

Refresher Course on

Data Science

Assignment : 5

1. **Dataset:** product-info.csv containing columns like Product ID, Product Name, Price, Quantity Sold, and Sales Date.

Problem: Write a Python program to calculate the total sales, average sales price, and the standard deviation of sales prices. Additionally, identify the product with the highest and lowest average sales price.

2. **Dataset:** A CSV file customer-satisfaction.csv with columns such as Customer ID, Satisfaction Rating (scale of 1-10), Age, and Gender.

Problem: Create a Python program to calculate the mean, median, and mode of the satisfaction ratings. Analyze the data by grouping it by age groups and gender to identify any trends.

3. **Dataset:** A CSV file house-info.csv containing columns such as House ID, Location, Number of Bedrooms, Square Footage, and Price.

Problem: Develop a Python program to calculate the first, second (median), and third quartiles of house prices. Also, compute the interquartile range (IQR) and identify any outliers in the dataset.

4. **Dataset:** A CSV file visitor-info.csv with columns such as Date, Number of Visitors, Average Session Duration, Bounce Rate, and Conversion Rate.

Problem: Create a Python program to analyze the average, minimum, and maximum number of visitors per day. Investigate the relationship between bounce rate and conversion rate using correlation analysis.

5. **Dataset:** A dataset with columns (purchase-info.csv) such as Customer ID, Purchase Amount, Purchase Date, and Category.

Problem: Develop a Python program to calculate the average purchase amount for each category. Also, find the most frequent purchase category and the customer who made the most purchases.

Dataset Location: <https://mitu.co.in/dataset>