

Assignment 3. Data Analysis.

You have the following datasets:

Dataset 1: Sales Data

Structure:

Columns: OrderID, ProductID, Quantity, PricePerUnit, OrderDate, CustomerID, Region

Dataset 2: Customer Feedback

Structure:

Columns: CustomerID, Feedback, Rating

Description:

This dataset contains customer feedback and ratings for products and services.

Dataset 3: Employee Performance

Structure:

Columns: EmployeeID, Department, Salary, PerformanceScore

Description:

This dataset contains information about employees, including their departments, salaries, and performance scores.

Dataset 4: Product Inventory

Structure:

Columns: ProductID, ProductName, StockQuantity, ReorderLevel

Description:

This dataset contains information about products, including their names, current stock quantities, and reorder levels.

Dataset 5: Customer Demographics

Structure:

Columns: CustomerID, Age, Gender, Income

Description:

This dataset contains demographic information about customers, including their age, gender, and income.

FOR EACH DATASET SIMULATE AT LEAST 3 ROWS.

Exercise 1: Hypothesis Testing

For the Sales Data dataset (Dataset 1), perform a hypothesis test to determine if there is a statistically significant difference in average product prices (PricePerUnit) between different regions (Region). Use a suitable statistical test, such as ANOVA or a t-test, and interpret the results.

Exercise 2: Probability Distributions

Calculate the probability distribution of customer ratings (Rating) in the Customer Feedback dataset (Dataset 2).

Exercise 3: Exploratory Data Analysis (EDA)

For the Employee Performance dataset (Dataset 3):

Conduct an EDA to identify outliers in the Salary column using appropriate visualization techniques.

Calculate the interquartile range (IQR) for salaries in each department and determine if there are significant differences.

Exercise 4: Data Manipulation and Joins

Combine the Sales Data dataset (Dataset 1) and the Product Inventory dataset (Dataset 4) using an appropriate join operation to create a new dataset that includes product sales and inventory information. Calculate the total revenue for each product sold and the profit margin based on the price per unit and cost per unit (assume a fixed cost per unit).

Exercise 5: Pivoting and Aggregation

Using the comprehensive customer-related dataset created in the previous exercise:

Pivot the data to create a table that shows the total sales revenue by customer and product category.

Calculate the average income by age group and gender from the Customer Demographics dataset (Dataset 5).