

EDA Practice Case Studies - Set 2

Case Study 1: URL Dataset (Wine Quality Dataset)

Dataset URL:

https://raw.githubusercontent.com/mwaskom/seaborn-data/master/wine_quality.csv

Problem Statement:

You are a data analyst analyzing wine quality data to understand what factors influence wine quality ratings.

Tasks:

1. Load dataset from URL using Pandas.
2. Identify numerical and categorical variables.
3. Check missing values and treat them.
4. Perform distribution analysis (alcohol, pH, sulphates).
5. Detect outliers using IQR method.
6. Create correlation heatmap.
7. Analyze relationship between alcohol and quality.
8. Provide key insights before modeling.

Case Study 2: SQL Dataset (Sales Analysis)

Table Name: sales

Columns:

order_id, customer_name, region, product_category,
sales_amount, quantity, order_date, profit

Problem Statement:

You are hired to analyze sales performance and profitability.

Tasks:

1. Create SQLite database and insert records.
2. Load data using `pd.read_sql()`.
3. Identify target variable (profit).
4. Check missing values and duplicates.
5. Perform distribution analysis (sales_amount, profit).
6. Detect outliers using IQR.
7. Analyze region-wise and category-wise profit.
8. Provide business recommendations.