Example Task: Downtown Dallas Coffee Shop

Final rating: 7 · B Much better

Deciding axis: Sourcing. Response B is preferred because sourcing is the highest-priority evaluation axis, and its claims have stronger citation support. Response A performs better on Accuracy because its evidence is more directly relevant to the requested business decision, but that advantage does not outweigh B's sourcing advantage under the rubric's stated priorities.

📝 Prompt

I’m opening a small coffee shop in Dallas, Texas, and I’ve narrowed it down to two locations, one near Main Street in Downtown Dallas and one in Bishop Arts District. I’m trying to decide which gives the business the better financial opportunity over the next several years, not simply which has the lower rent. Compare the two based on current rents, customer traffic, nearby growth, and recent performance of comparable businesses, and tell me which location you would choose.

⚖️ Axis by axis

Axis	Better	Why
Sourcing (highest priority)	✅ **B**	B has correct reliable sources. A attributes a claim to a source that doesn’t contain the sales data stated.
Accuracy	**A**	A more accurate conclusions based on the evidence that better evaluate the financial opportunity
Completeness	**A**	A provides specific figures for rents, customer traffic, and comparable coffee shop sales, addressing more of the requested factors
Clarity	**B**	B organizes its supporting evidence into distinct sections making the reasoning easier to follow

📝 Main Claim

❓**Main Ask (from Prompt)**: which location gives the coffee shop the better financial opportunity over the next several years

🅰️**Response A Main Claim (from Response)**: “I would choose Downtown Dallas because its higher pedestrian traffic and higher comparable coffee shop sales suggest stronger business performance that could outweigh the higher occupancy costs.”

🅱️**Response B Main Claim (from Response):** “I would choose Downtown Dallas because population growth, nearby national retailer openings, and rising household incomes suggest a stronger long-term market opportunity.”

🅰️ Critique of Response A

Sourcing: The claim, “Asking rents are 18% higher in Downtown Dallas” (Dallas Commercial Leasing Report, cited in the response), is supported by its cited source. This was selected because asking rents directly inform the comparison of occupancy costs, which affects whether higher sales could offset Downtown's higher costs. The claim, “Pedestrian traffic is 35% higher in Downtown Dallas” (Dallas Pedestrian Traffic Report, cited in the response), is also supported by its cited source. This was selected because pedestrian traffic indicates potential customer exposure, which may contribute to stronger business performance, although higher foot traffic alone does not establish actual customer demand or sales. The claim, “Average monthly sales at comparable coffee shops are 22% higher Downtown” (Dallas Commercial Leasing Report, cited in the response), is not supported by its cited source, which contains no coffee shop sales data. The figure was independently checked against the Dallas Coffee Sales Report (www.xyz.com), which confirms the 22% difference. The claim is independently substantiated, but the original citation is incorrect. This is material because comparable sales are central to Response A's argument that stronger business performance could outweigh higher occupancy costs. The cited Dallas municipal reports are reliable primary sources, while the Dallas Coffee Sales Report is a less authoritative secondary source. Overall, the evidence supports higher pedestrian traffic and comparable sales potential, but it does not establish that the additional sales would offset higher rents and other operating costs.

Accuracy: The recommendation that Downtown Dallas is the better location is partially supported by the reported 35% higher pedestrian traffic and 22% higher average monthly

sales at comparable coffee shops, which suggest greater customer exposure and stronger commercial performance. However, these figures do not establish that the additional revenue would offset Downtown's 18% higher asking rents or other operating costs. The response does not compare projected revenue against total operating expenses or establish whether the higher sales would translate into greater profitability. As a result, although the evidence supports Downtown's potential to generate more business, the conclusion that it offers the stronger overall financial opportunity remains insufficiently substantiated.

Completeness: Response A compares Downtown Dallas and Bishop Arts District and provides specific figures for rents, pedestrian traffic, and comparable coffee shop sales. However, its discussion of nearby growth is less developed, and its analysis of how the locations' financial opportunities may evolve over the next several years is limited. Overall, it addresses the main components of the prompt but lacks depth in its long-term comparison.

Clarity: Response A is concise and easy to follow, with a direct recommendation, “I would choose Downtown Dallas,” followed by supporting evidence in short paragraphs. However, the opening could state the primary reason for the recommendation alongside the decision, giving the reader the conclusion and its rationale immediately. Overall, the response is readable and direct, with only a minor opportunity to improve the presentation.

🅱️ Critique of Response B

Sourcing: The claim, “Downtown has had 12% population growth” (Dallas Neighborhood Growth Report, cited in the response), is supported by its cited source. This was selected because population growth provides evidence of an expanding potential customer base, supporting the response's recommendation that Downtown has a stronger long-term market opportunity. The claim, “Several national retailers have opened nearby” (Downtown Retail Development Report, cited in the response), is also supported by its cited source. This was selected because nearby retail development may indicate commercial investment and increased activity in the area, which could support the location's potential to attract customers. The claim, “The average household income in Downtown Dallas increased by 8% over the past three years” (Dallas Household Income Report, cited in the response), is supported by its cited source. This was selected because rising household income may indicate greater potential purchasing capacity among local consumers, which is relevant to assessing the market opportunity for a coffee shop. Together, these three checks provide traceable evidence supporting Downtown's potential for long-term growth and the rationale for choosing it. However, they do not directly establish coffee shop demand or profitability, so the recommendation is supported as a growth opportunity but not as a proven stronger financial choice. The cited municipal reports are reliable primary sources that provide direct evidence for the reported population growth, retail development, and household income trends.

Accuracy: The recommendation to choose Downtown Dallas is partially supported by the reported 12% population growth and 8% increase in average household income over the past three years, as well as the opening of several national retailers nearby. These indicators suggest a growing market with potential customer spending capacity, but they do not establish that more people will visit the coffee shop or make purchases. The response also does not demonstrate that these market trends would generate enough additional revenue to justify Downtown's 18% higher asking rents. Without evidence connecting the reported growth to expected coffee shop sales and profitability, the conclusion that Downtown offers the better long-term financial opportunity remains insufficiently substantiated.

Completeness: Response B discusses population growth, nearby retail development, and rising household incomes in Downtown Dallas. However, it does not compare customer traffic or recent performance among comparable businesses, two factors explicitly requested in the prompt. Its comparison of both locations' long-term financial opportunities is also limited, leaving important parts of the user's decision unanswered.

Clarity: Response B is organized into logical sections and uses bolding and bullet points to make information easier to scan. However, the paragraph beginning “The Dallas Neighborhood Growth Report…” groups population growth, household income, and retail development together rather than separating these findings into distinct points. Breaking up this paragraph would make the supporting evidence easier to review quickly.

🏁Overall Preference Rationale

Overall: Response B is “Much” better because sourcing is the highest-priority axis, and it is better on clarity. Although Response A performs better on Accuracy and Completeness, those advantages do not outweigh Response B's sourcing advantage.

Sourcing: Response B is stronger because all three claims about population growth, nearby national retailer openings, and rising household income are supported by municipal Dallas reports that are reliable sources such as “Pedestrian traffic is 35% higher in Downtown Dallas” (Fake Dallas Pedestrian Traffic Report). In contrast, Response A attributes the claim that comparable coffee shops generate 22% higher average monthly sales Downtown to the Dallas Commercial Leasing Report, which contains no coffee shop sales data. Although the figure was independently verified using the Dallas Coffee Sales Report, the original citation remains incorrect. Response B therefore has stronger sourcing because its cited sources support the claims attributed to them and are more reliable.

Accuracy: Response A is stronger because the pedestrian traffic and comparable coffee shop sales figures are more directly relevant to evaluating the financial opportunity, based on the main ask of the prompt, of the two locations than Response B's population growth, retail development, and household income indicators. However, neither response establishes that Downtown's potential revenue advantages would offset its 18% higher asking rents and other operating costs. Response A's evidence is more directly relevant to the decision of the user that is central to the prompt.

Completeness: Based on the prompt, Response A is stronger because it provides specific figures for rents, customer traffic, and comparable coffee shop sales, addressing more of the requested factors with directly relevant evidence. Response B discusses market growth

indicators but does not compare customer traffic or comparable-business performance and therefore lacking completeness.

Clarity: Response B is clearer than Response A because it organizes its supporting evidence into distinct sections such as “Downtown Offerings”, making the reasoning easier to follow. Response A is concise and direct, such as providing the recommendation “I would choose Downtown Dallas,” but B presents its supporting points more systematically.