PHASE-5 PROJECT DOCUMENTATION AND SUBMISSION

PRODUCT SALES ANALYSIS

INTRODUCTION:

In today's competitive business landscape, understanding customer behaviour and market trends is vital for any company striving to succeed. The ability to analyse sales data effectively not only provides valuable insights into consumer preferences but also helps businesses make informed decisions, optimize their strategies, and enhance overall profitability. This is precisely why sales analysis projects play a pivotal role in the modern business environment.

Our **Product Sales Analysis Project** aims to delve deep into the intricacies of sales data, patterns, identifying key drivers, and drawing meaningful conclusions. By leveraging advanced analytical tools and techniques, we will dissect sales figures, customer demographics, and purchasing habits, illuminating the path to smarter marketing, improved customer satisfaction, and increased revenue.

PROJECT OVERVIEW:

In this project, we will focus on a comprehensive analysis of our product sales data to achieve several critical objectives:

1. **Understanding Customer Behavior:**

We will examine customer preferences, purchasing frequency, and the factors influencing buying decisions. By identifying patterns, we can tailor our marketing efforts to meet specific customer needs effectively.

2. **Identifying Top-Selling Products**:

By analyzing sales data, we will pinpoint our best-performing products. Understanding what makes these products successful can guide inventory management and marketing strategies to optimize sales further.

3. Market Segmentation:

We will categorize customers into segments based on demographics, geography, or purchasing behavior. This segmentation will enable us to personalize marketing campaigns and enhance customer engagement.

4.Sales Trends and Forecasting:

Through historical sales data analysis, we will uncover trends and predict future sales patterns. Accurate forecasting is essential for inventory management and ensuring products are available when customers demand them.

5.Competitor Analysis:

Understanding how our products fare against competitors is crucial. By comparing sales data and market share, we can identify our competitive advantages and areas that require improvement.

6. Customer Satisfaction Analysis:

Utilizing customer feedback data, we can gauge customer satisfaction levels. This analysis can provide insights into areas needing improvement, leading to enhanced customer experiences and brand loyalty.

Significance of the Project:

The insights derived from this sales analysis project will serve as a guiding light for our business strategies. In an era where data-driven decision-making is paramount, this project holds immense significance. By harnessing the power of data analytics, we aim to not only boost sales but also foster lasting relationships with our customers, enhance operational efficiency, and secure a prominent position in the market.

DATASET LINK:

https://www.kaggle.com/datasets/ksabishek/product-sales-data

REC corp LTD. is small-scaled business venture established in India. They have been selling FOUR PRODUCTS for OVER TEN YEARS . The products are P1, P2, P3 and P4. They have collected data from their retail centers and organized it into a small csv file , which has been given to us.

The excel file contains about 8 numerical parameters :

- 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

PROBLEM DEFINITION AND DESIGN THINKING:

- **Understand Sales Performance :**

Gain a deep understanding of how each product in the company's portfolio is performing in terms of revenue, quantity sold, and profitability.

- **Customer Insights:**

  Analyse customer behaviour, including purchasing patterns, demographics, and preferences, to tailor marketing strategies and product offerings.

- **Market Trends**:

Identify and react to emerging market trends, competitor strategies, and external factors affecting sales performance.

- **Pricing Strategy:**

Analyse the impact of pricing changes on product sales and profitability. Determine optimal pricing strategies for each product.

- **Inventory Management:**

Use sales data to optimize inventory levels, reducing carrying costs while ensuring products are readily available to meet demand.

DESIGN THINKING:

- **Empathize**:

Understand the users' perspectives, needs, and experiences related to the product. This involves conducting interviews, surveys, and observations to gather insights.

- **Define**:

Clearly articulate the problem you're trying to solve based on the information gathered during the empathize stage. Define the user's needs, challenges, and aspirations in a specific manner.

- **Ideate:**

Generate a wide range of creative ideas to solve the defined problem. Encourage brainstorming sessions and explore different possibilities without judgment.

- **Prototype:**

Develop tangible representations of the ideas generated during the ideation phase. Prototypes can be sketches, wireframes, or even simple models that help visualize the solutions.

- **Test:**

Test the prototypes with real users to gather feedback. This step involves observing how users interact with the prototypes, understanding their reactions, and refining the designs based on the feedback received.

PREDICTIVE ANALYTICS:

Implementing the machine learning models to predict future sales trends based on historical data. This can help businesses proactively adjust their strategies. Here are some machine learning models and techniques such as linear regression, time series forecasting, random forest regression, gradient boosting, recurrent neural networks, convolution neural network, ensemble methods, evaluation metrics, hyperparameter tuning, feature engineering.

One popularly used machine learning algorithm for predicting future sales and trends is the Time Series analysis.

TIME SERIES ANALYSIS

Time series analysis is a statistical technique used to analyze and interpret data points collected or recorded at specific time intervals. It is widely used in various fields such as finance, economics, weather forecasting, and signal processing. Time series data consists of observations on a variable or several variables over time.

KEY COMPONENTS OF TIME SERIES ANALYSIS INCLUDE:

Trend:

The long-term movement in a time series, which can be upward , downward , or stable.

Seasonality:

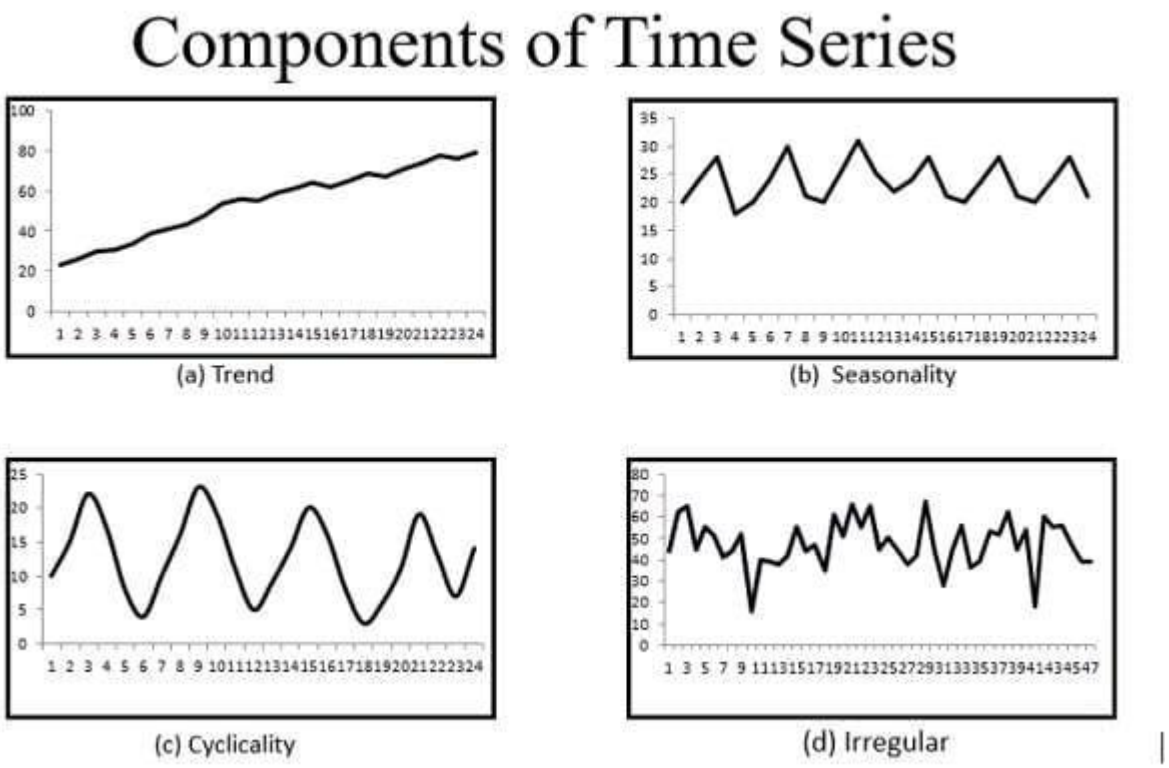
Patterns that occur regularly at specific intervals, often influenced by factors like seasons, months, days, or hours.

Cyclical Patterns:

Repeating up and down movements in the data that are not of fixed frequency like seasonality.

Noise:

Random variation or irregular movements in the data that cannot be attributed to the above components.



FEATURES OF TIME SERIES ANALYSIS:

- Time ordered data
- Trend
- Seasonality
- Noise
- Auto correlation
- Stationarity
- Forecasting
- Data decomposition
- Visualization

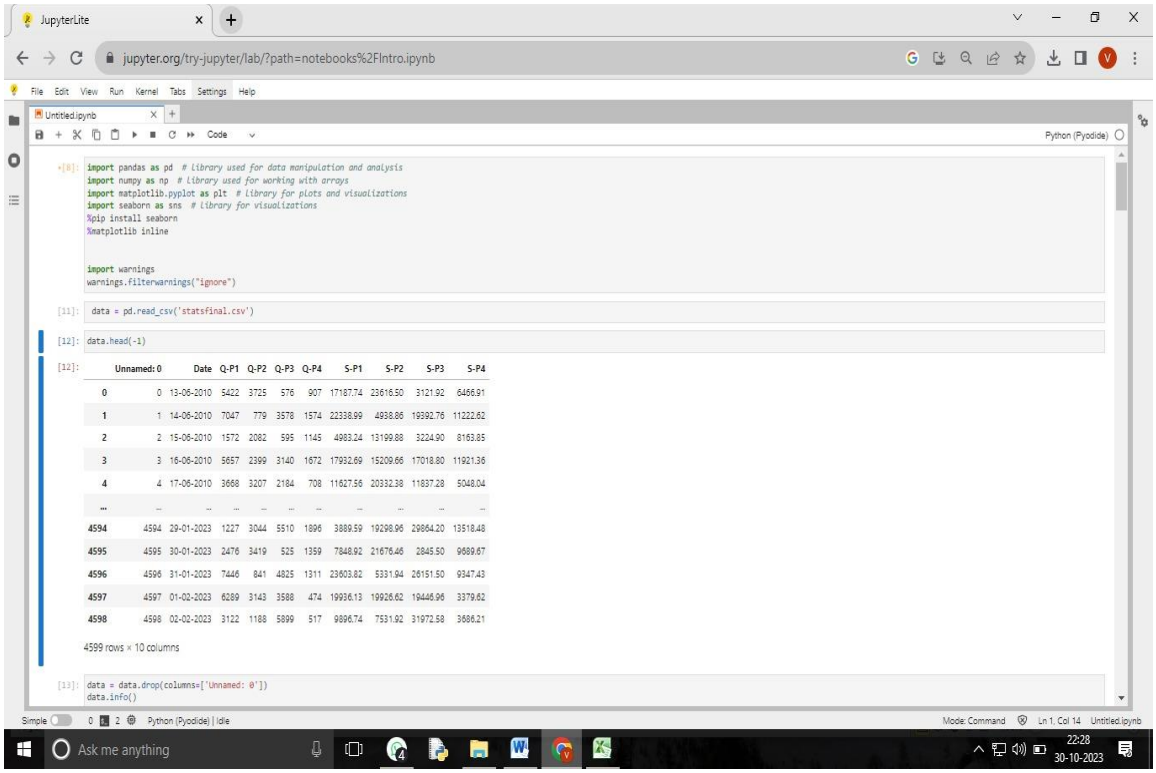
LOADING AND PREPROCESSING THE DATASET

To load and analyze a dataset for product sales using machine learning, follow these steps in Python:

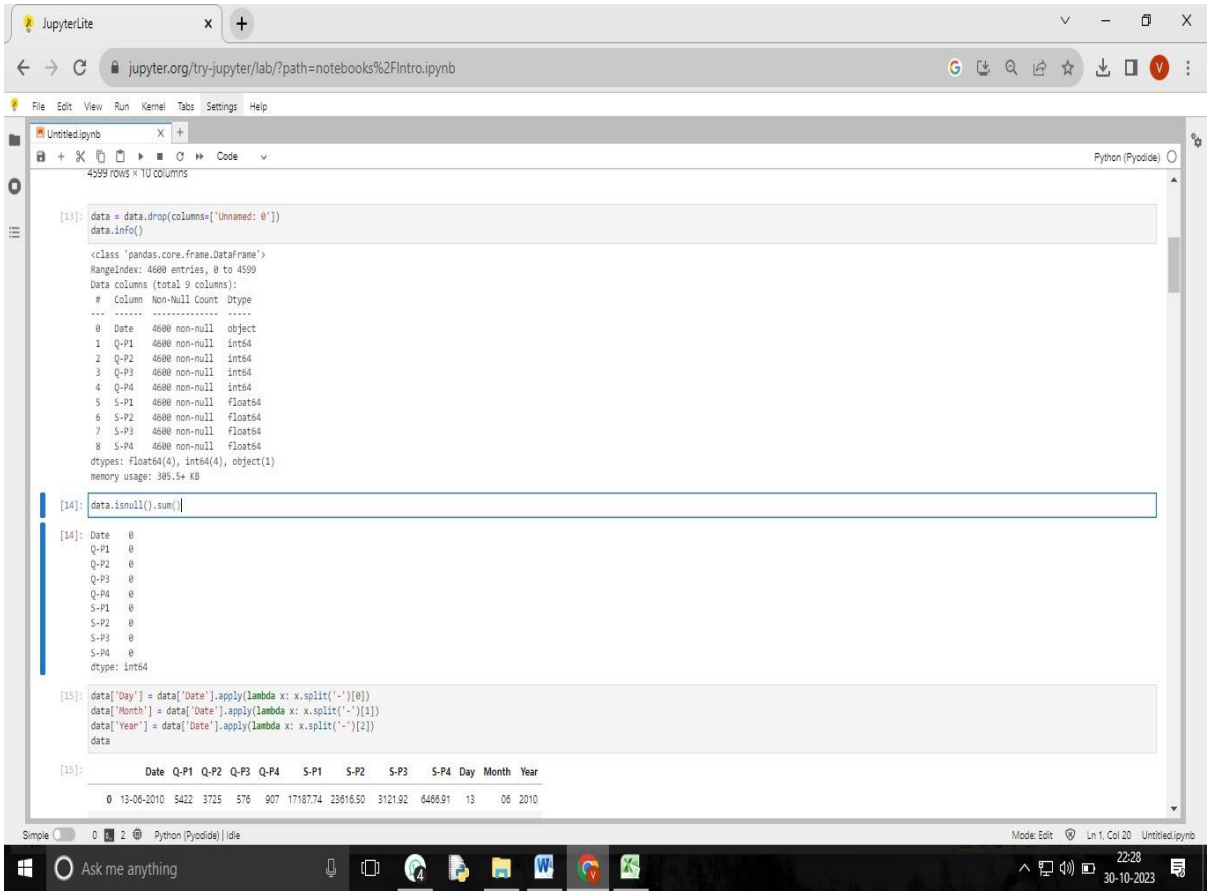
- ✓ Import Required Libraries

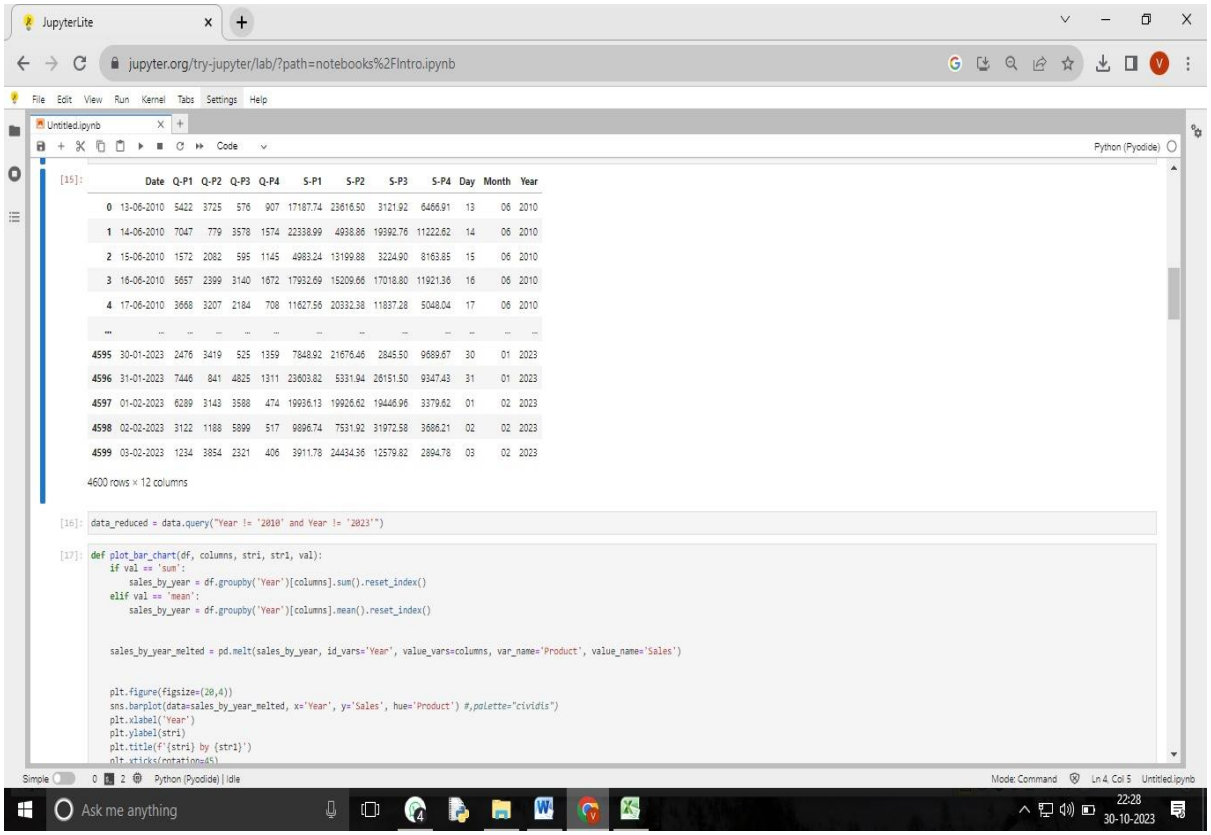
- ✓ Read the Datasets
- ✓ Data preprocessing
- ✓ Split the dataset
- ✓ Choose a machine learning model
- ✓ Train the model
- ✓ Evaluate the model
- ✓ Intrepret and visualize results
- ✓ Further Analysis

Here are some example for the above mentioned program steps,

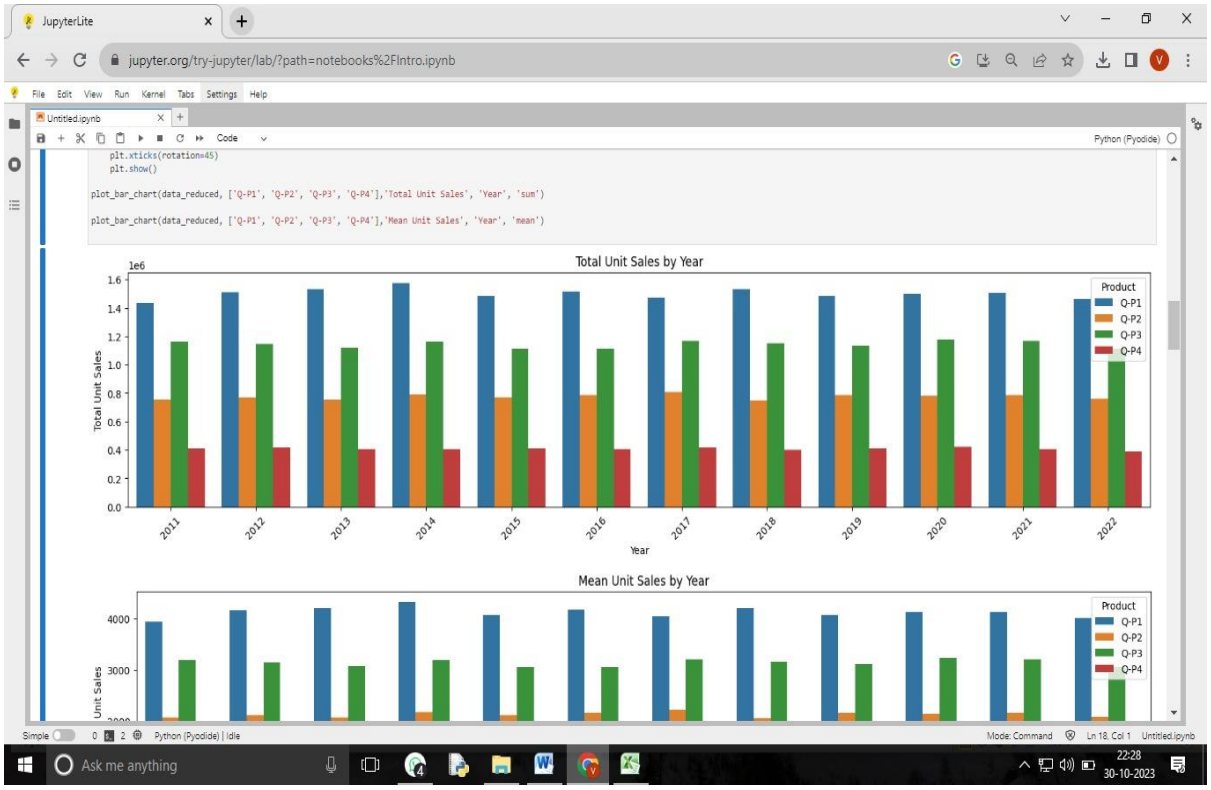


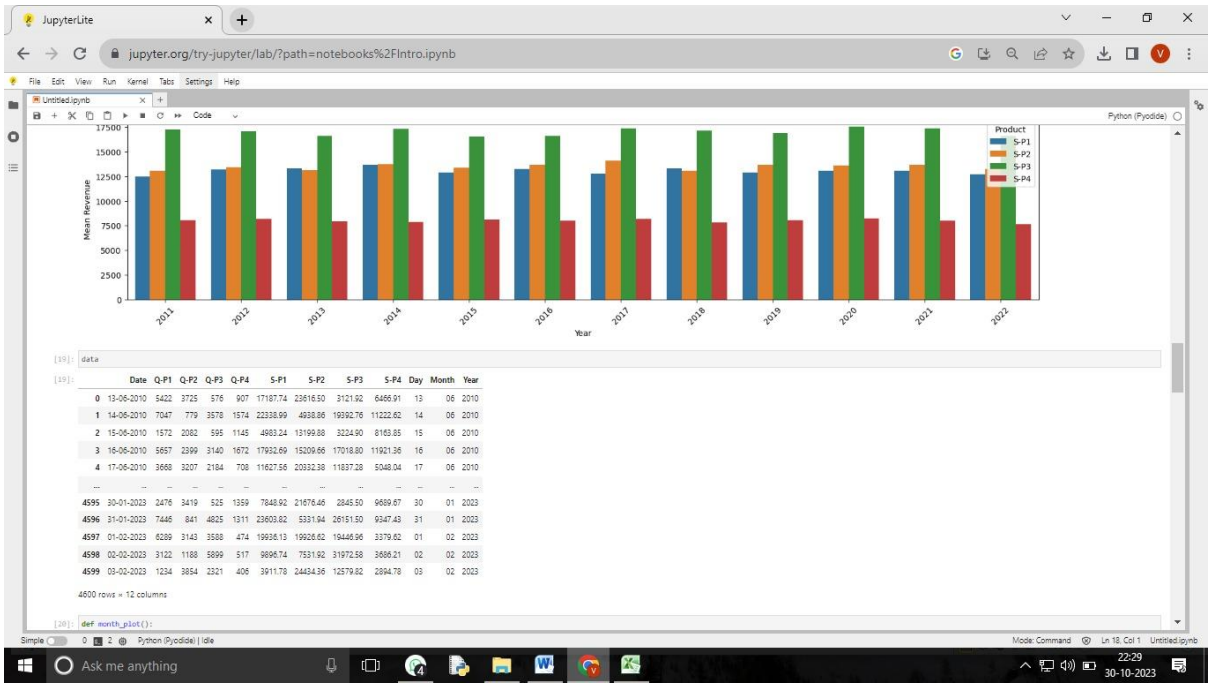
The above output of the program consists of datasets which contains 4599 rows and 10 columns. For dropping the columns , the following code will be executed.


JupyterLite
jupyter.org/try-jupyter/lab/?path=notebooks%2FIntro.ipynb
File Edit View Run Kernel Tabs Settings Help
Untitled.ipynb
Code
Python (Pyodide)
4599 rows × 10 columns
[13]:
data = data.drop(columns=['Unnamed: 0'])
data.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4600 entries, 0 to 4599
Data columns (total 9 columns):
Column Non-Null Count Dtype
--- ------ -------------- -----
0 Date 4600 non-null object
1 Q-P1 4600 non-null int64
2 Q-P2 4600 non-null int64
3 Q-P3 4600 non-null int64
4 Q-P4 4600 non-null int64
5 S-P1 4600 non-null float64
6 S-P2 4600 non-null float64
7 S-P3 4600 non-null float64
8 S-P4 4600 non-null float64
dtypes: float64(4), int64(4), object(1)
memory usage: 305.5+ KB
[14]:
data.isnull().sum()
[14]:
Date 0
Q-P1 0
Q-P2 0
Q-P3 0
Q-P4 0
S-P1 0
S-P2 0
S-P3 0
S-P4 0
dtype: int64
[15]:
data['Day'] = data['Date'].apply(lambda x: x.split('-')[0])
data['Month'] = data['Date'].apply(lambda x: x.split('-')[1])
data['Year'] = data['Date'].apply(lambda x: x.split('-')[2])
data
[15]:
Date Q-P1 Q-P2 Q-P3 Q-P4 S-P1 S-P2 S-P3 S-P4 Day Month Year
0 13-06-2010 5422 3725 576 907 17187.74 23616.50 3121.92 6466.91 13 06 2010
Simple 0 2 Python (Pyodide) | Idle
Mode: Edit Ln 1, Col 20 Untitled.ipynb
Ask me anything
22:28
30-10-2023

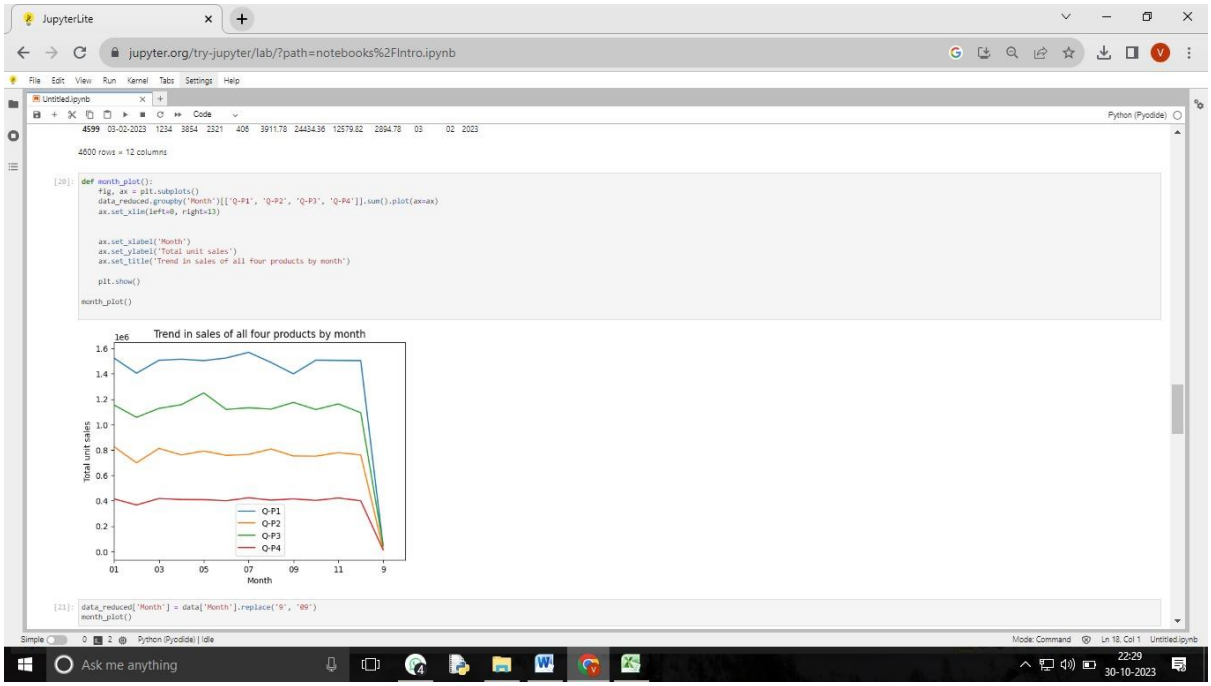

JupyterLite
jupyter.org/try-jupyter/lab/?path=notebooks%2FIntro.ipynb
File Edit View Run Kernel Tabs Settings Help
Untitled.ipynb
Code
Python (Pyodide)
[15]:
Date Q-P1 Q-P2 Q-P3 Q-P4 S-P1 S-P2 S-P3 S-P4 Day Month Year
0 13-06-2010 5422 3725 576 907 17187.74 23616.50 3121.92 6466.91 13 06 2010
1 14-06-2010 7047 779 3578 1574 22338.99 4938.86 19392.76 11222.62 14 06 2010
2 15-06-2010 1572 2082 595 1145 4983.24 13199.88 3224.90 8163.85 15 06 2010
3 16-06-2010 5657 2399 3140 1672 17932.69 15209.66 17018.80 11921.36 16 06 2010
4 17-06-2010 3668 3207 2184 708 11627.56 20332.38 11837.28 5048.04 17 06 2010
... ...
4595 30-01-2023 2476 3419 525 1359 7848.92 21676.46 2845.50 9689.67 30 01 2023
4596 31-01-2023 7446 841 4825 1311 23603.82 5331.94 26151.50 9347.43 31 01 2023
4597 01-02-2023 6289 3143 3588 474 19936.13 19926.62 19446.96 3379.62 01 02 2023
4598 02-02-2023 3122 1188 5899 517 9896.74 7531.92 31972.58 3686.21 02 02 2023
4599 03-02-2023 1234 3854 2321 406 3911.78 24434.36 12579.82 2894.78 03 02 2023
4600 rows × 12 columns
[16]:
data_reduced = data.query("Year != '2010' and Year != '2023'")
[17]:
def plot_bar_chart(df, columns, str1, str1, val):
if val == 'sum':
sales_by_year = df.groupby('Year')[columns].sum().reset_index()
elif val == 'mean':
sales_by_year = df.groupby('Year')[columns].mean().reset_index()
sales_by_year_melted = pd.melt(sales_by_year, id_vars='Year', value_vars=columns, var_name='Product', value_name='Sales')
plt.figure(figsize=(20,4))
sns.barplot(data=sales_by_year_melted, x='Year', y='Sales', hue='Product') #,palette="cividis")
plt.xlabel('Year')
plt.ylabel(str1)
plt.title(f'{str1} by {str1}')
Simple 0 2 Python (Pyodide) | Idle
Mode: Command Ln 4, Col 5 Untitled.ipynb
Ask me anything
22:28
30-10-2023

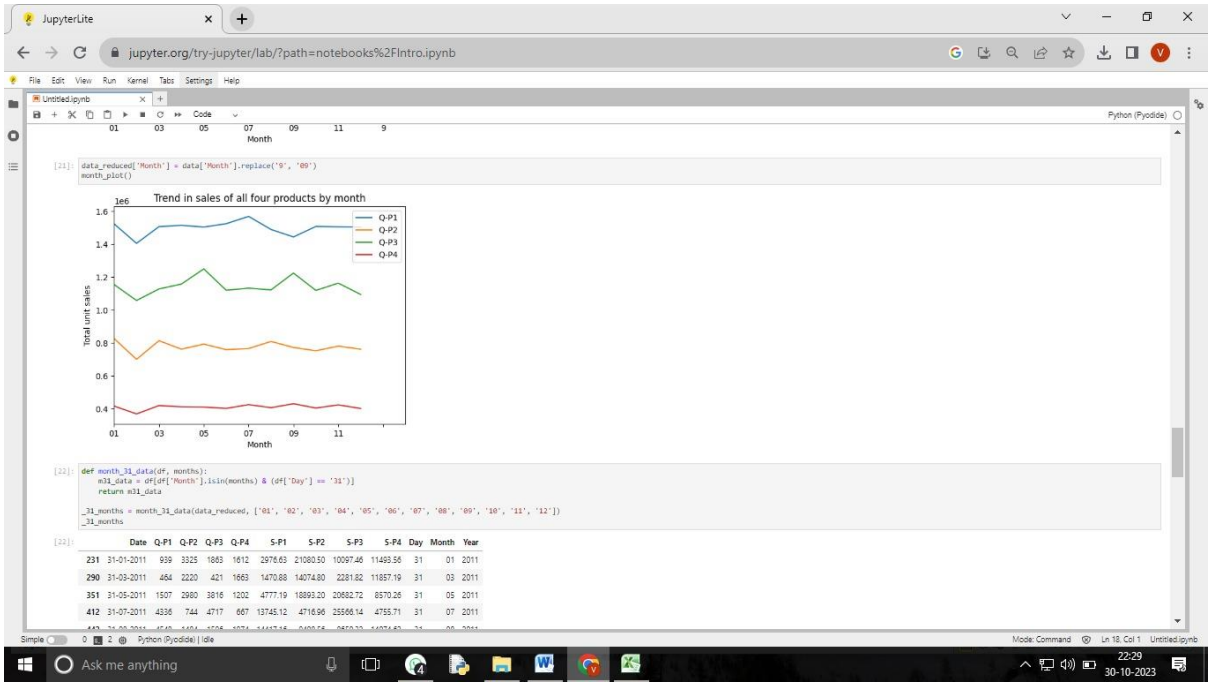
To understand the clear insights for product sales analysis, the following is the bar chart for visualization. It consists of total unit sales by year, mean unit sales by year and the datasets which contains the overall information of product sales analysis.

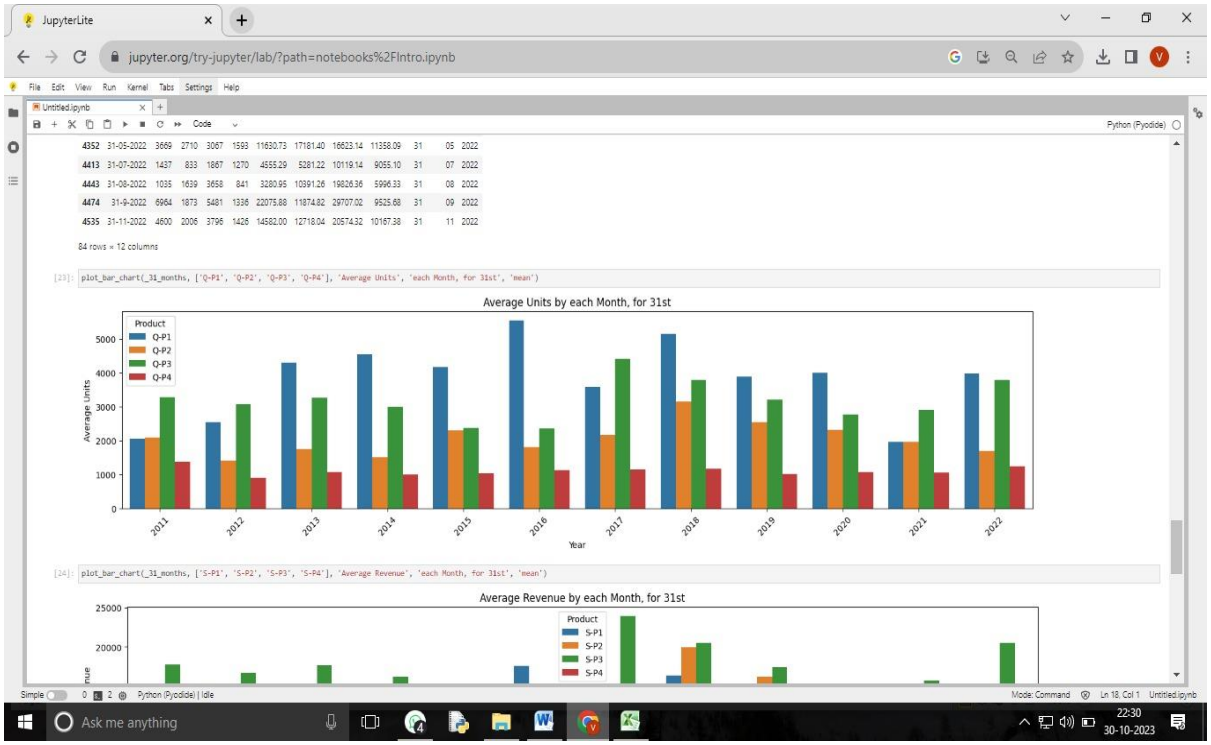


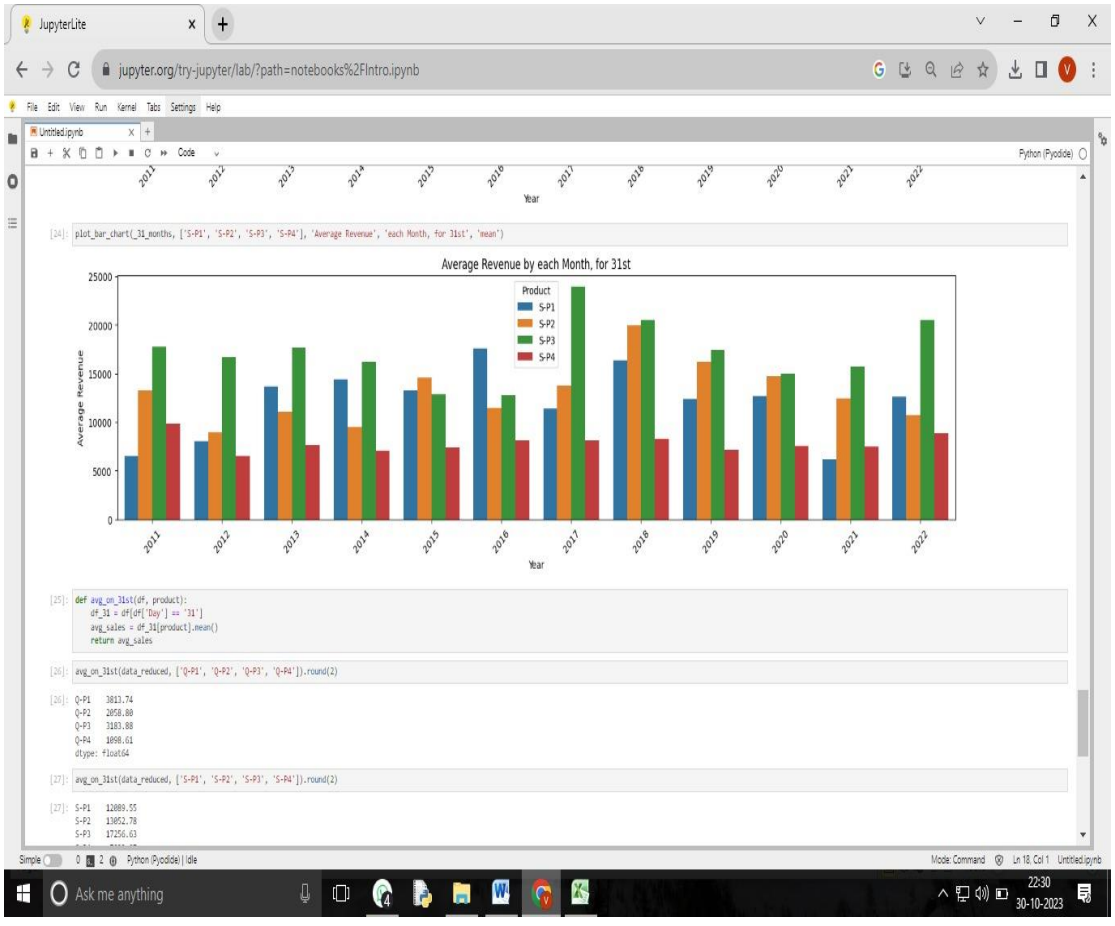


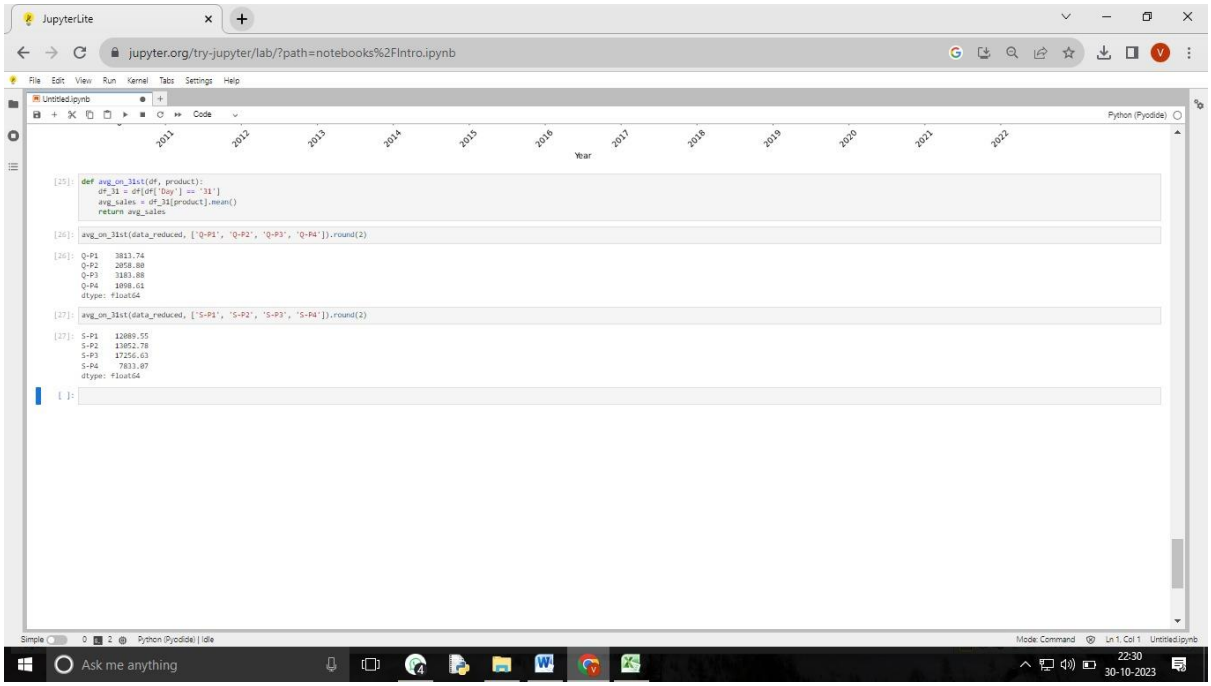
Here is the another plot which represents the trend in sales of all four products by month and a datasets which consists of 84 rows and 12 columns . A bar chart which represents an average units by each month for 31st and an average revenue by each month for 31st with visualization.






JupyterLite
jupyter.org/try-jupyter/lab/?path=notebooks%2FIntro.ipynb
File Edit View Run Kernel Tabs Settings Help
Untitled.ipynb
Code
Python (Pyodide)
4352 31-05-2022 3669 2710 3067 1593 11630.73 17181.40 16623.14 11358.09 31 05 2022
4413 31-07-2022 1437 833 1867 1270 4555.29 5281.22 10119.14 9055.10 31 07 2022
4443 31-08-2022 1035 1639 3658 841 3280.95 10391.26 19826.36 5996.33 31 08 2022
4474 31-9-2022 6964 1873 5481 1336 22075.88 11874.82 29707.02 9525.68 31 09 2022
4535 31-11-2022 4600 2006 3795 1426 14582.00 12718.04 20574.32 10167.38 31 11 2022
84 rows × 12 columns
[23]: plot_bar_chart(_31_months, ['Q-P1', 'Q-P2', 'Q-P3', 'Q-P4'], 'Average Units', 'each Month, for 31st', 'mean')
Average Units by each Month, for 31st
Product
Q-P1
Q-P2
Q-P3
Q-P4
Average Units
Year
[24]: plot_bar_chart(_31_months, ['S-P1', 'S-P2', 'S-P3', 'S-P4'], 'Average Revenue', 'each Month, for 31st', 'mean')
Average Revenue by each Month, for 31st
Product
S-P1
S-P2
S-P3
S-P4
Simple
Python (Pyodide) | Idle
Mode: Command
Ln 18, Col 1
Untitled.ipynb
Ask me anything
22:30
30-10-2023


JupyterLite
jupyter.org/try-jupyter/lab/?path=notebooks%2FIntro.ipynb
File Edit View Run Kernel Tabs Settings Help
Untitled.ipynb
Code
Python (Pyodide)
Year
[24]: plot_bar_chart(_31_months, ['S-P1', 'S-P2', 'S-P3', 'S-P4'], 'Average Revenue', 'each Month, for 31st', 'mean')
Average Revenue by each Month, for 31st
Product
S-P1
S-P2
S-P3
S-P4
Average Revenue
Year
[25]: def avg_on_31st(df, product):
df_31 = df[df['Day'] == '31']
avg_sales = df_31[product].mean()
return avg_sales
[26]: avg_on_31st(data_reduced, ['Q-P1', 'Q-P2', 'Q-P3', 'Q-P4']).round(2)
[26]: Q-P1 3813.74
Q-P2 2058.88
Q-P3 3183.88
Q-P4 1898.61
dtype: float64
[27]: avg_on_31st(data_reduced, ['S-P1', 'S-P2', 'S-P3', 'S-P4']).round(2)
[27]: S-P1 12089.55
S-P2 13052.78
S-P3 17256.63
Simple
Python (Pyodide) | Idle
Mode: Command
Ln 18, Col 1
Untitled.ipynb
Ask me anything
22:30
30-10-2023



BUILDING PRODUCT SALES ANALYSIS:

For building a product sales analysis , the process invlovedare feature engineering , model training and evaluation . Those process can involve data collection , data cleaning and preparation , analysis and insights , visualization and reporting , continuous improvement.

FEATURE ENGINEERING:

Feature engineering for product sales analysis involves creating meaningful data attributes to enhance predictive modeling and uncover insights. It involves process such as time-based features , lag features , moving averages , categorical encoding etc… Feature engineering transforms raw data into informative features, enhancing the performance of machine learning models and providing deeper insights into product sales trends.

MODEL TRAINING :

Model training for product sales analysis is the process of using historical sales data to build a predictive model. It involves data preparation, feature selection , data splitting , model selection , training , evaluation , tuning , and deployment.

EVALUATION:

Evaluating product sales analysis involves scrutinizing the results of your sales data investigation. It includes performance metrices , trends and patterns , customer segmentation , pricing and promotion , inventory management , marketing campaign and competitor comparison .

CONCLUSION:

In conclusion, the product sales analysis project provides valuable insights into our sales performance. Through rigorous data analysis, we have identified key trends, customer preferences, and market demand patterns. This analysis equips us with the knowledge to make informed decisions, optimize our product offerings, and enhance customer satisfaction. Moving forward, it is crucial to leverage these findings to formulate targeted marketing strategies, improve inventory management, and strengthen customer relationships. By aligning our efforts with the data-driven recommendations outlined in this report, we can drive sustainable growth, increase revenue, and maintain a competitive edge in the market.