EDA questions

🐍 PYTHON — FINAL 15 QUESTIONS (MASTER SET)

(Assume `Age_Bucket` and `City_Tier` already exist)

🔹 CUSTOMER SEGMENTATION (Core Understanding)

1. How are customers distributed across **Age_Bucket** and **City_Tier**, and which segments contribute the most revenue?
2. Compare **Average Order Value (AOV)** across Age_Buckets and City_Tiers.
3. Which Age_Bucket has the **highest percentage of returning customers**?
4. How do **session duration and pages viewed** differ across Age_Buckets?
5. Which **Product Categories** are preferred by different Age_Buckets?

🔹 SALES FORECASTING (TIME SERIES — No ML)

6. Convert order dates into a time series and analyze the **overall revenue trend** over time.
7. Aggregate sales monthly and identify any **seasonality or growth patterns**.
8. Plot a **rolling average of revenue** and interpret short-term future sales direction.
9. Compare revenue trends across **City_Tiers** over time.

🔹 MARKETING CAMPAIGN OPTIMIZATION

(Hypothesis Test #1)

10. Compare **AOV and quantity purchased** for discounted vs non-discounted orders.

11. Are discounts more effective for **new customers or returning customers**?

12. **Hypothesis Test:**
    Do discounted orders lead to a **significant increase in quantity purchased**?

🔹 PRICE OPTIMIZATION

(Hypothesis Test #3)

13. Analyze the relationship between **Unit Price and Quantity** across Product Categories.

14. **Hypothesis Test:**
    Does **Age_Bucket** have a significant impact on customer spending behavior?

🔹 DELIVERY & CUSTOMER EXPERIENCE

(Hypothesis Test #2)

15. **Hypothesis Test:**
    Does **delivery time significantly affect customer ratings**, and how strong is this relationship?

SQL insights

🔹 SQL — FINAL 20 MASTER QUESTIONS

1. For each customer, calculate their **total lifetime revenue**, total orders, and average order value.

2. Identify the **top 10% highest-value customers** based on lifetime spend.

3. Calculate **recency** for each customer as the number of days since their last order.

4. Classify customers into **New, Active, and Inactive** segments based on recency and order frequency.

5. Calculate the **retention rate** of customers who placed at least one repeat order.

6. For each customer's **first purchase month**, calculate how many customers returned in subsequent months (**cohort analysis**).

7. Compute **month-over-month revenue growth** using window functions.

8. Calculate a **rolling 3-month average of total revenue**.

9. Identify customers who made **only one purchase** and never returned.

10. For each customer, calculate the **time gap between consecutive orders** and identify customers with increasing gaps.

11. Recommend the **top product category** for each customer based on their historical purchases.

12. Identify **pairs of product categories** frequently purchased by the same customers.

13. For each city, calculate the **average delivery time** and the percentage of late deliveries.

14. Analyze the relationship between **delivery time and customer ratings** at a city level.

15. Identify cities with **high order volume but low average ratings**.

16. Calculate the **churn rate** based on customers inactive for more than a defined period.

17. Compare **CLV of returning vs non-returning customers**.

18. Rank customers within each city based on **lifetime value**.

19. Identify customers whose **spending trend is declining** over their recent orders.

20. Calculate the **share of total revenue contributed by the top 20% of customers**.