

# Retail Sales EDA – Complete Question Bank

## Section 1 — Data Understanding & Cleaning

- 1 How many rows and columns are present in the dataset?
- 2 What percentage of missing values exists in each column?
- 3 Which columns have the highest data quality issues?
- 4 How many duplicate order IDs are present?
- 5 How many invalid or negative quantities exist?
- 6 How many records have missing or incorrect dates?
- 7 List all inconsistent city name variations.
- 8 Which rows contain extreme price outliers?
- 9 How many rows have missing revenue values?
- 10 After cleaning, how many valid transactions remain?

## Section 2 — Descriptive EDA

- 1 What is the total revenue generated?
- 2 What is the average order value?
- 3 Which city generates the highest revenue?
- 4 Which region performs best overall?
- 5 What are the top 10 selling products by revenue?
- 6 Which product category contributes most to revenue?
- 7 What is the least performing category?
- 8 What is the median quantity per order?
- 9 Which payment method is used most frequently?
- 10 What percentage of sales comes from online vs in-store?

## Section 3 — Time - based Analysis

- 1 What is the monthly revenue trend?
- 2 Which month has the highest sales?
- 3 Which month has the lowest sales?
- 4 Is there any seasonality pattern in sales?
- 5 How does average order value change month - to - month?
- 6 Are weekend sales higher than weekday sales?
- 7 Which quarter contributes most revenue?

## Section 4 — Customer Behavior

- 1 Do returning customers spend more than new customers?
- 2 Which customer type generates more revenue share?
- 3 What is the average quantity per customer type?
- 4 Which city has the highest returning - customer ratio?
- 5 Are discounts more common for new customers?

## Section 5 — Pricing & Discount Insights

- 1 What is the average discount percentage overall?
- 2 Which category gives the highest discounts?
- 3 Do higher discounts increase revenue?
- 4 Find correlation between discount, price, and quantity.
- 5 Identify loss - making or suspicious transactions.

## Section 6 — Geographic Insights

- 1 Which top 5 cities contribute most revenue?
- 2 Which region has lowest performance?
- 3 Category preference differences across regions.
- 4 Which city buys most electronics?
- 5 Which city prefers fashion products?

## Section 7 — Outliers & Anomaly Detection

- 1 Detect price outliers using IQR or Z - score.
- 2 Detect abnormally high revenue orders.
- 3 Are outliers data errors or real premium sales?
- 4 Impact of removing outliers on average revenue.

## Section 8 — Visualization Tasks

- 1 Create monthly revenue line chart.
- 2 Create top products bar chart.
- 3 Create category revenue pie chart.
- 4 Create heatmap of correlations.
- 5 Build city vs category pivot visualization.

## Section 9 — Business Insight Questions

- 1 What are the top 5 actionable business insights?
- 2 Which city should receive more marketing budget?
- 3 Which category should the company expand?
- 4 Where is the biggest revenue leakage happening?
- 5 Should the company increase or reduce discounts?
- 6 Which sales channel should be prioritized?
- 7 What data quality improvements should the company implement?