

Exploratory Data Analysis (EDA) - Case Study Questions (Set 2)

Case Study 1: Online Retail Sales Analysis (CSV Dataset)

Dataset Name: Online_Retail_Sales.csv

Columns:

Order_ID, Order_Date, Customer_ID, Region, Product_Category, Quantity, Unit_Price, Total_Sales, Payment_Method, Returned (Yes/No)

Problem Statement:

You are hired as a Data Analyst for an e-commerce company.
Perform complete EDA before building a sales prediction model.

Tasks:

1. Load dataset using Pandas.
2. Identify rows, columns, and data types.
3. Identify target variable (Total_Sales).
4. Handle missing values and duplicates.
5. Perform distribution analysis for Quantity and Total_Sales.
6. Detect outliers using IQR method.
7. Perform categorical analysis (Region, Payment_Method vs Sales).
8. Extract Month and Year from Order_Date (Feature Engineering).
9. Analyze correlation between numerical variables.
10. Write business insights and recommendations.

Case Study 2: Student Academic Performance (SQL Dataset)

Table Name: student_performance

Columns:

student_id, age, gender, study_hours, attendance_percentage, assignment_score, exam_score, final_grade, passed (Yes/No)

Problem Statement:

You are working as a Data Analyst in an educational institution.
Perform EDA before building a student performance prediction model.

Tasks:

1. Create SQLite database and insert sample records.
2. Load data using `pd.read_sql()`.
3. Identify numerical and categorical variables.
4. Detect missing values and handle them.
5. Perform distribution analysis for `study_hours` and `exam_score`.
6. Detect outliers using Boxplot and IQR method.
7. Analyze relationship between `study_hours` and `exam_score`.
8. Perform correlation heatmap.
9. Check class imbalance in `passed` variable.
10. Provide academic improvement recommendations.