School: SOET Campus: Vzm

Academic Year: 2021/2025 Subject Name: DAVP Subject Code: Csbm1018

Semester: 7th Program: B.TECH Branch: ECE Specialization: ECE

Date:

Applied and Action Learning

(Learning by Doing and Discovery)

Name of the Experiement : Read a dataset containing sales data from csv file. perform data cleaning.

Coding Phase: Pseudo Code / Flow Chart / Algorithm

→ load the dataset from the csv file and inspect the first few rows.

→ clean the dataset by removing rows with missing values, dropping, duplicate rows, and ensuring columns have the correct data types.

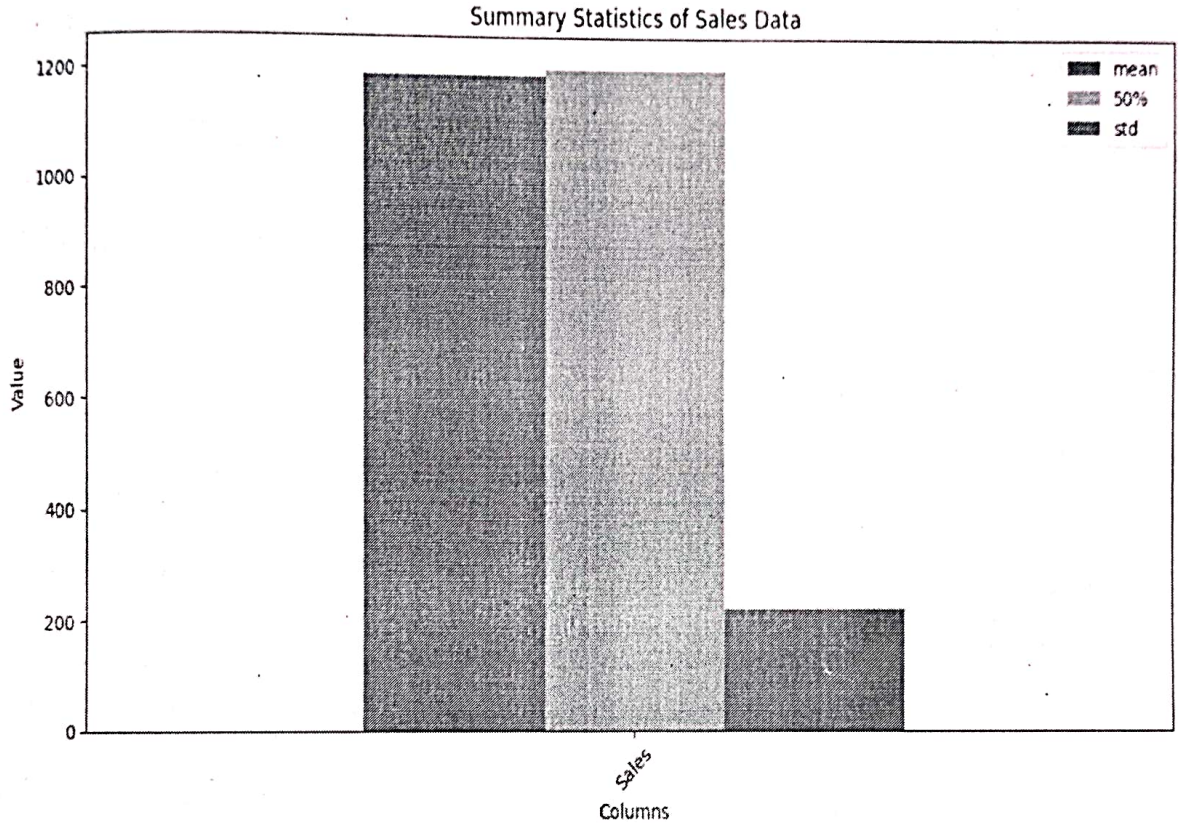
→ generate summary statistic like the mean, median, standard deviation for the cleaned dataset.

→ Visualize the summary statistics using a bar chart to display the mean, median (50%).

→ Display the cleaned dataset to verify the changes after data cleaning

Testing Phase: Compilation of Code (error detection)

Implementation Phase: Final Output (no error)



ASSESSMENT

Rubrics	Full Mark	Marks Obtained	Remarks
Concept	10		
Planning and Execution/ Practical Simulation/ Programming	10		
Result and Interpretation	10		
Record of Applied and Action Learning	10		
Viva	10		
Total	**50**		

Signature of the Student: G. Sumanth

Name : G. Sumanth

Regn. No. : 211801131001

Signature of the Faculty:

```
Import pandas as pd
Import Matplotlib.pyplot as plt
df=pd.read_csv('Sales_data.csv')
Print(df.head())
df_cleaned = df.dropna()
df_cleaned = df_cleaned.drop_duplicates()
df_cleaned['Sales'] = pd.to_numeric(df_cleaned['sales'], errors
Summary_stats = df_cleaned.describe()
Summary_stats.T[['mean', '50%', 'std']].plot(kind='bar', figsize=
                                              (10,6))
plt.title('Summary Statistics of Sales Data')
plt.ylabel ('value')
plt.xlabel ('columns')
plt.xticks (rotation=45)
plt.tight_layout()
plt.show()
Print (df_cleaned.head())
```