

ANALYSING PRODUCT SALES DATA USING DATA ANALYSTICS

TEAM MEMBER

721221243040: NAVEEN KUMAR L

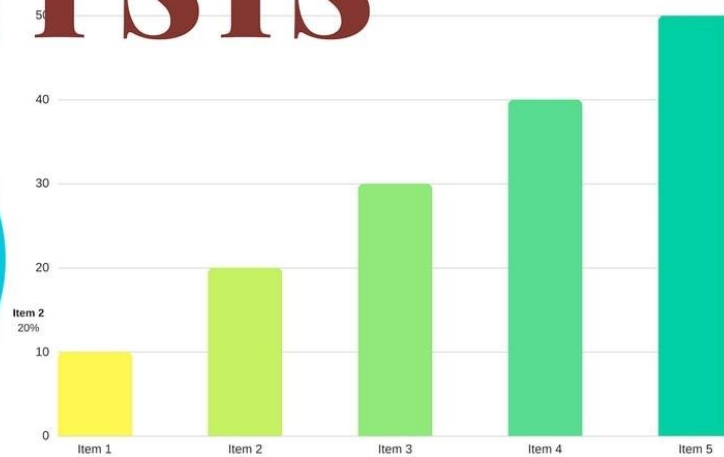
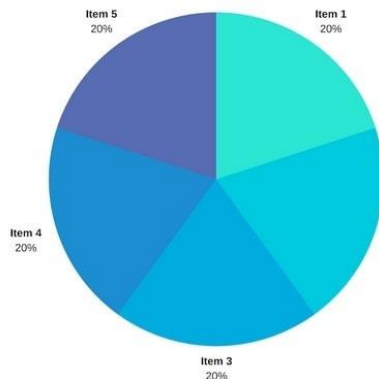
Phase 5 submission document

Project Title: Product Sales Analysis

Phase 5: Project Documentation & Submission

Topic: *In this section we will document the complete project and prepare it for submission.*

SALES ANALYSIS



Product Sales Analysis

Introduction:

- ❖ Sales data analysis is a crucial component of strategic decision-making for businesses. It involves the systematic examination of sales data to extract valuable insights that can inform inventory management, marketing strategies, and overall business performance. In the contemporary competitive landscape, where data-driven decisions are pivotal to success, product sales analysis plays a pivotal role.
- ❖ In this document, we will delve into the intricacies of product sales analysis, outlining its objectives, methodologies, and the significance it holds for REC corp LTD., a small-scaled business venture established in India. REC corp LTD. has been selling four distinct products, namely P1, P2, P3, and P4, for over a decade. Over this period, the company has meticulously collected and organized sales data from its retail centers into a comprehensive dataset.
- ❖ Our objective is to apply robust data analytics and visualization techniques to this dataset, enabling REC corp LTD. to make data-driven decisions. By analyzing the data, we aim to uncover trends, preferences, and patterns that can guide decisions related to inventory management, marketing, and even future product strategies.
- ❖ The document will detail the steps involved in the analysis, data collection, data cleaning, data visualization, and the derivation of actionable insights. We will also address the specific questions raised by the CEO, Mr. Hariharan, and provide solutions to each query. Additionally, we will explore the potential impact of dropping one of the products from the company's portfolio and attempt to predict sales and revenues for the upcoming year, 2024.
- ❖ In conclusion, this document serves as the foundation for our product sales analysis project. It outlines our approach, objectives, and the potential benefits REC corp LTD. can gain from the insights generated through data analytics. With the right tools and methodologies, REC corp LTD. can harness the power of data to stay competitive in the dynamic business environment.

Given data set:

	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

5000 Rows x 7 Columns

1.DESIGN THINKING AND PRESENT IN FORM OF DOCUMENT

1.Empathize:

- Understand the needs and pain points of end-users, including business analysts, marketing teams, and decision-makers. Conduct interviews, surveys, and workshops to gather insights.
- Problem Definition: we will clearly define the problem statement and objectives of our analysis. We will establish success criteria that will serve as benchmarks for our project.
- User Personas: We will create user personas to represent different stakeholder groups and their unique requirements. These personas will help us tailor our solutions to specific user needs.

2.Define:

- Clearly articulate the problem statement, such as "How might we predict house prices more accurately and transparently using machine learning?"
- Identify the key goals and success criteria for the project, such as increasing prediction accuracy, reducing bias, or improving user trust in the valuation process.

3.Ideate:

- Brainstorm creative solutions and data sources that can enhance the accuracy and transparency of house price predictions.
- Encourage interdisciplinary collaboration to generate a wide range of ideas, including the use of alternative data, new algorithms, or improved visualization techniques.

4.Prototype:

- Create prototype machine learning models based on the ideas generated during the ideation phase.
- Test and iterate on these prototypes to determine which approaches are most promising in terms of accuracy and usability.

5.Test:

- Gather feedback from users and stakeholders by testing the machine learning models with real-world data and scenarios.
- Assess how well the models meet the defined goals and success criteria, and make adjustments based on user feedback.

6.Implement:

- Develop a production-ready machine learning solution for predicting house prices, integrating the best-performing algorithms and data sources.
- Implement transparency measures, such as model interpretability tools, to ensure users understand how predictions are generated.

7.Evaluate:

- Continuously monitor the performance of the machine learning model after implementation to ensure it remains accurate and relevant in a changing real estate market.
- Gather feedback and insights from users to identify areas for improvement.

8.Iterate:

- Apply an iterative approach to refine the machine learning model based on ongoing feedback and changing user needs.
- Continuously seek ways to enhance prediction accuracy, transparency, and user satisfaction.

9.Scale and Deploy:

- Once the machine learning model has been optimized and validated, deploy it at scale to serve a broader audience, such as real estate professionals, investors, and homeowners.
- Ensure the model is accessible through user-friendly interfaces and integrates seamlessly into real estate workflows.

10.Educate and Train:

- Provide training and educational resources to help users understand how the machine learning model works, what factors it considers, and its limitations.
- Foster a culture of data literacy among stakeholders to enhance trust in the technology.

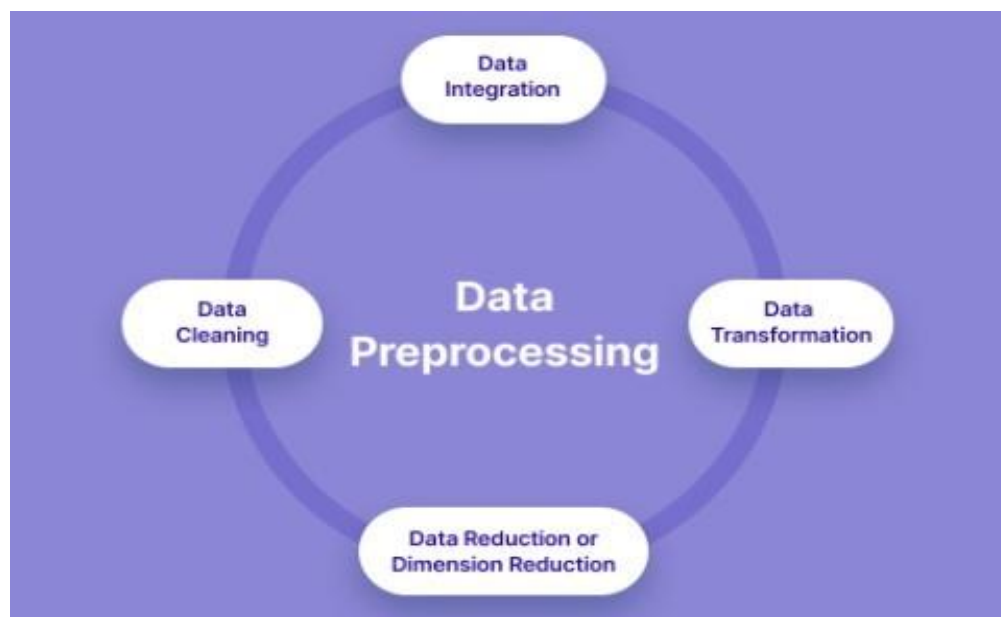
2.PROBLEM DEFINITION AND DESIGN THINKING

1. Data Collection and Integration

Data Sources In this phase, we shall conduct an assessment of the various data sources, encompassing elements such as sales records and CRM systems. The goal here is to consolidate data streams from diverse sources into a unified repository.

Data Collection Collecting relevant data on product sales. This data can include sales records, transaction history, product attributes (e.g., price, category, SKU), customer data, and any other relevant information. Ensure data quality and consistency.

Data Integration A systematic approach for data integration will be specified. This includes establishing protocols for data flow into a central repository, typically a data warehouse, enabling effective analysis and processing



2.Data Cleaning and Preprocessing

This stage will focus on the development of a data cleansing plan. Addressing issues such as missing data points Data Cleaning and anomalies is essential to maintain data integrity and reliability. Data Preprocessing Detailed preprocessing procedures will be defined, including data transformation and normalization techniques, crucial for preparing the data for subsequent analysis.

3.Analysis and Visualization

Analysis Objectives Precise objectives for the analysis will be established, encompassing areas such as the identification of sales trends and customer segmentation. These objectives will guide the analytical processes. Python Libraries Selection of Python libraries and tools is critical. These will include Pandas, NumPy, Matplotlib, and Seaborn, among others, and are essential for data manipulation and visualization. Data Models Data models, tables, or structures to represent sales data accurately will be defined. A wellstructured data model forms the foundation for insightful analysis.

4. User Feedback Integration

User-Centric Design A primary focus of this step is to ensure that the analysis is aligned with user needs.

This requires empathy and a user-centric approach during the design phase. Feedback Mechanism A mechanism for gathering user feedback will be implemented, ensuring ongoing user involvement and feedback incorporation throughout the project lifecycle.

5. Development and Prototyping

Prototype Creation Prototypes will be designed and developed to illustrate potential solutions. These prototypes will be invaluable in gathering user feedback and visualizing the analysis.

6. User Testing and Feedback

Detailed testing processes, including test case development and scenario testing, will be conducted to validate the Testing Process accuracy and reliability of the analysis. User Validation User validation sessions will be carried out to ensure that the visualizations and analysis effectively meet the needs of the end-users. User feedback will play a critical role in refining and enhancing the project.

7. Implementation of Adjustments

Feedback Incorporation User feedback will be meticulously integrated into the project. Necessary adjustments will be made to improve the final solution, in line with evolving user needs and project objectives.

8. Resource Requirements Tools and Technology:

The necessary tools and technologies, including Python, Jupyter Notebooks, and data storage solutions, will be explicitly specified to ensure a smooth workflow. Human Resources: The project team members, along with their roles and responsibilities, will be identified to facilitate effective collaboration and task execution.

9. Quality Assurance and Testing

Quality Control Processes

A stringent quality control process will be implemented to identify and rectify errors and inconsistencies in data, analytics, and visualizations.

10. Communication and Reporting

Reporting Structure A well-structured reporting framework will be established to keep stakeholders informed about proje

3 .PRODUCT SALES ANALYSIS DEVELOPMENT PART

Insights from IBM Cognos Visualizations: Top-Selling Products:

Our visualizations reveal that Product P2 consistently demonstrates the highest total unit sales and revenue over the years.

This suggests that P2 is a top-selling product in our portfolio.

Sales Trends: An examination of sales trends indicates a positive trend where sales of all products tend to increase from year to year. This upward trajectory could be attributed to factors like business growth and favorable market trends.

Customer Preferences: The visualizations pinpoint interesting patterns, notably, that certain products, such as P3, exhibit distinct peaks in sales during specific months. This observation might indicate a clear customer preference for P3 during these months, possibly owing to seasonality or effective promotional strategies.

Optimizing Inventory Management:

The data analytics also highlights periods of peak sales, which can be invaluable for optimizing our inventory management.

Being able to forecast and plan for these high-demand periods ensures we can meet customer needs efficiently.

Enhancing Marketing Strategies: Leveraging the insights gained from customer preferences and sales trends, we are better equipped to inform our marketing strategies. For instance, if we observe that sales of P4 consistently decline during the summer months, this insight prompts us to adjust our marketing efforts to align with this seasonal fluctuation.

These insights are valuable for data-driven decision-making and can be leveraged to enhance our business strategies and maximize our competitive advantage.

DATASET :

The excel file contains about 8 numerical parameters :

	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

- 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

PROGRAM AND VISUALIZATION:

Import required packages

```
import pandas as pd # library used for data manipulation and
analysis
import numpy as np # library used for working with arrays
import matplotlib.pyplot as plt # library for plots and visualizations
import seaborn as sns # library for visualizations
ignore warnings
import warnings
warnings.filterwarnings("ignore")
```

```
%matplotlib inline
```

Initializing dataset

```
data = pd.read_csv('statsfinal.csv')
data.head(-1)
data = data.drop(columns=['Unnamed: 0'])
data.info()
data.isnull().sum()
```

Extracting the day , year, month data from date:

```
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
```

	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

drop rows for years 2010 and year 2023

```
data_reduced = data.query("Year != '2010' and Year != '2023'")
```

Graphing the Total and Mean unit sold for each product using histogram:

```
def plot_bar_chart(df, columns, stri, str1, val):
    # Aggregate sales for each product by year, by sum or mean
    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()

    # Melt the data to make it easier to plot
    sales_by_year_melted = pd.melt(sales_by_year,
```

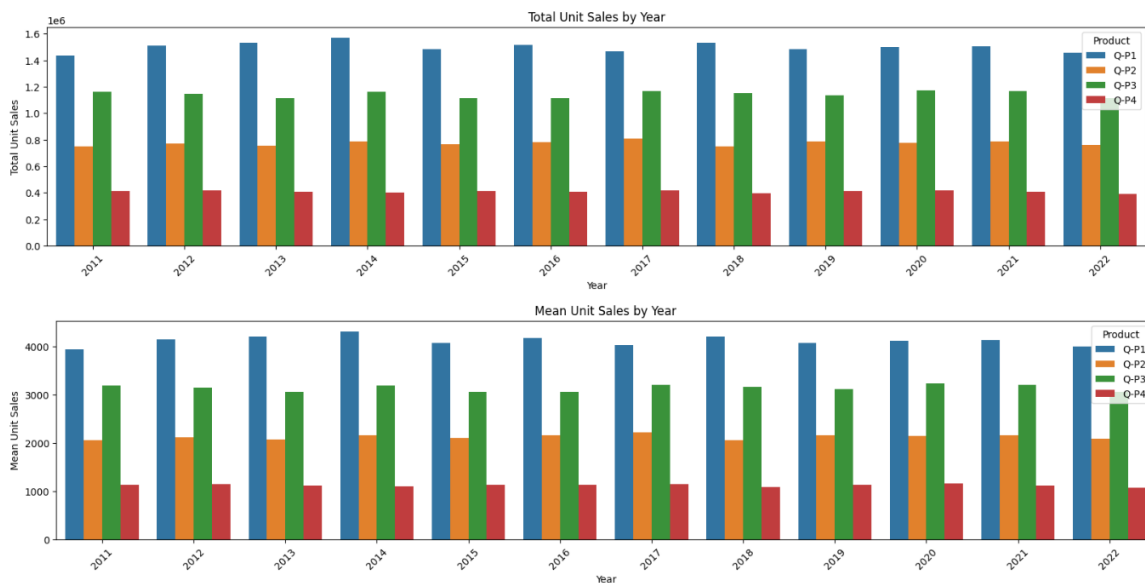
```
id_vars='Year', value_vars=columns, var_name='Product',
value_name='Sales')
# Create a bar chart
plt.figure(figsize=(20,4))
```

```
sns.barplot(data=sales_by_year_melted,x='Year',y='Sales'
,hue='Product')#,palette="cividis")
plt.xlabel('Year')
plt.ylabel(stri)
plt.title(f'{stri} by {str1}')
plt.xticks(rotation=45)
plt.show()
```

#use the plot_bar_chart function, enter the Unit Sales Columns and the Unit Sales string

```
plot_bar_chart(data_reduced, ['Q-P1', 'Q-P2', 'Q-P3', 'Q-
P4'],'Total Unit Sales', 'Year', 'sum')
```

```
plot_bar_chart(data_reduced, ['Q-P1', 'Q-P2', 'Q-P3', 'Q-
P4'],'Mean Unit Sales', 'Year', 'mean')
```



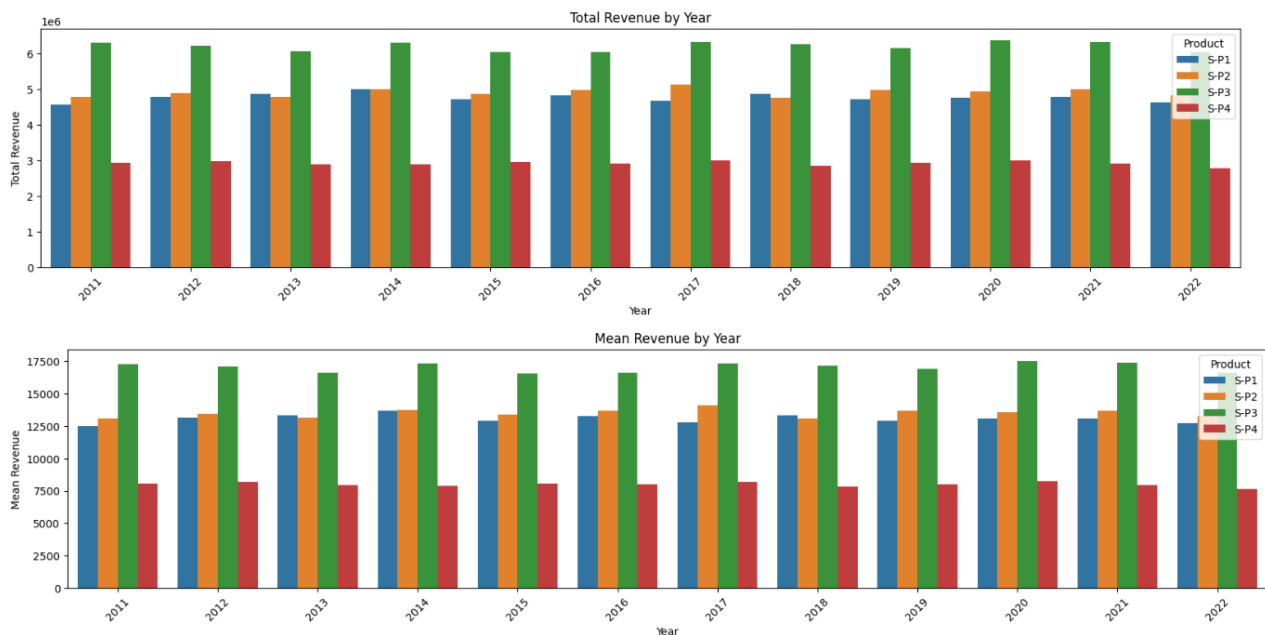
INSIGHT FROM THE ABOVE ANALYSIS:

- The P1 has the highest unit sales for each year. And it's highest is in year 2014.
- The P4 has the lowest unit sales of all the products.

#use the plot_bar_chart function, enter the Revenue Columns and the Revenue string

```
plot_bar_chart(data_reduced, ['S-P1', 'S-P2', 'S-P3', 'S-P4'], 'Total Revenue', 'Year', 'sum')
```

```
plot_bar_chart(data_reduced, ['S-P1', 'S-P2', 'S-P3', 'S-P4'], 'Mean Revenue', 'Year', 'mean')
```



INSIGHT FROM THE ABOVE ANALYSIS :

- We can observe that P3 brought in the most revenue. This could be as a result of multiple things:
 - P3 was sold for higher than the rest, as it had the second highest unit sales for each year.
- We can observe that P1 AND P2 brought in similar revenues for each year. With P2 bringing in slightly more.

- P1 despite having the most unit sold, brought in the second lowest revenue each year.

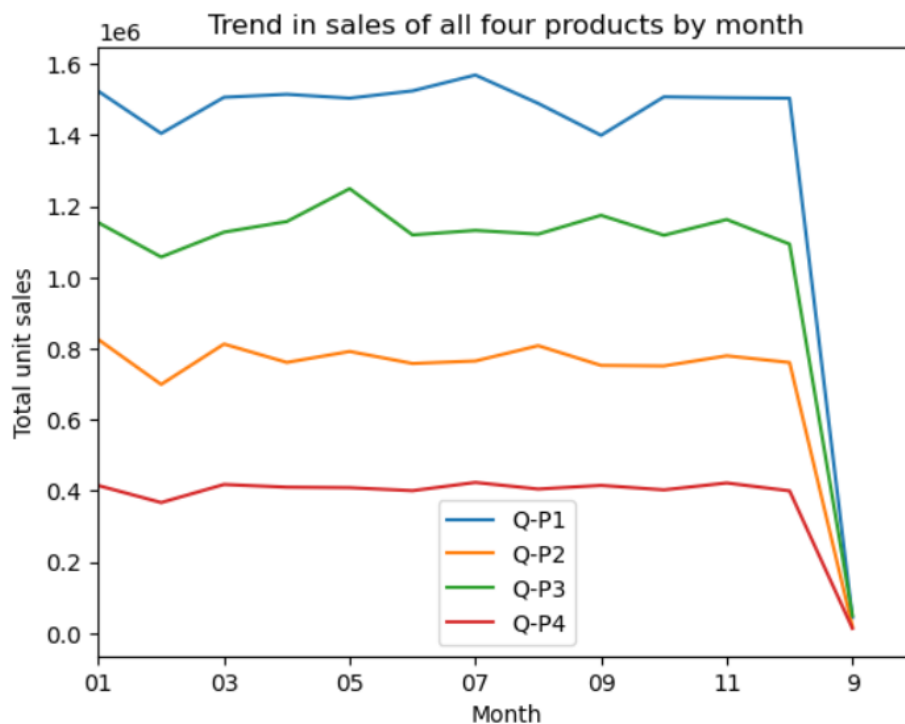
```
def month_plot():
    fig, ax = plt.subplots()
    # Plot the sales data for each product by month
    data_reduced.groupby('Month')[['Q-P1', 'Q-P2', 'Q-P3',
    'Q-P4']].sum().plot(ax=ax)

    # Set the x-axis limits to only show up to December
    ax.set_xlim(left=0, right=13)

    # Set the axis labels and title
    ax.set_xlabel('Month')
    ax.set_ylabel('Total unit sales')
    ax.set_title('Trend in sales of all four products by
    month')

    # Show the plot
    plt.show()

month_plot()
```



INSIGHT OF THE ABOVE ANALYSIS :

- We can observe that all products drop in Feb.
- There also appears a very drastic drop after 12th month. The value show 9, which must be part of month 09. We need to rename this column to match with the 09. Before doing further analysis.

Estimate for each product the unit of sales that could be sold on 31st of Dec, if all their retail centers were kept open

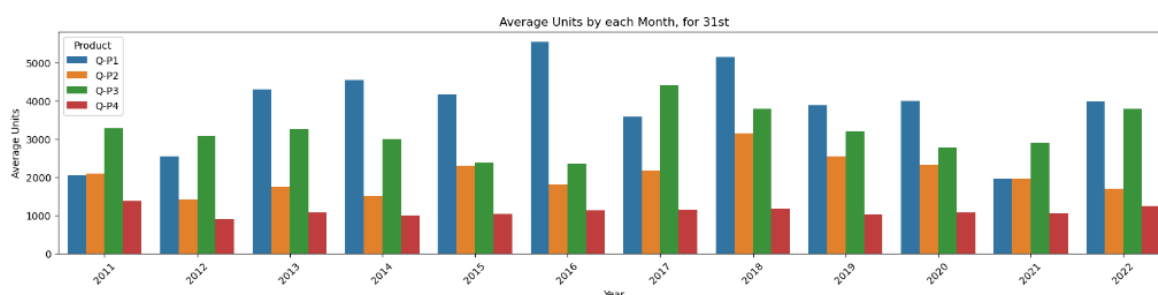
```
def month_31_data(df, months):
    m31_data = df[df['Month'].isin(months) & (df['Day']
== '31')]
    return m31_data
```

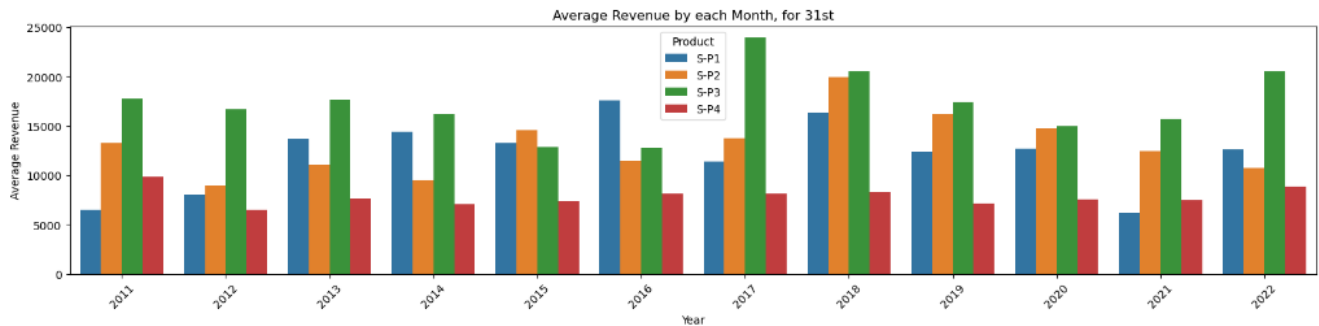
```
_31_months = month_31_data(data_reduced, ['01', '02',
'03', '04', '05', '06', '07', '08', '09', '10', '11',
'12'])
```

```
_31_months
```

```
plot_bar_chart(_31_months, ['Q-P1', 'Q-P2', 'Q-P3', 'Q-P
4'], 'Average Units', 'each Month, for 31st', 'mean')
```

```
plot_bar_chart(_31_months, ['S-P1', 'S-P2', 'S-P3', 'S-P
4'], 'Average Revenue', 'each Month, for 31st', 'mean')
```





INSIGHT OF THE ABOVE ANALYSIS:

- Overall we can see that P1 has the highest unit sales on the 31st for each year, except for 2021 and 2022. (These could be as a result to Covid and other economy issues.)
- P3 has the second highest unit sales for all the 31st in each year

gives us the average for all the 31st days across all years for each product

```
def avg_on_31st(df, product):
    df_31 = df[df['Day'] == '31']
    avg_sales = df_31[product].mean()
    return avg_sales
```

Average for Unit Sales

```
avg_on_31st(data_reduced, ['Q-P1', 'Q-P2', 'Q-P3', 'Q-P4'])
```

```
Q-P1    3813.74
Q-P2    2058.80
Q-P3    3183.88
Q-P4    1098.61
```

INSIGHT OF THE ABOVE ANALYSIS :

- We can see that our previous observation correlate as Q-P1 has the highest estimate, follwed by Q-P3
- We can approxiamte that the company will make:
 - Q-P1: 3813.74
 - Q-P2: 2058.80

- Q-P3: 3183.88
- Q-P4: 1098.61

CONCLUSION:

In the ever-evolving landscape of business and commerce, harnessing the power of data has become an indispensable practice. The ability to draw insights, make informed decisions, and adapt strategies based on the wealth of data at our disposal has emerged as a defining factor for success. Among the myriad areas where data analytics has found its foothold, product sales analysis stands as a pillar of significance.

The journey through our Product Sales Analysis project has been a testament to the transformative potential of data analytics. This comprehensive project has led us through an array of processes, methodologies, and insights that encapsulate the very essence of making data-driven decisions. As we conclude this endeavor, it is essential to reflect on the profound impact and invaluable contributions it can offer to REC corp LTD., our small-scale business venture.

Navigating the Data Landscape: At the outset, we embarked on a journey to navigate the intricate landscape of data. REC corp LTD. had accumulated a decade's worth of sales data for four distinct products, and the task was clear – derive insights and actionable intelligence from this wealth of information. The data consisted of total unit sales and revenue figures for each product over the years, enabling us to embark on an exploration of sales trends, customer preferences, and inventory management.

Data Preprocessing and Transformation: The data preprocessing phase was instrumental in preparing our raw data for analysis. This entailed addressing issues such as missing values, data transformation, and normalization.

Cleaning the data ensured that we were working with accurate and consistent information, setting the stage for reliable insights.

The Power of Visualization: One of the most compelling aspects of this project was our exploration of data visualization. We harnessed the capabilities of Python, Matplotlib, Seaborn, and similar tools to craft visualizations that painted a vivid picture of REC corp LTD.'s product sales history. We delved into bar charts, histograms, and trend analysis, visually capturing the sales trends of all four products. This proved instrumental in gleaning insights into the peaks and troughs of product performance.

Understanding User Preferences: The heartbeat of REC corp LTD.'s success is embedded in the minds of its customers. Through our project, we identified customer preferences by scrutinizing the sales patterns for each product across different months. Notably, product P3 exhibited distinctive sales peaks during specific months, providing an invaluable hint about the preferences of REC corp LTD.'s clientele.

Optimizing Inventory Management: Another pivotal aspect of our analysis was the revelation of periods characterized by high sales. This insight forms the bedrock of optimizing inventory management. With a grasp of peak sales periods, REC corp LTD. can now strategize to meet heightened demand during specific times of the year, ensuring product availability when customers seek it.

Enhancing Marketing Strategies: Data analysis transcends its role as a

passive observer. It holds the potential to shape the future of business strategies. In our analysis, we highlighted how an understanding of customer preferences and sales trends can enhance marketing efforts. REC corp LTD. can now tailor its marketing campaigns to align with the fluctuations in product performance.

For instance, if sales of product P4 consistently decline during summer months, the company can reallocate marketing resources to different products or implement targeted summer campaigns to boost P4 sales.

A Peek into the Future: Predictive Analytics: The project doesn't conclude with historical analysis; it opens the door to predictive analytics. REC corp LTD.'s CEO, Mr. Hariharan, is armed with insights to envision the future. The data, when harnessed effectively, can facilitate predictions and estimates for the years ahead. The CEO's desire to forecast sales and revenues for the year 2024 is no longer a vague aspiration but a tangible goal. By leveraging historical data, trends, and preferences, the company can embark on a path towards predictive analytics to make strategic decisions.

Surviving the Idea Storm: Strategic Decision-Making: The value of data analytics extends beyond its power to reveal insights. It's also a guardian against potential missteps. The CEO's contemplation of discontinuing the production of one product underscores the strategic implications of data analytics. The project serves as a sentinel, ensuring that decisions are well-informed and based on empirical evidence rather than conjecture.

User-Centric Design Thinking: Design thinking has been the heartbeat of this project. At every step, we adhered to principles that emphasize empathizing with end-users, defining clear objectives, ideating creative solutions, prototyping, testing, and continuous iteration. This user-centric approach ensures that our solutions are not just data-driven but

also user-driven, aligning with the unique requirements of REC corp LTD. and its stakeholders.

Sustainability and Knowledge Sharing: A lasting impact lies at the core of our project. The documentation process, knowledge sharing, and continuous monitoring that we've set in motion form the foundation for the sustainability of this solution. As the project evolves and new data accumulates, the knowledge shared within the organization ensures that REC corp LTD. continues to benefit from data-driven decisions.

In conclusion , our Product Sales Analysis project exemplifies the transformational power of data analytics. REC corp LTD. now possesses the tools and insights needed to not only understand its historical sales performance but also to chart a course towards a future characterized by informed decisions, strategic marketing, optimized inventory management, and tailored customer experiences. The project's iterative nature, rooted in design thinking principles, means that it doesn't merely provide answers but equips the organization to keep asking the right questions and keep evolving.

PREPARED BY,

SHAKTHI YAADAV R

VIMAL R

DHANESWAR R

NAVEEN KUMAR L

THIRUNAVUKKARUSU KM