



## DATAVISTA: Sales Data Analysis and Visualization

### Team:

Gagankumar Seth [gaganseth2102@gmail.com](mailto:gaganseth2102@gmail.com)

Himanshi Nagori [himanshinagori.wit@gmail.com](mailto:himanshinagori.wit@gmail.com)

Chavi Maru [chavimaru19@gmail.com](mailto:chavimaru19@gmail.com)

**Guide:** Mr. Vignesh Muthuvelan

Disclaimer: The content is curated for educational purposes only.

## OUTLINE

- Abstract of the Project
- Problem Statement
- Proposed Solution
- System Architecture
- Live Demo of the Project
- Embedded Video of Project
- Conclusion
- Future Scope

## Abstract

- The DataVista project analyses supermarket sales data to uncover patterns and insights, helping businesses optimize sales strategies. It leverages Python, Pandas, NumPy, Matplotlib, Seaborn, and Scikit-learn for data processing and visualization.

## Problem Statement

- Businesses struggle with sales trends, customer behaviour, product performance, and payment preferences, leading to ineffective strategies. Limited predictive analysis hampers sales forecasting and customer satisfaction improvements.

## Proposed Solution

- **Data Preprocessing:** Handling missing values, converting data types, and scaling numerical features.
- **Exploratory Data Analysis (EDA):** Using visualizations (bar plots, pie charts, line graphs) to uncover sales trends and customer behavior.
- **Statistical Analysis:** Conducting correlation and regression analysis to predict sales based on unit price and quantity sold.
- **Data Splitting & Scaling:** Standardizing data using Min-Max Scaling, Z-score normalization, and Decimal Scaling.

# System Architecture

- **Data Collection:** Sales data from supermarket transactions.
- **Preprocessing Module:** Cleansing, handling missing values, and transforming data.
- **EDA Module:** Graphical analysis to visualize sales insights.
- **Statistical Analysis Module:** Correlation, regression, and predictive modeling.
- **Visualization & Insights Dashboard:** Displaying key metrics like sales distribution, customer demographics, and gross income trends.



# Live Demo of Project

## IMPORTING LIBRARIES

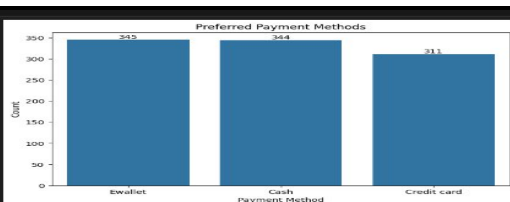
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import matplotlib.dates as mdates
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import MinMaxScaler, StandardScaler, MaxAbsScaler
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
```

## DATA PREPROCESSING

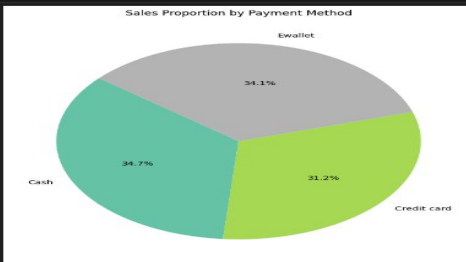
```
data = pd.read_csv('dataset.csv')
data.head()
```

```
data.shape
```

```
(1000, 17)
```

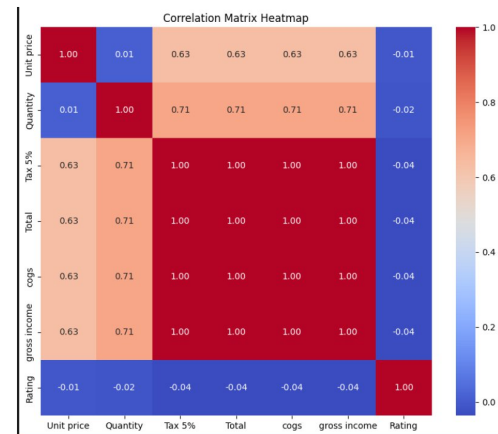
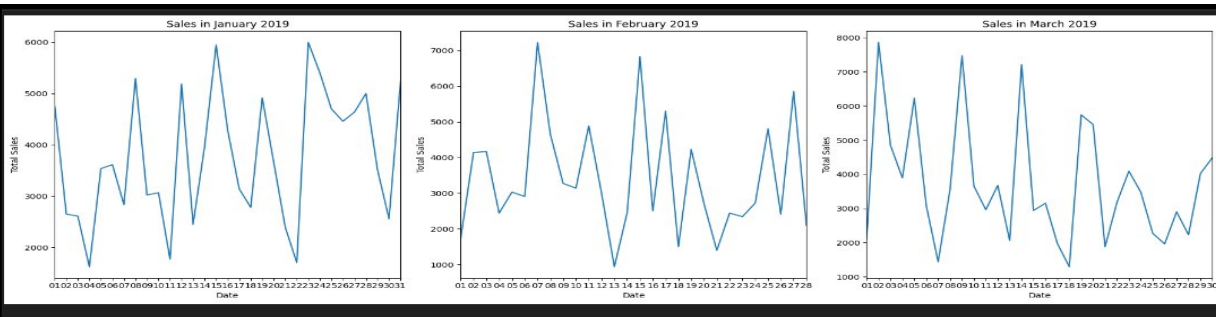


```
# Proportion of Payment Method by Sales
payment_sales = data.groupby('Payment')['Total'].sum()
plt.figure(figsize=(8, 6))
payment_sales.plot(kind='pie', autopct='%1.1f%%', startangle=140, cmap='Set2')
plt.title('Sales Proportion by Payment Method')
plt.show()
```



```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 17 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Invoice ID             1000 non-null  object
1   Branch                 1000 non-null  object
2   city                   1000 non-null  object
3   customer type          1000 non-null  object
4   gender                 1000 non-null  object
5   Product line           1000 non-null  object
6   Unit price             1000 non-null  float64
7   Quantity               1000 non-null  int64
8   Tax 5%                 1000 non-null  float64
9   Total                  1000 non-null  float64
10  Date                   1000 non-null  object
11  Time                   1000 non-null  object
12  Payment                1000 non-null  object
13  cogs                   1000 non-null  float64
14  gross margin percentage 1000 non-null  float64
15  gross income           1000 non-null  float64
16  Rating                 1000 non-null  float64
dtypes: float64(7), int64(1), object(9)
memory usage: 132.9+ KB
```



# Video of Project Demo

The screenshot shows a Jupyter Notebook interface for a project named 'DataVista'. The notebook is titled 'DataVista.ipynb' and is currently in the 'IMPORTING LIBRARIES' section. The table of contents on the left lists the following sections:

- IMPORTING LIBRARIES (1 cell hidden)
- DATA PREPROCESSING (9 cells hidden)
- DATA VISUALIZATION (15 cells hidden)
- DATA TRANSFORMATION (5 cells hidden)

The interface includes a top navigation bar with the file path 'C:\> Users > GAGANKUMAR > OneDrive > Desktop > T&P > Training > Capstone project > DataVista.ipynb'. The toolbar at the top right contains icons for 'Code', 'Markdown', 'Run All', 'Clear All Outputs', and 'Outline'. The bottom status bar shows 'Cell 1 of 34', 'Go Live', 'Kernels initialized!', and 'Server not selected'.



## Conclusion

- DataVista provides actionable insights into sales trends, customer behaviour, and revenue factors.
- Businesses can optimize inventory, pricing, and marketing strategies using data-driven decisions.
- The project enhances efficiency, profitability, and customer satisfaction through predictive analysis.

## Future Scope

- Implement real-time data analysis for immediate sales trend tracking.
- Develop an interactive dashboard with predictive analytics.
- Integrate AI-based recommendation systems for sales forecasting.
- Expand analysis to include external market factors like competitor pricing and economic trends.

**Thank you!**