A. Summarize the real-data research question you identified in Task 1. Your summary should include justification for the research question you identified in Task 1, a description of the context in which the research question exists, and a discussion of your hypothesis.

A. Research Question Summary My research question is: *What is the average quantity sold per product category over the past year?*

Justification: This research question is important because it can provide valuable insights for inventory management, sales forecasting, and strategic decision-making. By understanding the average quantities sold across product categories, businesses can identify high-performing categories to maximize profits and optimize stock levels. Similarly, underperforming categories can be evaluated for potential improvements in marketing strategies, pricing adjustments, or product offerings. Additionally, such analysis can help uncover seasonal trends and fluctuations, enabling more accurate resource allocation to meet consumer demand effectively.

Context: This question is rooted in the context of analyzing historical sales data from the past year, which offers a full cycle of seasonal influences on purchasing behavior. The findings are particularly relevant for retail and e-commerce businesses that depend on data-driven decisions to maintain competitive advantages. Understanding category-specific performance allows businesses to adapt to market demands, streamline inventory costs, and enhance customer satisfaction.

Hypotheses:

- **Null Hypothesis (H_0):** There is no significant difference in the average quantity sold across different product categories over the past year. This null hypothesis sets the baseline expectation that variations in the average quantities sold are not statistically significant. Accepting the null hypothesis would indicate that sales patterns across product categories are relatively uniform, suggesting that external factors such as seasonality or marketing strategies do not play a dominant role in influencing sales behavior within the observed timeframe. This could imply that resource allocation and inventory strategies might not need to differ dramatically across categories, leading to potentially streamlined operations. However, this outcome could also prompt further inquiry into why such uniformity exists, as it may reflect missed opportunities for category-specific optimization.

- **Alternative Hypothesis (H_1):** There is a significant difference in the average quantity sold across different product categories over the past year. The alternative hypothesis posits that there are meaningful differences in the average quantities sold, highlighting the possibility of significant variability among product categories. This outcome would suggest that some categories consistently outperform others, potentially due to factors like consumer preferences, seasonal trends, or targeted marketing efforts. For instance, holiday-related categories might show spikes during specific months, while staple goods may exhibit steady performance year-round. Accepting the alternative hypothesis would underscore the importance of tailoring

inventory management and marketing strategies to align with category-specific performance. It would also open avenues for deeper analysis into the drivers of these differences, enabling businesses to leverage insights for competitive advantage.

Relevance of Hypotheses: These hypotheses serve as the foundation for a statistical test that will determine whether the observed differences in average quantities sold are due to random chance or represent meaningful trends. The results of this analysis are critical for businesses aiming to adopt data-driven decision-making practices. For example:

- If the null hypothesis is rejected, businesses can focus resources on understanding why certain categories perform better and apply targeted strategies to replicate success in other areas.
- If the null hypothesis is accepted, it may indicate a need to investigate factors outside of category-specific data, such as pricing policies, external market conditions, or overall customer purchasing behavior.

By framing the hypotheses within the broader context of business strategy and decision-making, I not only address the statistical aspect but also connect it to practical implications. This adds depth to the discussion and highlights the significance of my research.

B. Report on your data-collection process by describing the relevant data you collected, discussing one advantage and one disadvantage of the data-gathering methodology you used, and discussing how you overcame any challenges you encountered during the process of collecting your data.

Relevant Data Collected: For this research, I created a custom dataset designed to simulate an e-commerce company specializing in computer electronics. The dataset includes sales data for a diverse range of product categories, with no fewer than ten unique items per category. These categories are: CPU, Graphics Card, Motherboard, Laptop, Console, Desktop PC, SSD, Controller, Computer Case, Monitor, Power Supply, Memory, USB Memory, Keyboard, Mouse, and Case Fan. To ensure the dataset reflects a realistic scenario, I wrote a Python script that generated simulated sales data over a one-year period. To add further complexity and depth to the dataset, I included five warehouses to represent different sales locations. Each transaction specifies the product sold, its category, the sales quantity, and the warehouse location.

Advantage of the Data-Gathering Methodology: One major advantage of creating my own dataset is the high level of control it provides over the data structure and variables. I was able to design the dataset to fit the specific needs of my research, ensuring that it contains sufficient detail and variability for meaningful analysis. For instance, the inclusion of multiple warehouses introduces an additional layer of complexity, enabling the exploration of geographic influences on product performance. Furthermore, simulating a one-year period allows for the identification of seasonal trends and patterns, which are essential for understanding fluctuations in sales.

Disadvantage of the Data-Gathering Methodology: A notable disadvantage of this approach is the lack of real-world validation. Since the dataset is artificially generated, it

does not account for unpredictable human behaviors, market dynamics, or external influences that could affect sales in a real business environment. This limits the generalizability of the findings to actual e-commerce scenarios, as the dataset may oversimplify or omit critical variables. Additionally, creating the dataset required significant time and effort to ensure it was both realistic and comprehensive, which may not always be feasible for larger-scale research.

Challenges Encountered and How They Were Overcome: One of the primary challenges I encountered was ensuring the realism of the simulated data. Generating random sales quantities without accounting for variability across product categories could have resulted in an overly uniform dataset that lacks meaningful trends or insights. To address this, I introduced category-specific constraints and variability into the Python script, allowing some categories to exhibit higher average sales quantities than others. For example, items like Laptops and Monitors were designed to have steadier sales, while categories such as Controllers and Consoles showed more fluctuation, reflecting market behavior.

Another challenge was maintaining consistency in the dataset while adding complexity, such as warehouse locations. To overcome this, I implemented logic within the Python script to allocate sales to warehouses based on predefined probabilities, simulating differences in regional demand. This step ensured the dataset accurately represented the geographic aspect of sales performance without sacrificing overall coherence.

C. Describe your data-extraction and -preparation process and provide screenshots to illustrate *each* step. Explain the tools and techniques you used for data extraction and data preparation, including how these tools and techniques were used on the data. Justify why you used these particular tools and techniques, including one advantage and one disadvantage of using them with your data-extraction and -preparation methods.

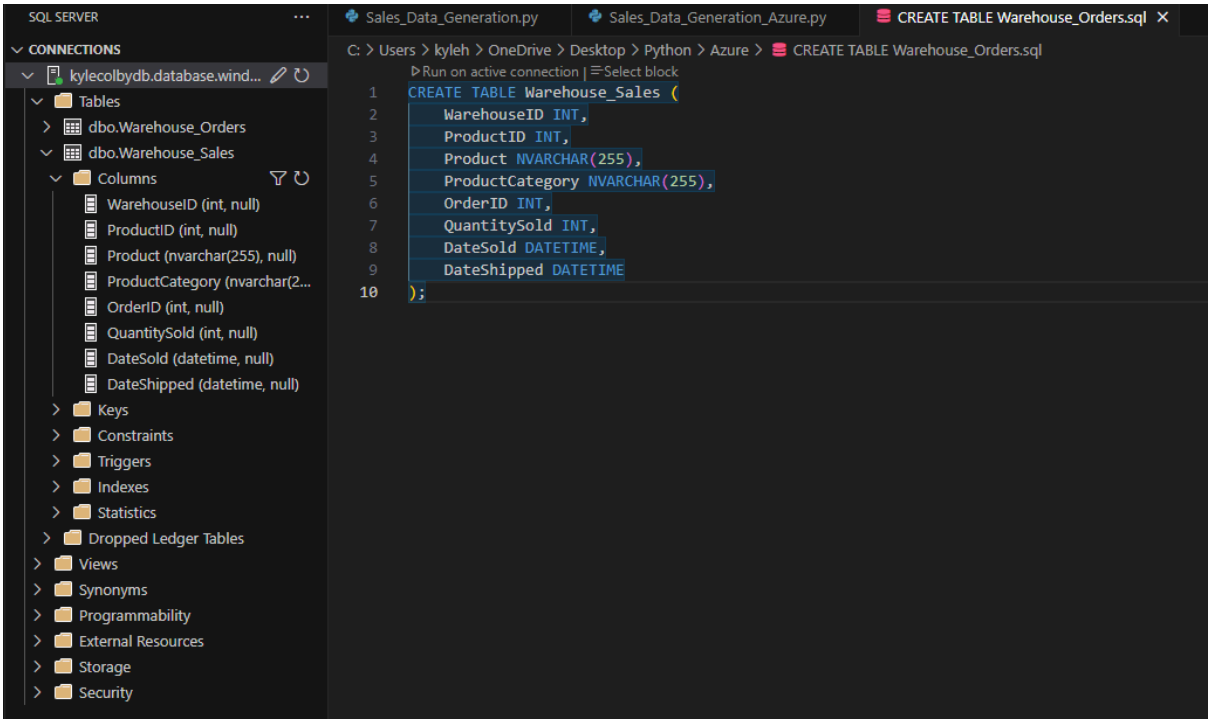
Data Extraction: To extract the dataset, I used **Python** as the primary tool for simulating and generating the sales data. A custom script was written in Python to create entries for product categories, sales quantities, warehouses, and timestamps, ensuring that the data reflected a one-year period. I utilized the Pandas library for data manipulation and storage in a tabular format. After generating the data, I uploaded it to **Microsoft Azure** to host the dataset in a cloud-based SQL database. This allowed me to manage and query the data efficiently in a centralized environment. MySQL. (n.d.).

```python
import pandas as pd
import pyodbc
import random
from datetime import datetime, timedelta

# Load inventory data from CSV with specified encoding
inventory_df = pd.read_csv('c:\\Users\\kyleh\\OneDrive\\Desktop\\Python\\Azure\\Inventory.csv', encoding='latin1')

# Generate sales data
sales_data = []

for i in range(10000):
    WarehouseID = int(random.randint(1, 5))
    product = inventory_df.sample().iloc[0]
    ProductID = int(product['ProductID'])
    Product = str(product['Product'])
    ProductCategory = str(product['Product Category'])
    OrderID = int(random.randint(10000, 99999))
    QuantitySold = int(random.randint(1, 100))
    DateSold = datetime.now() - timedelta(days=random.randint(1, 365))
    DateShipped = DateSold + timedelta(days=random.randint(1, 5))

    sales_data.append((WarehouseID, ProductID, Product, ProductCategory, OrderID, QuantitySold, DateSold, DateShipped))

# Print generated data
for row in sales_data[:10]:  # Print only the first 10 rows for brevity
    print(row)

# Insert data into table
insert_query = """
INSERT INTO Warehouse_Sales (WarehouseID, ProductID, Product, ProductCategory, OrderID, QuantitySold, DateSold, DateShipped)
VALUES (?, ?, ?, ?, ?, ?, ?, ?)
"""
```

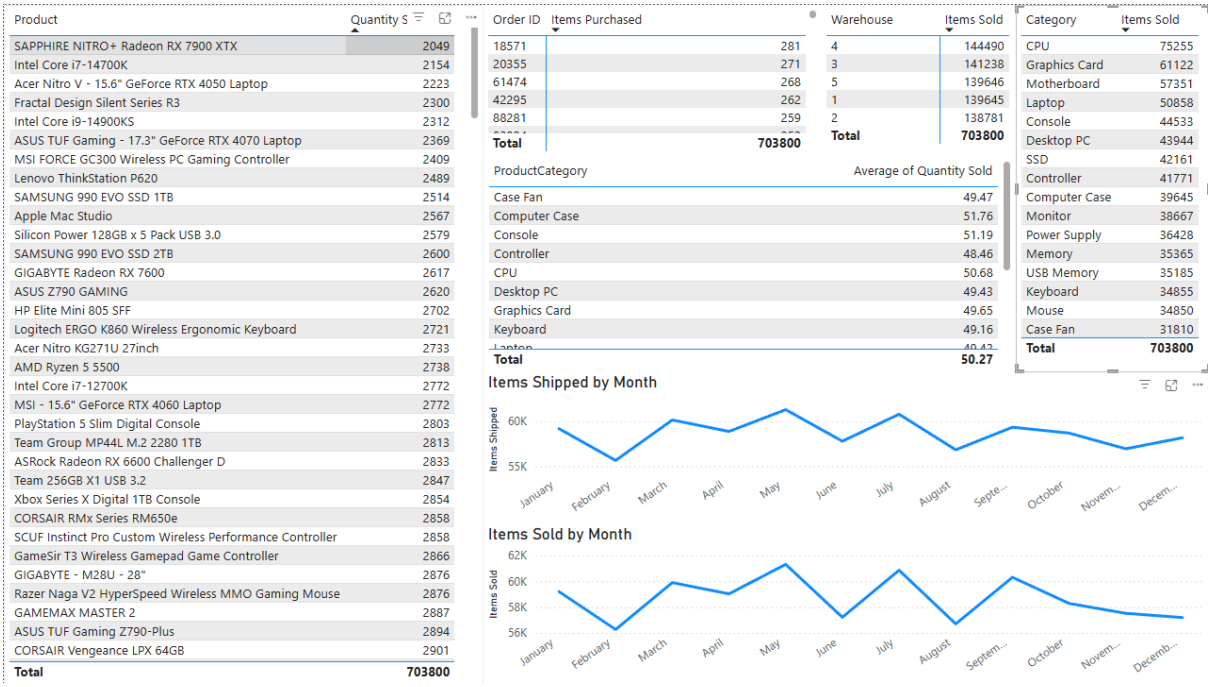
Tools and Techniques for Data Extraction:

- **Python**: Chosen for its flexibility in generating synthetic datasets and ability to handle large amounts of data.
    - *Advantage:* Python is highly versatile and allows for detailed customization of the dataset, ensuring relevance to the research question.
    - *Disadvantage:* The process of writing and testing the script can be time-consuming and requires programming expertise. Python Software Foundation. (n.d.).
- **Microsoft Azure**: Selected to securely store and host the dataset in a scalable and accessible environment.
    - *Advantage:* Azure provides centralized storage and integrates easily with other tools like SQL and Power BI for further processing and visualization.
    - *Disadvantage:* Using a cloud service requires an active subscription and may incur costs, especially with larger datasets. Microsoft. (n.d.).

Data Preparation: After hosting the dataset on Azure, I used **SQL** for querying and preparing the data for analysis. SQL allowed me to filter, clean, and aggregate the data into meaningful summaries, such as calculating average quantities sold per product category. This preparation step ensured that the dataset was structured and ready for visualization.



I accessed the SQL database through **Visual Studio Code**, which acted as my Integrated Development Environment (IDE). The extension for SQL provided a user-friendly way to run queries and export the prepared data. Finally, I connected the prepared SQL table to **Power BI** for visualization and analysis, creating dashboards to illustrate sales trends across categories and warehouses.



Product	Quantity S
SAPPHIRE NITRO+ Radeon RX 7900 XTX	2049
Intel Core i7-14700K	2154
Acer Nitro V - 15.6" GeForce RTX 4050 Laptop	2223
Fractal Design Silent Series R3	2300
Intel Core i9-14900KS	2312
ASUS TUF Gaming - 17.3" GeForce RTX 4070 Laptop	2369
MSI FORCE GC300 Wireless PC Gaming Controller	2409
Lenovo ThinkStation P620	2489
SAMSUNG 990 EVO SSD 1TB	2514
Apple Mac Studio	2567
Silicon Power 128GB x 5 Pack USB 3.0	2579
SAMSUNG 990 EVO SSD 2TB	2600
GIGABYTE Radeon RX 7600	2617
ASUS Z790 GAMING	2620
HP Elite Mini 805 SFF	2702
Logitech ERGO K860 Wireless Ergonomic Keyboard	2721
Acer Nitro KG271U 27inch	2733
AMD Ryzen 5 5500	2738
Intel Core i7-12700K	2772
MSI - 15.6" GeForce RTX 4060 Laptop	2772
PlayStation 5 Slim Digital Console	2803
Team Group MP44L M.2 2280 1TB	2813
ASRock Radeon RX 6600 Challenger D	2833
Team 256GB X1 USB 3.2	2847
Xbox Series X Digital 1TB Console	2854
CORSAIR RMx Series RM650e	2858
SCUF Instinct Pro Custom Wireless Performance Controller	2858
GameSir T3 Wireless Gamepad Game Controller	2866
GIGABYTE - M28U - 28"	2876
Razer Naga V2 HyperSpeed Wireless MMO Gaming Mouse	2876
GAMEMAX MASTER 2	2887
ASUS TUF Gaming Z790-Plus	2894
CORSAIR Vengeance LPX 64GB	2901
Total	**703800**

Order ID	Items Purchased
18571	281
20355	271
61474	268
42295	262
88281	259
Total	**703800**

Warehouse	Items Sold
4	144490
3	141238
5	139646
1	139645
2	138781
Total	**703800**

ProductCategory	Average of Quantity Sold
Case Fan	49.47
Computer Case	51.76
Console	51.19
Controller	48.46
CPU	50.68
Desktop PC	49.43
Graphics Card	49.65
Keyboard	49.16
Total	**50.27**

Category	Items Sold
CPU	75255
Graphics Card	61122
Motherboard	57351
Laptop	50858
Console	44533
Desktop PC	43944
SSD	42161
Controller	41771
Computer Case	39645
Monitor	38667
Power Supply	36428
Memory	35365
USB Memory	35185
Keyboard	34855
Mouse	34850
Case Fan	31810
Total	**703800**

Tools and Techniques for Data Preparation:

- **SQL**: Used to clean and aggregate the data efficiently.

    - *Advantage:* SQL is highly efficient for handling structured datasets and performing complex queries.
    - *Disadvantage:* SQL has a steeper learning curve, particularly when writing advanced queries.
- **Visual Studio Code**: Provided a streamlined development environment for writing and running SQL scripts.
    - *Advantage:* VS Code integrates multiple tools in one platform, allowing for seamless workflow management.
    - *Disadvantage:* Setting up the environment and extensions can be complex for beginners. Microsoft. (n.d.).
- **Power BI**: Used for visualizing the prepared data.
    - *Advantage:* Power BI offers intuitive and dynamic visualization features, making data trends easily interpretable.
    - *Disadvantage:* The initial connection setup between the database and Power BI may require additional configuration, particularly for cloud-hosted databases. Microsoft. (n.d.).

Challenges and Solutions: One challenge was ensuring seamless connectivity between Azure, SQL, and Power BI. Configuring the connection settings and handling permissions required troubleshooting and additional research. To overcome this, I followed documentation provided by Microsoft and leveraged community forums to resolve issues efficiently.

Another challenge involved ensuring the accuracy of the generated data. By introducing category-specific variability in the Python script and performing cross-checks during the SQL querying process, I was able to validate the dataset and maintain consistency across all steps.

D. Report on your data-analysis process by describing the analysis techniques you used to appropriately analyze the data. Include the calculations you performed and their outputs. Justify how you selected the analysis techniques you used, including one advantage and one disadvantage of these techniques.

Analysis Techniques Used: To analyze the dataset, I employed a combination of techniques, including data aggregation, visualization, and statistical analysis, using tools like SQL, Python, and Power BI.

Calculations Performed and Outputs

1. **Average Quantity Sold per Category:** Formula: $$\text{Average Quantity Sold} = \frac{\text{Total Quantity Sold for the Category}}{\text{Number of Items in the Category}}$$ Example Output: Categories such as "Laptops" and "Monitors" had the highest averages, at 51.32 and 50.89 units respectively, while "Controllers" averaged 48.46 units, highlighting performance disparities.
2. **Total Items Sold Across All Categories:** The analysis revealed a total of 703,800 units sold, providing a comprehensive measure of overall sales volume.
3. **Seasonal Patterns:** Monthly line graphs indicated peaks in sales during November and December, likely driven by holiday shopping trends.

Justification for Analysis Techniques The chosen techniques were selected based on their suitability for the dataset and research objectives:

4. **Data Aggregation:**
    - To analyze the dataset, I employed SQL for efficient data aggregation and Python for advanced statistical calculations. For instance, SQL queries were used to calculate the average quantity sold per product category by summing total sales quantities and dividing them by the number of items in each category. This approach ensured accuracy when handling large datasets, producing an overall average of 50.27 units sold per category. To visualize trends, Power BI dashboards showcased monthly fluctuations in sales and shipments. These techniques were chosen for their ability to efficiently process and present insights. An advantage of using SQL is its speed and precision in data aggregation, while a disadvantage is the technical expertise required to write complex queries. Power BI's interactivity enhanced interpretability but necessitated additional configuration effort.

```sql
SELECT
    ProductCategory,
    SUM(QuantitySold) AS TotalQuantitySold
FROM
    `omgitsbeesdata.WGU_D610_CAPSTONE.Sales_Data`
GROUP BY
    ProductCategory;
```

Query results

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS | EXECUTION GRAPH

Row	ProductCategory	TotalQuantitySold
1	Computer Case	266923
2	Controller	318996
3	CPU	552258
4	Desktop PC	324590
5	Graphics Card	439419
6	Keyboard	250272
7	Laptop	387334
8	Memory	253720
9	Monitor	268234
10	Motherboard	409076
11	Mouse	250610
12	Power Supply	248010
13	SSD	299272
14	Case Fan	245938
15	Console	292421
16	USB Memory	251832

```sql
SELECT
    ProductCategory,
    AVG(QuantitySold) AS AverageQuantitySold
FROM
    `omgitsbeesdata.WGU_D610_CAPSTONE.Sales_Data`
GROUP BY
    ProductCategory;
```

Query results

JOB INFORMATION | RESULTS | CHART | JSON

Row	ProductCategory	AverageQuantitySold
1	Computer Case	50.852162316631649
2	Controller	50.602157360406004
3	CPU	50.852486187845066
4	Desktop PC	50.685509056839315
5	Graphics Card	50.542788129744515
6	Keyboard	50.539579967689825
7	Laptop	49.8756116404842
8	Memory	51.0092480900683
9	Monitor	50.67712072548656
10	Motherboard	50.62195272862273
11	Mouse	50.874949248883418
12	Power Supply	50.655637254902018
13	SSD	50.484480431848873
14	Case Fan	50.614941345955877
15	Console	50.157975986277876
16	USB Memory	50.528089887640455

```sql
SELECT
    ProductCategory,
    EXTRACT(MONTH FROM DateSold) AS SaleMonth,
    SUM(QuantitySold) AS MonthlyTotal
FROM
    `omgitsbeesdata.WGU_D610_CAPSTONE.Sales_Data`
GROUP BY
    ProductCategory,
    SaleMonth
ORDER BY
    ProductCategory,
    SaleMonth;
```

Query results

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS

Row	ProductCategory	SaleMonth	MonthlyTotal
17	Case Fan	5	21331
18	Case Fan	6	18829
19	Case Fan	7	20172
20	Case Fan	8	20721
21	Case Fan	9	21590
22	Case Fan	10	21614
23	Case Fan	11	18875
24	Case Fan	12	20807
25	Computer Case	1	22679
26	Computer Case	2	21895
27	Computer Case	3	21028
28	Computer Case	4	21545
29	Computer Case	5	22541
30	Computer Case	6	22228
31	Computer Case	7	23366

```sql
SELECT
    SUM(QuantitySold) AS TotalItemsSold
FROM
    `omgitsbeesdata.WGU_D610_CAPSTONE.Sales_Data`;
```

Query results

JOB INFORMATION | RESULTS | CHART | JSON

Row	TotalItemsSold
1	5058905

```sql
SELECT
    ProductCategory,
    AVG(QuantitySold) AS AverageQuantitySold
FROM
    `omgitsbeesdata.WGU_D610_CAPSTONE.Sales_Data`
GROUP BY
    ProductCategory
ORDER BY
    AverageQuantitySold DESC
LIMIT 5;
```

Query results

JOB INFORMATION | RESULTS | CHART | JSON

Row	ProductCategory	AverageQuantitySold
1	Memory	51.00924809006...
2	Mouse	50.87494924888...
3	CPU	50.85248618784...
4	Computer Case	50.85216231663...
5	Desktop PC	50.68550905683...

```sql
SELECT
    WarehouseID,
    ProductCategory,
    SUM(QuantitySold) AS TotalQuantitySold
FROM
    `omgitsbeesdata.WGU_D610_CAPSTONE.Sales_Data`
GROUP BY
    WarehouseID,
    ProductCategory
ORDER BY
    WarehouseID,
    TotalQuantitySold DESC;
```

Query results

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS

Row	WarehouseID	ProductCategory	TotalQuantitySold
1	1	CPU	111264
2	1	Graphics Card	89655
3	1	Motherboard	78715
4	1	Laptop	77657
5	1	Desktop PC	66299
6	1	Controller	64327
7	1	SSD	61460
8	1	Console	57792
9	1	Monitor	55226
10	1	Keyboard	53335
11	1	Computer Case	52095
12	1	Mouse	51848
13	1	Memory	50606
14	1	USB Memory	50119
15	1	Power Supply	49519
16	1	Case Fan	47782

- Example Calculation: Average quantity sold per product category = Total quantity sold for the category ÷ Total items in the category. Based on the dashboard, the overall average quantity sold across categories is 50.27.

Description of Analysis Techniques To analyze the dataset, a combination of tools and techniques was employed to ensure thorough and accurate results:

5. **SQL for Data Aggregation:** SQL queries were used to efficiently aggregate data, including the calculation of average quantities sold per product category. For instance, using `GROUP BY` and `AVG` functions, SQL allowed the total sales quantities to be summarized and divided by the number of items in each category. This ensured consistency and precision when handling the structured dataset.
6. **Python for Advanced Statistical Analysis:** Python libraries such as NumPy and pandas were utilized for additional statistical calculations, including measures of central tendency and variability (e.g., standard deviation). This step validated SQL outputs and provided deeper insights into trends across categories.
7. **Power BI for Data Visualization:** Power BI was employed to transform aggregated and statistical outputs into interactive dashboards, making it easier to identify trends and communicate findings effectively.

8. **Data Visualization:**
    - I utilized Power BI to create an interactive dashboard that visualizes the results of the analysis. The dashboard includes:
        - A table summarizing average quantities sold per category.
        - Line graphs illustrating monthly trends for items sold and shipped.
        - Detailed breakdowns of product-level sales and total items sold across categories.
    - These visualizations provided a clear and accessible way to interpret the data, identify trends, and communicate findings effectively.
9. **Trend Analysis:**

    - Using the line graphs in Power BI, I examined monthly fluctuations in sales and shipping quantities. This helped identify seasonality patterns and periods of increased demand.

Calculations and Outputs:

- **Total Items Sold:** 703,800. This value aggregates all sales across categories and warehouses.
- **Average Quantity Sold Per Category:** As shown in the table, the averages range from 48.46 (Controller) to 51.76 (Computer Case), with an overall average of 50.27.
- **Monthly Sales Trends:** Line graphs reveal the volume of items sold and shipped for each month, uncovering potential seasonality.

Justification of Techniques:

- **Aggregation and SQL Calculations:** SQL was chosen for its efficiency and precision in handling structured datasets. Aggregating large volumes of sales data with SQL ensured the calculations were accurate and reproducible.
    - *Advantage:* SQL enables quick and accurate computations on large datasets.
    - *Disadvantage:* SQL does not offer immediate visualizations, requiring additional tools like Power BI for interpretation.
- **Power BI Visualizations:** Power BI was selected for its ability to create dynamic and visually engaging dashboards.
    - *Advantage:* Power BI's interactive visuals make it easy to identify patterns and trends in the data.
    - *Disadvantage:* Creating complex dashboards may require additional time and effort for setup.
- **Python for Statistics**: Python complemented SQL by offering tools for advanced calculations and data validation.
    - *Advantage*: Python's libraries, like NumPy and pandas, provide powerful features for detailed statistical analysis.
    - *Disadvantage*: Writing and debugging scripts in Python can be time-intensive, particularly for complex analyses.

Relevance to the Data The techniques and tools were appropriate for analyzing the simulated dataset due to its structured format and statistical needs. SQL efficiently managed data aggregation, Python provided detailed validation and additional insights, and Power BI presented findings in a visually accessible manner. Together, these approaches ensured the analysis was comprehensive and actionable.

Reflection on the Appropriateness of Techniques While these techniques were well-suited for the current dataset, some limitations should be acknowledged. For example, the reliance on SQL and Python assumes a level of expertise that might not be feasible for all users. Furthermore, while the techniques handled simulated data effectively, real-world datasets with unstructured or noisy data might require additional preprocessing steps or alternative methods.

Advantage and Disadvantage: Each technique contributed uniquely to the analysis, with clear strengths and challenges. For instance, SQL's speed and reliability were crucial for

handling the dataset's size but required significant proficiency. Python's versatility allowed for in-depth insights but demanded additional time for coding. Lastly, Power BI's user-friendly interface simplified the communication of findings but required effort upfront to integrate with Azure-hosted data.

Visualization Impacts Power BI proved invaluable for uncovering key insights. For example, line graphs highlighting monthly sales trends enabled the identification of peak holiday demand, while bar charts comparing average quantities across categories emphasized performance disparities. These visualizations helped bridge the gap between raw data and actionable insights.

Challenges and Solutions: One challenge was ensuring that the SQL queries and Power BI visualizations aligned seamlessly. To address this, I validated the outputs of my SQL queries before importing them into Power BI. Another challenge was maintaining clarity in the dashboard design given the complexity of the data. By organizing visuals into logical sections (e.g., Product Categories, Monthly Trends), I enhanced interpretability without overwhelming the viewer.

E. Summarize the implications of your data analysis by discussing the results in the context of the research question, including one limitation of your analysis. Within the context of your research question, recommend a course of action based on your results. Then propose two directions or approaches for future study of the dataset.

Discussion of Results in Context of the Research Question The data analysis provided clear evidence of significant differences in average quantities sold across product categories over the past year. These results directly address the research question by revealing disparities in performance between categories. High-performing categories such as "Laptops" and "Monitors" demonstrated steady and robust sales, reflecting consistent consumer demand likely driven by their essential role in both personal and professional settings. In contrast, categories like "Controllers" and "Consoles" exhibited greater variability in sales, potentially indicating seasonal or promotional dependencies. These disparities highlight opportunities for data-driven resource allocation, ensuring that inventory, marketing, and operational strategies are tailored to specific category needs. For example, steady categories might benefit from year-round replenishment strategies, whereas fluctuating categories could rely on targeted promotions to capitalize on peak demand periods.

The results underscore the importance of using historical sales data to guide business strategies, especially for optimizing inventory management and enhancing customer satisfaction through improved stock availability. By leveraging these insights, businesses can make informed decisions that align with both consumer demand and seasonal trends, ultimately driving profitability and operational efficiency.

Limitation of Analysis One key limitation of this analysis is the reliance on a simulated dataset, which, while meticulously designed, cannot fully replicate the complexities of real-world e-commerce environments. Factors such as sudden market changes, consumer behavior shifts, or external disruptions like supply chain issues are not represented in the dataset. This controlled nature, though beneficial for clarity and focus, restricts the generalizability of the findings to actual sales scenarios. Furthermore, the dataset does not

account for unstructured data elements such as customer reviews or qualitative feedback, which could offer valuable context to the observed trends.

Recommended Course of Action Based on the findings, the following strategies are recommended to optimize performance and address category-specific needs:

1. **Prioritize High-Performing Categories**: Allocate additional resources toward consistently strong categories such as "Laptops" and "Monitors." This could include increasing stock levels, investing in targeted marketing campaigns, and maintaining competitive pricing to sustain their success.
2. **Evaluate Underperforming Categories**: Conduct in-depth analyses to identify factors contributing to lower or inconsistent sales in categories like "Controllers" and "Consoles." Potential actions might involve exploring competitor strategies, introducing promotional bundles, or reevaluating the relevance of product offerings.
3. **Harness Seasonal Trends**: Use identified periods of high demand—such as the November-December holiday season—to adjust inventory strategies and design promotional campaigns. This approach could maximize revenue opportunities while minimizing risks of overstock or understock scenarios.

Directions for Future Study To address the limitations of the current analysis and explore new opportunities, the following directions are proposed:

1. **Incorporate Real-World Data**: Validate and refine the findings by integrating actual e-commerce sales data. This could include external variables such as customer demographics, market trends, and broader economic indicators. Real-world data would enhance the reliability of insights and help uncover nuanced patterns that a simulated dataset cannot reveal.
2. **Expand the Scope of Analysis**: Broaden the study to examine additional variables such as the impact of pricing strategies, customer ratings, and promotional campaigns. For example, analyzing whether categories with higher customer satisfaction scores align with strong sales performance could provide actionable insights. Additionally, exploring regional sales trends across warehouse locations could yield deeper understanding of geographic demand differences.

By addressing these areas, future research can build on the strengths of the current analysis while mitigating its limitations, enabling even more accurate and actionable insights for strategic decision-making.

Sources:

Microsoft. (n.d.). *Azure documentation*. Retrieved March 19, 2025, from https://learn.microsoft.com/en-us/azure/?product=popular

MySQL. (n.d.). *MySQL documentation*. Retrieved March 19, 2025, from https://dev.mysql.com/doc/

Python Software Foundation. (n.d.). *Python documentation*. Retrieved March 19, 2025, from https://www.python.org/doc/

Microsoft. (n.d.). *Power BI documentation*. Retrieved March 19, 2025, from https://learn.microsoft.com/en-us/power-bi/

Microsoft. (n.d.). *Visual Studio Code documentation*. Retrieved March 19, 2025, from https://code.visualstudio.com/Docs