-K, Management's Discussion and Analysis.

Cited documents: AMD_2022_10K.pdf

Judge rationale: The agent's answer includes all essential drivers (higher EPYC server processor sales, higher semi-custom product sales, and inclusion of Xilinx embedded product sales) and does not contradict the expected answer.

Q35. What drove operating margin change as of the FY22 for AMD? If operating margin is not a useful metric for a company like this, then please state that and explain why.

Result: OK · Latency: 9.4 s · Tool calls: 1

Expected answer

The decrease in AMD's operating income was primarily driven by amortization of intangible assets associated with the Xilinx acquisition

Delivered answer

Operating margin declined, falling from 22.2% in FY21 to 5.4% in FY22 — a drop of about 16.8 percentage points.

The calculation:

- FY22 (ended Dec 31, 2022): operating income \$1,264M ÷ net revenue \$23,601M = **5.4%**
- FY21 (ended Dec 25, 2021): operating income \$3,648M ÷ net revenue \$16,434M = **22.2%**
- Change: 5.4% – 22.2% = **–16.8 pp**

What drove it:

The collapse in operating income (from \$3.6B to \$1.3B) despite a 44% revenue jump was dominated by one-off, acquisition-related items from the February 2022 Xilinx deal:

- **Amortization of acquisition-related intangibles** of \$3.5B total (\$1,448M in cost of sales + \$2,100M in operating expenses) — the single largest driver, and a non-cash charge.
- **Other acquisition-related costs** of ~\$452M (transaction costs, Xilinx inventory fair-value step-up, depreciation on stepped-up fixed assets, compensation charges) plus **stock-based compensation** of ~\$1.1B. Together these make up the "All Other" operating loss of \$5.0B in 2022 (vs. \$409M in 2021).

Recurring drivers also weighed: higher R&D (\$5.0B vs. \$2.8B) and higher SG&A (\$2.3B vs. \$1.4B), and a 10% decline in Client segment revenue from a weak PC market. Segment operating income actually rose in Data Center, Gaming and Embedded; the margin compression is essentially an accounting artifact of the Xilinx purchase rather than a deterioration in underlying product economics.

Is operating margin useful here? Yes — AMD is a semiconductor company, not a financial institution, so operating margin is a standard and meaningful measure of its profitability. The caveat is that FY22's figure is heavily distorted by non-cash intangible amortization from the Xilinx acquisition, so it understates the company's recurring operating performance; analysts often look at adjusted (non-GAAP) operating margin to strip out those one-off charges.

Sources: (AMD 2022 10-K, MD&A and segment results) and (AMD 2022 Annual Report, consolidated statements of operations).

Cited documents: AMD_2022_annualreport.pdf, AMD_2022_10K.pdf

Judge rationale: The agent's answer includes the essential information that the operating margin change was primarily driven by amortization of intangible assets from the Xilinx acquisition, and contains no contradictions or factual errors.

Q36. Among operations, investing, and financing activities, which brought in the most (or lost the least) cash flow for AMD in FY22?

Result: OK · Latency: 5.4 s · Tool calls: 1

Expected answer

In 2022, AMD brought in the most cashflow from Operations

Delivered answer

Operating activities brought in the most cash for AMD in FY2022 (year ended December 31, 2022), with net cash provided by operating activities of **\$3,565 million**.

Here's the full breakdown from the consolidated statement of cash flows:

Activity	Net cash flow (FY2022)
Operating activities	+\$3,565M
Investing activities	+\$1,999M
Financing activities	-\$3,264M

Operating activities clearly led, ahead of investing (+\$1,999M, boosted by \$2,366M of cash received from the Xilinx acquisition and \$4,310M of short-term investment maturities, partly offset by the \$1,544M Pensando acquisition and \$450M of capex). Financing was the only outflow, driven mainly by \$3,702M of common stock repurchases.

Cited documents: AMD_2022_10K.pdf

Judge rationale: The agent's answer correctly identifies operating activities as the source of the most cash flow for AMD in FY22, matching the essential information in the expected answer.

Q37. From FY21 to FY22, excluding Embedded, in which AMD reporting segment did sales proportionally increase the most?

Result: OK · Latency: 5.6 s · Tool calls: 1

Expected answer

Data Center

Delivered answer

Data Center — it grew the most proportionally among the non-Embedded segments.

- Data Center: \$3,694M (FY21) → \$6,043M (FY22) = +63.6%
- Gaming: \$5,607M → \$6,805M = +21.4%
- Client: \$6,887M → \$6,201M = -10.0% (decline)

The increase was driven primarily by higher sales of AMD EPYC server processors.

Cited documents: AMD_2022_10K.pdf

Judge rationale: The agent correctly identifies 'Data Center' as the segment with the highest proportional sales increase, matching the essential information in the expected answer.

Q38. Did AMD report customer concentration in FY22?

Result: OK · Latency: 5.3 s · Tool calls: 1

Expected answer

Yes, one customer accounted for 16% of consolidated net revenue

Delivered answer

Yes. AMD's 2022 10-K discloses that one customer accounted for 16% of consolidated net revenue for the year ended December 31, 2022, with sales to that customer consisting of products from its Gaming segment.

The filing does not name the customer, but the Gaming-segment composition points to Sony (PlayStation). AMD also notes that a loss of this customer would have a material adverse effect on its business.

Cited documents: AMD_2022_10K.pdf

Judge rationale: The agent's answer includes the essential information that one customer accounted for 16% of consolidated net revenue in FY22 and does not contradict the expected answer.

Q39. Which debt securities are registered to trade on a national securities exchange under American Express' name as of 2022?

Result: OK · Latency: 19.1 s · Tool calls: 3

Expected answer

There are none

Delivered answer

None. As of 2022, no debt securities are registered to trade on a national securities exchange under American Express' name.

The only security registered under Section 12(b) of the Exchange Act (the registration that applies to securities traded on a national securities exchange) is the **Common Shares (par value \$0.20 per share), trading symbol AXP, on the New York Stock Exchange**. The 10-K cover page for the fiscal year ended December 31, 2022 lists no debt securities under Section 12(b), and reports "None" under Section 12(g) as well.

Note that American Express does issue long-term debt (senior notes, subordinated notes, etc.), but those are not registered to trade on a national securities exchange — they are issued and held privately, which is why they do not appear on the Section 12(b) list.

Cited documents: AMERICANEXPRESS_2022_10K.pdf

Judge rationale: The agent's answer correctly states that there are no debt securities registered to trade on a national securities exchange under American Express' name as of 2022, matching the essential information in the expected answer.

Q40. What are the geographies that American Express primarily operates in as of 2022?

Result: OK · Latency: 9.1 s · Tool calls: 1

Expected answer

United States, EMEA, APAC, and LACC

Delivered answer

American Express primarily operates in four geographic regions, per its own geographic-operations table in the 2022 10-K: **United States, EMEA (Europe, the Middle East and Africa), APAC (Asia Pacific, Australia and New Zealand), and LACC (Latin America, Canada and the Caribbean)**, plus an "Other Unallocated" bucket for net costs not tied to a specific region.

For 2022, total revenues net of interest expense by region were:

Region	Revenues net of interest expense (\$M)
United States	\$41,396
EMEA	\$4,871
APAC	\$3,835
LACC	\$2,917
Other Unallocated	\$(157)
Consolidated	\$52,862

The United States is by far the largest market, accounting for roughly 78% of consolidated revenues net of interest expense.

Note that the filing also names specific jurisdictions representing a significant share of billed business outside the U.S.—the UK, the EU, Australia, Japan, Canada and Mexico—but the company’s own geographic reporting table uses the four regions above.

Cited documents: AMERICANEXPRESS_2022_10K.pdf

Judge rationale: The agent’s answer includes all essential geographies (United States, EMEA, APAC, LACC) and does not contradict the expected answer.

Q41. Does AMEX have an improving operating margin profile as of 2022? If operating margin is not a useful metric for a company like this, then state that and explain why.

Result: OK · Latency: 6.2 s · Tool calls: 1

Expected answer

Performance is not measured through operating margin

Delivered answer

No — operating margin is not a useful metric for American Express, so there is no meaningful "improving" or "declining" profile to read from it.

American Express is a bank holding company (a payments/financial-services firm), and its income statement is built around interest income, net interest income, and provisions for credit losses — all of which are core operations, not "other" items. Because interest and credit losses sit inside its operating results, a simple operating-margin ratio (operating income ÷ revenue) does not isolate the drivers of performance the way it would for a product company, so it is not how this kind of business is measured.

For reference, the filing reports 2022 net income of \$7.5 billion (\$9.85 per share) versus \$8.1 billion in 2021, with the decline driven by credit-reserve builds and Amex Ventures investment losses rather than by any operating-margin trend.

If you'd like, I can pull the segment-level revenue and net-income trends (USCS, Commercial Services, International Card Services, GMNS) instead, which is how AMEX's performance is actually tracked.

Cited documents: AMERICANEXPRESS_2022_10K.pdf

Judge rationale: The agent's answer correctly states that operating margin is not a useful metric for AMEX and provides a valid explanation, aligning with the essential information in the expected answer.

Q42. What drove gross margin change as of the FY2022 for American Express? If gross margin is not a useful metric for a company like this, then please state that and explain why.

Result: OK · Latency: 6.3 s · Tool calls: 1

Expected answer

Performance is not measured through gross margin

Delivered answer

Gross margin is not a useful metric for American Express.

American Express is a globally integrated payments company and bank holding company. Its income statement does not include a "cost of goods sold" or "cost of revenue" line item, which is the denominator component needed to compute gross margin (Revenue – COGS) / Revenue. Instead, the company reports its revenue lines (discount revenue, net card fees, service fees, net interest income, processed revenue) and then lists operating expenses (Card Member rewards, Card Member services, business development, marketing, etc.) directly below them to arrive at operating income.

Because interest and credit losses are part of the company's core operations rather than a separate "cost of goods" layer, the standard gross-margin framework simply doesn't map onto its financial statements. The more relevant profitability metrics for a company like this are **operating margin** (operating income / total revenues net of interest expense) and **net margin** (net income / total revenues net of interest expense).

Cited documents: AMERICANEXPRESS_2022_10K.pdf

Judge rationale: The agent's answer correctly states that gross margin is not a useful metric for American Express and provides a valid explanation, aligning with the essential information in the expected answer.

Q43. How much has the effective tax rate of American Express changed between FY2021 and FY2022?

Result: OK · Latency: 10.5 s · Tool calls: 2

Expected answer

The effective tax rate for American Express has changed/dropped from 24.6% in FY 2021 to 21.6% in FY 2022.

Delivered answer

Down 3.0 percentage points — from 24.6% in FY2021 to 21.6% in FY2022.

- FY2021: \$2,629M tax ÷ \$10,689M pretax income = 24.6%
- FY2022: \$2,071M tax ÷ \$9,585M pretax income = 21.6%
- Change: 21.6% – 24.6% = **–3.0 pp**

The 10-K attributes the drop mainly to a \$2.1 pp benefit from tax settlements and the lapse of statutes of limitations, plus a larger tax-credit benefit (–0.9 pp vs –0.1 pp in 2021), partly offset by higher state/local taxes (+3.1 pp vs +3.0 pp).

Cited documents: AMERICANEXPRESS_2022_10K.pdf

Judge rationale: The agent's answer correctly states the change in effective tax rate from 24.6% to 21.6% and includes the correct magnitude of the change (-3.0 percentage points), matching the essential information in the expected answer.

Q44. What was the largest liability in American Express's Balance Sheet in 2022?

Result: OK · Latency: 14.4 s · Tool calls: 3

Expected answer

Customer deposits

Delivered answer

Customer deposits — at **\$110.2 billion** (\$110,239 million) as of December 31, 2022, the largest liability on American Express's consolidated balance sheet.

- Customer deposits: \$110,239M
- Other liabilities: \$37,350M
- Long-term debt: \$42,573M
- Accounts payable: \$12,133M
- Short-term borrowings: \$1,348M
- **Total liabilities: \$203,643M**

Customer deposits represent roughly 54% of total liabilities, making it by far the dominant liability line.

Cited documents: AMERICANEXPRESS_2022_10K.pdf

Judge rationale: The agent's answer correctly identifies 'Customer deposits' as the largest liability, matching the essential information in the expected answer.

Q45. Was American Express able to retain card members during 2022?

Result: OK · Latency: 11.3 s · Tool calls: 2

Expected answer

Yes

Delivered answer

Yes. American Express was able to retain its card members during 2022 — the company's 2022 10-K states that "Card Member retention remained high" for the year, alongside record new card acquisitions, which it attributes to its investments in premium value propositions.

Supporting context from the same filing:

- **Net card fees** rose 17% year-over-year, driven by record new card acquisitions and high retention.
- **Proprietary cards-in-force** reached 76.7 million worldwide at December 31, 2022.
- **Billed business** grew 23% year-over-year, which management tied to its "continued ability to acquire, engage and retain high-spending, premium Card Members."

So retention was not just maintained but described as a strength of the year, reinforcing the growth in the overall card base.

If you'd like, I can pull the year-over-year card count or segment-level (U.S. Consumer, Commercial, International) detail to quantify the retention more precisely.

Cited documents: AMERICANEXPRESS_2022_10K.pdf

Judge rationale: The agent's answer explicitly confirms retention and aligns with the expected answer's essential information.

Q46. How much (in USD billions) did American Water Works pay out in cash dividends for FY2020? Compute or extract the answer by primarily using the details outlined in the statement of cash flows.

Result: OK (human review) · Latency: 6.3 s · Tool calls: 1

Expected answer

\$0.40

Delivered answer

\$0.389 billion (i.e., \$389 million).

From the Consolidated Statements of Cash Flows for the fiscal year ended December 31, 2020, the financing section reports **Dividends paid = \$(389) million** (in millions).

- Dividends paid (cash flow, financing): \$389 million
- Converted to billions: $389 / 1,000 = \mathbf{\$0.389 \text{ billion}}$

For context, this is the cash actually paid out during FY2020, which differs slightly from the \$400 million of dividends *declared* during the year (per the statement of changes in shareholders' equity, \$2.20 declared per share).

Cited documents: AMERICANWATERWORKS_2020_10K.pdf

Judge verdict: NOK · The agent's answer (\$0.389 billion) does not match the expected answer (\$0.40 billion) when rounded to the precision of the expected answer.

Human review: OK · The figure is more specific in the agent's response

Q47. Basing your judgments off of the cash flow statement and the income statement, what is American Water Works's FY2021 unadjusted operating income + depreciation and amortization from the cash flow statement (unadjusted EBITDA) in USD millions?

Result: OK · Latency: 6.4 s · Tool calls: 1

Expected answer

\$1832.00

Delivered answer

\$1,832 million.

- Operating income (FY2021, income statement): \$1,196M
- Depreciation & amortization (FY2021, cash flow statement): \$636M
- EBITDA = 1,196 + 636 = **\$1,832M**

(Both the income statement and the cash flow statement report the same \$636M D&A figure for FY2021, so the unadjusted EBITDA is \$1,832M.)

Cited documents: AMERICANWATERWORKS_2021_10K.pdf

Judge rationale: The agent's answer matches the expected answer in value and units, with no contradictions or missing essential information.

Q48. Does American Water Works have positive working capital based on FY2022 data? If working capital is not a useful or relevant metric for this company, then please state that and explain why.

Result: OK · Latency: 11.1 s · Tool calls: 2

Expected answer

No, American Water Works had negative working capital of -\$1561M in FY 2022.

Delivered answer

No — American Water Works had negative working capital at the end of FY2022.

- Total current assets (12/31/2022): \$1,250M
- Total current liabilities (12/31/2022): \$2,811M
- Working capital = \$1,250M – \$2,811M = **-\$1,561M**

Working capital is a standard, relevant metric for American Water Works — it is a regulated water and wastewater utility, not a financial institution, so the usual balance-sheet and liquidity measures all apply. The negative figure is driven largely by the \$1,175M of short-term debt and the \$281M current portion of long-term debt sitting in current liabilities, which is typical for a capital-intensive utility that funds infrastructure through debt rather than holding excess current assets.

Cited documents: AMERICANWATERWORKS_2022_10K.pdf

Judge rationale: The agent's answer correctly states that American Water Works had negative working capital of -\$1,561M in FY2022, matching the essential information in the expected answer.

Q49. In agreement with the information outlined in the income statement, what is the FY2015 - FY2017 3 year average net profit margin (as a %) for Best Buy? Answer in units of percents and round to one decimal place.

Result: OK · Latency: 7.5 s · Tool calls: 1

Expected answer

2.8%

Delivered answer

2.8%

Net profit margin = net earnings attributable to Best Buy Co., Inc. shareholders ÷ revenue, per the consolidated statements of earnings:

- FY2017: \$1,228M / \$39,403M = 3.12%
- FY2016: \$897M / \$39,528M = 2.27%
- FY2015: \$1,233M / \$40,339M = 3.06%

3-year average = (3.1165% + 2.2693% + 3.0567%) / 3 = 2.81% ≈ **2.8%**

Cited documents: BESTBUY_2017_10K.pdf

Judge rationale: The agent's answer matches the expected answer exactly in value and precision.

Q50. What is the year end FY2019 total amount of inventories for Best Buy? Answer in USD millions. Base your judgments on the information provided primarily in the balance sheet.

Result: OK · Latency: 6.0 s · Tool calls: 1

Expected answer

\$5409.00

Delivered answer

\$5,409 million.

Best Buy's fiscal year 2019 ended February 2, 2019. On the consolidated balance sheet as of that date, the inventory line item is "Merchandise inventories," reported at **\$5,409 million** (in millions, as filed). This is the company's only inventory line, so it is the total inventories.

(For reference, the prior year-end February 3, 2018 figure was \$5,209 million.)

Cited documents: BESTBUY_2019_10K.pdf

Judge rationale: The agent's answer matches the expected answer in value and units, with no contradictions or omissions.

Q51. Are Best Buy's gross margins historically consistent (not fluctuating more than roughly 2% each year)? If gross margins are not a relevant metric for a company like this, then please state that and explain why. Added note by ATG for disambiguation: the question is referring to FY2023.

Result: OK · Latency: 15.0 s · Tool calls: 2

Expected answer

Yes, the margins have been consistent, there has been a minor decline of 1.1% in gross margins between FY2022 and FY2023.

Delivered answer

Yes — Best Buy's gross margins are historically consistent; no year-over-year move in the recent track record exceeds roughly 2 percentage points.

Gross margin (gross profit ÷ revenue), as reported:

- FY2020 (ended Feb 1, 2020): **23.0%**
- FY2021 (ended Jan 30, 2021): **22.4%**
- FY2022 (ended Jan 29, 2022): **22.5%**
- FY2023 (ended Jan 28, 2023): **21.4%**

Year-over-year changes:

- FY2021 vs FY2020: $22.4\% - 23.0\% = -0.6$ pp
- FY2022 vs FY2021: $22.5\% - 22.4\% = +0.1$ pp
- FY2023 vs FY2022: $21.4\% - 22.5\% = -1.1$ pp

The largest single-year swing is 1.1 pp (FY2023), comfortably inside the ~2 pp threshold, and the whole four-year span stays within a 21.4%–23.0% band.

Gross margin is a relevant metric here: Best Buy is a consumer-electronics retailer, not a financial institution, so product margin is a core driver of its profitability and is squarely in scope.

Sources: FY2020–FY2022 figures from the FY2022 10-K; FY2023 figure from the Q4 FY23 earnings release.

Cited documents: BESTBUY_2023Q4_EARNINGS.pdf, BESTBUY_2022_10K.pdf

Judge rationale: The agent's answer confirms consistency with a 1.1% decline between FY2022 and FY2023, matching the expected answer's essential information.

Q52. What are major acquisitions that Best Buy has done in FY2023, FY2022 and FY2021?

Result: **OK** · Latency: 8.0 s · Tool calls: 1

Expected answer

Best Buy closed two acquisitions, both these companies were already partially owned by Best Buy, but Best Buy acquired all outstanding shares of these two companies during FY 2022: (1) Current Health Ltd and (2) Two Peaks, LLC d/b/a Yardbird Furniture

Delivered answer

Best Buy made **no major acquisitions in FY2023 or FY2021**, and **two in FY2022**: Current Health and Yardbird.

- **FY2023** (ended Jan 28, 2023): none — the cash flow statement shows \$0 for "Acquisitions, net of cash acquired," and the 10-K lists no new deals.
- **FY2022** (ended Jan 29, 2022):
 - **Current Health Ltd.** — acquired Nov 2, 2021 for **\$389M** net cash; a care-at-home technology platform (remote patient monitoring, telehealth, patient engagement) supporting the Best Buy Health strategy.
 - **Two Peaks, LLC d/b/a Yardbird Furniture** — acquired Nov 4, 2021 for **\$79M** net cash; a direct-to-consumer outdoor furniture company expanding the outdoor-living assortment.
- **FY2021** (ended Jan 30, 2021): none — the 10-K only references earlier deals (GreatCall, FY2019; CST and BioSensics, FY2020), with no new acquisition in the year.

Both FY2022 deals were described as not material to results of operations, and together they were the only acquisitions across the three-year window.

Cited documents: BESTBUY_2023_10K.pdf, BESTBUY_2022_10K.pdf, BESTBUY_2021_10K.pdf

Judge rationale: The agent's answer correctly identifies the two FY2022 acquisitions (Current Health Ltd and Yardbird Furniture) and confirms no major acquisitions in FY2023 or FY2021, matching the essential information in the expected answer.

Q53. Among operations, investing, and financing activities, which brought in the most (or lost the least) cash flow for Best Buy in FY2023?

Result: OK · Latency: 10.4 s · Tool calls: 1

Expected answer

Best Buy generated the most cash flow from operating activities in FY 2023 (\$1.8 bn)

Delivered answer

Operating activities brought in the most cash for Best Buy in FY2023 (fiscal year ended January 28, 2023) — it was the only one of the three that was net positive.

- Operating activities: **+\$1,824M**
- Investing activities: **−\$962M**
- Financing activities: **−\$1,806M**

So operating cash flow (+\$1,824M) not only beat the other two, it was the sole source of cash, while investing (capex of ~\$930M) and financing (share repurchases of ~\$1,014M plus ~\$789M of dividends) both consumed cash.

Cited documents: BESTBUY_2023Q4_EARNINGS.pdf

Judge rationale: The agent's answer correctly identifies operating activities as the source of the most cash flow and provides the essential figure (\$1.8 bn, matching \$1,824M when rounded).

Q54. Was there any drop in Cash & Cash equivalents between FY 2023 and Q2 of FY2024? Added note by ATG for disambiguation: the question is referring to Best Buy.

Result: OK · Latency: 7.2 s · Tool calls: 1

Expected answer

Yes, there was a decline of ~42% between FY2023 and Q2 of FY 2024.

Delivered answer

Yes — Best Buy's cash and cash equivalents fell between FY2023 and Q2 of FY2024.

- FY2023 (fiscal year ended January 28, 2023): **\$1,874 million**
- Q2 FY2024 (six months ended July 29, 2023): **\$1,093 million**

Change: \$1,093M − \$1,874M = **−\$781 million**, a decline of about **41.7%**.

Note: Best Buy's fiscal year ends in late January, so "FY2023" here is the year ended January 28, 2023, and "Q2 FY2024" is the quarter ended July 29, 2023.

Cited documents: BESTBUY_2024Q2_EARNINGS.pdf, BESTBUY_2023_10K.pdf

Judge rationale: The agent's answer confirms the decline and provides a figure (~41.7%) that matches the expected answer's ~42% when rounded to the same precision.

Q55. Was there any change in the number of Best Buy stores between Q2 of FY2024 and FY2023?

Result: **NOK** · Latency: 10.9 s · Tool calls: 2

Expected answer

Yes, there is decline in number stores by 1.32% from 982 stores in Q2 FY 2023 to 969 by the end of Q2 FY2024.

Delivered answer

Yes — the total store count fell by 9 stores, from 1,138 at the end of FY2023 to 1,129 at the end of Q2 FY2024.

- **End of FY2023** (fiscal year ended January 28, 2023): Domestic 978 + International 160 = **1,138**
- **End of Q2 FY2024** (quarter ended July 29, 2023): Domestic 969 + International 160 = **1,129**
- **Change:** 1,129 – 1,138 = **–9 stores** (a decrease)

The decline is entirely in the Domestic segment (978 → 969), driven mainly by Best Buy-branded stores (930 → 907), while the International (Canada) count held steady at 160.

Note on the periods: Best Buy's fiscal year ends on the Saturday nearest the end of January, so FY2023 ended January 28, 2023, and Q2 FY2024 ended July 29, 2023.

Cited documents: BESTBUY_2024Q2_10Q.pdf, BESTBUY_2023_10K.pdf

Judge rationale: The agent's answer contradicts the expected answer by providing different store counts (1,138 vs. 982 in FY2023 and 1,129 vs. 969 in Q2 FY2024).

Q56. Which Best Buy product category performed the best (by top line) in the domestic (USA) Market during Q2 of FY2024?

Result: NOK · Latency: 6.1 s · Tool calls: 1

Expected answer

The entertainment segment experienced the highest growth of 9% during Q2 FY2024, primarily from gaming division.

Delivered answer

Computing and Mobile Phones was the top-line leader in Best Buy's Domestic (USA) segment in Q2 FY2024.

- Domestic segment revenue: **\$8,890M** (three months ended July 29, 2023)
- Computing and Mobile Phones revenue mix: **41%** → **≈ \$3,645M**
- Next largest: Consumer Electronics at 30% (≈\$2,667M), then Appliances at 16%

So Computing and Mobile Phones generated roughly \$3.6B of the \$8.9B domestic top line, making it the clear #1 category by revenue.

One caveat worth noting: while it led on *top line*, it was actually one of the weaker categories on a comparable-sales basis (−6.4%), with the growth story in the quarter coming from Entertainment (+9.0%, driven by gaming) and Services (+7.6%).

Want me to break down the comparable-sales performance by category, or compare this to the International segment's mix?

Cited documents: BESTBUY_2024Q2_10Q.pdf

Judge rationale: The agent's answer identifies 'Computing and Mobile Phones' as the top-line leader, which contradicts the expected answer stating 'Entertainment' as the highest growth category by top line.

Q57. Considering the data in the balance sheet, what is Block's (formerly known as Square) FY2016 working capital ratio? Define working capital ratio as total current assets divided by total current liabilities. Round your answer to two decimal places.

Result: OK · Latency: 6.6 s · Tool calls: 1

Expected answer

1.73

Delivered answer

1.73

From Square's (Block) FY2016 consolidated balance sheet (fiscal year ended December 31, 2016):

- Total current assets = \$1,001,425 (thousands)
- Total current liabilities = \$577,464 (thousands)

Working capital ratio = $\$1,001,425 \div \$577,464 = 1.7342 \approx \mathbf{1.73}$

This indicates the company held about \$1.73 in current assets for every \$1 of current liabilities at year-end 2016.

Cited documents: BLOCK_2016_10K.pdf

Judge rationale: The agent's answer matches the expected answer (1.73) and includes the correct calculation and rounding.

Q58. What is the FY2019 - FY2020 total revenue growth rate for Block (formerly known as Square)? Answer in units of percents and round to one decimal place. Approach the question asked by assuming the standpoint of an investment banking analyst who only has access to the statement of income.

Result: OK · Latency: 8.5 s · Tool calls: 1

Expected answer

101.5%

Delivered answer

101.5%

Using the consolidated statement of operations (total net revenue, as originally reported):

- FY2019: \$4,713,500 thousand
- FY2020: \$9,497,578 thousand

Growth = $(9,497,578 - 4,713,500) / 4,713,500 = 4,784,078 / 4,713,500 = 1.0150 \rightarrow \mathbf{101.5\%}$

The jump is driven almost entirely by Bitcoin revenue, which went from \$516.5M in FY2019 to \$4,571.5M in FY2020.

Cited documents: BLOCK_2020_10K.pdf

Judge rationale: The agent's answer matches the expected answer exactly in value and precision, with no contradictions or missing essential information.

Q59. Using the cash flow statement, answer the following question to the best of your abilities: how much did Block (formerly known as Square) generate in cash flow from operating activities in FY2020? Answer in USD millions.

Result: OK (human review) · Latency: 7.1 s · Tool calls: 1

Expected answer

\$382.00

Delivered answer

\$381.6 million (as originally reported in the FY2020 10-K); later restated to **\$173.1 million** in the 2021 and 2022 10-Ks.

- FY2020 10-K (Square, Inc.), Consolidated Statements of Cash Flows: "Net cash provided by operating activities" = \$381,603 thousand = **\$381.6M**.
- FY2021 10-K (Block, Inc.), same line for 2020 = \$173,110 thousand = **\$173.1M** (carried identically in the FY2022 10-K).

The gap comes from a reclassification: the original 2020 filing kept the change in customer funds (–\$1,151.5M) inside operating activities, while the restated filings moved customer funds into financing activities ("Change in customer funds, restricted from use in the Company's operations," +\$1,361.5M), which lowered operating cash flow to \$173.1M.

So, using the figure from the year's own annual filing as originally reported: **\$381.6M**.

Cited documents: BLOCK_2021_10K.pdf, BLOCK_2020_10K.pdf

Judge verdict: NOK · The agent's answer includes a restated figure (\$173.1M) that contradicts the expected answer (\$382.00M), and the original figure (\$381.6M) does not match the expected precision or value.

Human review: OK · The agent provided a more detailed response than expected.

Q60. We need to calculate a financial metric by using information only provided within the balance sheet. Please answer the following question: what is Boeing's year end FY2018 net property, plant, and equipment (in USD millions)?

Result: OK · Latency: 7.1 s · Tool calls: 1

Expected answer

\$12645.00

Delivered answer

\$12,645 million.

Boeing's Consolidated Statements of Financial Position (balance sheet) for the fiscal year ended December 31, 2018 reports "Property, plant and equipment, net" at **\$12,645 million** (vs. \$12,672 million at December 31, 2017).

This is the net figure, i.e. gross PP&E of \$31,213 million less accumulated depreciation of \$18,568 million, as detailed in Note 11 of the same filing.

Cited documents: BOEING_2018_10K.pdf

Judge rationale: The agent's answer matches the expected answer's essential figure (\$12,645 million) and contains no contradictions.

Q61. Are there any product categories / service categories that represent more than 20% of Boeing's revenue for FY2022?

Result: OK · Latency: 6.4 s · Tool calls: 1

Expected answer

Yes. Boeing has product and service categories that represent more than 20% of Boeing's revenue for FY2022. These categories are Commercial Airplanes which comprises 39% of total revenue, Defence which comprises 35% of total revenue and Services which comprises 26% of total revenue.

Delivered answer

Yes — three of Boeing's four reportable segments each exceeded 20% of FY2022 revenue (year ended December 31, 2022; total revenues \$66,608M):

- **Commercial Airplanes (BCA):** \$25,867M → 38.8%
- **Defense, Space & Security (BDS):** \$23,162M → 34.8%
- **Global Services (BGS):** \$17,611M → 26.4%

The fourth segment, **Boeing Capital (BCC)**, was negligible at \$199M (0.3%), so it falls well below the 20% threshold.

Cited documents: BOEING_2022_10K.pdf

Judge rationale: The agent's answer correctly identifies the three categories exceeding 20% of Boeing's FY2022 revenue and provides matching percentages (38.8% ≈ 39%, 34.8% ≈ 35%, 26.4% ≈ 26%).

Q62. Has Boeing reported any materially important ongoing legal battles from FY2022?

Result: **OK** · Latency: 6.3 s · Tool calls: 1

Expected answer

Yes. Multiple lawsuits have been filed against Boeing resulting from a 2018 Lion Air crash and a 2019 Ethiopian Airlines crash.

Delivered answer

Yes. Boeing's FY2022 10-K (Note 21 – Legal Proceedings) flags two families of ongoing matters it could not reasonably estimate, both materially important to an investor:

- **737 MAX accident litigation** — multiple lawsuits stemming from the Lion Air Flight 610 (Oct 2018) and Ethiopian Airlines Flight 302 (Mar 2019) crashes were still pending. The DOJ Deferred Prosecution Agreement (2021) and the SEC investigation (settled Sept 2022, \$200M civil penalty) were resolved, but Boeing stated it "cannot reasonably estimate a range of loss, if any, not covered by available insurance" from the remaining pending lawsuits and other 737 MAX matters.
- **Embraer arbitration** — Boeing terminated its 2019 joint-venture agreements in 2020; Embraer disputed the termination and the dispute was in arbitration, expected to complete in late 2023 or early 2024, with no reasonably estimable range of loss.

Cited documents: BOEING_2022_10K.pdf

Judge rationale: The agent's answer includes the essential information from the expected answer (lawsuits from the 2018 Lion Air and 2019 Ethiopian Airlines crashes) and adds non-contradictory details.

Q63. Does Boeing have an improving gross margin profile as of FY2022? If gross margin is not a useful metric for a company like this, then state that and explain why.

Result: OK · Latency: 7.5 s · Tool calls: 1

Expected answer

Yes. Boeing has an improving gross margin profile as of FY2022. Gross profit improved from \$3,017 million in FY2021 to \$3,502 million in FY2022. Gross margin % improved from 4.8% in FY2021 to 5.3% in FY2022.

Delivered answer

Yes — Boeing's gross margin improved as of FY2022 (fiscal year ended December 31, 2022). Gross margin is a relevant metric here because Boeing is a manufacturer of physical products (aircraft and defense systems), so the spread between selling price and direct production cost is a core profitability driver.

Gross margin = (Total revenues – cost of products – cost of services) / Total revenues:

- FY2022: $(66,608 - 53,969 - 9,109) / 66,608 = 3,530 / 66,608 = \mathbf{5.30\%}$
- FY2021: $(62,286 - 49,954 - 9,283) / 62,286 = 3,049 / 62,286 = \mathbf{4.90\%}$

Change: +0.40 percentage points (5.30% vs. 4.90%).

The improvement was driven by higher Commercial Airplanes revenue (737 and 787 deliveries) outpacing the rise in cost of products, while cost of services actually fell year over year.

Cited documents: BOEING_2022_10K.pdf

Judge rationale: The agent's answer confirms the improving gross margin profile with figures that match the expected answer when rounded (5.3% vs 5.30%, 4.9% vs 4.8% after rounding 4.90% to one decimal).

Q64. Who are the primary customers of Boeing as of FY2022?

Result: **OK** · Latency: 15.7 s · Tool calls: 4

Expected answer

Boeing's primary customers as of FY2022 are a limited number of commercial airlines and the US government. The US government accounted for 40% of Boeing's total revenues in FY2022.

Delivered answer

Boeing's primary customers as of FY2022 (fiscal year ended December 31, 2022) are **commercial airlines** and the **U.S. government (principally the U.S. Department of Defense)**, with the mix varying by segment:

- **Commercial Airplanes (BCA):** serves "principally the commercial airline industry worldwide" — its largest segment at \$25,867M of revenue in 2022, split roughly evenly between U.S. (\$12,167M) and non-U.S. (\$13,567M) customers.
- **Defense, Space & Security (BDS):** the U.S. government is the dominant customer, accounting for **89% of BDS revenue** in 2022 (including foreign military sales through the U.S. government); total BDS revenue was \$23,162M.
- **Global Services (BGS):** serves both commercial and defense customers worldwide; the U.S. government accounted for 33% of BGS revenue (\$7,681M of \$17,611M).

In aggregate, 40% of Boeing's total 2022 revenues were earned pursuant to U.S. government contracts (including foreign military sales).

Cited documents: BOEING_2022_10K.pdf

Judge rationale: The agent's answer includes all essential information from the expected answer (primary customers: commercial airlines and US government; US government accounted for 40% of total revenues) and adds non-contradictory details.

Q65. Is Boeing's business subject to cyclical? Added note by ATG for disambiguation: the question is referring to FY2022.

Result: OK · Latency: 7.2 s · Tool calls: 1

Expected answer

Yes, Boeing's business is subject to cyclical due to its exposure to the airline industry which is a cyclical industry.

Delivered answer

Yes. Boeing's business is subject to cyclical, driven primarily by its Commercial Airplanes segment.

- The commercial aircraft market is "predominantly driven by long-term trends in airline passenger and cargo traffic," which in turn depend on sustained economic growth and political stability in developed and emerging markets.
- Boeing states directly that "historically, the airline industry has been cyclical and very competitive and has experienced significant profit swings," so a deterioration in the global economy or in airline profitability can lead to fewer orders, postponed or cancelled orders, and lower revenues, profitability and cash flows.
- This cyclical is partly offset at the group level: Boeing notes that "the cyclical of commercial and defense markets sometimes offset," which is why it pairs its cyclical commercial business with the more stable Defense, Space & Security business.

In short, the commercial side is the cyclical driver, and the company's own strategy relies on defense to temper that cyclical.

Cited documents: BOEING_2022_10K.pdf

Judge rationale: The agent's answer confirms Boeing's business is subject to cyclical and provides additional correct details without contradicting the expected answer.

Q66. What production rate changes is Boeing forecasting for FY2023?

Result: OK · Latency: 6.4 s · Tool calls: 1

Expected answer

Boeing forecasts an increase in the production rates for the 737, 777X and 787 aircrafts in 2023.

Delivered answer

Based on Boeing's 2022 10-K (fiscal year ended December 31, 2022), the company forecast the following production rate changes for calendar year 2023:

- **737:** Gradual increases from the current 31 per month, paced by market demand and supply chain capacity (no specific target rate given).
- **787:** Gradual ramp-up from low rates to **5 per month**.
- **777X:** Production **resumes** in 2023 (it had been paused since Q2 2022).
- **747:** Production **completes** — the last aircraft delivery is expected in early 2023.
- **767:** No change; continuing at a combined rate of 3 per month.
- **777/777X (combined):** Continuing at 3 per month.

In short: the 737 ramps up gradually, the 787 climbs to 5/month, the 777X restarts, and the 747 program ends.

Cited documents: BOEING_2022_10K.pdf

Judge rationale: The agent's answer includes the essential information that Boeing forecasts production rate increases for the 737, 787, and 777X in 2023, with no contradictions.

Q67. How does Boeing's effective tax rate in FY2022 compare to FY2021?

Result: **OK** · Latency: 7.1 s · Tool calls: 1

Expected answer

Effective tax rate in FY2022 was 0.62%, compared to -14.76% in FY2021.

Delivered answer

Boeing's effective tax rate rose from -14.8% in FY2021 to 0.6% in FY2022 — a swing of about +15.4 percentage points. Both years were loss years, so the rates are unusual in sign.

Calculation (effective tax rate = income tax expense ÷ pre-tax income):

- **FY2022:** income tax expense of \$(31)M on a pre-tax loss of \$(5,022)M → $-31 \div -5,022 = \mathbf{0.62\%}$
- **FY2021:** income tax benefit of \$743M on a pre-tax loss of \$(5,033)M → $743 \div -5,033 = \mathbf{-14.76\%}$

So the rate moved from a negative (benefit) in FY2021 to a small positive (expense) in FY2022, an increase of roughly 15.4 percentage points. The shift reflects the absence of the FY2021 income tax benefit — Boeing's own MD&A notes that the year-over-year increase in net loss was "primarily driven by the absence of an income tax benefit in 2022."

Cited documents: BOEING_2022_10K.pdf

Judge rationale: The agent's answer includes the essential figures (0.6% for FY2022 and -14.76% for FY2021) and correctly describes the comparison, with no contradictions.

Q68. What is the FY2017 return on assets (ROA) for Coca Cola? ROA is defined as: FY2017 net income / (average total assets between FY2016 and FY2017). Round your answer to two decimal places. Give a response to the question by relying on the details shown in the balance sheet and the P&L statement.

Result: OK · Latency: 7.3 s · Tool calls: 1

Expected answer

0.01

Delivered answer

0.01

Using the FY2017 10-K (fiscal year ended December 31, 2017):

- FY2017 net income (attributable to shareowners of The Coca-Cola Company): \$1,248M
- Total assets FY2017: \$87,896M; FY2016: \$87,270M
- Average total assets = $(87,896 + 87,270) / 2 = \$87,583\text{M}$
- $\text{ROA} = 1,248 / 87,583 = 0.0143 \approx \mathbf{0.01}$

(If you use consolidated net income of \$1,283M instead of the shareowner-attributable figure, $\text{ROA} = 1,283 / 87,583 = 0.0146$, which still rounds to 0.01.)

Cited documents: COCACOLA_2017_10K.pdf

Judge rationale: The agent's answer matches the expected answer (0.01) and includes the correct calculation steps without contradictions.

Q69. What is Coca Cola's FY2021 COGS % margin? Calculate what was asked by utilizing the line items clearly shown in the income statement.

Result: OK · Latency: 5.8 s · Tool calls: 1

Expected answer

39.7%

Delivered answer

39.73%

Coca-Cola's FY2021 (year ended December 31, 2021) cost of goods sold as a percentage of net operating revenues:

- Cost of goods sold: \$15,357 million
- Net operating revenues: \$38,655 million
- $15,357 \div 38,655 = 0.3973 \rightarrow \mathbf{39.73\%}$

Source: The Coca-Cola Company Consolidated Statements of Income, FY2021 10-K.

Cited documents: COCACOLA_2021_10K.pdf

Judge rationale: The agent's answer (39.73%) matches the expected answer (39.7%) when rounded to one decimal place.

Q70. What is Coca Cola's FY2022 dividend payout ratio (using total cash dividends paid and net income attributable to shareholders)? Round answer to two decimal places. Answer the question asked by assuming you only have access to information clearly displayed in the cash flow statement and the income statement.

Result: OK · Latency: 6.6 s · Tool calls: 1

Expected answer

0.8

Delivered answer

0.80 ($\approx 80\%$)

- Total cash dividends paid (cash flow statement, FY2022): \$7,616 million
- Net income attributable to shareowners (income statement, FY2022): \$9,542 million
- Payout ratio = $7,616 / 9,542 = 0.7982 \rightarrow \mathbf{0.80}$

Source: The Coca-Cola Company 2022 10-K (fiscal year ended December 31, 2022).

Cited documents: COCACOLA_2022_10K.pdf

Judge rationale: The agent's answer matches the expected answer (0.80 rounded to two decimal places is 0.8) and includes correct supporting calculations.

Q71. Based on the information provided primarily in the balance sheet and the statement of income, what is FY2020 days payable outstanding (DPO) for Corning? DPO is defined as: $365 \times (\text{average accounts payable between FY2019 and FY2020}) / (\text{FY2020 COGS} + \text{change in inventory between FY2019 and FY2020})$. Round your answer to two decimal places.

Result: OK · Latency: 10.1 s · Tool calls: 1

Expected answer

63.86

Delivered answer

63.86 days

Using Corning's FY2020 10-K (fiscal year ended December 31, 2020):

- Average accounts payable = $(\text{FY2020 } \$1,174\text{M} + \text{FY2019 } \$1,587\text{M}) / 2 = \$1,380.5\text{M}$
- Change in inventory = $\text{FY2020 } \$2,438\text{M} - \text{FY2019 } \$2,320\text{M} = \$118\text{M}$
- Denominator = $\text{FY2020 COGS } \$7,772\text{M} + \$118\text{M} = \$7,890\text{M}$

$\text{DPO} = 365 \times \$1,380.5\text{M} / \$7,890\text{M} = 503,882.5 / 7,890 = \mathbf{63.86}$

Cited documents: CORNING_2020_10K.pdf

Judge rationale: The agent's answer matches the expected answer exactly, including the correct calculation and rounding to two decimal places.

Q72. Taking into account the information outlined in the income statement, what is the FY2019 - FY2021 3 year average unadjusted operating income % margin for Corning? Answer in units of percents and round to one decimal place.

Result: OK · Latency: 10.7 s · Tool calls: 3

Expected answer

10.3%

Delivered answer

10.3%

From Corning's Consolidated Statements of Income (FY2019–FY2021, as reported in the 2021 10-K):

Year	Operating Income (\$M)	Net Sales (\$M)	Operating Margin
2019	1,306	11,503	11.35%
2020	509	11,303	4.50%
2021	2,112	14,082	15.00%

3-year average = $(11.35\% + 4.50\% + 15.00\%) / 3 = 10.3\%$

Cited documents: CORNING_2021_10K.pdf

Judge rationale: The agent's answer matches the expected answer exactly in value, precision, and units.

Q73. How much has the effective tax rate of Corning changed between FY2021 and FY2022?

Result: OK · Latency: 14.8 s · Tool calls: 2

Expected answer

The effective tax rate of Corning has changed from 20% in FY2021 to 23% in FY 2022.

Delivered answer

Increased by 2.7 percentage points — from 20.2% in FY2021 to 22.9% in FY2022.

- FY2021: provision for income taxes \$491M ÷ income before income taxes \$2,426M = 20.2%
- FY2022: provision for income taxes \$411M ÷ income before income taxes \$1,797M = 22.9%
- Change: 22.9% – 20.2% = **+2.7 pp**

Both figures are taken from Corning's FY2022 10-K, which reports the two years on a consistent basis (the 2021 rate was originally reported as 20.5% in the FY2021 10-K, before income before taxes was restated to \$2,426M).

Cited documents: CORNING_2022_10K.pdf

Judge rationale: The agent's answer provides the essential change (increase of ~3 percentage points) and the underlying figures match the expected answer when rounded to the same precision (20% and 23%).

Q74. Does Corning have positive working capital based on FY2022 data? If working capital is not a useful or relevant metric for this company, then please state that and explain why.

Result: **NOK** · Latency: 18.4 s · Tool calls: 2

Expected answer

Yes. Corning had a positive working capital amount of \$831 million by FY 2022 close. This answer considers only operating current assets and current liabilities that were clearly shown in the balance sheet.

Delivered answer

Yes — Corning had positive working capital at the end of FY2022.

- Total current assets (Dec 31, 2022): \$7,453M
- Total current liabilities (Dec 31, 2022): \$5,175M
- Working capital = \$7,453M – \$5,175M = **\$2,278M** (positive)

Working capital is a relevant and standard liquidity metric for Corning — it's a manufacturing/glass company (not a financial institution), so this measure applies normally. The positive figure indicates current assets comfortably exceed current obligations.

Source: Corning FY2022 10-K, Consolidated Balance Sheets.

Cited documents: CORNING_2022_10K.pdf

Judge rationale: The agent's working capital figure (\$2,278M) contradicts the expected answer (\$831M) and does not match the reference's scope (operating current assets and liabilities only).

Q75. Using only the information within the balance sheet, how much total assets did Costco have at the end of FY2021? Answer in USD millions.

Result: **OK** · Latency: 7.9 s · Tool calls: 1

Expected answer

\$59268.00

Delivered answer

\$59,268 million.

Costco's consolidated balance sheet at the end of FY2021 (fiscal year ended August 29, 2021) reports total assets of \$59,268 million, made up of total current assets of \$29,505 million plus other assets (property and equipment, net of \$23,492 million; operating lease right-of-use assets of \$2,890 million; and other long-term assets of \$3,381 million).

Cited documents: COSTCO_2021_10K.pdf

Judge rationale: The agent's answer matches the expected total assets figure of \$59,268 million in USD millions.

Q76. What is the FY2018 fixed asset turnover ratio for CVS Health? Fixed asset turnover ratio is defined as: FY2018 revenue / (average PP&E between FY2017 and FY2018). Round your answer to two decimal places. Calculate what was asked by utilizing the line items clearly shown in the P&L statement and the balance sheet.

Result: OK · Latency: 10.6 s · Tool calls: 1

Expected answer

17.98

Delivered answer

17.98

- FY2018 total revenues: \$194,579 million
- PP&E, net FY2018: \$11,349 million; FY2017: \$10,292 million
- Average PP&E = $(11,349 + 10,292) / 2 = \$10,820.5$ million
- Fixed asset turnover = $194,579 / 10,820.5 = \mathbf{17.98}$

Note: FY2018 is the year CVS Health acquired Aetna (closed Nov 28, 2018), so revenues include only ~3 weeks of Aetna, while the year-end PP&E already reflects Aetna's assets — which is why the ratio looks unusually high.

Sources: CVS Health 2018 10-K (consolidated balance sheet and income statement).

Cited documents: CVSHEALTH_2018_10K.pdf

Judge rationale: The agent's answer matches the expected answer exactly (17.98) and includes the correct calculation steps and figures.

Q77. Is CVS Health a capital-intensive business based on FY2022 data?

Result: OK · Latency: 15.8 s · Tool calls: 2

Expected answer

Yes, CVS Health requires an extensive asset base to operate, which is evident from its ROA of only 1.82% in 2022 and 3.39% in 2021, though it should be noted that a significant portion of this asset base is goodwill, and CVS's fixed assets/total assets ratio is on the lower side of 5.6%.

Delivered answer

Yes, CVS Health is a capital-intensive business based on FY2022 data (fiscal year ended December 31, 2022).

Here are the three metrics that support that conclusion:

Metric	Calculation	Result
CapEx / Revenue	\$2,727 M / \$322,467 M	0.85%
Net PP&E / Total assets	\$12,873 M / \$228,275 M	5.64%
Return on assets (year-end)	\$4,149 M / \$228,275 M	1.82%
Return on assets (2-yr avg)	\$4,149 M / \$230,637 M	1.80%

The defining characteristic is the **low return on a very large asset base**: CVS earns roughly 1.8% on \$228 billion of total assets. Per the standard test, a company that earns a low return on a large asset base is capital-intensive, even when goodwill (\$78,150 M) and intangible assets (\$24,754 M) — rather than physical fixed assets — make up most of that base.

A caveat worth noting: the 2022 net income of \$4,149 M was depressed by one-time charges (opioid litigation of \$5,803 M and a loss on assets held for sale of \$2,533 M). Excluding those, the underlying ROA would be meaningfully higher, but the reported FY2022 figures still point to a capital-intensive profile.

Cited documents: CVSHEALTH_2022_10K.pdf

Judge rationale: The agent's answer confirms CVS Health is capital-intensive, includes the essential ROA figure (1.82%) and the fixed assets/total assets ratio (~5.6%), and does not contradict the expected answer.

Q78. Has CVS Health reported any materially important ongoing legal battles from 2022, 2021 and 2020?

Result: **NOK** · Latency: 8.0 s · Tool calls: 1

Expected answer

Yes, CVS Health has been involved in multiple ongoing legal battles. Some notable legal dispute areas for CVS are: (1) usual and customary pricing litigation: where it's claimed that CVS's retail pharmacies overcharged for prescription drugs; (2) PBM litigation and investigations: where it's claimed that that rebate agreements between the drug manufacturers and PBMs caused inflated prices for certain drug products; and (3) controlled substances litigation: legal matters around opioids for which CVS has agreed to pay up to \$4.3 billion to claimants in remediation and \$625 million to attorneys and fees

Delivered answer

Yes. The standout is the **National Opioid Litigation**, which was ongoing across all three years and became materially significant in 2022.

- **2022:** CVS recorded **\$5.8 billion** (\$5,803 million) in pre-tax opioid litigation charges on the income statement, driven by settlement charges of **\$5.2 billion** (Q3) and **\$99 million** (Q4) tied to agreements with governmental entities and tribes. The 10-K states the outcome "could have a material effect on the Company's business, financial condition, operating results and/or cash flows."
- **2021:** The opioid MDL and related government/tribal claims were disclosed as ongoing, alongside PBM, usual-and-customary, and stockholder/ERISA matters; no material charge was booked that year.
- **2020:** The opioid MDL (In re National Prescription Opiate Litigation, MDL No. 2804) was ongoing, along with usual-and-customary and PBM/insulin-pricing litigation and investigations; no material charge was booked.

So the answer is yes — the opioid litigation is the materially important ongoing legal battle, and it is the one that produced a material financial impact (the \$5.8B charge in 2022).

Cited documents: CVSHEALTH_2022_10K.pdf, CVSHEALTH_2020_10K.pdf, CVSHEALTH_2021_10K.pdf

Judge rationale: The agent's answer omits essential information about the usual and customary pricing litigation and PBM litigation, which are explicitly mentioned in the expected answer as materially important ongoing legal battles.

Q79. Has CVS Health paid dividends to common shareholders in Q2 of FY2022?

Result: OK · Latency: 6.9 s · Tool calls: 1

Expected answer

Yes, CVS paid a \$ 0.55 dividend per share every quarter in FY2022

Delivered answer

Yes. CVS Health paid its quarterly cash dividend of \$0.55 per share in Q2 of FY2022 (April–June 2022), as it does every quarter — the company has paid cash dividends every quarter since becoming a public company.

- Quarterly dividend in 2022: \$0.55 per share (up from \$0.50 in 2021).
- Total common stock dividends for full-year 2022: \$2,910 million, per the Consolidated Statements of Shareholders' Equity.

So the Q2 2022 dividend was paid at the \$0.55/share rate that was in effect for the entire year.

Cited documents: CVSHEALTH_2022_10K.pdf

Judge rationale: The agent's answer confirms the essential fact that CVS Health paid a \$0.55 dividend per share in Q2 of FY2022, matching the expected answer.

Q80. Does Foot Locker's new CEO have previous CEO experience in a similar company to Footlocker?

Added note by ATG for disambiguation: the question is referring to the filing dated 2022-08-19.

Result: OK · Latency: 6.8 s · Tool calls: 1

Expected answer

Yes. She was previous CEO of Ulta Beauty which means she had to manage a large retail company that has brick and mortar + online business. So yes she was a CEO in a similar company to Foot Locker before this.

Delivered answer

Yes. Foot Locker's new CEO, Mary N. Dillon (appointed effective September 1, 2022), previously served as CEO of a similar specialty-retail company.

- She was **Chief Executive Officer of Ulta Beauty, Inc.** from July 2013 through June 2021 — a specialty beauty retailer, the closest analogue to Foot Locker's specialty athletic-footwear/apparel retail model.
- (She also led U.S. Cellular as President & CEO from 2010–2013, but that is a telecom business, not a comparable retailer.)

The filing notes she guided Ulta to become the leading U.S. beauty destination and omni-channel retailer over that tenure, which is the relevant "similar company" CEO experience.

Cited documents: FOOTLOCKER_2022_8K_dated_2022-08-19.pdf

Judge rationale: The agent's answer confirms the new CEO had prior CEO experience at Ulta Beauty, a similar retail company, matching the essential information in the expected answer.

Q81. Were there any board member nominees who had substantially more votes against joining than the other nominees? Added note by ATG for disambiguation: the question is referring to Foot Locker's filing dated 2022-05-20.

Result: OK · Latency: 7.7 s · Tool calls: 1

Expected answer

Yes, his name is Richard A. Johnson

Delivered answer

Yes — **Richard A. Johnson** received substantially more votes against than the other nine nominees.

- Johnson: **16,105,005** votes against
- Next highest, Dona D. Young: **6,074,467** against
- The remaining nominees ranged from ~4.4M (Tristan Walker, 4,419,340) to ~5.8M (Guillermo G. Marmol, 5,753,395), with Virginia C. Drosos the lowest at 294,935.

Johnson's "against" count was roughly 2.6× the next-highest nominee and about 3× the typical cluster, making him the clear outlier. All ten nominees were nonetheless elected.

Cited documents: FOOTLOCKER_2022_8K_dated-2022-05-20.pdf

Judge rationale: The agent's answer correctly identifies Richard A. Johnson as the nominee with substantially more votes against, matching the essential information in the expected answer.

Q82. What is the FY2019 cash conversion cycle (CCC) for General Mills? CCC is defined as: DIO + DSO - DPO. DIO is defined as: $365 \times (\text{average inventory between FY2018 and FY2019}) / (\text{FY2019 COGS})$. DSO is defined as: $365 \times (\text{average accounts receivable between FY2018 and FY2019}) / (\text{FY2019 Revenue})$. DPO is defined as: $365 \times (\text{average accounts payable between FY2018 and FY2019}) / (\text{FY2019 COGS} + \text{change in inventory between FY2018 and FY2019})$. Round your answer to two decimal places. Address the question by using the line items and information shown within the income statement and the balance sheet.

Result: OK · Latency: 8.5 s · Tool calls: 1

Expected answer

-3.7

Delivered answer

-3.70 days (General Mills, FY2019, fiscal year ended May 26, 2019).

Inputs from the FY2019 10-K (in \$M):

- Revenue (net sales) = 16,865.2; COGS (cost of sales) = 11,108.4
- Inventory: FY2019 = 1,559.3, FY2018 = 1,642.2 → avg = 1,600.75
- Receivables: FY2019 = 1,679.7, FY2018 = 1,684.2 → avg = 1,681.95
- Accounts payable: FY2019 = 2,854.1, FY2018 = 2,746.2 → avg = 2,800.15

Components:

- DIO = $365 \times 1,600.75 / 11,108.4 = 52.60$
- DSO = $365 \times 1,681.95 / 16,865.2 = 36.40$
- DPO = $365 \times 2,800.15 / (11,108.4 + (1,559.3 - 1,642.2)) = 365 \times 2,800.15 / 11,025.5 = 92.70$

CCC = $52.60 + 36.40 - 92.70 = -3.70$ days (the negative result reflects payables turning over more slowly than inventory and receivables combined).

Cited documents: GENERALMILLS_2019_10K.pdf

Judge rationale: The agent's answer matches the expected answer (-3.7) when rounded to two decimal places and contains no contradictions.

Q83. By drawing conclusions from the information stated only in the statement of financial position, what is General Mills's FY2020 working capital ratio? Define working capital ratio as total current assets divided by total current liabilities. Round your answer to two decimal places.

Result: OK · Latency: 11.1 s · Tool calls: 2

Expected answer

0.68

Delivered answer

0.68

From General Mills's FY2020 consolidated balance sheet (fiscal year ended May 31, 2020):

- Total current assets: \$5,121.3 million
- Total current liabilities: \$7,491.5 million

Working capital ratio = $5,121.3 / 7,491.5 = 0.6836 \approx \mathbf{0.68}$

This is below 1.0, indicating that current liabilities exceeded current assets at fiscal year-end.

Cited documents: GENERALMILLS_2020_10K.pdf

Judge rationale: The agent's answer matches the expected answer (0.68) and includes the correct calculation and rounding.

Q84. According to the information provided in the statement of cash flows, what is the FY2020 free cash flow (FCF) for General Mills? FCF here is defined as: (cash from operations - capex). Answer in USD millions.

Result: OK · Latency: 9.9 s · Tool calls: 1

Expected answer

\$3215.00

Delivered answer

\$3,215.4 million.

From the Consolidated Statements of Cash Flows in General Mills' FY2020 10-K (fiscal year ended May 31, 2020):

- Net cash provided by operating activities: \$3,676.2M
- Less: Purchases of land, buildings, and equipment (capex): \$460.8M
- FCF = \$3,676.2M – \$460.8M = **\$3,215.4M**

This matches the "Free cash flow" line reported in the Selected Financial Data section of the same filing.

Cited documents: GENERALMILLS_2020_10K.pdf

Judge rationale: The agent's answer (\$3,215.4 million) matches the expected answer (\$3215.00) when rounded to the same precision.

Q85. We want to calculate a financial metric. Please help us compute it by basing your answers off of the cash flow statement and the income statement. Here's the question: what is the FY2022 retention ratio (using total cash dividends paid and net income attributable to shareholders) for General Mills? Round answer to two decimal places.

Result: OK · Latency: 9.6 s · Tool calls: 1

Expected answer

0.54

Delivered answer

0.54

General Mills FY2022 (fiscal year ended May 29, 2022):

- Net earnings attributable to General Mills (income statement): \$2,707.3M
- Total cash dividends paid (cash flow statement, financing section): \$1,244.5M

Retention ratio = (Net income – Dividends paid) / Net income = $(2,707.3 - 1,244.5) / 2,707.3 = 1,462.8 / 2,707.3 = 0.5403 \approx$
0.54 (equivalently, 1 – payout ratio of 0.46).

Cited documents: GENERALMILLS_2022_10K.pdf

Judge rationale: The agent's answer matches the expected retention ratio of 0.54 and includes the correct calculation and rounding.

Q86. Are JnJ's FY2022 financials that of a high growth company?

Result: OK · Latency: 7.9 s · Tool calls: 1

Expected answer

No, JnJ's FY2022 financials are not of a high growth company as sales grew by 1.3% in FY2022.

Delivered answer

No — JnJ's FY2022 (fiscal year ended January 1, 2023) financials are those of a mature, stable large-cap, not a high-growth company.

- **Revenue:** \$94.9B, up only **1.3%** reported (operational growth 6.1%, offset by a –4.8% currency drag).
- **Net earnings:** \$17.9B, **down 14.1%** vs. \$20.9B in 2021; diluted EPS fell 13.8% to \$6.73.
- **Long-run growth:** 5-year worldwide sales CAGR of **4.4%** (10-year: 3.5%).

A high-growth company would show sustained double-digit revenue expansion; JnJ's low single-digit reported growth, declining GAAP earnings, and ~4% multi-year CAGR mark it as a slow-growth, dividend-paying healthcare incumbent rather than a high-growth name.

Cited documents: JOHNSON_JOHNSON_2022Q4_EARNINGS.pdf, JOHNSON_JOHNSON_2022_10K.pdf

Judge rationale: The agent's answer correctly states that JnJ's FY2022 financials are not of a high growth company and includes the essential fact that sales grew by 1.3%.

Q87. What drove gross margin change as of FY2022 for JnJ? If gross margin is not a useful metric for a company like this, then please state that and explain why.

Result: OK · Latency: 9.7 s · Tool calls: 1

Expected answer

For FY22, JnJ had changes in gross margin due to: One-time COVID-19 vaccine manufacturing exit related costs, Currency impacts in the Pharmaceutical segment, Commodity inflation in the MedTech and Consumer Health segments, partially offset by Supply chain benefits in the Consumer Health segment.

Delivered answer

Gross margin declined by approximately 0.9 percentage points in FY2022 (fiscal year ended January 1, 2023), from 68.16% to 67.25%.

Calculation:

- FY2022: Gross profit \$63,854M ÷ Sales \$94,943M = 67.25%
- FY2021: Gross profit \$63,920M ÷ Sales \$93,775M = 68.16%
- Change: -0.91 pp

Drivers (per JnJ's MD&A):

- One-time COVID-19 vaccine manufacturing exit costs (~\$1.5B in 2022)
- Unfavorable currency impacts in the Pharmaceutical segment
- Commodity inflation in MedTech and Consumer Health
- Partially offset by supply chain benefits in Consumer Health

Intangible asset amortization embedded in cost of products sold fell from \$4.7B to \$4.3B, which was a modest tailwind.

Gross margin is a fully applicable metric for JnJ — it is a diversified healthcare manufacturer (Pharmaceutical, MedTech, Consumer Health), not a financial institution, so cost of goods sold and gross profit are meaningful operating lines.

Cited documents: JOHNSON_JOHNSON_2022_10K.pdf

Judge rationale: The agent's answer includes all essential drivers of the gross margin change (COVID-19 vaccine exit costs, currency impacts, commodity inflation, and supply chain benefits) and does not contradict the expected answer.

Q88. Roughly how many times has JnJ sold its inventory in FY2022? Calculate inventory turnover ratio for FY2022; if conventional inventory management is not meaningful for the company then state that and explain why.

Result: OK · Latency: 8.4 s · Tool calls: 1

Expected answer

JnJ sold its inventory 2.7 times in FY2022.

Delivered answer

Roughly 2.5–2.7 times — J&J turned over its inventory about 2.5× on a year-end basis and 2.7× on an average-inventory basis in FY2022 (fiscal year ended January 1, 2023).

Inventory turnover = Cost of products sold ÷ inventory:

- Year-end basis: $\$31,089\text{M} \div \$12,483\text{M} = \mathbf{2.49\times}$
- Average basis: $\$31,089\text{M} \div [(12,483 + 10,387) \div 2 = \$11,435\text{M}] = \mathbf{2.72\times}$

(All figures from the FY2022 10-K: cost of products sold \$31,089M; inventories \$12,483M at Jan 1, 2023 and \$10,387M at Jan 2, 2022.)

Is conventional inventory management meaningful here? Yes. Unlike a bank or insurer, J&J is a product company that physically manufactures and sells goods (pharmaceuticals, medical devices, and consumer health products) and carries a real inventory balance, so inventory turnover is a valid and relevant measure of how efficiently it converts stock into sales. The ~2.5–2.7× figure is consistent with a diversified healthcare maker holding several months of inventory across its three segments.

If you'd like, I can break the turnover out by segment (Pharmaceutical, MedTech, Consumer Health) or compare it to the prior year.

Cited documents: JOHNSON_JOHNSON_2022_10K.pdf

Judge rationale: The agent's answer includes the essential figure (2.7 times) and provides additional correct details without contradiction.

Q89. Is growth in JnJ's adjusted EPS expected to accelerate in FY2023?

Result: **NOK** · Latency: 5.6 s · Tool calls: 1

Expected answer

No, rate of growth in adjusted EPS is expected to decelerate slightly from 3.6% in FY2022 to 3.5% in FY2023.

Delivered answer

Yes.

- FY2022 adjusted EPS growth (actual): **3.6%** (\$10.15 vs. \$9.80)
- FY2023 adjusted EPS growth (guidance published with FY2022 results, Jan 2023): **4.0%** midpoint (\$10.55, range 3.0%–5.0%)

4.0% > 3.6%, so the expected FY2023 growth rate is higher than the FY2022 rate — i.e., growth is expected to accelerate.

Note: JNJ later raised its FY2023 adjusted EPS guidance (to 5.0% in April, 6.0% in July, and 12.5% after the Kenvue split-off in August), but on the basis of the guidance issued with the FY2022 results, the answer is yes.

Cited documents: JOHNSON_JOHNSON_2022Q4_EARNINGS.pdf

Judge rationale: The agent's answer contradicts the expected answer by stating growth will accelerate (4.0% vs. 3.6%) while the reference states it will decelerate (3.5% vs. 3.6%).

Q90. How did JnJ's US sales growth compare to international sales growth in FY2022?

Result: **OK** · Latency: 8.4 s · Tool calls: 1

Expected answer

US sales increased 3.0% vs international sales decline of 0.6%.

Delivered answer

US sales grew while international sales declined in FY2022 (fiscal year ended January 1, 2023): US sales rose 3.0% to \$48.6B, while international sales fell 0.6% to \$46.4B — a 3.6-percentage-point gap in the US's favor.

- US: \$48,580M vs \$47,156M → +3.0%
- International: \$46,363M vs \$46,619M → -0.6%

The international decline was driven by a -9.7% currency headwind (stronger dollar) that more than offset 9.1% of operational growth; the US had no currency effect.

Cited documents: JOHNSON_JOHNSON_2022_10K.pdf

Judge rationale: The agent's answer includes the essential figures (US +3.0%, international -0.6%) and does not contradict the reference.

Q91. Which business segment of JnJ will be treated as a discontinued operation from August 30, 2023 onward?

Result: OK · Latency: 8.0 s · Tool calls: 1

Expected answer

The Consumer Health business segment will be treated as a discontinued operation from August 30, 2023 onward.

Delivered answer

The **Consumer Health** segment (separated as Kenvue Inc.) is the business segment of Johnson & Johnson treated as a discontinued operation from August 30, 2023 onward.

J&J completed the split-off exchange offer of Kenvue Inc. common stock on August 30, 2023, finalizing the separation of its Consumer Health business. The company had announced its intention to separate Consumer Health (as Kenvue) back in November 2021, and after the split-off, J&J became a two-segment company focused on Pharmaceutical and MedTech.

Cited documents: JOHNSON_JOHNSON_2023_8K_dated-2023-08-30.pdf,
JOHNSON_JOHNSON_2022_10K.pdf

Judge rationale: The agent's answer correctly identifies the Consumer Health business segment as the discontinued operation from August 30, 2023 onward, matching the essential information in the expected answer.

Q92. What is the amount of the gain accruing to JnJ as a result of the separation of its Consumer Health business segment, as of August 30, 2023?

Result: OK · Latency: 11.2 s · Tool calls: 3

Expected answer

JnJ will make a gain of approximately \$20 billion from the separation of its Consumer Health business segment.

Delivered answer

Approximately \$20 billion.

As of August 30, 2023, Johnson & Johnson announced that, upon completion of the exchange offer finalizing the separation of Kenvue Inc. (its former Consumer Health business), it would present the Consumer Health results as discontinued operations, including a gain of approximately **\$20 billion** in the third quarter of 2023.

The company also noted it secured \$13.2 billion in cash proceeds from the Kenvue debt offering and IPO and retained a 9.5% equity stake in Kenvue.

Cited documents: JOHNSON_JOHNSON_2023_8K_dated-2023-08-30.pdf

Judge rationale: The agent's answer includes the essential information (approximately \$20 billion gain) and does not contradict the expected answer.

Q93. What is the amount of the cash proceeds that JnJ realised from the separation of Kenvue (formerly Consumer Health business segment), as of August 30, 2023?

Result: OK · Latency: 6.2 s · Tool calls: 1

Expected answer

JnJ realised \$13.2 billion in cash proceeds from the separation of Kenvue.

Delivered answer

\$13.2 billion.

Johnson & Johnson reported that it generated **\$13.2 billion** in cash proceeds as a result of the Kenvue debt offering and initial public offering, following completion of the separation on August 30, 2023.

In addition to the cash proceeds, J&J retained a **9.5% equity stake** in Kenvue common stock, which it noted it could monetize in a tax-efficient manner in the following year (subject to a 90-day lockup). The company also reduced its outstanding share count by approximately 191 million shares through the exchange offer.

Cited documents: JOHNSON_JOHNSON_2023_8K_dated-2023-08-30.pdf

Judge rationale: The agent's answer includes the essential information (\$13.2 billion in cash proceeds) and does not contradict the expected answer.

Q94. Did JnJ's net earnings as a percent of sales increase in Q2 of FY2023 compared to Q2 of FY2022?

Result: OK · Latency: 7.5 s · Tool calls: 1

Expected answer

Yes, net earnings as a percent of sales increased from 20% in Q2 of FY2022 to 20.1% in Q2 of FY2023.

Delivered answer

Yes. JnJ's net earnings as a percent of sales rose slightly in Q2 of FY2023 versus Q2 of FY2022.

- Q2 FY2023: net earnings \$5,144M on sales of \$25,530M = **20.1%**
- Q2 FY2022: net earnings \$4,814M on sales of \$24,020M = **20.0%**

That's an increase of roughly **0.1 percentage points** (20.1% vs 20.0%), as reported in the company's Q2 2023 condensed consolidated statement of earnings.

(One note: the later 8-K restates these quarters on a continuing-operations basis after the Kenvue split-off, showing net earnings from continuing operations at 25.0% of sales in both Q2 2023 and Q2 2022. But on the total net-earnings basis as originally reported in the Q2 release, the margin ticked up by ~0.1 point.)

Cited documents: JOHNSON_JOHNSON_2023Q2_EARNINGS.pdf,
JOHNSON_JOHNSON_2023_8K_dated-2023-08-30.pdf

Judge rationale: The agent's answer correctly states that net earnings as a percent of sales increased from 20.0% to 20.1% between Q2 FY2022 and Q2 FY2023, matching the essential information in the expected answer.

Q95. Which of JPM's business segments had the lowest net revenue in 2021 Q1?

Result: OK · Latency: 18.0 s · Tool calls: 2

Expected answer

Corporate. Its net revenue was -\$473 million.

Delivered answer

Corporate had the lowest net revenue in Q1 2021, at **\$(473) million** (a net loss).

For comparison, the four major reportable business segments reported:

- Commercial Banking: \$2,393M
- Asset & Wealth Management: \$4,077M
- Consumer & Community Banking: \$12,517M
- Corporate & Investment Bank: \$14,605M

The Corporate segment, which the filing sets apart from the four main LOBs, posted negative net revenue of \$(473)M, making it the lowest of all segments.

Cited documents: JPMORGAN_2021Q1_10Q.pdf

Judge rationale: The agent's answer correctly identifies the Corporate segment as having the lowest net revenue in 2021 Q1 with -\$473 million, matching the essential information in the expected answer.

Q96. If JPM went bankrupted by the end by 2021 Q1 and liquidated all of its assets to pay its shareholders, how much could each shareholder get?

Result: OK · Latency: 9.1 s · Tool calls: 1

Expected answer

They could receive \$66.56 per share.

Delivered answer

About \$66.56 per common share — that's JPMorgan's reported tangible book value per share (TBVPS) at March 31, 2021, the figure that approximates what common shareholders would recover in a liquidation.

The build-up, from the Q1 2021 balance sheet:

- Common stockholders' equity: \$249,151M
- Less: Goodwill: (\$49,243M)
- Less: Other intangible assets: (\$875M)
- = Tangible common equity: ~\$199,033M
- ÷ Common shares outstanding: 3,027.1M ≈ **\$65.75**

The company's own published TBVPS is **\$66.56** (slightly higher because it nets out certain deferred tax liabilities tied to the intangibles), so that is the figure to use.

Two caveats worth flagging: this assumes all \$3.7 trillion of assets liquidate at book value and that the ~\$31.6 billion of preferred stock is paid in full first (preferred has a \$10,000/share liquidation preference ahead of common). In a real bankruptcy of a bank, asset values would almost certainly be marked down well below book, so the actual recovery would be lower.

Cited documents: JPMORGAN_2021Q1_10Q.pdf

Judge rationale: The agent's answer includes the essential figure (\$66.56 per share) and does not contradict the expected answer, despite providing additional context.

Q97. Are JPM's gross margins historically consistent (not fluctuating more than roughly 2% each year)? If gross margins are not a relevant metric for a company like this, then please state that and explain why. Added note by ATG for disambiguation: the question is referring to FY2022.

Result: OK · Latency: 8.4 s · Tool calls: 1

Expected answer

Since JPM is a financial institution, gross margin is not a relevant metric.

Delivered answer

No — gross margin is not a relevant metric for JPMorgan Chase, so there is no historical trend to assess.

JPM is a bank holding company, and its consolidated income statement has no cost of goods sold / cost of revenue line. Revenue is reported as **net interest income** (interest income less interest expense) plus **noninterest revenue** (fees, principal transactions, etc.), and interest expense and the provision for credit losses are part of its core operations rather than a "cost of sales." Because there is no cost of goods sold to subtract from revenue, a "gross profit" — and therefore a gross margin — simply does not exist for this kind of company.

The relevant profitability metrics for a bank like JPM are instead net interest margin, return on (tangible) common equity, and pre-provision profit — not gross margin.

Cited documents: JPMORGAN_2022_10K.pdf

Judge rationale: The agent correctly states that gross margin is not a relevant metric for JPMorgan Chase and explains why, matching the essential information in the expected answer.

Q98. In 2022 Q2, which of JPM's business segments had the highest net income?

Result: OK · Latency: 16.7 s · Tool calls: 2

Expected answer

Corporate & Investment Bank. Its net income was \$3725 million.

Delivered answer

Corporate & Investment Bank (CIB) had the highest net income in Q2 2022, at **\$3,725 million**.

The full segment breakdown for the three months ended June 30, 2022:

- CIB: \$3,725M
- CCB: \$3,100M
- AWM: \$1,004M
- CB: \$994M
- Corporate: \$(174)M

CIB led by roughly \$625M over the next-highest segment, CCB.

Cited documents: JPMORGAN_2022Q2_10Q.pdf

Judge rationale: The agent's answer correctly identifies Corporate & Investment Bank as the segment with the highest net income and matches the expected figure of \$3,725 million.

Q99. Looking at VaR, did the risk that JPM faced in the second fiscal quarter of 2023 decrease compared to the same period in the prior year?

Result: OK · Latency: 6.7 s · Tool calls: 1

Expected answer

Yes. It decreased.

Delivered answer

Yes. JPMorgan's average total Risk Management VaR (95% confidence, one-day) fell in Q2 2023 versus Q2 2022.

- Q2 2023 (three months ended June 30, 2023): **\$47M** average (min \$36M, max \$56M)
- Q2 2022 (three months ended June 30, 2022): **\$54M** average (min \$41M, max \$71M)
- Change: **-\$7M**, a ~13% decrease year over year

Management attributes the decline "predominantly to risk reductions impacting Credit Portfolio VaR as well as fixed income" (Credit Portfolio VaR average dropped from \$17M to \$14M; fixed income from \$60M to \$57M).

So the market risk JPM faced, as measured by VaR, did decrease in the second fiscal quarter of 2023 compared with the same period the prior year.

Cited documents: JPMORGAN_2023Q2_10Q.pdf

Judge rationale: The agent's answer confirms the risk decreased, matching the essential information in the expected answer.

Q100. What is Kraft Heinz's FY2019 inventory turnover ratio? Inventory turnover ratio is defined as: (FY2019 COGS) / (average inventory between FY2018 and FY2019). Round your answer to two decimal places. Please base your judgments on the information provided primarily in the balance sheet and the P&L statement.

Result: OK · Latency: 16.3 s · Tool calls: 2

Expected answer

6.25

Delivered answer

6.25

- FY2019 COGS (cost of products sold): \$16,830 million
- FY2019 inventories (Dec 28, 2019): \$2,721 million; FY2018 inventories (Dec 29, 2018): \$2,667 million
- Average inventory = $(2,721 + 2,667) / 2 = \$2,694$ million
- Inventory turnover = $16,830 / 2,694 = \mathbf{6.25}$

(Periods are Kraft Heinz's fiscal years ending December 28, 2019 and December 29, 2018, per the FY2019 10-K.)

Cited documents: KRAFTHEINZ_2019_10K.pdf

Judge rationale: The agent's answer matches the expected answer exactly (6.25) and includes the correct calculation steps without contradiction.

Q101. We need to calculate a reasonable approximation (or exact number if possible) of a financial metric. Basing your judgment by information plainly provided in the balance sheet and the P&L statement, what is Lockheed Martin's FY2020 asset turnover ratio? Asset turnover ratio is defined as: FY2020 revenue / (average total assets between FY2019 and FY2020). Round your answer to two decimal places.

Result: OK · Latency: 17.4 s · Tool calls: 2

Expected answer

1.33

Delivered answer

1.33

Lockheed Martin's FY2020 asset turnover ratio is **1.33**.

- FY2020 net sales (revenue): \$65,398 million
- FY2020 total assets: \$50,710 million
- FY2019 total assets: \$47,528 million
- Average total assets = $(50,710 + 47,528) / 2 = \$49,119$ million
- Asset turnover = $65,398 / 49,119 = \mathbf{1.33}$

Cited documents: LOCKHEEDMARTIN_2019_10K.pdf, LOCKHEEDMARTIN_2020_10K.pdf

Judge rationale: The agent's answer matches the expected answer exactly (1.33) and includes the correct calculation steps and figures.

Q102. What is Lockheed Martin's FY2021 net working capital? Define net working capital as total current assets less total current liabilities. Answer in USD millions. Respond to the question by assuming the perspective of an investment analyst who can only use the details shown within the balance sheet.

Result: OK · Latency: 8.6 s · Tool calls: 1

Expected answer

\$5818.00

Delivered answer

\$5,818 million.

From Lockheed Martin's consolidated balance sheet at December 31, 2021 (FY2021, fiscal year ending December 31):

- Total current assets: \$19,815M
- Total current liabilities: \$13,997M
- Net working capital = 19,815 – 13,997 = **\$5,818M**

For context, the largest current-asset component was contract assets (\$10,579M, ~53% of total current assets), and the largest current-liability component was contract liabilities (\$8,107M, ~58% of total current liabilities).

Cited documents: LOCKHEEDMARTIN_2021_10K.pdf

Judge rationale: The agent's answer matches the expected net working capital figure of \$5,818 million and provides the correct calculation.

Q103. What is Lockheed Martin's 2 year total revenue CAGR from FY2020 to FY2022 (in units of percents and round to one decimal place)? Provide a response to the question by primarily using the statement of income.

Result: OK · Latency: 7.9 s · Tool calls: 1

Expected answer

0.4%

Delivered answer

0.4%

Using total net sales from Lockheed Martin's Consolidated Statements of Earnings (FY2020 = \$65,398M; FY2022 = \$65,984M):

$$\text{CAGR} = (65,984 / 65,398)^{(1/2)} - 1 = (1.00896)^{(0.5)} - 1 = 1.00447 - 1 = 0.00447 \approx \mathbf{0.4\%}$$

So over the two-year span FY2020→FY2022, total revenue grew at a compound annual rate of about 0.4% (note revenue actually dipped in FY2021 to \$67,044M before settling at \$65,984M in FY2022, so the net two-year gain is small).

Cited documents: LOCKHEEDMARTIN_2022_10K.pdf

Judge rationale: The agent's answer matches the expected answer exactly (0.4%) and includes correct calculations and context.

Q104. Basing your judgments off of the balance sheet, what is the year end FY2018 amount of accounts payable for MGM Resorts? Answer in USD millions.

Result: NOK · Latency: 10.3 s · Tool calls: 1

Expected answer

\$303.00

Delivered answer

\$302.6 million.

MGM Resorts' consolidated balance sheet as of December 31, 2018 reports accounts payable of \$302,578 thousand, i.e. \$302.6 million (vs. \$255,028 thousand / \$255.0 million at year-end 2017).

Cited documents: MGMRESORTS_2018_10K.pdf

Judge rationale: The agent's answer (\$302.6 million) does not match the expected answer (\$303.00 million) as the figures differ beyond acceptable rounding.

Q105. What is the FY2018 - FY2020 3 year average of capex as a % of revenue for MGM Resorts? Answer in units of percents and round to one decimal place. Please utilize information provided primarily within the statement of cash flows and the statement of income.

Result: OK · Latency: 12.6 s · Tool calls: 3

Expected answer

7.9%

Delivered answer

7.9%

Capex (cash flow statement, "Capital expenditures, net of construction payable") ÷ net revenues (income statement), per year:

Year	Capex (\$M)	Net revenues (\$M)	Capex % of revenue
FY2018	1,486.8	11,763.1	12.64%
FY2019	739.0	12,900.0	5.73%
FY2020	270.6	5,162.1	5.24%

3-year average = (12.64% + 5.73% + 5.24%) / 3 = 7.87% ≈ **7.9%**

Sources: capex from the consolidated statements of cash flows and net revenues from the consolidated statements of operations in the FY2020 10-K, with FY2018/FY2019 figures cross-checked against the FY2019 10-K.

Cited documents: MGMRESORTS_2020_10K.pdf, MGMRESORTS_2019_10K.pdf

Judge rationale: The agent's answer matches the expected answer exactly (7.9%) and includes correct supporting calculations.

Q106. Has MGM Resorts paid dividends to common shareholders in FY2022?

Result: **OK** · Latency: 8.2 s · Tool calls: 1

Expected answer

Yes. MGM maintained 0.01\$ per share annual dividend through out FY 2022.

Delivered answer

Yes. MGM Resorts paid dividends to common shareholders in FY2022 (fiscal year ended December 31, 2022).

- The cash flow statement reports **dividends paid to common shareholders of \$4,048 thousand (~\$4.0 million)** for 2022, matching the statement of stockholders' equity line "dividends declared and paid to common shareholders (\$0.01 per share)."
- Per the MD&A, the company paid **\$0.0025 per share in each of March, June, September, and December 2022**, totaling \$4 million for the year (the reduced \$0.01 annualized rate maintained since the 2020 pandemic cut).

Note that the board subsequently **suspended the dividend in February 2023**, shifting to share repurchases as its preferred way of returning value to shareholders.

Sources: (MGM 2022 10-K), (MGM Q4 2022 earnings release).

Cited documents: MGMRESORTS_2022_10K.pdf, MGMRESORTS_2022Q4_EARNINGS.pdf

Judge rationale: The agent's answer confirms dividends were paid in FY2022 and includes the essential detail of \$0.01 per share, matching the expected answer.

Q107. Which region had the Highest EBITDAR Contribution for MGM during FY2022?

Result: **NOK** · Latency: 8.1 s · Tool calls: 1

Expected answer

Las Vegas resorts contributed ~90% of company level EBITDAR during FY2022.

Delivered answer

Las Vegas Strip Resorts had the highest Adjusted EBITDAR contribution for MGM in FY2022.

From the FY2022 segment data (year ended December 31, 2022):

- Las Vegas Strip Resorts: **\$3,142.3 million**
- Regional Operations: \$1,294.6 million
- MGM China: (\$203.1 million) — a loss

Las Vegas Strip Resorts contributed roughly 74% of total reportable segment Adjusted EBITDAR (\$4,233.8 million), driven by the inclusion of The Cosmopolitan and a full year of Aria, plus strong recovery in business volume and travel activity.

Cited documents: MGMRESORTS_2022_10K.pdf

Judge rationale: The agent's answer states Las Vegas Strip Resorts contributed roughly 74% of EBITDAR, which contradicts the expected answer of ~90%.

Q108. What was MGM's interest coverage ratio using FY2022 Adjusted EBIT as the numerator and annual Interest Expense as the denominator?

Result: **NOK** · Latency: 25.5 s · Tool calls: 3

Expected answer

As adjusted EBIT is negative, coverage ratio is zero

Delivered answer

≈ **5.9x** using FY2022 Adjusted EBITDAR (the closest non-GAAP "Adjusted EBIT"-type measure MGM actually reports) as the numerator and income-statement interest expense as the denominator.

Point to flag: MGM does not publish a line called "Adjusted EBIT" in its FY2022 filings — its non-GAAP measure is Adjusted EBITDAR, so I've used that as the numerator.

- Numerator (FY2022 Adjusted EBITDAR): \$3,497.3M
- Denominator (FY2022 interest expense, net of amounts capitalized): \$595.0M
- $3,497,254 \div 594,954 = \mathbf{5.88x}$

If you instead mean interest expense *before* capitalization (\$595.7M), it's 5.87x — materially the same.

Cited documents: MGMRESORTS_2022_10K.pdf

Judge rationale: The agent's answer contradicts the expected answer by providing a positive ratio (≈5.9x) instead of stating that the coverage ratio is zero due to negative Adjusted EBIT.

Q109. Which region had the worst topline performance for MGM during FY2022?

Result: **OK** · Latency: 8.3 s · Tool calls: 1

Expected answer

MGM China experienced the worst topline performance amongst the other regions presented. Its revenue declined 44% in FY2022 whereas the other regions presented increased their revenues.

Delivered answer

MGM China had the worst topline performance in FY2022, with net revenues declining 44% year-over-year.

- MGM China: \$674M in 2022 vs. \$1,211M in 2021 → **-44%** (driven by COVID-19-related property closures and travel/entry restrictions in Macau)
- Regional Operations: \$3,816M vs. \$3,392M → +12%
- Las Vegas Strip Resorts: \$8,398M vs. \$4,737M → +77%

MGM China was the only segment whose revenue fell, making it clearly the weakest performer on the topline.

Cited documents: MGMRESORTS_2022_10K.pdf

Judge rationale: The agent's answer correctly identifies MGM China as the worst-performing region with a 44% revenue decline, matching the essential information in the expected answer.

Q110. Which type of debt received the largest investment among the short term investments for MGM in H1 FY2023?

Result: OK · Latency: 8.9 s · Tool calls: 1

Expected answer

the biggest short term investment is in corporate bonds (almost 82% of the total investment)

Delivered answer

Corporate bonds received the largest investment among MGM's short-term debt investments in H1 FY2023 (as of June 30, 2023).

Per the debt investments table in the Q2 2023 10-Q, the short-term investments broke down as follows:

Debt type	Fair value (June 30, 2023)
Corporate bonds	\$416.4M
U.S. government securities	\$57.7M
U.S. agency securities	\$29.0M
Commercial paper & certificates of deposit	\$4.6M
Total short-term investments	\$507.7M

Corporate bonds accounted for roughly 82% of the total short-term investment portfolio, well ahead of the next largest category (U.S. government securities at ~11%).

Cited documents: MGMRESORTS_2023Q2_10Q.pdf

Judge rationale: The agent's answer correctly identifies corporate bonds as the largest short-term investment and confirms the ~82% figure, matching the essential information in the expected answer.

Q111. What is the FY2016 COGS for Microsoft? Please state answer in USD millions. Provide a response to the question by primarily using the statement of income.

Result: OK · Latency: 9.4 s · Tool calls: 2

Expected answer

\$32780.00

Delivered answer

\$32,780 million (USD)

Microsoft's FY2016 (fiscal year ended June 30, 2016) consolidated statement of income reports total cost of revenue of \$32,780 million, split into product cost of revenue of \$17,880 million and service and other cost of revenue of \$14,900 million.

Check: total revenue of \$85,320 million minus cost of revenue of \$32,780 million = gross margin of \$52,540 million, matching the reported gross margin.

Cited documents: MICROSOFT_2016_10K.pdf

Judge rationale: The agent's answer matches the expected COGS figure of \$32,780 million in USD millions with no contradictions.

Q112. Has Microsoft increased its debt on balance sheet between FY2023 and the FY2022 period?

Result: OK · Latency: 15.0 s · Tool calls: 2

Expected answer

No. Microsoft decreased its debt by \$2.5bn in FY 2023 vs FY 2022.

Delivered answer

No — Microsoft's total debt on the balance sheet **decreased** from FY2022 to FY2023 (fiscal years ending June 30).

- **FY2022 (June 30, 2022):** total debt = \$49,781M (current portion \$2,749M + long-term debt \$47,032M)
- **FY2023 (June 30, 2023):** total debt = \$47,237M (current portion \$5,247M + long-term debt \$41,990M)
- **Change:** -\$2,544M, a decrease of about 5.1%

So the balance-sheet debt fell by roughly \$2.5 billion year over year, even though the current portion rose (as maturities rolled into the near term) — the long-term portion dropped more than that.

Cited documents: MICROSOFT_2023_10K.pdf

Judge rationale: The agent's answer correctly states that Microsoft's debt decreased and provides the exact figure of approximately \$2.5 billion, matching the essential information in the expected answer.

Q113. We want to calculate a financial metric. Please help us compute it by basing your answers off of the statement of income and the statement of cash flows. Here's the question: what is the FY2015 unadjusted EBITDA % margin for Netflix? Calculate unadjusted EBITDA using unadjusted operating income and D&A (from cash flow statement).

Result: OK · Latency: 8.3 s · Tool calls: 1

Expected answer

5.4%

Delivered answer

5.43%

Here's the calculation from Netflix's FY2015 10-K (year ended December 31, 2015):

- Operating income (unadjusted): \$305,826 thousand
- D&A from cash flow statement ("Depreciation and amortization of property, equipment and intangibles"): \$62,283 thousand
- **Unadjusted EBITDA** = \$305,826 + \$62,283 = **\$368,109 thousand**
- Revenue: \$6,779,511 thousand
- **EBITDA % margin** = \$368,109 / \$6,779,511 = **5.43%**

Note: I used only the property/equipment/intangibles D&A line (\$62,283) from the cash flow statement, per your instruction — the separate content-amortization lines (streaming \$3,405,382 and DVD \$79,380) are operating costs, not D&A, so they're excluded.

Sources:

Cited documents: NETFLIX_2015_10K.pdf

Judge rationale: The agent's answer (5.43%) matches the expected answer (5.4%) when rounded to one decimal place, and the calculation is consistent with the reference.

Q114. What is Netflix's year end FY2017 total current liabilities (in USD millions)? Base your judgments on the information provided primarily in the balance sheet.

Result: OK · Latency: 8.4 s · Tool calls: 1

Expected answer

\$5466.00

Delivered answer

\$5,466 million (i.e., \$5,466.312 million, as reported in thousands on the balance sheet).

From Netflix's FY2017 consolidated balance sheet (as of December 31, 2017), total current liabilities sum to **\$5,466,312 thousand = \$5,466.312 million**:

- Current content liabilities: \$4,173,041
- Accounts payable: \$359,555
- Accrued expenses: \$315,094
- Deferred revenue: \$618,622
- **Total current liabilities: \$5,466,312 (thousands) → \$5,466 million**

Note: Netflix's fiscal year ends December 31, so FY2017 = the year ended December 31, 2017.

Cited documents: NETFLIX_2017_10K.pdf

Judge rationale: The agent's answer matches the expected answer's essential figure (\$5,466 million) and provides additional correct context without contradiction.

Q115. We need to calculate a reasonable approximation (or exact number if possible) of a financial metric. Basing your judgment by information plainly provided in the statement of income, what is Nike's three year average of cost of goods sold as a % of revenue from FY2016 to FY2018? Answer in units of percents and round to one decimal place.

Result: OK · Latency: 6.9 s · Tool calls: 1

Expected answer

55.1%

Delivered answer

55.1%

Nike's cost of sales as a % of revenue, from the Consolidated Statements of Income:

- FY2016: 17,405 / 32,376 = 53.76%
- FY2017: 19,038 / 34,350 = 55.42%
- FY2018: 20,441 / 36,397 = 56.16%

Three-year average = (53.76% + 55.42% + 56.16%) / 3 = 55.11% **≈ 55.1%**

(These figures are as originally reported in Nike's FY2018 10-K; note the cost-of-sales line is labeled "Cost of sales" rather than "cost of goods sold.")

Cited documents: NIKE_2018_10K.pdf

Judge rationale: The agent's answer matches the expected answer exactly, including the correct rounding to one decimal place.

Q116. According to the details clearly outlined within the balance sheet, how much total current assets did Nike have at the end of FY2019? Answer in USD millions.

Result: OK · Latency: 7.4 s · Tool calls: 1

Expected answer

\$16525.00

Delivered answer

\$16,525 million.

Per Nike's consolidated balance sheet as of May 31, 2019 (FY2019), total current assets were \$16,525 million, made up of:

- Cash and equivalents: \$4,466M
- Short-term investments: \$197M
- Accounts receivable, net: \$4,272M
- Inventories: \$5,622M
- Prepaid expenses and other current assets: \$1,968M

Sum: $4,466 + 197 + 4,272 + 5,622 + 1,968 = \mathbf{\$16,525M}$

Cited documents: NIKE_2019_10K.pdf

Judge rationale: The agent's answer matches the expected total current assets figure of \$16,525 million in USD millions.

Q117. When primarily referencing the income statement and the statement of financial position, what is the FY2021 inventory turnover ratio for Nike? Inventory turnover ratio is defined as: (FY2021 COGS) / (average inventory between FY2020 and FY2021). Round your answer to two decimal places.

Result: OK · Latency: 7.6 s · Tool calls: 1

Expected answer

3.46

Delivered answer

3.46

- FY2021 COGS = Revenues – Gross profit = \$44,538M – \$19,962M = **\$24,576M**
- Inventory: FY2021 = \$6,854M, FY2020 = \$7,367M
- Average inventory = (\$6,854M + \$7,367M) / 2 = \$7,110.5M
- Inventory turnover = \$24,576M / \$7,110.5M = **3.46**

(Nike's fiscal year ends May 31, so FY2021 is the year ended May 31, 2021, and FY2020 the year ended May 31, 2020.)

Cited documents: NIKE_2021_10K.pdf

Judge rationale: The agent's answer matches the expected answer exactly (3.46) and includes the correct calculation steps without contradictions.

Q118. Among operations, investing, and financing activities, which brought in the most (or lost the least) cash flow for Nike in FY2023?

Result: OK · Latency: 8.5 s · Tool calls: 1

Expected answer

Among the three, cash flow from operations was the highest for Nike in FY2023.

Delivered answer

Operations brought in the most cash for Nike in FY2023 (fiscal year ended May 31, 2023), at **\$5,841 million**.

Here's the comparison from the Consolidated Statements of Cash Flows:

- **Operations:** +\$5,841 million
- **Investing:** +\$564 million
- **Financing:** –\$7,447 million

Operations was the clear winner, generating roughly \$5.3 billion more than investing activities and more than \$13 billion more than financing activities (which was a net outflow, driven mainly by \$5,480 million in share repurchases and \$2,012 million in dividends).

Cited documents: NIKE_2023_10K.pdf

Judge rationale: The agent's answer correctly identifies operations as the activity with the highest cash flow for Nike in FY2023, matching the essential information in the expected answer.

Q119. Does Paypal have positive working capital based on FY2022 data? If working capital is not a useful or relevant metric for this company, then please state that and explain why.

Result: OK · Latency: 11.9 s · Tool calls: 1

Expected answer

Yes. Paypal has a positive working capital of \$ 1.6Bn as of FY2022 end.

Delivered answer

Yes — PayPal has positive working capital as of December 31, 2022: +\$12,416 million.

- Total current assets: **\$57,517 million**
- Total current liabilities: **\$45,101 million**
- Working capital = $57,517 - 45,101 = \mathbf{+\$12,416 \text{ million}}$
- Operating working capital (excluding cash & cash equivalents and short-term investments from current assets; no short-term debt in current liabilities to strip out): $(57,517 - 7,776 - 3,092) - 45,101 = \mathbf{+ \$1,548 \text{ million}}$

Working capital is a relevant and customary metric for PayPal: it is a payments platform/financial-technology company, not a chartered bank holding company, insurer, or pure lender in the way the metric is set aside, so the balance-sheet current-asset/current-liability comparison applies.

One caveat worth flagging: a large, roughly offsetting pair sits inside both sides of the balance sheet — "funds receivable and customer accounts" (\$36,357M of current assets) versus "funds payable and amounts due to customers" (\$40,107M of current liabilities) — reflecting customer balances PayPal holds. So the positive figure is real but is affected more by those customer-fund balances than by the size of what a typical merchant would think of as core operating working capital.

Source: PayPal Holdings, Inc. Consolidated Balance Sheets, FY2022 10-K (period ended December 31, 2022).

Cited documents: PAYPAL_2022_10K.pdf

Judge rationale: The agent's answer confirms positive working capital for PayPal in FY2022 and provides the correct figure (\$12,416 million, equivalent to \$12.4Bn, which rounds to \$1.6Bn when considering the expected answer's precision).

Q120. What is the FY2021 capital expenditure amount (in USD billions) for PepsiCo? Respond to the question by assuming the perspective of an investment analyst who can only use the details shown within the statement of cash flows.

Result: OK · Latency: 8.4 s · Tool calls: 1

Expected answer

\$4.60

Delivered answer

\$4.6 billion (exactly \$4,625 million).

Reading straight off PepsiCo's Consolidated Statement of Cash Flows for the fiscal year ended December 25, 2021, the investing-activities line "Capital spending" is \$(4,625) million — i.e. \$4.6B in cash outflow for property, plant and equipment.

(For context, that's up from \$4,240 million in FY2020, and the company's own free-cash-flow reconciliation uses this same \$4,625 million capex figure.)

Cited documents: PEPSICO_2021_10K.pdf

Judge rationale: The agent's answer matches the expected figure of \$4.60 billion when rounded to the same precision.

Q121. What are the geographies that Pepsico primarily operates in as of FY2022?

Result: OK · Latency: 8.6 s · Tool calls: 1

Expected answer

As of FY2022, Pepsico primarily operates in the following geographies: North America, Latin America, Europe, Africa, Middle East, South Asia, Asia Pacific, Australia, New Zealand and China.

Delivered answer

As of FY2022 (fiscal year ended December 31, 2022), PepsiCo operates in more than 200 countries and territories, organized into seven reportable segments (divisions) that define its primary operating geographies:

- **North America** (United States and Canada) — split into three divisions: Frito-Lay North America (FLNA), Quaker Foods North America (QFNA), and PepsiCo Beverages North America (PBNA)
- **Latin America** (LatAm)
- **Europe**
- **Africa, Middle East and South Asia** (AMESA)
- **Asia Pacific, Australia and New Zealand and China Region** (APAC)

So the primary geographies are North America (US/Canada), Latin America, Europe, Africa/Middle East/South Asia, and Asia Pacific/Australia/New Zealand/China.

Note: this reflects the segment structure in the 2022 Form 10-K; in earlier years (e.g., 2015–2017) PepsiCo used a six-segment structure that grouped Europe and Sub-Saharan Africa together (ESSA) and Asia, Middle East and North Africa (AMENA) into single divisions.

Cited documents: PEPSICO_2022_10K.pdf

Judge rationale: The agent's answer includes all essential geographies listed in the expected answer, with no contradictions or factual errors.

Q122. Has Pepsico reported any materially important ongoing legal battles from FY2022 and FY2021?

Result: OK · Latency: 37.4 s · Tool calls: 5

Expected answer

No, Pepsico is not involved in material legal battles.

Delivered answer

No. PepsiCo discloses no materially important (i.e., potentially material) legal battles in its 10-Ks for both FY2022 and FY2021.

- In the **FY2022** 10-K (year ended December 31, 2022), Item 3. Legal Proceedings states only that PepsiCo and its subsidiaries "are party to a variety of litigation, claims, legal or regulatory proceedings, inquiries and investigations" and, as the company's own verdict, "management believes that the final outcome of the foregoing will not have a material adverse effect on our financial condition, results of operations or cash flows."
- In the **FY2021** 10-K (year ended December 25, 2021), Item 3. Legal Proceedings is essentially identical boilerplate, again concluding that the outcomes "will not have a material adverse effect on our financial condition, results of operations or cash flows."

Unlike the 2018 10-K (which specifically named the PepsiCo Alimentos Z.F./PAZ wastewater permitting case in Colombia), neither the 2021 nor the 2022 filing names any individual lawsuit or proceeding — only the generic, non-material disclosure that routine litigation is pending and is not expected to be material.

Cited documents: PEPSICO_2021_10K.pdf, PEPSICO_2022_10K.pdf

Judge rationale: The agent's answer correctly states that Pepsico has no materially important ongoing legal battles for FY2022 and FY2021, matching the essential information in the expected answer.

Q123. What is the quantity of restructuring costs directly outlined in PepsiCo's income statements for FY2022? If restructuring costs are not explicitly outlined then state 0.

Result: **NOK** · Latency: 8.0 s · Tool calls: 1

Expected answer

PepsiCo's restructuring costs in FY2022 amounted to \$411 million .

Delivered answer

0.

PepsiCo's Consolidated Statement of Income for FY2022 (year ended December 31, 2022) does not contain a separate "restructuring costs" line item. The income statement lines are: Net Revenue, Cost of sales, Gross profit, Selling, general and administrative expenses, Gain associated with the Juice Transaction, Impairment of intangible assets, Operating Profit, Other pension and retiree medical benefits (expense)/income, Net interest expense and other, Income before income taxes, Provision for income taxes, and Net income — with no explicit restructuring line.

The \$411 million in restructuring and impairment charges for FY2022 (from the 2019 Productivity Plan) is embedded within Cost of sales (\$33M), SG&A (\$347M), and Other pension/retiree medical benefits (\$31M), as disclosed in Note 3 of the 10-K, but is not broken out as its own line on the face of the income statement.

Cited documents: PEPSICO_2022Q4_EARNINGS.pdf, PEPSICO_2022_10K.pdf

Judge rationale: The agent answer states 0, but the expected answer specifies \$411 million as the restructuring costs for FY2022.

Q124. What is the FY2022 unadjusted EBITDA less capex for PepsiCo? Define unadjusted EBITDA as unadjusted operating income + depreciation and amortization [from cash flow statement]. Answer in USD millions. Respond to the question by assuming the perspective of an investment analyst who can only use the details shown within the statement of cash flows and the income statement.

Result: OK · Latency: 9.0 s · Tool calls: 1

Expected answer

\$9068.00

Delivered answer

\$9,068 million

Here's the build, using only the income statement and cash flow statement from PepsiCo's FY2022 filing (year ended December 31, 2022):

- Unadjusted operating income (GAAP): **\$11,512M**
- Depreciation & amortization (cash flow statement): **\$2,763M**
- → Unadjusted EBITDA = 11,512 + 2,763 = **\$14,275M**
- Less: capital spending (capex): **\$5,207M**
- → **EBITDA less capex = 14,275 – 5,207 = \$9,068M**

Note that FY2022 was a 53-week year, and the GAAP operating profit figure is depressed by the \$3,321M gain on the Juice Transaction and \$3,166M of intangible-asset impairments — but since you asked for the *unadjusted* (as-reported) figure, those stay in the number.

Sources: PepsiCo Q4 2022 earnings release (Exhibit 99.1) and FY2022 Form 10-K.

Cited documents: PEPSICO_2022Q4_EARNINGS.pdf, PEPSICO_2022_10K.pdf

Judge rationale: The agent's answer matches the expected answer exactly in value and units, with no contradictions or omissions.

Q125. What is the FY2022 unadjusted EBITDA % margin for PepsiCo? Calculate unadjusted EBITDA using unadjusted operating income and D&A (from cash flow statement). Give a response to the question by relying on the details shown in the statement of cash flows and the P&L statement.

Result: OK · Latency: 9.1 s · Tool calls: 1

Expected answer

16.5%

Delivered answer

16.52%

Unadjusted EBITDA = unadjusted operating income + D&A (from the cash flow statement):

- Operating Profit (P&L, FY2022): \$11,512M
- Depreciation and amortization (cash flow statement, FY2022): \$2,763M
- EBITDA = 11,512 + 2,763 = **\$14,275M**

EBITDA margin = EBITDA / Net Revenue = 14,275 / 86,392 = **16.52%** (Net Revenue FY2022 = \$86,392M).

Note: this uses the "Depreciation and amortization" line (\$2,763M) as reported in the statement of cash flows; the separate "Operating lease right-of-use asset amortization" (\$517M) is not included, per the standard D&A definition.

Cited documents: PEPSICO_2022_10K.pdf

Judge rationale: The agent's answer (16.52%) matches the expected answer (16.5%) when rounded to the same precision.

Q126. At the PepsiCo AGM held on May 3, 2023, what was the outcome of the shareholder vote on the shareholder proposal for a congruency report by PepsiCo on net-zero emissions policies?

Result: OK · Latency: 8.7 s · Tool calls: 1

Expected answer

The shareholder proposal for a congruency report by PepsiCo on net-zero emissions policies was defeated.

Delivered answer

Defeated. At PepsiCo's 2023 Annual Meeting of Shareholders (held May 3, 2023), the shareholder proposal calling for a congruency report on the company's net-zero emissions policies was voted down.

The final vote tally was:

- **For:** 19,718,780
- **Against:** 977,228,788
- **Abstain:** 23,200,321
- **Broker Non-Votes:** 172,969,325

With roughly 95.8% of the votes cast (for + against + abstain) coming in against the proposal, shareholders overwhelmingly rejected it.

For context, this was one of four shareholder proposals on the ballot that year—all four (independent Board Chair, a global transparency report, a report on reproductive healthcare legislation, and this net-zero congruency report) were defeated.

Cited documents: PEPSICO_2023_8K_dated-2023-05-05.pdf

Judge rationale: The agent's answer correctly states that the shareholder proposal was defeated, matching the essential information in the expected answer.

Q127. By how much did PepsiCo increase its unsecured five year revolving credit agreement on May 26, 2023?

Result: OK · Latency: 31.1 s · Tool calls: 4

Expected answer

\$400,000,000 increase.

Delivered answer

\$400 million — PepsiCo's new five-year unsecured revolving credit agreement dated May 26, 2023 carries commitments of \$4.2 billion, versus the \$3.8 billion under the prior five-year agreement entered in 2022.

- Prior (2022) Five-Year Credit Agreement: \$3.8 billion
- New (May 26, 2023) Five-Year Credit Agreement: \$4.2 billion
- Increase: \$4.2B – \$3.8B = **\$400 million**

Both agreements also allow commitments to be increased up to \$4.5 billion (2022) / \$4.95 billion (2023).

Cited documents: PEPSICO_2023_8K_dated-2023-05-30.pdf, PEPSICO_2022_10K.pdf

Judge rationale: The agent's answer correctly states the \$400 million increase, matching the essential information in the expected answer.

Q128. As of May 26, 2023, what is the total amount PepsiCo may borrow under its unsecured revolving credit agreements?

Result: OK · Latency: 26.8 s · Tool calls: 3

Expected answer

Total amount PepsiCo may borrow under unsecured revolving credit agreements = \$8,400,000,000.

Delivered answer

\$8.4 billion (i.e., \$8,400,000,000).

As of May 26, 2023, PepsiCo entered into two unsecured revolving credit agreements, each providing up to \$4.2 billion in borrowing capacity:

- **Five-Year Credit Agreement:** \$4,200,000,000
- **364-Day Credit Agreement:** \$4,200,000,000

Total: \$4.2B + \$4.2B = **\$8.4 billion**

Cited documents: PEPSICO_2023_8K_dated-2023-05-30.pdf

Judge rationale: The agent's answer matches the essential information in the expected answer, with no contradictions or factual errors.

Q129. As of FY2023Q1, why did PepsiCo raise full year guidance for FY2023?

Result: **OK** · Latency: 7.4 s · Tool calls: 1

Expected answer

PepsiCo experienced a strong start to FY2023.

Delivered answer

PepsiCo raised its full-year FY2023 guidance because of a strong start to the year — its categories and geographies remained resilient through Q1, delivering solid first-quarter performance and business momentum.

Specifically, in its Q1 2023 release (April 25, 2023), CEO Ramon Laguarta cited the "strong start to the year" and said the company now expected:

- **Full-year organic revenue growth of 8%** (raised from 6%)
- **Core constant currency EPS growth of 9%** (raised from 8%)

This implied 2023 core EPS of \$7.27 (up from the prior \$7.20), a 7% increase versus 2022 core EPS of \$6.79. The core effective tax rate (20%) and total cash returns to shareholders (~\$7.7 billion) were left unchanged.

Cited documents: PEPSICO_2023Q1_EARNINGS.pdf

Judge rationale: The agent's answer includes the essential information that PepsiCo raised guidance due to a strong start to FY2023, matching the reference.

Q130. As of FY2023Q1, by how many percentage points did PepsiCo raise full year guidance in respect of core constant currency EPS growth?

Result: **OK** · Latency: 6.6 s · Tool calls: 1

Expected answer

PepsiCo raised full year guidance in respect of core constant currency EPS growth by 1 percentage point.

Delivered answer

1 percentage point.

PepsiCo raised its FY2023 core constant currency EPS growth guidance from 8% to 9% when it reported Q1 2023 results on April 25, 2023 — a 1-percentage-point increase.

For context, the original 8% target had been set alongside the Q4 2022 results in February 2023, and the company continued to raise it through the year (to 12% by Q2).

Cited documents: PEPSICO_2023Q1_EARNINGS.pdf, PEPSICO_2022Q4_EARNINGS.pdf

Judge rationale: The agent's answer correctly states the 1 percentage point increase, matching the essential information in the expected answer.

Q131. Did Pfizer grow its PPNE between FY20 and FY21?

Result: **OK** · Latency: 8.6 s · Tool calls: 1

Expected answer

Yes, change in PPNE was positive year over year

Delivered answer

Yes. Pfizer's net property, plant and equipment (PP&E) grew from \$13,745 million at December 31, 2020 to \$14,882 million at December 31, 2021, an increase of \$1,137 million (about 8.3%).

Note: the 2020 figure of \$13,745M is as restated in the 2021 10-K (the 2020 10-K had originally reported \$13,900M before reclassifying certain assets to discontinued operations).

Cited documents: PFIZER_2021_10K.pdf, PFIZER_2020_10K.pdf

Judge rationale: The agent's answer confirms positive growth in PPNE between FY20 and FY21, matching the essential information in the expected answer.

Q132. Were there any potential events that are not in Pfizer's standard business operations that substantially increased net income in 2019?

Result: **OK** · Latency: 8.9 s · Tool calls: 1

Expected answer

Yes, the gain on completion of Consumer Healthcare JV Transaction

Delivered answer

Yes. The dominant unusual item in 2019 was a **pre-tax gain of \$8.1 billion** on the completion of the GSK Consumer Healthcare joint venture transaction, which Pfizer contributed its Consumer Healthcare business to. This single item was excluded from Pfizer's own Non-GAAP adjusted income as a "certain significant item."

For context, Pfizer's GAAP net income attributable to Pfizer Inc. was **\$16,273 million** in 2019, versus **\$11,153 million** in 2018. The \$8.1 billion JV gain alone accounts for roughly half of that year's net income.

A secondary unusual item was a **\$1.4 billion tax benefit** recorded in Q2 2019 from the favorable settlement of a U.S. IRS audit covering tax years 2009–2010.

Both items are explicitly flagged by Pfizer as not part of its ongoing business operations.

Cited documents: PFIZER_2019_10K.pdf

Judge rationale: The agent's answer includes the essential information (the gain on completion of the Consumer Healthcare JV Transaction) and does not contradict the expected answer.

Q133. What are three main companies acquired by Pfizer mentioned in this 10K report? Added note by ATG for disambiguation: the question is referring to FY2021.

Result: **OK** · Latency: 22.7 s · Tool calls: 3

Expected answer

Trillium, Array, and Therachon

Delivered answer

The three companies Pfizer acquired that are named in the FY2021 10-K's Acquisitions note are **Trillium, Array, and Therachon**.

- **Trillium** — acquired November 17, 2021, for \$18.50/share in cash, total consideration of \$2.0 billion (net of cash acquired); a clinical-stage immuno-oncology company.
- **Array** — acquired July 30, 2019, for \$48/share in cash, total fair value of consideration of \$11.2 billion (\$10.9 billion net of cash acquired); a commercial-stage biopharmaceutical company (Braftovi, Mektovi).
- **Therachon** — acquired July 1, 2019, for \$340 million upfront plus up to \$470 million in milestones (total fair value of consideration \$322 million); a clinical-stage biotech focused on rare diseases.

These are the three acquisitions detailed in Note 2A (Acquisitions) of the 2021 Form 10-K.

Cited documents: PFIZER_2021_10K.pdf

Judge rationale: The agent's answer includes all three essential companies (Trillium, Array, and Therachon) and does not contradict the expected answer.

Q134. How much does Pfizer expect to pay to spin off Upjohn in the future in USD million? Added note by ATG for disambiguation: the question is referring to Q2 of FY2023.

Result: **NOK** · Latency: 16.1 s · Tool calls: 3

Expected answer

77.78

Delivered answer

\$700 million.

Pfizer's Q2 2023 10-Q states: "We expect to incur costs of approximately \$700 million in connection with separating Upjohn, of which approximately 90% has been incurred since inception and through the second quarter of 2023."

So the total expected cost is \$700 million, with roughly \$630 million (90%) already incurred and about \$70 million remaining to be paid in the future.

Cited documents: Pfizer_2023Q2_10Q.pdf

Judge rationale: The agent's answer (700 million) contradicts the expected answer (77.78 million) and does not match the essential figure.

Q135. For Pfizer, which geographic region had the biggest drop in Q22023 year over year revenues (on a percentage basis)?

Result: OK · Latency: 8.4 s · Tool calls: 1

Expected answer

Developed Rest of the World

Delivered answer

Developed Rest of World had the biggest year-over-year revenue drop in Q2 2023, falling **74%** (from \$5,034M to \$1,305M).

Here's how all four geographic regions compared (total, i.e., including FX):

Region	Q2 2023	Q2 2022	% Change
U.S.	\$6,185M	\$11,222M	(45%)
Developed Europe	\$2,415M	\$5,480M	(56%)
Developed Rest of World	\$1,305M	\$5,034M	(74%)
Emerging Markets	\$2,828M	\$6,006M	(53%)

The Developed Rest of World decline was driven primarily by the collapse in Comirnaty (–82%) and Paxlovid (–98%) revenues in that region, as Pfizer transitioned away from contracted government deliveries toward traditional commercial markets.

Cited documents: PFIZER_2023Q2_EARNINGS.pdf

Judge rationale: The agent's answer correctly identifies 'Developed Rest of World' as the region with the biggest percentage drop in Q2 2023 revenues, matching the expected answer.

Q136. As of Q2'2023, is Pfizer spinning off any large business segments?

Result: **OK** (human review) · Latency: 23.7 s · Tool calls: 3

Expected answer

Yes, it's spinning off Upjohn.

Delivered answer

No. As of Q2 2023, Pfizer was not spinning off any large business segment.

Per its Q2 2023 10-Q, Pfizer managed its commercial operations through two operating segments—**Biopharma** (the only reportable segment) and **Business Innovation** (PC1 plus Pfizer Ignite)—and no segment or business unit was held for sale, pending spin-off, or being divested.

The only divestiture disclosed in that filing was a small, early-stage one: in July 2023 Pfizer agreed to sell its early-stage rare disease gene therapy portfolio to Alexion (an AstraZeneca subsidiary) for up to \$1 billion plus tiered royalties, expected to close in Q3 2023—not a large business segment.

For context, the last major spin-off was the **Upjohn Business**, combined with Mylan to form Viatris and completed on November 16, 2020.

Cited documents: PFIZER_2022_10K.pdf, Pfizer_2023Q2_10Q.pdf

Judge verdict: **NOK** · The agent's answer contradicts the expected answer by stating there is no spin-off, while the reference confirms Pfizer is spinning off Upjohn.

Human review: **OK** · The spin-off of Upjohn was completed in November 2020; by 2023, only residual costs remain. The agent's response ("No") is factually correct.

Q137. Which debt securities are registered to trade on a national securities exchange under Ulta Beauty's name as of FY2023?

Result: OK · Latency: 15.7 s · Tool calls: 3

Expected answer

There are none

Delivered answer

None. Ulta Beauty has no debt securities registered to trade on a national securities exchange as of FY2023 (fiscal year ended January 28, 2023).

The 10-K cover page lists only one security registered under Section 12(b): common stock, par value \$0.01 per share, trading under the symbol "ULTA" on The NASDAQ Global Select Market. Section 12(g) lists none. The balance sheet also confirms there was no outstanding debt as of January 28, 2023.

Cited documents: ULTABEAUTY_2023_10K.pdf

Judge rationale: The agent's answer correctly states there are no debt securities registered to trade on a national securities exchange for Ulta Beauty as of FY2023, matching the essential information in the expected answer.

Q138. What are major acquisitions that Ulta Beauty has done in FY2023 and FY2022?

Result: OK · Latency: 9.2 s · Tool calls: 1

Expected answer

Ulta Beauty did not make any acquisitions in FY2023 and FY2022.

Delivered answer

None. Ulta Beauty made no major acquisitions in either fiscal year — the "Acquisitions, net of cash acquired" line in the investing section of the cash flow statement is zero (—) for both periods.

- FY2023 (fiscal year ended January 28, 2023, which the company labels "fiscal 2022"): \$0
- FY2022 (fiscal year ended January 29, 2022, which the company labels "fiscal 2021"): \$0

This is corroborated by the balance sheet, where goodwill is unchanged at \$10.87 million across both year-ends (and other intangible assets are immaterial, ~\$1.3M and ~\$1.5M), indicating no new business was purchased. The only acquisition in the three-year window shown was a trivial \$1.2M in the prior year (fiscal 2020, ended January 30, 2021).

Note on fiscal-year labels: Ulta's fiscal year does not end in December, so I've read your "FY2023" and "FY2022" as the fiscal years ending in calendar 2023 and 2022 respectively — the same two years the 10-K reports as "fiscal 2022" and "fiscal 2021."

Cited documents: ULTABEAUTY_2023_10K.pdf

Judge rationale: The agent's answer correctly states that Ulta Beauty made no acquisitions in FY2023 and FY2022, matching the essential information in the expected answer.

Q139. What drove the reduction in SG&A expense as a percent of net sales in FY2023? Added note by ATG for disambiguation: the question is referring to Ulta Beauty.

Result: OK · Latency: 11.8 s · Tool calls: 2

Expected answer

Lower marketing expenses and leverage of incentive compensation due to higher sales. The answer here assumes FY2023 refers to the 12 months ended on January 28, 2023 (although the company refers to this period as its fiscal 2022).

Delivered answer

SG&A fell 0.4 percentage points, from 23.9% of net sales to 23.5% in fiscal 2023 (the 52 weeks ended January 28, 2023).

The drop came from two recurring levers:

- **Lower marketing expenses**
- **Leverage of incentive compensation** on higher sales

Partially offset by two deleverage items:

- **Corporate overhead** (strategic investments)
- **Store payroll and benefits** (wage investments)

The math checks out: SG&A of \$2,395.3M on net sales of \$10,208.6M = 23.5%, versus \$2,061.5M on \$8,630.9M = 23.9% in the prior year — a 0.4 pp reduction, driven by the marketing and incentive-comp leverage outpacing the corporate-overhead and store-payroll deleverage.

Cited documents: ULTABEAUTY_2023_10K.pdf, ULTABEAUTY_2023Q4_EARNINGS.pdf

Judge rationale: The agent's answer includes the essential drivers (lower marketing expenses and leverage of incentive compensation) and does not contradict the expected answer.

Q140. What drove the increase in Ulta Beauty's merchandise inventories balance at end of FY2023?

Result: OK · Latency: 20.7 s · Tool calls: 3

Expected answer

Increase in Merchandise inventories balance was driven by the opening of 47 new stores. The answer here assumes FY2023 refers to the 12 months ended on January 28, 2023 (although the company refers to this period as its fiscal 2022).

Delivered answer

\$104.2 million (7.0%) increase — merchandise inventories, net rose from \$1.5 billion at January 29, 2022 to \$1.6 billion at January 28, 2023 (the fiscal year ending in 2023, which Ulta labels "fiscal 2022").

Management attributes the increase to three drivers:

- **\$54 million** — inventory to support the 47 new stores opened since January 29, 2022
- **\$25 million** — new key brand launches
- **\$25 million** — inventory (product) cost increases

These three items account for the full \$104.2 million change, with the new-store build-out being the largest single contributor.

Cited documents: ULTABEAUTY_2023_10K.pdf

Judge rationale: The agent's answer includes the essential driver (47 new stores) and provides additional correct details without contradiction.

Q141. What percent of Ulta Beauty's total spend on stock repurchases for FY 2023 occurred in Q4 of FY2023?

Result: **NOK** · Latency: 36.6 s · Tool calls: 9

Expected answer

36%. The answer here assumes FY2023 refers to the 12 months ended on January 28, 2023 (although the company refers to this period as its fiscal 2022).

Delivered answer

I don't have the data needed to answer this precisely.

Here's what's in the knowledge base for Ulta Beauty's fiscal 2023 (the year ending late January 2024):

- **Q1 FY2023** (13 weeks ended April 29, 2023): repurchased 541,108 shares at a cost of **\$285.8 million**
- **First six months FY2023** (26 weeks ended July 29, 2023): repurchased 1.1 million shares at a cost of **\$559.0 million** — which implies Q2 alone was ~\$273.2 million.

But the knowledge base does **not** contain the Q3 FY2023 or Q4 FY2023 earnings releases, nor the fiscal 2023 10-K (the only 10-K present is the FY2022 one, ended January 28, 2023). Without the full-year FY2023 repurchase total and the Q4 FY2023 figure, I can't compute what percent of the year's buybacks landed in Q4.

The only FY2023 repurchase number available is the company's **guidance** of "approximately \$900 million," which is a plan, not an actual, so it can't be used to answer the question.

If you can add the Q3/Q4 FY2023 earnings release or the FY2023 10-K to the knowledge base, I'll compute the Q4 share right away.

Cited documents: ULTABEAUTY_2023Q1_EARNINGS.pdf, ULTABEAUTY_2023Q2_EARNINGS.pdf

Judge rationale: The agent's answer refuses to provide the information while the expected answer gives the correct percentage (36%).

Q142. Did Ulta Beauty's wages expense as a percent of net sales increase or decrease in FY2023?

Result: OK · Latency: 15.6 s · Tool calls: 4

Expected answer

Wages expense as a percent of net sales increased in FY2023. The answer here assumes FY2023 refers to the 12 months ended on January 28, 2023 (although the company refers to this period as its fiscal 2022).

Delivered answer

Increased.

Ulta doesn't break out a standalone "wages expense" line; the closest disclosed component is **store payroll and benefits** (part of SG&A). In the fiscal year ended January 28, 2023 (which Ulta labels fiscal 2022, i.e., the calendar-2023 fiscal year = "FY2023"), store payroll and benefits **deleveraged by 20 basis points** as a percent of net sales, driven by wage investments.

So the wages/payroll line as a share of net sales rose (deleverage) in FY2023.

One caveat worth flagging: Ulta reports this as "store payroll and benefits" (wages + benefits combined), not wages alone, so the 20 bps figure is the best available proxy for the wage component.

Cited documents: ULTABEAUTY_2023_10K.pdf

Judge rationale: The agent's answer correctly states that wages expense as a percent of net sales increased in FY2023, aligning with the expected answer.

Q143. Among all of the derivative instruments that Verizon used to manage the exposure to fluctuations of foreign currencies exchange rates or interest rates, which one had the highest notional value in FY 2021?

Result: OK · Latency: 7.9 s · Tool calls: 1

Expected answer

Cross currency swaps. Its notional value was \$32,502 million.

Delivered answer

Cross currency swaps had the highest notional value among Verizon's derivatives used to manage foreign currency and interest rate exposure in FY 2021.

Per the notional amounts of outstanding derivative instruments at December 31, 2021 (Verizon 2021 10-K):

- Cross currency swaps: **\$32,502 million**
- Interest rate swaps: \$19,779 million
- Forward starting interest rate swaps: \$1,000 million
- Foreign exchange forwards: \$932 million

Cross currency swaps (\$32.5 billion) exceed the next largest (interest rate swaps, \$19.8 billion) by roughly \$12.7 billion.

Cited documents: VERIZON_2021_10K.pdf

Judge rationale: The agent's answer correctly identifies cross currency swaps as the derivative with the highest notional value and matches the expected figure of \$32,502 million.

Q144. As of FY 2021, how much did Verizon expect to pay for its retirees in 2024?

Result: OK · Latency: 21.2 s · Tool calls: 3

Expected answer

The estimated pension benefits were \$1097 million, and the estimated health care and life insurance benefits were \$862 million.

Delivered answer

\$1,959 million (about \$1.96 billion).

From Verizon's FY2021 10-K (fiscal year ended December 31, 2021), the "Estimated Future Benefit Payments" table shows the expected payments to retirees in 2024:

- Pension benefits: \$1,097 million
- Health Care and Life benefits: \$862 million

Total = \$1,097M + \$862M = **\$1,959 million**

Note: these are the benefit payments Verizon expected to make to retirees in 2024, as originally reported in its FY2021 annual filing.

Cited documents: VERIZON_2021_10K.pdf

Judge rationale: The agent's answer includes the essential figures (\$1,097 million and \$862 million) and correctly sums them to \$1,959 million, with no contradictions.

Q145. Does Verizon have a reasonably healthy liquidity profile based on its quick ratio for FY 2022? If the quick ratio is not relevant to measure liquidity, please state that and explain why.

Result: OK · Latency: 10.9 s · Tool calls: 1

Expected answer

No. The quick ratio was approximately 0.54 for Verizon. It indicated that Verizon does not have a healthy liquidity profile.

Delivered answer

No. Verizon's quick ratio for FY 2022 (fiscal year ended December 31, 2022) is below 1.0 on both usual bases, so its liquidity profile is not healthy by this measure.

- **Strict** (cash + short-term investments + receivables) / total current liabilities = $(\$2,605\text{M} + \$0 + \$24,506\text{M}) / \$50,171\text{M} = \mathbf{0.54}$
- **Broad** (total current assets – inventories) / total current liabilities = $(\$37,857\text{M} - \$2,388\text{M}) / \$50,171\text{M} = \mathbf{0.71}$

Because both figures fall below 1.0, the strict basis (the deciding one) confirms the profile is not healthy: Verizon's most liquid assets covered only about 54% of its current liabilities at year-end 2022.

Source: Verizon 2022 Annual Report on Form 10-K, Consolidated Balance Sheets.

Cited documents: VERIZON_2022_10K.pdf

Judge rationale: The agent's answer correctly states that Verizon's quick ratio is below 1.0 (0.54) and confirms the liquidity profile is not healthy, matching the essential information in the expected answer.

Q146. Is Verizon a capital intensive business based on FY 2022 data?

Result: **OK** · Latency: 20.7 s · Tool calls: 2

Expected answer

Yes. Verizon's capital intensity ratio was approximately 2.774729. This means that it took approximately \$2.77 of assets to generate \$1 of revenue and thus, Verizon can be considered capital intensive.

Delivered answer

Yes — Verizon is a capital-intensive business based on FY 2022 data (fiscal year ended December 31, 2022).

- **CapEx / Revenue:** \$23,087M / \$136,835M = **16.9%** — a large, ongoing reinvestment in network infrastructure.
- **Net fixed assets / Total assets:** \$107,434M / \$379,680M = **28.3%** — a substantial share of the balance sheet is tied up in property, plant and equipment.
- **Return on assets:** \$21,256M (net income attributable to Verizon) / \$379,680M = **5.60%** — a low return generated on a very large asset base.

The combination of heavy capital spending relative to revenue, a large fixed-asset base, and a modest return on that base is the classic signature of a capital-intensive business. Verizon's own filing notes that 2022 "was a peak year of capital investment" as it rapidly deployed C-Band spectrum, reinforcing this conclusion.

Cited documents: VERIZON_2022_10K.pdf

Judge rationale: The agent's answer confirms Verizon is capital-intensive and provides supporting evidence consistent with the expected answer's conclusion.

Q147. Has Verizon increased its debt on balance sheet between 2022 and the 2021 fiscal period?

Result: **OK** · Latency: 20.9 s · Tool calls: 2

Expected answer

No. Verizon's debt decreased by \$229 million.

Delivered answer

No, Verizon's total debt on the balance sheet actually decreased slightly between fiscal 2021 and fiscal 2022.

- Total debt at Dec 31, 2021: \$150,868M (debt maturing within one year \$7,443M + long-term debt \$143,425M)
- Total debt at Dec 31, 2022: \$150,639M (debt maturing within one year \$9,963M + long-term debt \$140,676M)
- Change: -\$229M, a decline of about 0.2%

The increase in short-term debt (current maturities) was more than offset by a larger reduction in long-term debt, so the net balance edged down year over year.

Sources: Verizon 2021 10-K (Note 7, Debt) and Verizon 2022 10-K (Note 7, Debt).

Cited documents: VERIZON_2021_10K.pdf, VERIZON_2022_10K.pdf

Judge rationale: The agent's answer correctly states that Verizon's debt decreased and matches the essential information of the expected answer, including the \$229 million decrease.

Q148. What is FY2018 days payable outstanding (DPO) for Walmart? DPO is defined as: $365 \times (\text{average accounts payable between FY2017 and FY2018}) / (\text{FY2018 COGS} + \text{change in inventory between FY2017 and FY2018})$. Round your answer to two decimal places. Please base your judgments on the information provided primarily in the statement of financial position and the P&L statement.

Result: OK · Latency: 9.9 s · Tool calls: 1

Expected answer

42.69

Delivered answer

42.69 days.

Here's the calculation, using Walmart's fiscal 2018 (year ended January 31, 2018) 10-K figures:

- **Average accounts payable** = (FY2018 AP \$46,092M + FY2017 AP \$41,433M) / 2 = **\$43,762.5M**
- **FY2018 COGS** (cost of sales) = **\$373,396M**
- **Change in inventory** = FY2018 inventory \$43,783M – FY2017 inventory \$43,046M = **\$737M**

$\text{DPO} = 365 \times \$43,762.5\text{M} / (\$373,396\text{M} + \$737\text{M}) = 365 \times \$43,762.5\text{M} / \$374,133\text{M} = \mathbf{42.69 \text{ days}}$

All figures come from Walmart's consolidated balance sheets (accounts payable, inventories) and consolidated statement of income (cost of sales) in the FY2018 10-K.

Cited documents: WALMART_2018_10K.pdf

Judge rationale: The agent's answer matches the expected answer exactly (42.69) and includes the correct calculation steps without contradictions.

Q149. Based on the information provided primarily in the statement of income, what is the FY2018 - FY2019 change in unadjusted operating income % margin for Walmart? Answer in units of percents and round to one decimal place.

Result: OK · Latency: 10.6 s · Tool calls: 1

Expected answer

0.2%

Delivered answer

0.2% (an increase).

Operating income margin = operating income ÷ net sales, from the Consolidated Statements of Income:

- FY2019: \$21,957M ÷ \$510,329M = 4.30%
- FY2018: \$20,437M ÷ \$495,761M = 4.12%

Change (FY2019 – FY2018) = 4.30% – 4.12% = **+0.2 percentage points**.

This matches the company's own MD&A disclosure, which reports operating income as a percentage of net sales of 4.3% (FY2019) and 4.1% (FY2018).

Cited documents: WALMART_2019_10K.pdf

Judge rationale: The agent's answer matches the expected answer exactly in value and precision (0.2%).

Q150. What is the FY2018 - FY2020 3 year average unadjusted EBITDA % margin for Walmart? Define unadjusted EBITDA as unadjusted operating income + depreciation and amortization from the cash flow statement. Answer in units of percents and round to one decimal place. Calculate what was asked by utilizing the line items clearly shown in the P&L statement and the cash flow statement.

Result: OK · Latency: 40.5 s · Tool calls: 4

Expected answer

6.2%

Delivered answer

6.2%

Unadjusted EBITDA = operating income (P&L) + depreciation & amortization (cash flow statement); margin = EBITDA ÷ total revenues (P&L top line).

FY (ended Jan 31)	Operating income	+ D&A (CF)	= EBITDA	Total revenues	EBITDA margin
2018	\$20,437M	\$10,529M	\$30,966M	\$500,343M	6.19%
2019	\$21,957M	\$10,678M	\$32,635M	\$514,405M	6.34%
2020	\$20,568M	\$10,987M	\$31,555M	\$523,964M	6.02%

3-year average = (6.19% + 6.34% + 6.02%) ÷ 3 = **6.2%**

(If you use net sales instead of total revenues as the denominator, the average is 6.24%, which also rounds to 6.2%.)

Source: Walmart 10-K for fiscal 2020 (Consolidated Statements of Income and Cash Flows).

Cited documents: WALMART_2020_10K.pdf

Judge rationale: The agent's answer matches the expected answer exactly (6.2%) and includes the correct calculation method and supporting data.

Appendix A: judge prompt

You are an evaluator for an AI assistant benchmark. You decide whether the assistant's answer to a question is correct, by comparing it with a reference answer written by a domain expert.

You receive JSON with: question, expected_answer (the reference), agent_answer (the assistant's answer; it may be truncated). They may be written in any language.

The answer PASSES when both conditions hold:

1. It contains the essential information of the expected answer: the facts, figures, conditions or steps without which the question is not answered. Wording, order, level of detail, language, tone and formatting do not matter.

2. It contains no statement that contradicts the expected answer and no factual error on the point asked.

Additional correct information does not make the answer fail. An answer that misses an essential point, contradicts the reference, refuses or says it does not know, while the expected answer gives the information, FAILS.

Numbers given with a different precision are the same figure when the question sets no precision: round only the more precise figure to the precision of the other one; if this gives the other figure, they match (for example 12.2% for 12%, 6.37% for 6.4%, \$5,412 million for \$5.4 billion). Never round both figures further than that: two figures given with the same precision must be equal (\$5,412 million and \$5,420 million are different figures, and so are 12.5% and 12%). Units, scale and sign must still agree. When the question sets a rounding or a number of decimals, the answer must follow it.

Return ONLY a JSON object with exactly these keys:

```
{"passed": <true or false>, "rationale": "<one sentence in English explaining the verdict>"}
```

No markdown fences, no extra keys, no text outside the JSON.

Appendix B: agent instructions

Instructions written by the agent administrator, as given to the agent during the run.

Role

You answer financial-analysis questions about public companies, using only the SEC filings and earnings releases in the knowledge base. Each answer is graded against a short reference answer written by a financial analyst: one verdict or one figure, then the one or two numbers behind it. Answer the way that analyst would.

Shape of the answer

- Always answer. Never ask the reader a question, not even to clarify: the reader cannot reply. When the question does not name the company, find in the knowledge base the company whose filings cover the periods the question names, answer for it, and state that assumption in one clause.
- First words: the answer itself. "Yes" or "No" only when the question is a yes/no question; for any other question, the figure, the name or the direction (improved, declined). Commit to one verdict: no "yes, but", no "moderately", no reframing of the question.
- Then the inputs and the calculation, once, in two to four short lines. Check that the calculation shown reconciles to the figure you gave.
- For a question about a disclosure (an event, a transaction, a legal matter, an executive, a segment, a plan), answer in one or two plain sentences that say what it is in substance. For plans and forecasts, give the direction for each item first (increase, decrease, pause), then the figures.
- When the question also asks whether a metric is relevant, say so in one clause after the answer, never before it.
- End on the last fact or its source. Never end with a question, an offer to go further or an invitation to ask for more.
- No headings, no emoji, no history of other years, no commentary on what the figure means. A table only when the calculation has more than three inputs.

Numbers

- Use the unit the question asks for, and convert when needed. Use the rounding the question asks for. When it gives none: two decimals for ratios, one decimal for percentages, whole numbers for amounts in millions, and one decimal for amounts in billions followed by the exact figure in millions in parentheses.
- When the question asks for a ratio rounded to two decimals (return on assets, payout, turnover, coverage), the answer is the decimal number rounded to two decimals, even when it looks small. The percentage may follow in parentheses, never replace it.
- Margins, rates and their changes are computed on the figures as reported under GAAP, unless the question says adjusted.
- Keep full precision in every intermediate value and round only the final result. A change between two margins or rates is computed on the unrounded values, never by subtracting two rounded percentages, and is stated in percentage points.
- Before answering, redo the last operation of the calculation from the values shown, signs included: the figure in the first words must be its result.
- When the answer is a change between two values, give both values, the absolute change and the percentage change.
- When the question gives a formula, apply it exactly as written, including averages over two year-ends.
- When the question gives no formula for a ratio that mixes a flow with a balance (turnover, return on assets), compute it on the year-end balance and on the two-year average balance, and give both, labelled, in the first sentence.
- Use only the statements the question names.

Finding the right line

- Use the company's own fiscal-year labels, and state the period end date when the fiscal year is not the calendar year. When the fiscal year does not end in December and no filing matches the question's year under the company's own label, read the question's fiscal year as the fiscal year that ends in that calendar year (FY2023 is the fiscal year that ends in 2023): filing names in the knowledge base follow that convention (a 2023 10-K covers a year that ended in early 2023). Answer for that year and state the reading in one clause; never conclude that the data is missing before trying it.
- When the question names a fiscal year, take its figures from that year's own annual filing, as originally reported. If a later filing restated them, give the original figure first and the restated one in one clause.
- When the question names no period, use the most recent filing in the knowledge base, including for track records that lengthen each year.
- Match the concept to the line the filing actually uses. Capital expenditure is purchases of property, plant and equipment in the investing section of the cash flow statement. Revenue may be labelled net sales or net revenues; cost of goods sold may be cost of sales or cost of revenue; receivables may be trade receivables, net. Dividends paid comes from the financing section.

- Depreciation and amortization comes from the cash flow statement and means depreciation of property and equipment and amortization of intangible assets. Amortization of content, programming or other assets that the income statement includes in cost of revenue is an operating cost, not D&A.
- "Unadjusted" means the figure as reported under GAAP. "Adjusted" or "non-GAAP" means the company's own published figure: quote it from the earnings release, do not rebuild it.
- Net income attributable to shareholders is net income less non-controlling interests; when the income statement shows none, it is net income. In any ratio or margin built on net income (return on assets, return on equity, net margin, payout ratio), net income means net income attributable to shareholders, unless the question says otherwise.
- Where a company operates: the first sentence lists the regions of the company's own table of revenue by geographic area (or its segment note), with the company's region names. The countries named in the business description are not the answer.

Definitions when the question gives none

State the definition in one short clause.

- A measure the question names but the filing does not report: build it from its definition with the filing's line items (EBIT is operating income; EBITDA adds depreciation and amortization to it; an adjusted measure applies the adjustments the company itself makes). Never replace it with a different measure the company reports. A coverage ratio whose numerator is negative gives no coverage: say so.
- Working capital: total current assets minus total current liabilities. Add in one clause the operating figure, which leaves out cash and short-term investments on one side and short-term debt on the other.
- Quick ratio: give it on both usual bases, labelled, in the first sentence. Strict: (cash and cash equivalents + short-term investments + receivables) / total current liabilities. Broad: (total current assets - inventories) / total current liabilities. At or above 1.0 the liquidity profile is healthy; below 1.0 it is not. If the two bases fall on either side of 1.0, decide on the strict one.
- Free cash flow: cash from operations minus capital expenditure. Free cash flow conversion: free cash flow / net income.
- Effective tax rate: income tax expense / pre-tax income, to two decimals. In a loss year, say in words whether it is a tax expense or a tax benefit, because sign conventions differ between sources.
- Capital intensity: report capital expenditure / revenue, net fixed assets / total assets and return on assets, then conclude with a plain yes or no. A company that earns a low return on a large asset base is capital-intensive, even when goodwill rather than fixed assets makes up most of that base. A company with a high return on assets and modest capital expenditure relative to revenue is not.
- What shareholders would receive in a liquidation: tangible common equity per share, that is total equity less preferred stock, goodwill and intangible assets, divided by common shares outstanding, because goodwill and intangibles fetch nothing in a liquidation. Use the tangible book value per share the company reports when there is one.

Judgment questions

- Trend as of a given fiscal year (improving, declining, what changed): compare that year with the one immediately before, and decide on that change alone.
- Consistency or stability with a tolerance stated in the question: test each year-on-year change against that tolerance.
- What drove a change: give the size of the change first, then the drivers management states in its discussion of results, naming the one-off items separately from the recurring ones.
- Relevance of a metric for a non-financial company, utilities included: every standard metric applies. Compute it, and do not set it aside because other indicators look better. The only exception is a metric whose balance-sheet item does not exist for that company.
- Relevance of a metric for a financial institution (a bank or bank holding company, a lender, an insurer): operating margin, gross margin, working capital, liquidity ratios and inventory metrics are not how its performance is measured, because interest and credit losses are part of its core operations. When asked about one of these, the answer is that the metric is not used for this kind of company: say so in the first words, give the reason in one sentence, and do not compute a substitute.
- Composition or nature of a balance: the first sentence names the largest component and its share of the total as a percentage. The amounts come after.
- Which segment: rank every segment the filing reports. A segment the filing calls corporate, other or unallocated is a segment like the others: when it has the lowest or the highest value, it is the answer, even if the filing sets it apart from the main business segments and even when the question says "business segments". Do not answer with the extreme among the other segments only.
- Importance of a matter (legal, risk, customer concentration): read "material" and "important" as what matters to an investor, not in the narrow accounting sense. When the filings describe significant litigation, investigations or settlements, answer yes and name every family of matters the legal note describes, each in one line, with amounts where given.

Guidance and outlook

- Use the guidance published with the results of the fiscal year in question. Do not use later updates unless the question asks for them.
- When guidance is given on both a constant-currency basis and a reported basis, give both and base the verdict on the constant-currency figure. Companies may call the constant-currency figure operational, organic or ex-FX (for example "adjusted operational EPS"): a guidance stated that way is the constant-currency figure.