04-08-25

Expt No 2 Upload and Analyze the data set given in csv format and perform data preprocessing and visualization

Q

Visualize the following.

1. Sales over the product.
2. Sales over time.
3. Display the correlation matrix.

Description: Use sample data set sales-data.csv.

Soln:

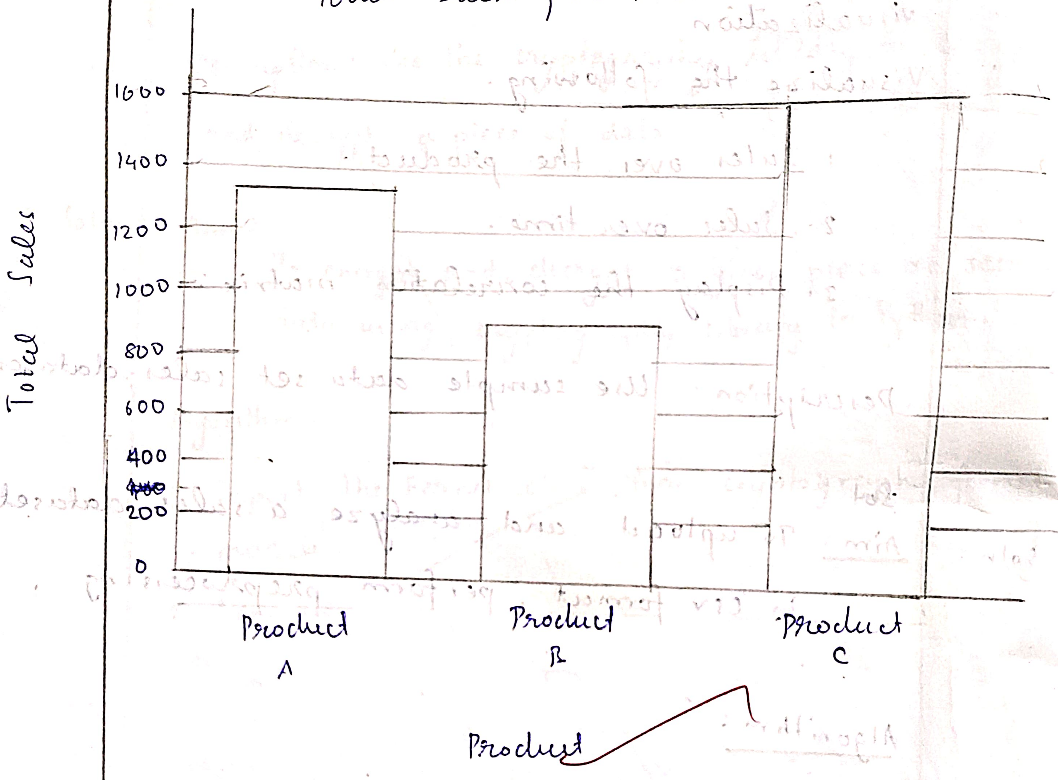
Aim: To upload and analyze a sales dataset in csv format, perform preprocessing.

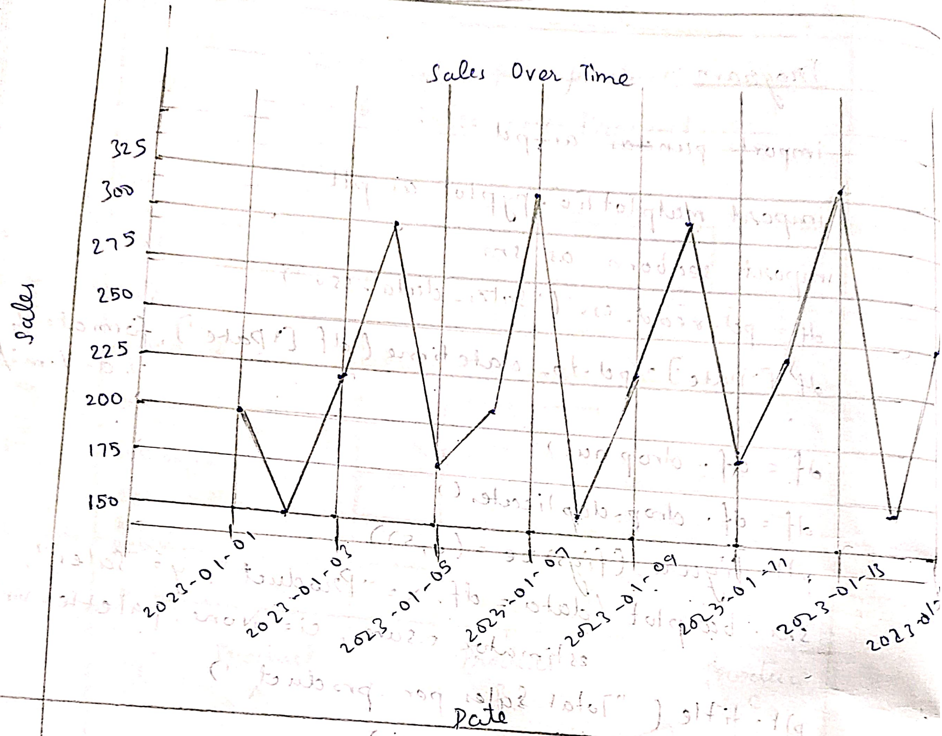
Algorithm:

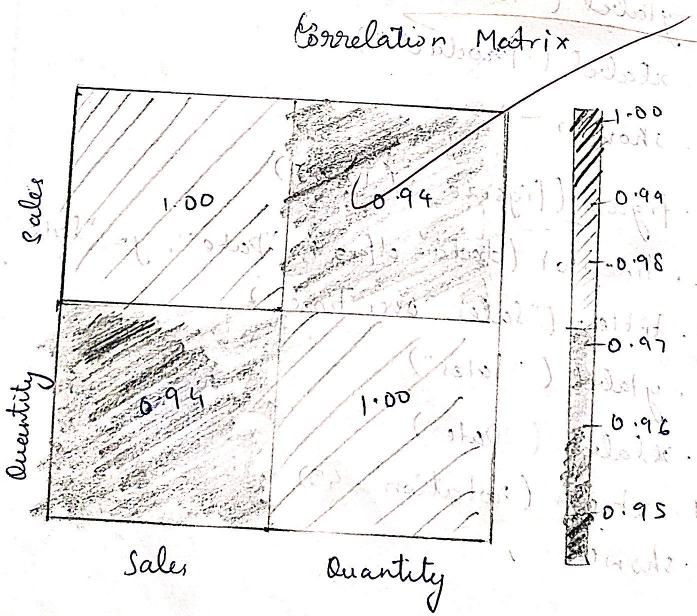
```
(i) Start
(ii) Import libraries: pandas, matplotlib, seaborn.
(iii) Load dataset: Read the csv file (sales_data.csv)
(iv) Preprocessing:
    - Convert Date column to date time format.
    . Handle missing values (drop or fill)
    . Remove duplicates if any.
(v) Visualization:
    . Plot bar chart of Total Sales per produc
    . Plot line chart of sales over time,
    . Generate heatmap of correlation matrix
(vi) End.
```

~~Total Sales Per product~~

Total Sales per Product




Sales Over Time
Sales
325
300
275
250
225
200
175
150
2023-01-01
2023-01-03
2023-01-05
2023-01-07
2023-01-09
2023-01-11
2023-01-13
2023-01-1
Date


Correlation Matrix
Sales
Quantity
1.00
0.94
0.94
1.00
Sales
Quantity
1.00
0.99
0.98
0.97
0.96
0.95

Program :

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
df = pd.read_csv ("Sales_data.csv")
df ['Date'] = pd.to_datetime (df ['Date'], format= %
                                        %d-%m-%Y)

df = df.dropna()
df = df.drop_duplicates()
plt.figure (figsize = (8,5))
sns.barplot (data = df, x = "Product", y = "Sales",
                estimator = sum, ci = None, palette = "viridis")
plt.title ( "Total Sales per product ")
plt.ylabel ("Total Sales")
plt.xlabel ("Product")
plt.show()
plt.figure (figsize = (10,5))
sns.lineplot (data = df, x = "Date", y= "Sales", marker="o")
plt.title ("Sales Over Time")
plt.ylabel ( "Sales")
plt.xlabel ( "Date")
plt.xticks (rotation = 45)
plt.show()
plt.figure(figsize=(6,4))
corr = df.corr(numeric_only = True)
sns.heatmap(corr, annot = True, cmap = "coolwarm",
                                        fmt = ".2f")
plt.title("Correlation Matrix")
plt.show()
```

Result: Thus, this is the python program ~~for the~~ to analyze the data set given in CSV format and perform data preprocessing and visualization executed successfully.