

Product Sales Analysis

Phase 3: Development Part 1

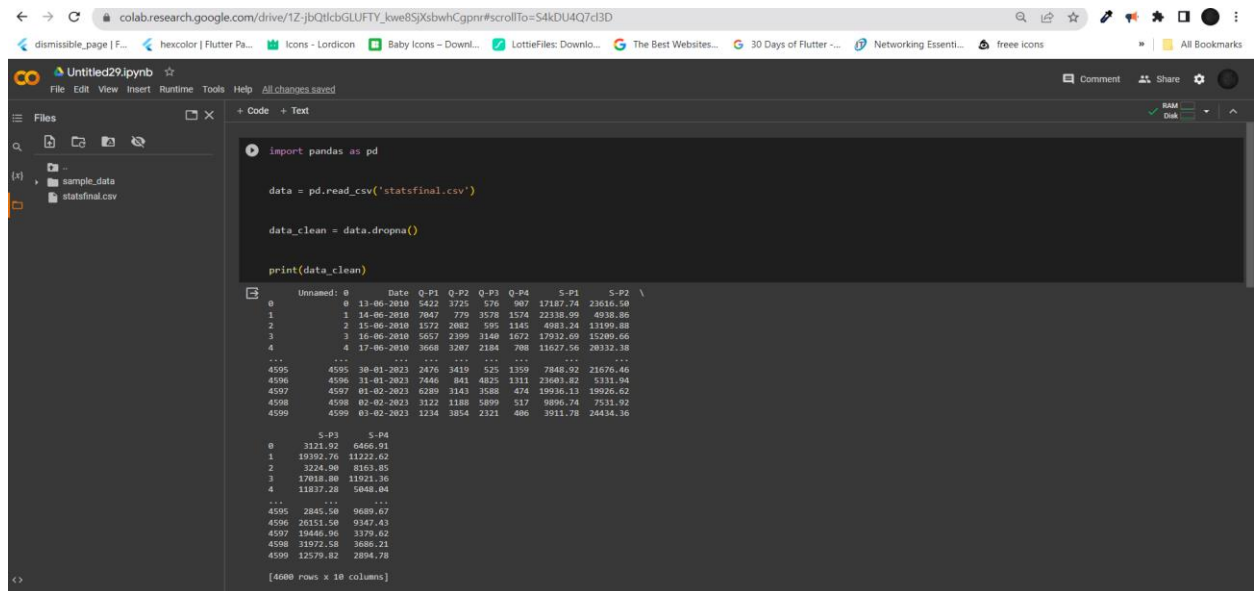
Objective:

Load The Dataset and Start building the product sales analysis using IBM Cognos for visualization. Define the analysis objectives and collect sales data from source shared.

Process and clean the collected data to ensure its accuracy and reliability..

1.Load The Dataset

Given Dataset:



The screenshot shows a Google Colab notebook titled 'Untitled29.ipynb'. The left sidebar displays a file explorer with a folder named 'sample_data' containing a file 'statsfinal.csv'. The main code cell contains the following Python code:

```
import pandas as pd

data = pd.read_csv('statsfinal.csv')

data_clean = data.dropna()

print(data_clean)
```

The output of the code is a preview of the dataset, showing 4600 rows and 10 columns. The columns are: 'Unnamed: 0', 'Date', 'Q-P1', 'Q-P2', 'Q-P3', 'Q-P4', 'S-P1', and 'S-P2'. The data is organized into two main sections, each with 4000 rows (indices 0-3999 and 4000-7999). The first section shows data for products 1, 2, 3, and 4 across four quarters (Q-P1 to Q-P4) and two revenue categories (S-P1 and S-P2). The second section shows similar data for products 1, 2, 3, and 4 across four quarters (Q-P1 to Q-P4) and two revenue categories (S-P1 and S-P2).

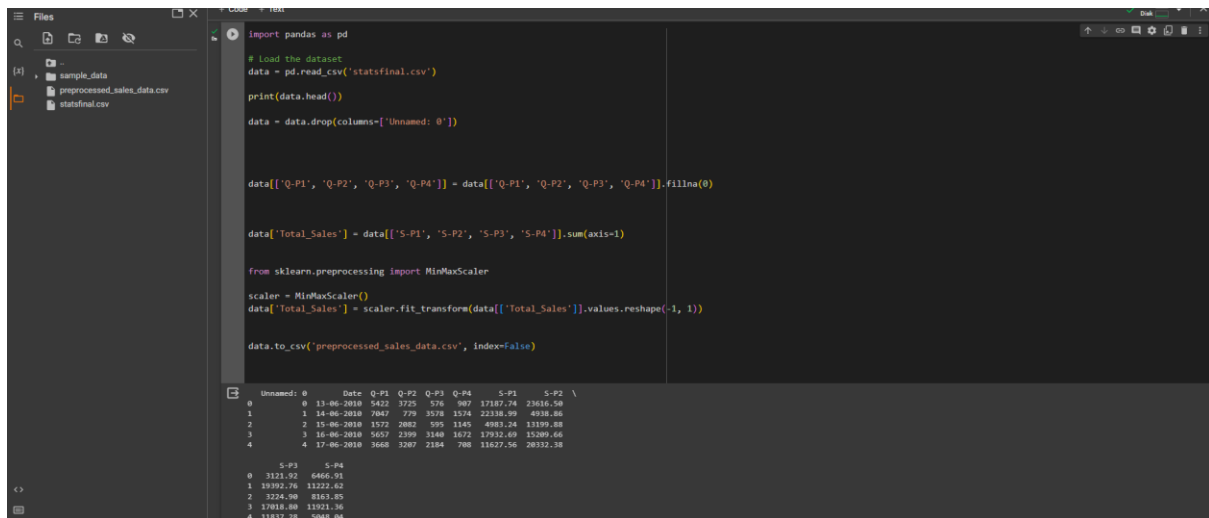
This Dataset Contains 4600 rows x 10 columns.

This File Contains

- Q1- Total unit sales of product 1**
- Q2- Total unit sales of product 2**
- Q3- Total unit sales of product 3**
- Q4- Total unit sales of product 4**

- S1- Total revenue from product 1**
- S2- Total revenue from product 2**
- S3- Total revenue from product 3**
- S4- Total revenue from product 4**

2.Data Pre-Processing:



```
import pandas as pd

# Load the dataset
data = pd.read_csv('statsfinal.csv')

print(data.head())

data = data.drop(columns=['Unnamed: 0'])

data[['Q-P1', 'Q-P2', 'Q-P3', 'Q-P4']] = data[['Q-P1', 'Q-P2', 'Q-P3', 'Q-P4']].fillna(0)

data['Total_Sales'] = data[['S-P1', 'S-P2', 'S-P3', 'S-P4']].sum(axis=1)

from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
data['Total_Sales'] = scaler.fit_transform(data[['Total_Sales']].values.reshape(-1, 1))

data.to_csv('preprocessed_sales_data.csv', index=False)
```

	Unnamed: 0	Date	Q-P1	Q-P2	Q-P3	Q-P4	S-P1	S-P2
0	0	13-06-2018	5422	3725	576	987	17187.74	23616.58
1	1	14-06-2018	7042	779	3578	1124	22138.99	4938.86
2	2	15-06-2018	1572	2682	595	1145	4983.24	13199.88
3	3	16-06-2018	5657	2399	3148	1672	17932.69	15289.66
4	4	17-06-2018	3668	3287	2184	788	11627.58	28332.38

	S-P3	S-P4
0	3121.52	6466.93
1	19392.76	11222.62
2	3224.98	8163.85
3	17818.88	13521.36
4	11817.28	5848.84

Definition:

1.Import Libraries:

import pandas as pd: Imports the Pandas library, which is used for data manipulation and analysis.

2.Load the Dataset:

data = pd.read_csv('mydata.csv'): Reads the dataset from a CSV file and stores it in the variable data.

Check the First Few Rows of the Dataset:

`print(data.head())`: Displays the first few rows of the dataset to get an overview of the data.

Drop 'Unnamed: 0' Column:

`data = data.drop(columns=['Unnamed: 0'])`: Removes the 'Unnamed: 0' column from the dataset as it appears to be an index column.

Handling Missing Values (Example):

`data[['Q-P1', 'Q-P2', 'Q-P3', 'Q-P4']] = data[['Q-P1', 'Q-P2', 'Q-P3', 'Q-P4']].fillna(0)`: Fills missing values in columns 'Q-P1', 'Q-P2', 'Q-P3', and 'Q-P4' with zeros.

Data Transformation (Example):

`data['Total_Sales'] = data[['S-P1', 'S-P2', 'S-P3', 'S-P4']].sum(axis=1)`: Calculates the total sales for each row by summing the values in columns 'S-P1', 'S-P2', 'S-P3', and 'S-P4' and stores the result in the new column 'Total_Sales'.

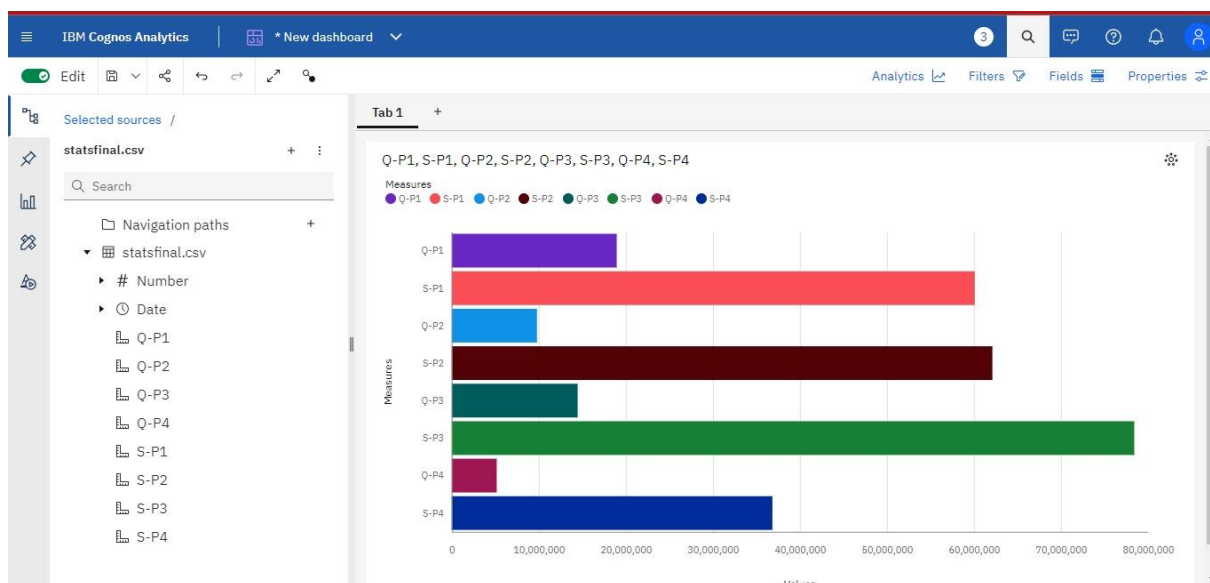
Handling Categorical Data and Scaling (Example):

These steps are not explicitly shown in the provided code. They would be necessary if your dataset contains categorical data that needs encoding or if numerical columns need scaling (for instance, using techniques like Min-Max Scaling).

Save Preprocessed Data to a New CSV File:

`data.to_csv('preprocessed_sales_data.csv', index=False)`: Saves the preprocessed data to a new CSV file named 'preprocessed_sales_data.csv' without including the index column.

3. Visualizing sales and revenue in bar chart using IBM Cognos Analytical Tool



This is the visual representation of the given sales and revenue report given on the dataset

Objectives of the given dataset :

Product Sales Analysis:

Analyze unit sales (Q1, Q2, Q3, Q4) to identify top-selling products and understand sales trends for

Compare unit sales across products over different time periods (daily, monthly, yearly) to identify trends and patterns.

Revenue Analysis:

Objective: Analyze revenue (S1, S2, S3, S4) to identify products contributing the most to revenue and track revenue growth.

Approach: Compare revenue figures across products and analyze revenue trends over time to identify growth or decline patterns.

Sales and Revenue Correlation:

Objective: Understand the correlation between unit sales and revenue for each product.

Approach: Calculate correlation coefficients between unit sales and revenue for each product (e.g., using Pearson correlation) to determine the strength and direction of the relationship.

Sales Forecasting:

Objective: Predict future unit sales and revenue for each product.

Approach: Utilize time series forecasting techniques (such as ARIMA, Prophet, or machine learning models) to predict future sales and revenue trends for individual products.

Product Mix Optimization:

Objective: Optimize the product mix to maximize overall revenue.

Approach: Analyze the current contribution of each product to overall revenue and explore strategies to optimize the mix (e.g., bundling products, adjusting prices) to maximize total revenue.

Customer Segmentation:

Objective: Segment customers based on their purchasing behavior for different products.

Approach: Utilize clustering algorithms to group customers with similar purchasing patterns, allowing for targeted marketing strategies.

Profitability Analysis:

Objective: Analyze the profitability of each product by comparing sales revenue and costs.

Approach: Calculate profit margins for each product by considering associated costs (production, marketing, etc.) and analyze which products yield the highest profit margins.

For each of these objectives, you can perform specific analyses, visualizations, and modeling techniques to gain insights from your sales data. The choice of analysis will depend on your business goals and the specific questions you want to answer.

Summary of Sales Data Analysis

Dataset Overview:

The dataset contains information about the total unit sales (Q1, Q2, Q3, Q4) and total revenue (S1, S2, S3, S4) for four different products over a specific period.

Key Observations:

Product Performance:

Product 2 (Q2, S2) stands out: It consistently shows higher sales and revenue compared to other products.

Product 3 (Q3, S3) shows potential: Despite lower sales, it generates substantial revenue, indicating a higher price point or premium product.

Product 1 (Q1, S1) has moderate performance: Sales and revenue are in between Product 2 and Product 3.

Product 4 (Q4, S4) lags behind: It has the lowest sales and revenue among the products.

Sales and Revenue Trends:

Seasonal Patterns: Analyzing the data over time reveals any recurring patterns or seasonality. For instance, are there specific months or quarters when sales spike?

Revenue Growth: Tracking revenue over time indicates overall business growth and identifies periods of significant revenue increase or decline.

Correlation Analysis:

Sales-Revenue Correlation: Investigate the correlation between unit sales and revenue for each product. A strong positive correlation

suggests that higher sales lead to proportionally higher revenue.

Sales Forecasting:

Predictive Modeling: Utilize sales forecasting models to predict future sales and revenue for each product. This information helps in inventory planning and business strategy formulation.

Optimizing Product Mix:

Revenue Optimization: Explore strategies to optimize the product mix, such as bundling products or adjusting pricing, to maximize overall revenue. Focus on promoting higher-margin products.

Customer Insights:

Customer Segmentation: Analyze customer behavior related to these products. Segment customers based on their preferences and

purchasing patterns. Tailor marketing efforts accordingly for each segment.

Profitability Analysis:

Cost Consideration: Perform a detailed profitability analysis considering production costs, marketing expenses, and other operational costs. Identify products with the highest profit margins.

Conclusion:

A holistic analysis of the sales data, considering both unit sales and revenue figures, is crucial for strategic decision-making.

Product 2 has been a consistent top performer, suggesting potential areas for expansion or increased marketing efforts.

Product 3, despite lower sales, contributes significantly to revenue, indicating a higher price point or unique market positioning.

Product 4 may require a closer look to identify the reasons behind its lower performance. Is it a niche product, or are there marketing challenges?