KGiSL Institute Of Technology

NAAN MUDHALVAN

Project Title:

Project Sales Analysis

Team Members:

1.Ashika.S

2.Dharshana.V

3.Aswitha.M

4.Dinesh Vishnu.S

PROJECT DESCRIPTION:

PHASE-3 : Analysing On The Product-sales DataSet

OBJECTIVE:

STEPS:

IN GOOGLE COLAB NOTEBOOK:

- Mount the GoogleDrive
- Loading the Dataset to GoogleDrive
- Processing and cleansing the dataset
- Accuracy

#Mount:

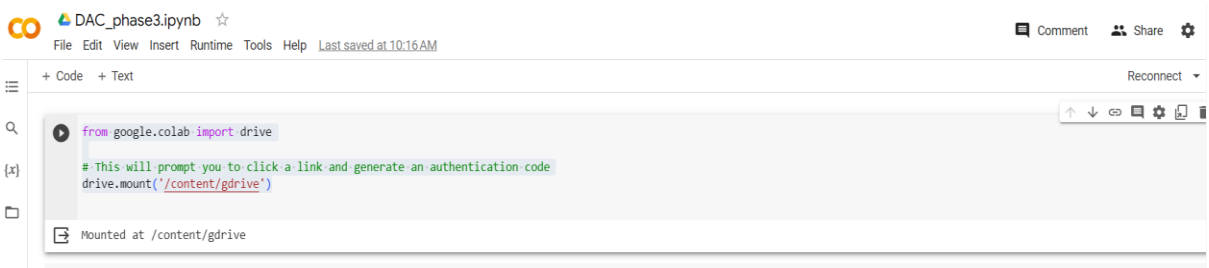
CODE:

```
from google.colab import drive

# This will prompt you to click a link and generate an authentication code

drive.mount('/content/gdrive')
```

OUTPUT:



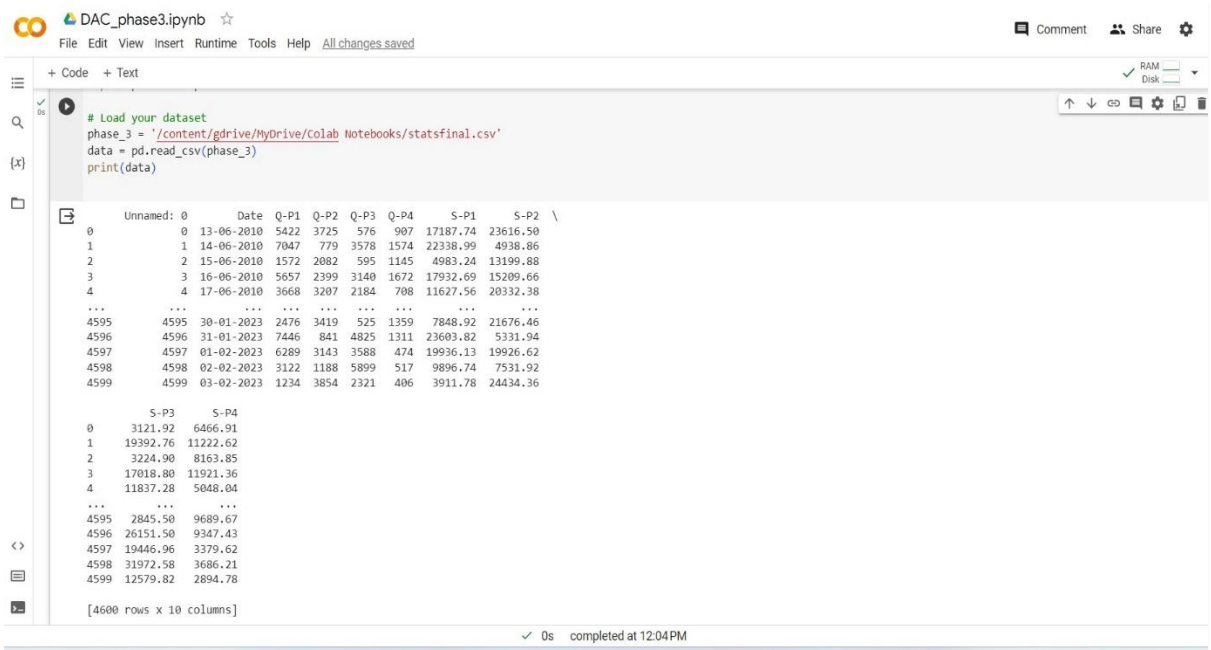
#LOADING THE DATASET:

CODE:

```
import pandas as pd

# Load your dataset
phase_3 = '/content/gdrive/MyDrive/Colab Notebooks/statsfinal.csv'
data = pd.read_csv(phase_3)
```

OUTPUT:

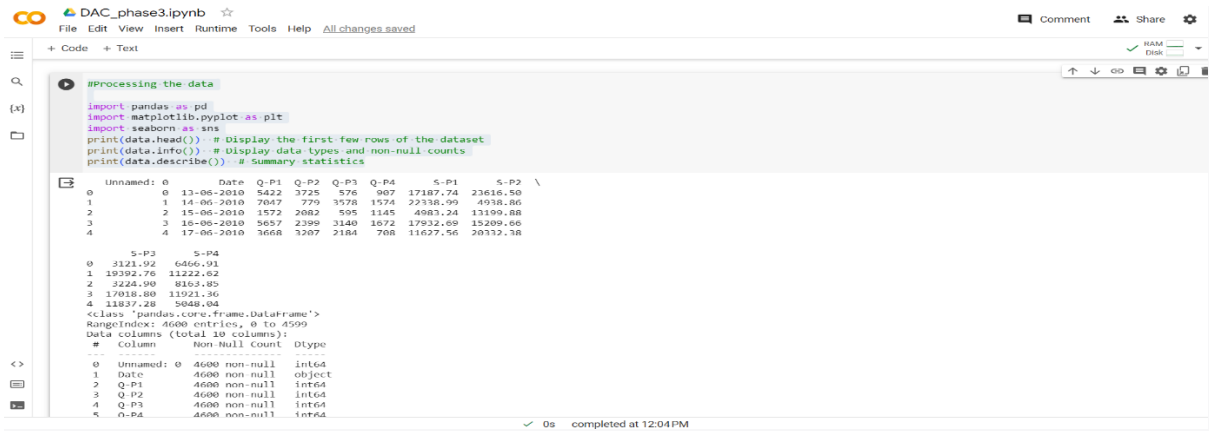


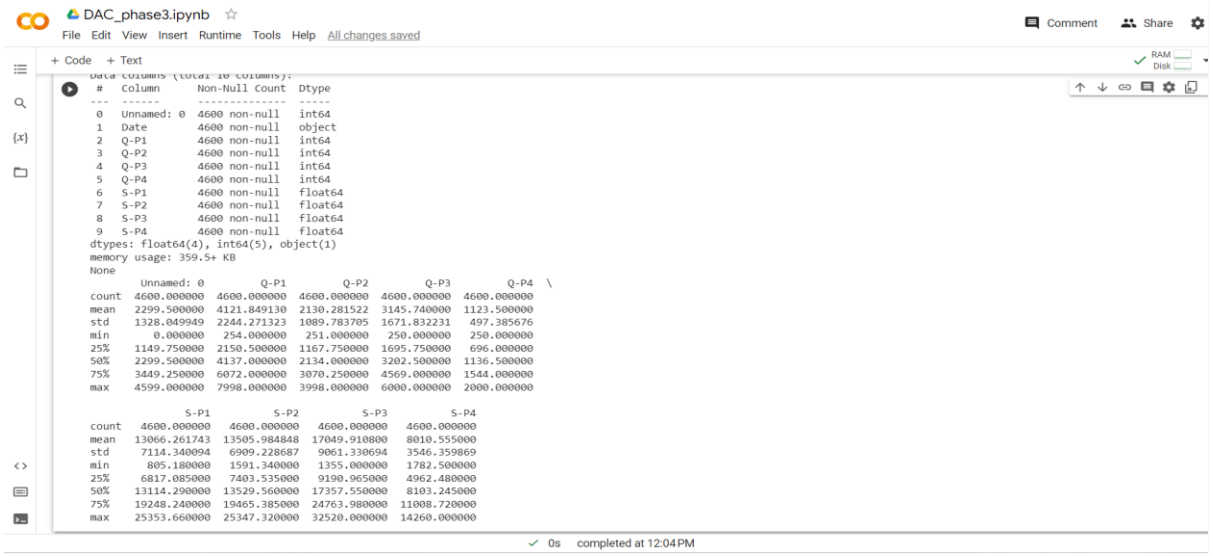
#PROCESSING AND CLEANSING THE DATASET:

CODE:

#PROCESSING THE DATA

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
print(data.head())  # Display the first few rows of the dataset
print(data.info())  # Display data types and non-null counts
print(data.describe())  # Summary statistics
```

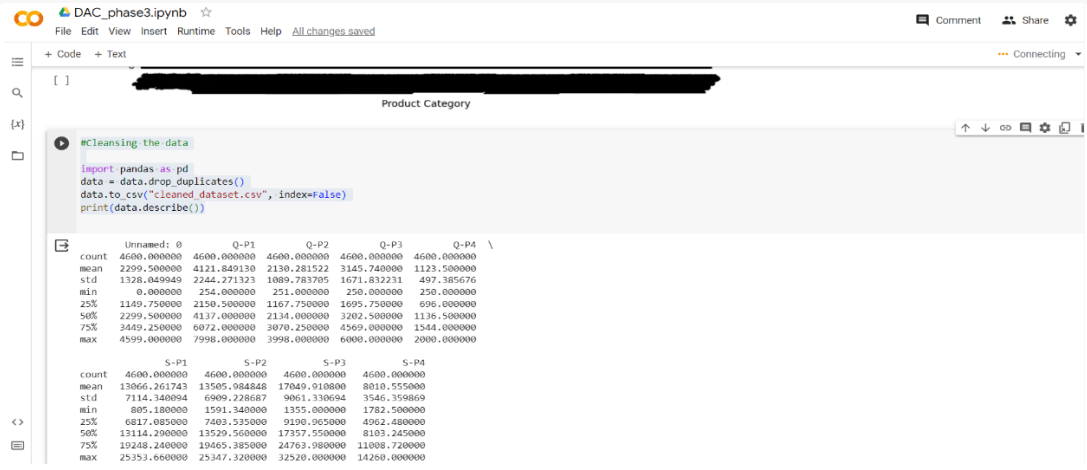




#CLEANSING THE DATA

```
import pandas as pd
data = data.drop_duplicates()
data.to_csv("cleaned_dataset.csv", index=False)
print(data.describe())
```

OUTPUT:



#ANALYSIS ON DATASET:

CODE:

```
plt.figure(figsize=(10, 6))
sns.barplot(x='Q-P1', y='S-P1', data=data)
plt.title("Sales by Product Category")
plt.xlabel("Product Category")
plt.ylabel("Sales")
plt.xticks(rotation=45)
plt.show()
```

OUTPUT:

ACCURACY:

CODE:

```
#LOGISTIC REGRESSION
from sklearn.datasets import make_classification
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
from sklearn.linear_model import LogisticRegression
nb_samples = 1000
x, y = make_classification(n_samples=nb_samples, n_features=2,
n_informative=2, n_redundant=0, n_clusters_per_class=1)
xtrain, xtest, ytrain, ytest = train_test_split(x, y, test_size=0.2,
random_state=42)
model = LogisticRegression()
```

```
model.fit(xtrain, ytrain)
```

OUTPUT:

```
max    25353.660000  25347.320000  32520.000000  14260.000000
```

```
#logistic regression
from sklearn.datasets import make_classification
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
from sklearn.linear_model import LogisticRegression
nb_samples = 1000
x, y = make_classification(n_samples=nb_samples, n_features=2, n_informative=2, n_redundant=0, n_clusters_per_class=1)
xtrain, xtest, ytrain, ytest = train_test_split(x, y, test_size=0.2, random_state=42)
model = LogisticRegression()
model.fit(xtrain, ytrain)
```

```
LogisticRegression()
```

CODE:

```
#ACCURACY OF THE DATA

print(accuracy_score(ytest, model.predict(xtest)))
```

OUTPUT:

```
#Accuracy of the data
print(accuracy_score(ytest, model.predict(xtest)))
```

```
0.88
```