COM 624 WEEK 2 LAB

Lab Guide: Descriptive Data Analysis for AI Using Python

Learning Outcomes

By the end of this lab, students will:

- Apply Python to perform descriptive data analysis.

- Calculate key statistical measures (mean, median, mode, standard deviation).

- Visualise data using Matplotlib and Seaborn.

- Interpret insights from a real-world dataset.

- Write a reflective report on findings and learning.

Lab Setup

Tools Required

- Python 3.x (Anaconda or standard installation)

- Jupyter Notebook or VS Code

- Libraries: pandas, numpy, matplotlib, seaborn, statistics, scipy

Dataset

Use the dataset here. It contains 700 samples with features like:

Download the Dataset here

- User ID

- Screen-On Time (hours/day)

- App Usage Time (minutes/day)

- Battery Drain (mAh/day)

- Number of Apps Installed

- Operating System (Android/iOS)

- Age, Gender, and User Behavior Class

Step-by-Step Instructions

- ◆ **Step 1: Import Libraries**
- ◆ **Step 2: Load and Clean Dataset**
- ◆ **Step 3: Descriptive Statistics**
- ◆ **Step 4: Visualisations**
- ◆ **Step 5: Grouped Analysis**
- ◆ **Step 6: Summary Table**
- ◆ **Step 7: Report Writing**

Example Python Code

```python
# Step 1: Import Libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import statistics
from scipy import stats


# Step 2: Load Dataset
df = pd.read_csv("mobile_device_usage.csv")  # Rename as needed
df.dropna(inplace=True)
df = df[df['Screen_On_Time'] > 0]


# Step 3: Descriptive Statistics
mean_screen = df['Screen_On_Time'].mean()
median_screen = df['Screen_On_Time'].median()
mode_screen = df['Screen_On_Time'].mode()[0]
std_screen = df['Screen_On_Time'].std()
range_screen = df['Screen_On_Time'].max() - df['Screen_On_Time'].min()
```

```python
print(f"Mean: {mean_screen:.2f}, Median: {median_screen}, Mode: {mode_screen}")

print(f"Standard Deviation: {std_screen:.2f}, Range: {range_screen}")


# Step 4: Visualisations

plt.figure(figsize=(8,5))

sns.histplot(df['Screen_On_Time'], bins=20, kde=True)

plt.title("Distribution of Screen-On Time")

plt.xlabel("Hours per Day")

plt.ylabel("Frequency")

plt.show()


plt.figure(figsize=(8,5))

sns.boxplot(x=df['App_Usage_Time'])

plt.title("Boxplot of App Usage Time")

plt.xlabel("Minutes per Day")

plt.show()


plt.figure(figsize=(10,6))

sns.heatmap(df.corr(), annot=True, cmap='coolwarm')

plt.title("Correlation Between Variables")

plt.show()


# Step 5: Grouped Analysis

os_group = df.groupby('Operating_System')['Screen_On_Time'].mean()

os_group.plot(kind='bar', color=['skyblue', 'salmon'])

plt.title("Average Screen-On Time by Operating System")

plt.ylabel("Hours per Day")

plt.show()


# Step 6: Summary Table
```

```
print(df.describe())
```

Student Report Template

Title: Descriptive Data Analysis of Mobile Device Usage

1. Introduction

Explain the purpose of the lab and its relevance to AI.

2. Methodology

Describe the dataset, tools used, and steps followed.

3. Results

Include:

- Summary statistics

- Graphs and visualisations

- Observations (e.g., which OS users spend more time on screens?)

4. Discussion

Interpret findings:

- What patterns emerged?

- Any surprising insights?

- How could this data inform AI-driven recommendations?

5. Reflection

What did you learn about data analysis and its role in AI?

6. Conclusion

Summarise key takeaways and future applications.