

PHASE 5

PROJECT DOCUMENTATION AND SUBMISSION

Date	01-11-2023
Team ID	715
Project Name	Product Sales Analysis

PROJECT TITLE: PRODUCT SALES ANALYSIS

S.NO	TABLE OF CONTENT	PG.NO
1	Project Overview	2
2	Problem Statement	2
3	Problem Identified	2
4	Introduction	2
5	Problem Objective	3
6	Literature Survey	4
7	Design Thinking Approach	7
8	Tools and Libraries	9
9	Design and Innovation Strategies	10
10	Data Visualization	12
11	Data Visualization With IBM Cognos	15
12	How Insight Improve User Experience:	21
13	Conclusion	22

1. PROJECT OVERVIEW:

The goal is to analyze sales data for insights such as identifying top-selling products, sales trends, and customer preferences. Data will be collected from various sources, integrated, and visualized using IBM Cognos. The derived insights will inform inventory management and marketing strategies, optimizing stock levels, pricing, and product development to maximize profitability and customer satisfaction.

2. PROBLEM STATEMENT:

Objective: The objectives of this project are to identify top-selling products, pinpoint peak sales periods, analyze customer preferences, optimize inventory management, and enhance marketing strategies to improve business performance.

Data: The project involves analyzing sales data using IBM Cognos, identifying top-selling products, peak sales periods, and customer preferences, and designing relevant visualizations.

3. PROBLEM IDENTIFIED:

The problem identified is a lack of informed decision-making in inventory management and marketing strategies. This results from the absence of systematic analysis of sales data to identify top-selling products, peak sales periods, and customer preferences. Without these insights, businesses may struggle with inefficient resource allocation and suboptimal marketing efforts, potentially leading to missed opportunities and decreased profitability.

4. INTRODUCTION:

Product sales analysis is a vital tool for businesses in today's competitive market. It helps identify top-performing products, understand sales trends, and understand customer preferences. This data-driven approach aligns with market dynamics and consumer expectations, enabling businesses to take precise actions. A robust product sales analysis balances supply and demand; ensuring products are readily available and preventing overstocking. It also allows businesses to craft compelling campaigns, fostering customer engagement and driving sales growth. In conclusion, product sales analysis is a key component of contemporary business strategy, enabling businesses to adapt, innovate, and make strategic decisions that strengthen their competitive advantage in a constantly changing marketplace.

The project uses IBM Cognos to analyze sales data to identify top-selling products, peak sales periods, and customer preferences. The aim is to improve inventory management and marketing strategies, reduce overstocking, and reduce customer dissatisfaction. The structured approach includes data collection, visualization design, and actionable insights. This data is crucial for businesses in the retail and e-commerce industry to understand and optimize sales

performance. By analyzing top-selling products, peak sales periods, and customer preferences, businesses can shape inventory management strategies, drive more effective marketing efforts, make informed decisions, allocate resources efficiently, and stay ahead of the competition.

5. PROJECT OBJECTIVES:

The project objectives for the problem statement "Analyzing sales data to identify top-selling products, peak sales periods, and customer preferences, aiding in inventory management and marketing strategies" are as follows:

Identify Top-Selling Products:

- This objective involves determining which products in our inventory are the most popular in terms of sales. You want to identify the items that contribute the most to our overall revenue or have the highest sales volume. This information helps you focus on resources and marketing efforts on the products that are performing well.

Determine Peak Sales Periods:

- The goal here is to pinpoint the specific time periods when sales experience a significant increase or "peak." This could be on a daily, weekly, monthly, or seasonal basis. Understanding these peak periods helps you plan for increased demand, adjust staffing levels, and optimize inventory to avoid stockouts or overstock situations.

Analyze Customer Preferences:

- This objective is all about gaining insights into what customers prefer. It involves analyzing customer behavior, demographics, and buying patterns. For example, you might discover that certain customer segments prefer particular product categories, price points, or delivery options. This data helps you tailor products and services to better meet customer expectations.

Inform Marketing Strategies:

- Marketing strategies can be fine-tuned with insights from the sales data analysis. For example, if you know which products are top sellers and when they sell best, you can create targeted marketing campaigns around those products during peak periods. This can lead to higher conversion rates and increased sales.

Improve Decision-Making:

- The overall goal of the project is to improve the decision-making processes within business. By providing data-driven insights, you enable decision-makers to make more

informed choices regarding inventory levels, product offerings, and marketing strategies. This leads to more efficient and profitable operations.

Enhance Business Competitiveness:

- Understanding customer preferences and responding to peak sales periods can give business a competitive edge. Being able to anticipate and meet customer demands more effectively than competitors can set business apart in the market.

Increase Sales and Revenue:

- Ultimately, the primary aim of this project is to contribute to the growth of business by increasing sales and revenue. By using the insights from the analysis to optimize product offerings, marketing strategies, and inventory management, you can drive higher sales and, in turn, generate more revenue.

The objectives outline a roadmap for sales analysis project, guiding the collection, cleaning, and analysis of sales data to extract valuable insights for informed business decisions and strategies, thereby maximizing business's potential and maintaining market competitiveness.

6. LITERATURE SURVEY:

Title 1: "An Analysis and Forecasts of Online Product Sales Based on BP Neural Network and Pearson Coefficient"

This paper presents a comprehensive analysis and forecasting approach for online product sales using the Back propagation (BP) Neural Network and the Pearson Coefficient. The study aims to improve our understanding of factors influencing online sales and develop a predictive model for future sales trends. BP Neural Networks are used for pattern recognition and prediction, capturing complex relationships in sales data. The Pearson Coefficient is introduced as a measure of linear correlation between variables, assessing the strength and direction of the relationship between various sales-related variables. The research methodology involves collecting historical sales data, including product price, advertising spending, seasonality, and customer demographics, to train the BP Neural Network to recognize patterns and relationships. The results of the analysis and forecasting model demonstrate how the combined use of BP Neural Networks and Pearson Coefficient can provide valuable insights into online product sales, leading to more accurate sales predictions and helping businesses optimize their strategies and resource allocation. The paper highlights the potential of combining BP Neural Networks and the Pearson Coefficient to gain a deeper understanding of online product sales and make informed forecasts.

Title 2: "Planning Product Configurations Based on Sales Data"

This paper presents a methodology for planning product configurations based on sales data analysis, aiming to help businesses optimize their offerings by tailoring them to meet customer demand and preferences effectively. The paper emphasizes the importance of offering products that align with customer preferences and market trends, and the challenges companies face in determining the most attractive configurations. To address these challenges, a data-driven approach is introduced, leveraging historical sales data and customer behavior analysis. The core of the paper introduces a planning framework using statistical and machine learning techniques to analyze sales data, identifying trends, patterns, and correlations that inform product configuration decisions. Data visualization and reporting tools are also discussed, aiding in identifying popular product features and optimal price points. The paper also discusses how to adapt product configurations based on changing market conditions and customer preferences, exploring dynamic pricing and agile product development approaches. The benefits of planning product configurations based on sales data include improved customer satisfaction, increased sales, and better resource allocation.

Title 3: "Developing a Product Pricing and Marketing Analysis System on the Website"

This paper presents a system for analyzing product pricing and marketing strategies on a website, aiming to optimize online product offerings and marketing efforts. It emphasizes the importance of pricing and marketing strategies in online business success and the challenges companies face in setting competitive prices and designing effective marketing campaigns. The system combines data analytics, web development, and user interface design to collect and analyze data related to product pricing, customer behavior, website traffic, and marketing campaigns. The technical aspects of the system include programming languages, databases, and data collection methods. The system tracks product sales, monitors customer interactions, and captures user feedback. The paper also discusses algorithms and models used for pricing analysis and marketing campaign assessment, covering aspects such as price elasticity, demand forecasting, customer segmentation, and marketing channel effectiveness. User interface design and data visualization are crucial components of the system, providing insights and recommendations to business users in a user-friendly format. The paper highlights the potential benefits of the developed system for businesses operating in the online marketplace, such as enhancing competitiveness, increasing sales, and improving customer satisfaction.

Title 4: "Survey on Customer Centric Sales Analysis and Prediction"

This paper offers a comprehensive analysis of customer-centric sales analysis and prediction, focusing on strategies and tools used by businesses to understand and serve their customers. It highlights the shift towards customer-centric approaches, emphasizing the importance of leveraging data and analytics to gain deeper insights into customer behavior. The paper explores various methods and technologies used in customer-centric sales analysis and prediction, including data collection through customer touch points, data integration, and data quality management. It also discusses the role of CRM systems, data analytics platforms, and machine learning models in extracting valuable insights from customer data. The paper also discusses the different aspects of customer-centric sales analysis, such as segmentation, customer lifetime value analysis, and churn prediction, and how businesses can use these techniques to identify valuable customers, personalize marketing efforts, and prevent customer attrition. It also discusses ethical considerations related to customer data privacy and compliance with data protection regulations, such as GDPR. The paper emphasizes the significance of customer-centric sales analysis and prediction in the modern business landscape, highlighting how understanding and anticipating customer needs can lead to higher sales, customer loyalty, and a competitive edge.

Title 5: "Real-Time Sales Data Analysis"

The paper discusses the significance of real-time sales data analysis in today's rapidly changing business environment. It discusses the technical and analytical aspects of this process, including data collection methods, data streaming, and the use of advanced analytics tools. The benefits of real-time sales data analysis include immediate insights into customer behavior, sales trends, inventory management, and predictive analytics for forecasting sales, optimizing pricing strategies, and personalizing marketing campaigns. The paper also discusses various use cases across industries, such as e-commerce tracking website traffic and monitoring product sales, and brick-and-mortar retailers managing inventory and staffing levels. The paper emphasizes the transformative potential of real-time sales data analysis for businesses, highlighting how timely insights can lead to improved decision-making, enhanced customer experiences, and increased revenue.

7. DESIGN THINKING APPROACH:

Empathize:

- ✓ Start by empathizing with the businesses and stakeholders involved. Understand their pain points, needs, and goals. Conduct interviews, surveys, and workshops with key stakeholders like sales teams, marketing teams, and inventory managers to gain deep insights into their challenges.

Actions:

- ✓ Conduct interviews with key stakeholders including sales teams, marketing teams, and inventory managers. Understand their pain points, challenges, and objectives related to sales data analysis.
- ✓ Observe daily operations to identify how sales data is collected and utilized. Note any bottlenecks or inefficiencies.
- ✓ Develop user personas based on characteristics and behaviors of stakeholders.

Define:

- ✓ Clearly define the problem statement based on the insights gathered during the empathize phase. In this case, it's the need to improve inventory management and marketing strategies through the analysis of sales data using IBM Cognos. Define the scope, objectives, and success criteria for the project.

Objectives:

- ✓ Clearly define the project's scope, objectives, and success criteria.
- ✓ Create a problem statement that highlights the need to enhance inventory management and marketing strategies through IBM Cognos-based sales data analysis.
- ✓ Establish key performance indicators (KPIs) for measuring project success.

Ideate:

- ✓ Encourage brainstorming and idea generation to explore potential solutions. Consider various approaches to utilizing IBM Cognos for analyzing sales data. Invite diverse perspectives to foster creative thinking. Generate ideas on how to extract actionable insights from the data.

Actions:

- ✓ Brainstorm various ways in which IBM Cognos can be used to extract insights from sales data.
- ✓ Explore creative approaches to data visualization and analysis techniques.
- ✓ Encourage collaboration among team members to generate innovative ideas for solving the defined problem.

Prototype:

- ✓ Create a prototype or preliminary plan for the project. Outline the steps involved, including data collection methods, data sources, and potential visualizations. Develop a rough sketch or mockup of how the insights might be presented to stakeholders through IBM Cognos.

Actions:

- ✓ Create a high-level project plan outlining the major steps and milestones.
- ✓ Develop a preliminary data collection and analysis framework.
- ✓ Share the initial project plan with stakeholders for feedback.

Test:

- ✓ Conduct small-scale tests or pilots to validate the chosen approach. Use sample data to test data collection, visualization design, and analysis techniques within IBM Cognos. Gather feedback from stakeholders and make necessary adjustments to the project plan.

Actions:

- ✓ Conduct a pilot analysis using a subset of data to validate the chosen approach.
- ✓ Gather feedback from stakeholders and make necessary adjustments to the project plan and analysis techniques.
- ✓ Ensure that the chosen visualizations effectively convey insights.

Implement:

- ✓ Present the final actionable insights to stakeholders in a clear and understandable manner. Provide recommendations for improving inventory management and marketing strategies based on the insights derived from IBM Cognos analysis.

Actions:

- ✓ Present the final actionable insights to stakeholders in a clear and understandable manner.
- ✓ Provide concrete recommendations for inventory management and marketing strategies based on the insights derived from IBM Cognos analysis.

Iterate:

- ✓ Continuously gather feedback from stakeholders and refine the project based on their input. Be open to adjusting analysis objectives, data collection methods, or visualization designs as new insights emerge or business needs change.

Actions:

- ✓ Continuously update the sensor network and application for reliability.
- ✓ Adapt mitigation measures based on ongoing data analysis.
- ✓ Adjust transportation route optimization strategies as traffic patterns change.

8. TOOLS AND LIBRARIES:**Python:**

- Python is a versatile, high-level programming language known for its simplicity and readability. It's widely used in data analysis, machine learning, and scientific computing due to its extensive libraries and frameworks.

Pandas:

- Pandas is a Python library for data manipulation and analysis. It provides data structures like DataFrames and Series, making it easy to work with structured data. Pandas is essential for data loading, cleaning, and transformation.

NumPy:

- NumPy is another Python library that focuses on numerical computing. It provides support for large, multi-dimensional arrays and matrices, as well as a variety of mathematical functions to operate on these arrays. It's fundamental for numerical data processing.

Matplotlib:

- Matplotlib is a popular data visualization library in Python. It allows you to create static, animated, or interactive visualizations in a wide range of formats, including line plots, bar charts, scatter plots, and more. It's excellent for visualizing data and trends.

Seaborn:

- Seaborn is a Python data visualization library built on top of Matplotlib. It provides a high-level interface for creating attractive and informative statistical graphics. Seaborn simplifies the creation of complex visualizations and is often used for creating aesthetically pleasing charts.

Scikit-Learn:

- Scikit-Learn, also known as sklearn, is a powerful machine learning library in Python. It offers a wide range of tools for data preprocessing, model selection, training, and evaluation. It's especially useful for building predictive models and performing machine learning tasks.

9. DESIGN AND INNOVATION STRATEGIES:

Define Analysis Objectives:

- Start by clearly defining the specific objectives of analysis, which include identifying top-selling products, peak sales periods, and customer preferences. Having well-defined goals will guide data collection and analysis.

Collect Sales Data:

- Gather the necessary sales data from a trusted source. This source can be organization's sales database, data provider, or any other reliable data repository. Ensure that the data includes relevant fields, such as product information, sales dates, quantities sold, prices, and customer details.

Import Libraries:

- Start by importing the required libraries. In this case, you'll use Pandas for data manipulation.

Import Libraries

```
# import the important 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

%matplotlib inline
```

Load the Dataset:

- This step involves loading Product sales dataset into Python environment. The dataset should be in a format that Pandas can easily handle, such as a CSV file.

Loading Dataset

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

Dataset Link: <https://www.kaggle.com/datasets/ksabishek/product-sales-data>

Data Preprocessing:

- Data preprocessing is crucial for ensuring the accuracy and reliability of analysis. In this phase, you should:

Data Cleaning:

- Remove or handle missing values, duplicates, and any inconsistencies in the dataset. This step is essential for data integrity.

```
data.isnull().sum()
```

Data Transformation:

- Standardize and format data elements as needed. For instance, ensure that date and time formats are consistent and consider encoding categorical data into numerical values for analysis.

Data Integration:

- If sales data is spread across multiple sources or tables, integrate them into a single dataset using common identifiers, like product IDs and customer IDs.

10. DATA VISUALIZATION:

The products are P1, P2, P3 and P4.

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

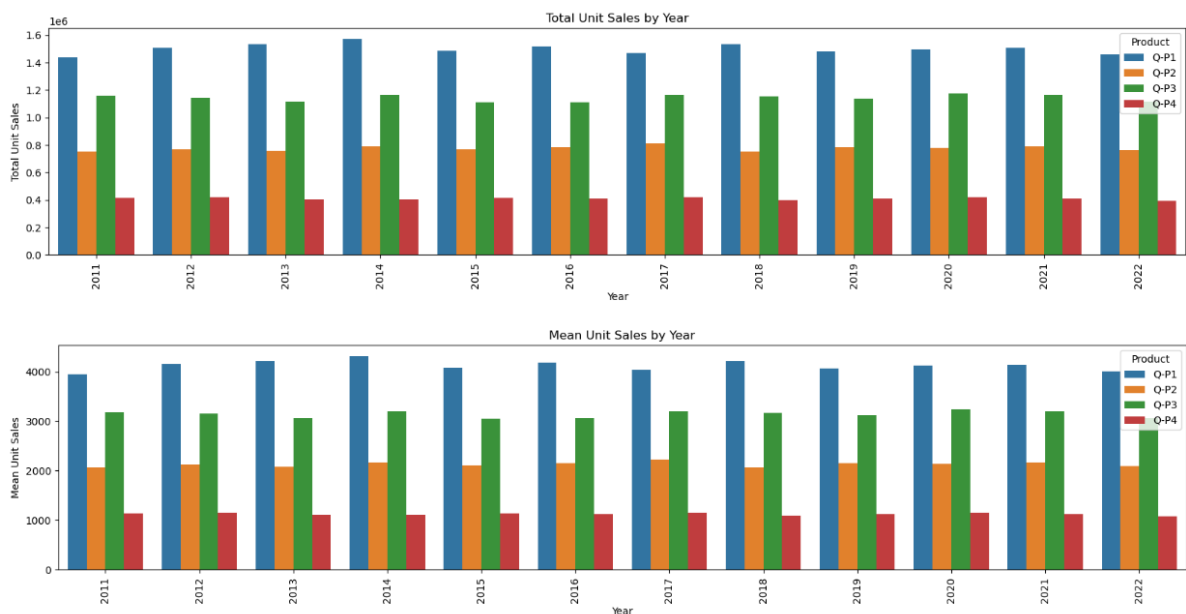
S1- Total revenue from product 1

S2- Total revenue from product 2

S3- Total revenue from product 3

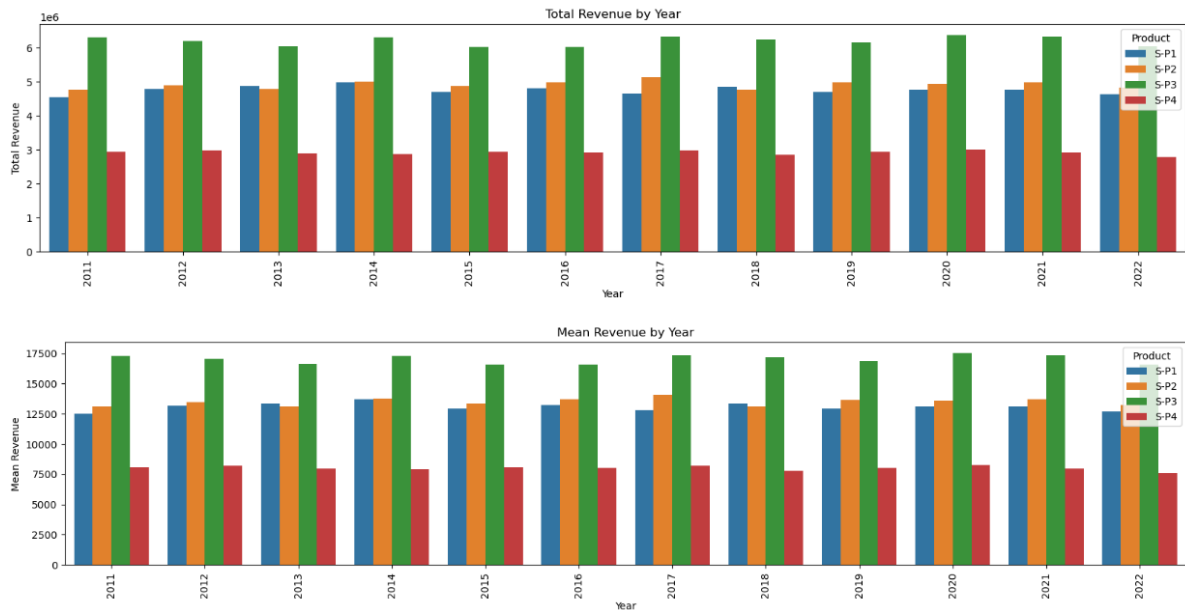
S4- Total revenue from product 4

```
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')
```



'Graph our TOTAL & MEAN revenue of sales for each product using a histogram.'

```
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')
```



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

```
avg_on_31st(data_reduced, ['Q-P1', 'Q-P2', 'Q-P3', 'Q-P4']).round(2)
```

```
Q-P1    3813.74
Q-P2    2058.80
Q-P3    3183.88
Q-P4    1098.61
dtype: float64
```

Observation for unit sales:

- We can see that our previous observation correlate as Q-P1 has the highest estimate, followed by Q-P3
- We can approximate that the company will make:

Q-P1: 3813.74

Q-P2: 2058.80

Q-P3: 3183.88

Q-P4: 1098.61

```
avg_on_31st(data_reduced, ['S-P1', 'S-P2', 'S-P3', 'S-P4']).round(2)
```

```
S-P1    12089.55  
S-P2    13052.78  
S-P3    17256.63  
S-P4     7833.07  
dtype: float64
```

Observation for Revenue from product:

- We can see that our previous observation correlate as S-P3 has the highest estimate, followed by S-P2
- We can approximate that the company will make:

S-P1: 12089.55

S-P2: 13052.78

S-P3: 17256.63

S-P4: 7833.07

11. DATA VISUALIZATION WITH IBM COGNOS:

- IBM Cognos is a powerful tool for data visualization and reporting. Use it to create visually appealing and informative representations of sales data. This phase includes:

Loading Data:

- Import the preprocessed sales data into IBM Cognos.

Data Selection:

- Choose the dataset or data source you want to visualize. You can use the data you've previously prepared or connect to live data sources.

Creating Visualizations:

- Use IBM Cognos to generate charts, graphs, dashboards, and reports to visualize the sales data. For instance, you can create bar charts to showcase top-selling products and line charts to display sales trends over time.

Interactive Exploration:

- To explore and analyze data, use IBM Cognos' interactive features.
- Apply filters and drill down into specific aspects of the data to gain insights.

Drag and Drop Data Elements:

- Drag and drop the data elements (dimensions and measures) into the appropriate areas of our chosen visualization. Dimensions are typically on the x and y-axes, while measures determine the size, color, or other properties of data points

Filters and Interactivity:

- Enhance your visualizations by adding filters and interactivity. Allow users to explore the data dynamically by setting up filter controls and enabling drill-through functionality.

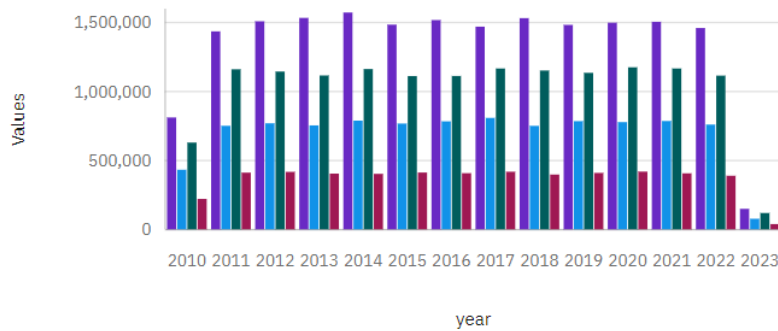
PRODUCT SALE ANALYSIS

Total unit sales of product 1,2,3,4 by Year

Q-P1, Q-P2, Q-P3 and Q-P4 by year

Measures

Q-P1 Q-P2 Q-P3 Q-P4



Month

Select...

01

02

03

04

05

06

Total revenue from product 1,2,3,4 by year

S-P1, S-P2, S-P3 and S-P4 by year

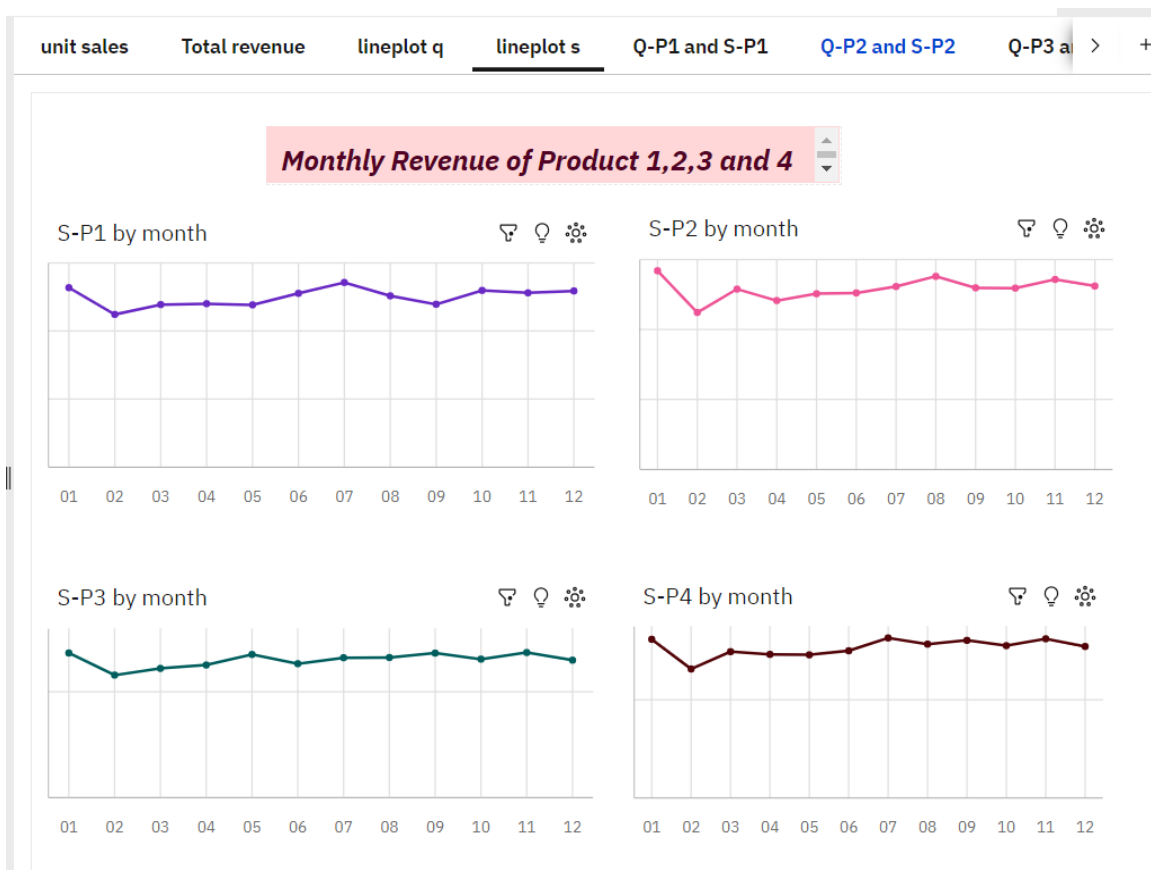
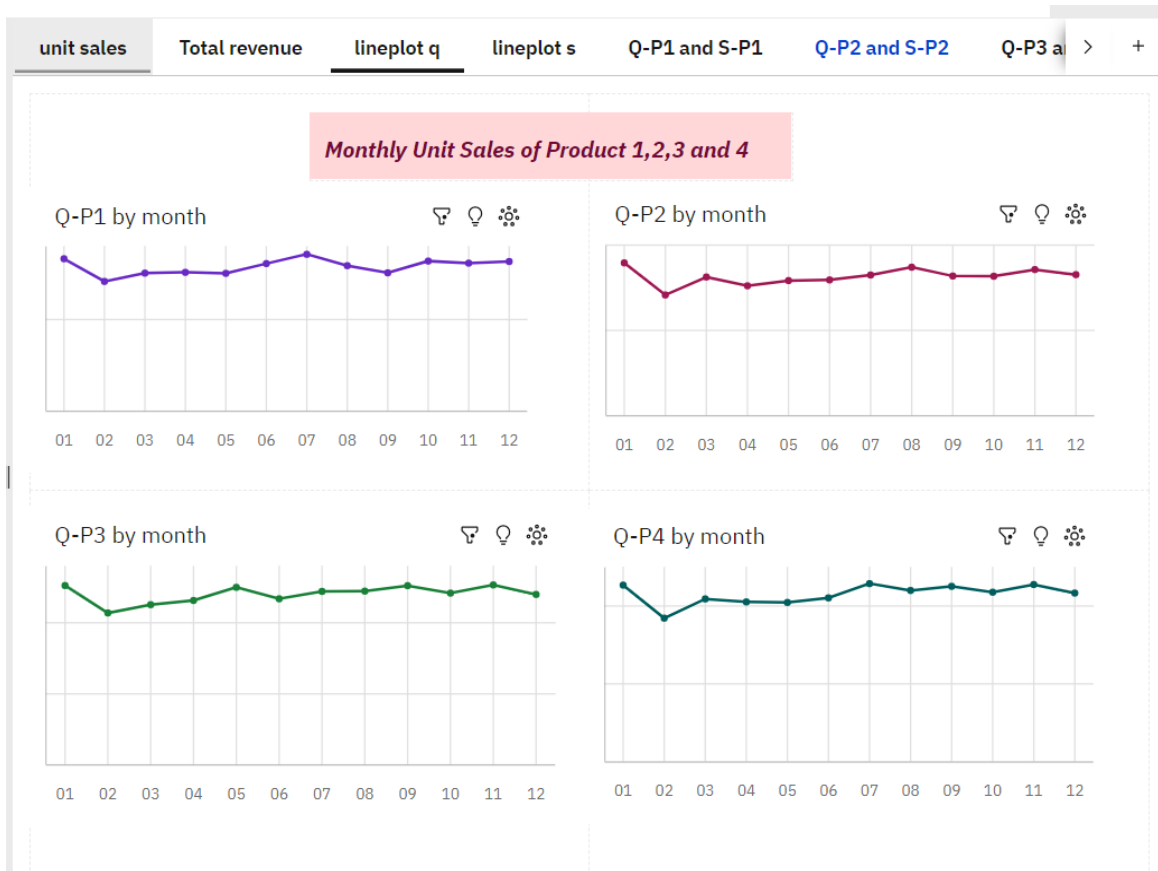
Measures

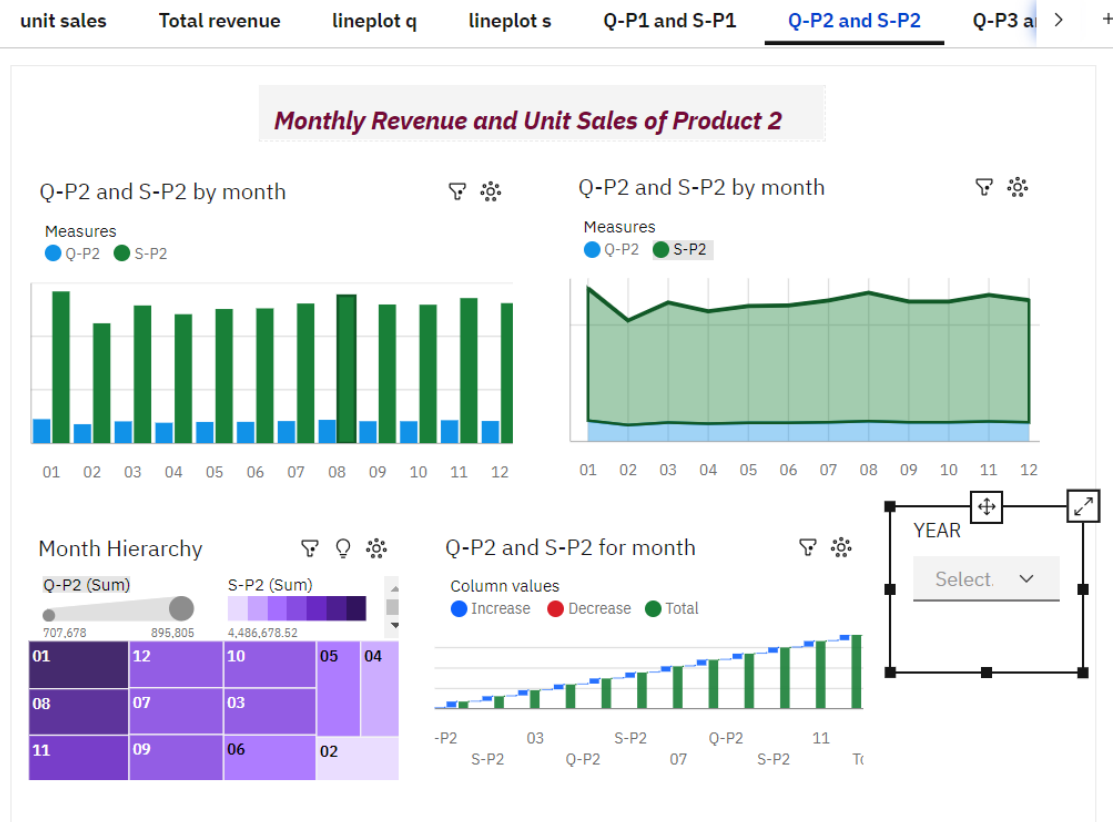
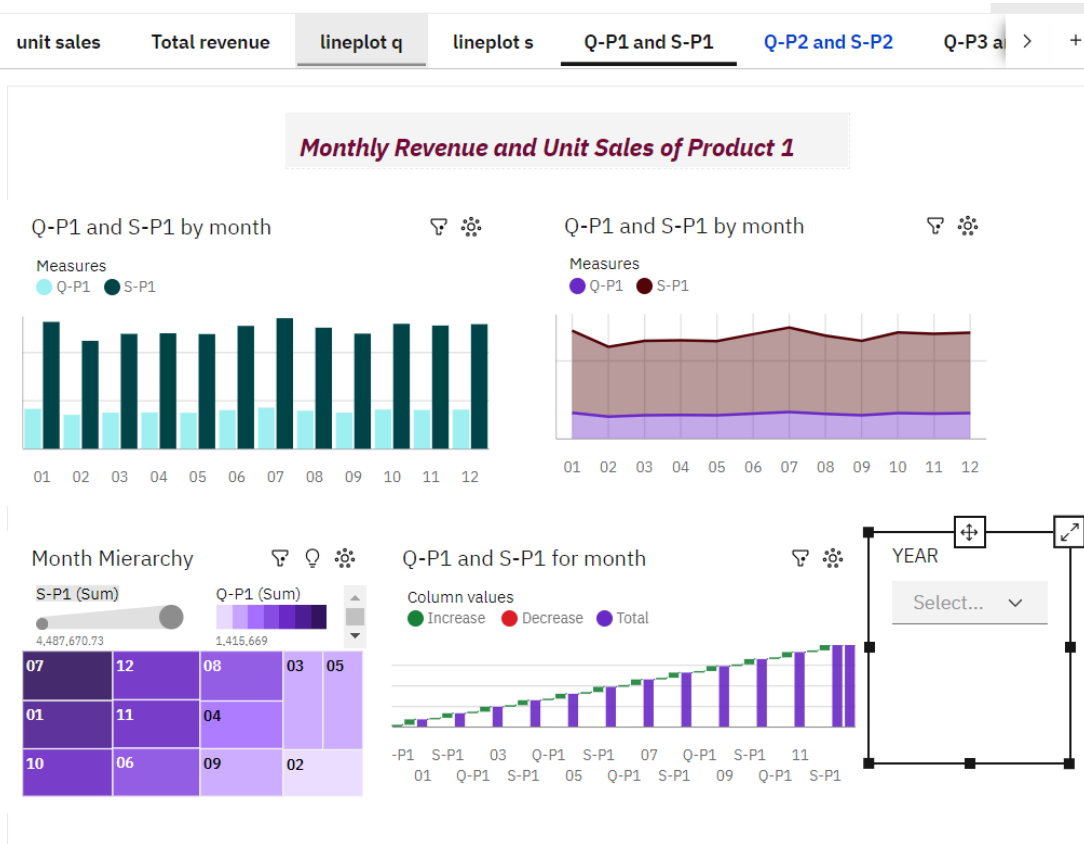
S-P1 S-P2 S-P3 S-P4

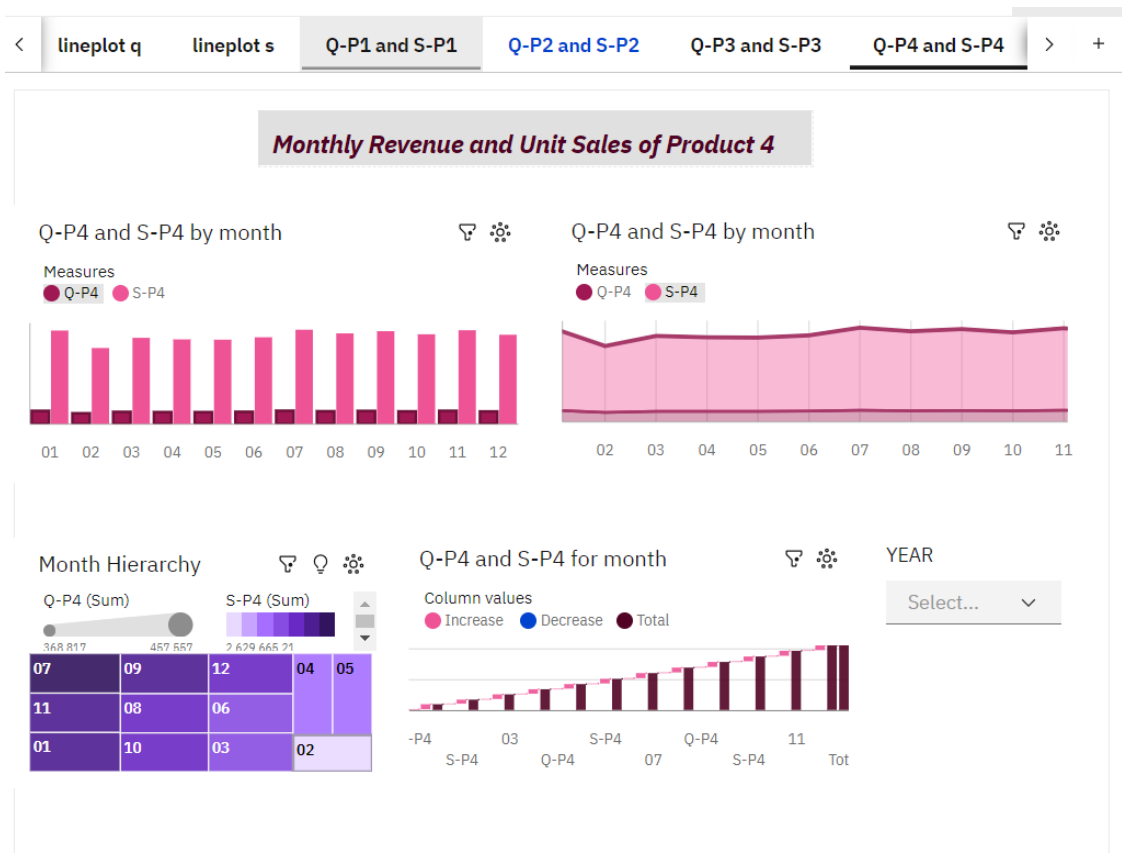


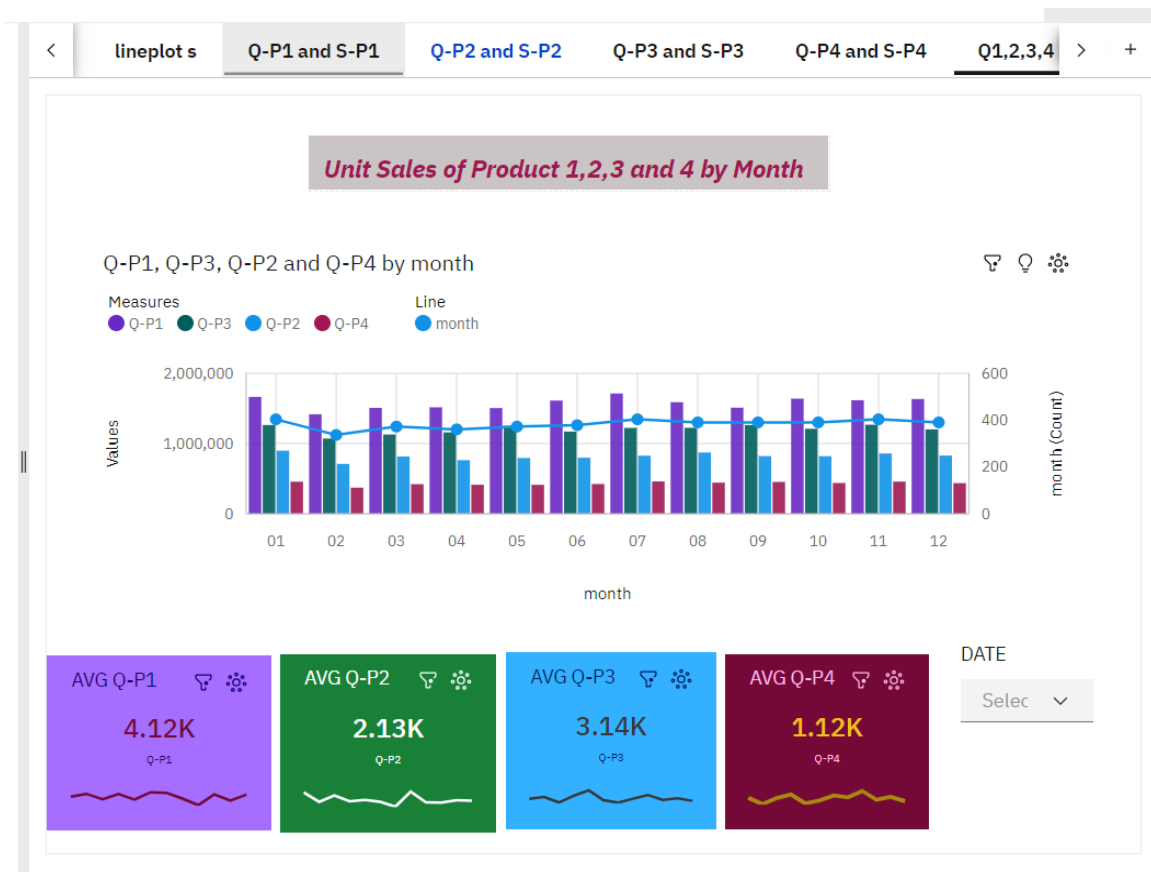
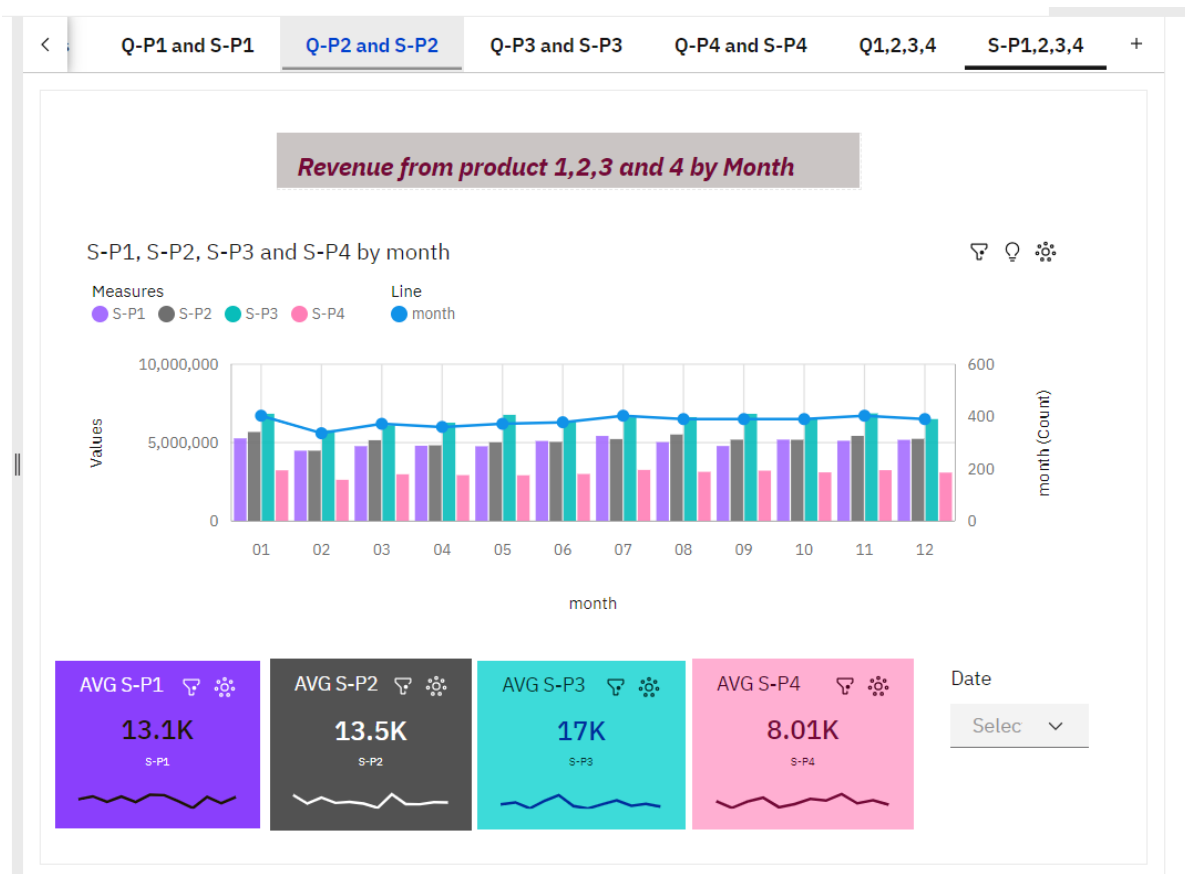
Month

Select...









Using IBM Cognos visualization, we can identify the top-selling products and customer preferences.

- Identifying top-selling products is crucial for retail or e-commerce businesses, as it optimizes inventory management, marketing strategies, and overall operations. Monitoring these products ensures effective customer demand meeting, avoids stockouts, and allocates resources effectively. Data-driven decision-making leads to efficient inventory management, targeted marketing efforts, and increased profitability. Q-P1 is the highest-selling product.
- Sales data analysis helps businesses understand customer preferences, enabling them to personalize their approach, provide products and experiences that align with their interests and needs. This leads to enhanced shopping experiences, increased satisfaction, and repeat business. Successful sales data analysis fosters long-term brand loyalty and revenue growth. S-P3 has the highest customer preference.

12. HOW INSIGHT IMPROVE USER EXPERIENCE:

Inventory Management:

- Identify top-selling products to prioritize stock levels and reordering.
- Use peak sales period insights to plan inventory for increased demand.
- Recognize slow-moving items and reduce overstock, optimizing inventory costs.
- Plan safety stock levels based on demand variability.
- Enhance supplier relationships to ensure timely inventory replenishment.
- Evaluate inventory turnover for efficiency.

Marketing Strategies:

- Target marketing campaigns toward top-selling products.
- Plan promotions during peak sales periods for maximum impact.
- Offer personalized product recommendations based on customer preferences.
- Bundle complementary products for cross-selling.
- Special promotions can be used to move slow-moving inventory.
- Segment customers for tailored marketing strategies.
- Continuously monitor customer feedback for campaign improvement.

In summary, insights guide inventory management by optimizing stock levels and costs, while also informing marketing strategies to drive sales, engagement, and customer satisfaction.

Unit Sales 2011 – 2022

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

Revenues 2011 - 2022

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 than 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.

Average Month Sales 2011 - 2022

- We can observe that all Products unit sales drop in Feb.
- We can observe that Feb and Dec have the lowest sales for each product
- For P1 We can observe Mar - Jul having the highest unit sales
- For P2 We can observe Jan, Mar - Aug having the highest unit sales
- For P3 We can observe May & Sep having the highest unit sales
- For P4 We can observe uniform sales from Jan - Dec

Estimated Unit Sales for 31st of Dec

- This value cannot be properly estimated with out Machine Learning. Currently we used the average for all the 31st days across all years for each product.
- 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.

13. CONCLUSION:

The objectives of this project encompass a holistic approach to improving business performance. By identifying top-selling products, the company can focus on what's driving revenue. Pinpointing peak sales periods allows for better resource allocation. Analyzing customer preferences provides insights for tailored offerings. Optimizing inventory management can reduce costs and enhance efficiency. Lastly, enhancing marketing strategies can boost customer engagement and overall business success. These objectives, when achieved, will contribute to a more successful and competitive business model.