

# DATA ANALYTICS WITH COGNOS-GROUP 2

## PROJECT: Product Sales Analysis

Phase5: Project Documentation and Submission

### PROJECT DOCUMENTATION

#### 1. Problem Definition and Design Thinking (Phase1)

##### PROBLEM STATEMENT:

The main challenge of this project is using IBM Cognos to efficiently analyse detailed sales data.

##### OBJECTIVES OF THE PROJECT:

###### Objective 1: Identify the best-selling products in the product category

In this value chain, the goal is to find consistently excellent products in sales.

##### Important:

Identifying the best-selling products is like identifying the MVPs of your product line. This is important because it helps you focus on the things that make you the most money. This knowledge guides your marketing strategies and inventory management, ensuring that you always have your best employees in stock.

**Analysis Approach:** To accomplish this, you probably analysed your sales data for all items with unit sales or gross revenue. This includes analysing each product's sales figures to see which ones outshine others at all times.

## Objective 2: To observe sales trends over time including peak sales periods

This goal involves understanding the ups and downs in sales, such as reading the seasons of your business:

**Important:** Like the changing seasons, sales have their own schedule. Seeing these patterns helps you make informed decisions. You can pick high and low seasons, the months when customers are most active, and even predict when you might need additional inventory.

**Methodology:** To accomplish this goal, you probably used time series analysis. It's like watching your sales data trend. Graphics such as line charts or time-based graphs can reveal peaks and bonuses in sales, show trends over time and help identify those peak sales periods.

## Objective 3: To understand consumer preferences and buying behaviour

This goal includes knowing your customers and focusing on what makes them tick:

**Important:** Think of this as the secret ingredient in the recipe for success. Understanding what your customers like and how they behave is essential to creating personalized experiences. It's like knowing your audience inside out, so you can tailor your offering to their interests.

**Analysis Methodology:** To achieve this goal, you probably analysed customer data in depth if available. This is similar to tracking the footprints left by your customers. You may have used techniques such as customer segmentation to group similar customers, revealing patterns in their preferences and behaviours. This valuable insight allows you to optimize your marketing and product recommendations for a more pleasant user experience.

## **Underlying Objectives:**

At its core, the project wants to deliver actionable insights that enable companies to make data-driven decisions, enhancing two key aspects of their operations:

### **1.Inventory Management Improvements:**

By identifying the best-selling items and understanding peak sales times, businesses can improve the quality of their inventories. This reduces the risk of overstock or oversupply, leading to cost savings and greater customer satisfaction.

### **2.Market strategy refinement:**

Insights into consumer preferences and behaviour enable businesses to refine their marketing strategies. This includes efforts to align advertising with consumer interests and optimize advertising spend to maximize impact.

# **DESIGN THINKING FOR PRODUCT SALES ANALYSIS:**

## **1. Empathize: Understand the needs of the stakeholders**

Start by empathizing with stakeholders, including business owners, marketing teams and inventory managers. Understand their pain points, goals, and the specific challenges they face in managing content and sales.

## **2. Define: Define the problem clearly**

Based on our sympathetic side, define the problem clearly. Make sure we have an accurate understanding of the insights stakeholders need from sales data analysis.

## **3. Ideate: Provide solutions and approaches**

Encourage brainstorming to generate ideas on how to solve a defined problem. This can include data collection techniques, visualization techniques, and potential insights that can be obtained.

## **4. Prototype: Create a strategic plan**

Create a model or action plan that shows how you want to conduct the research. This includes defining the objectives of the study, methods of data collection, tools to be used (such as IBM Cognos), and preliminary graphical design.

## **5. Test: Validate the process**

Before we get into the full analysis, do a small experiment to confirm our approach. Ensure that the data collection process is robust and that the chosen perspective effectively delivers insight.

## 6. Repeat: Clean and improve

Based on the information from the experiment, restate your plan. Make improvements as needed, whether in data collection methods, visualization methods, or analytical methods.

## 7. Implement: Do the research

Perform comprehensive analytics using IBM Cognos, control your sophisticated system. It involves collecting, preparing and analysing sales data to gain meaningful insights.

## 8. Interpret: Get the insight

Interpret the results of your research by extracting actionable insights about best-selling products, sales trends, and customer preferences.

## 9. Deliver: Share your findings

Create a clear and concise report or presentation to communicate your findings to stakeholders. Use data-driven visualizations and presentations to make insights easy and actionable.

## 10. Feedback and changes: Continuous improvement

Present and analyse your findings feedback from stakeholders. Understand how insights have influenced decision making and be open to revising your approach for future research.

## 2. Innovation (Phase2)

Design Thinking:

Empathize:

- In this phase, we will engage with the business or client to deeply understand their challenges. This could involve discussing their struggles with inventory management inefficiencies, marketing strategies that may not be working optimally, or their need for data-driven insights.
- We will conduct interviews, surveys, and discussions to gather insights into their specific pain points and objectives related to sales data analysis. This empathetic understanding will guide our project's direction.

Define (purpose and scope of project):

- Clearly explain the purpose of our project, which is to analyse sales data using IBM Cognos to extract specific insights.
- Determine the scope of the analysis. This includes issues such as deadlines for sales data, metrics to analyse (e.g., revenue, product sales, number of customers), and key questions we aim to answer (e.g., What are the best selling products? When do sales peak? What do customers want?).

Once these aspects are defined, we establish the boundaries and goals of our project.

## Ideate:

In the idea phase, we encourage creative thinking in our team. We can explore new ways of analysing sales data that go beyond traditional methods.

- Consider using IBM Cognos to deliver insights in unique ways. Let's consider other data visualization techniques, predictive analytics, or machine learning models that can add value to our research.
- The goal is to generate new and innovative ideas for our project.

## Prototype:

- Create prototype or sample analysis reports using IBM Cognos. This is a useful snapshot of our mind and consciousness.
- Develop a mock idea or prototype of how we plan to present research findings. This helps stakeholders visualize and to take action on the end result and to provide feedback.

## Test (collect and edit answers):

- We should share our model with relevant stakeholders, colleagues, or potential end users. We collect data on the clarity, applicability, and effectiveness of the model.

- We can use this feedback to refine our analysis methods, visualization techniques and overall structure of the paper.

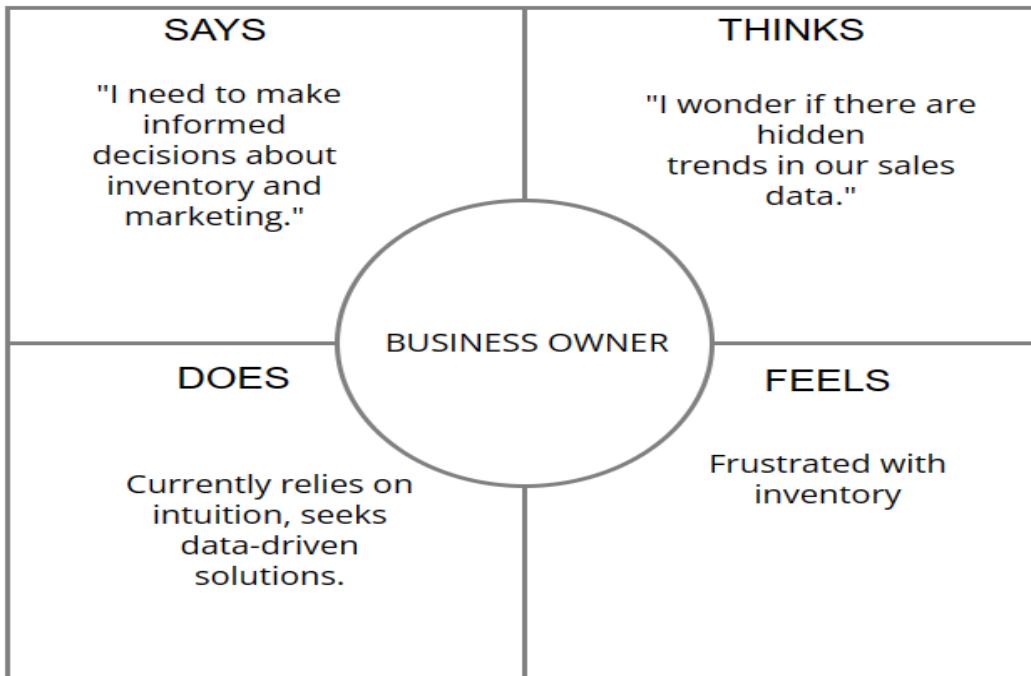
## Implement:

- Once our approach is refined and validated, let's move on to more accurate sales data analytics using IBM Cognos.
- Use selected methodologies and tools to extract insights into best-selling products, peak seasons and customer preferences. This phase involves working with real data to gain actionable insights.

## Evaluate:

- Once we have implemented the assessment, we need to evaluate its impact on performance. Let's see if the insights have helped improve inventory and marketing strategies.
- Let's reflect on the overall success of our project, what worked well and consider areas where we can apply the lessons learned to future projects.

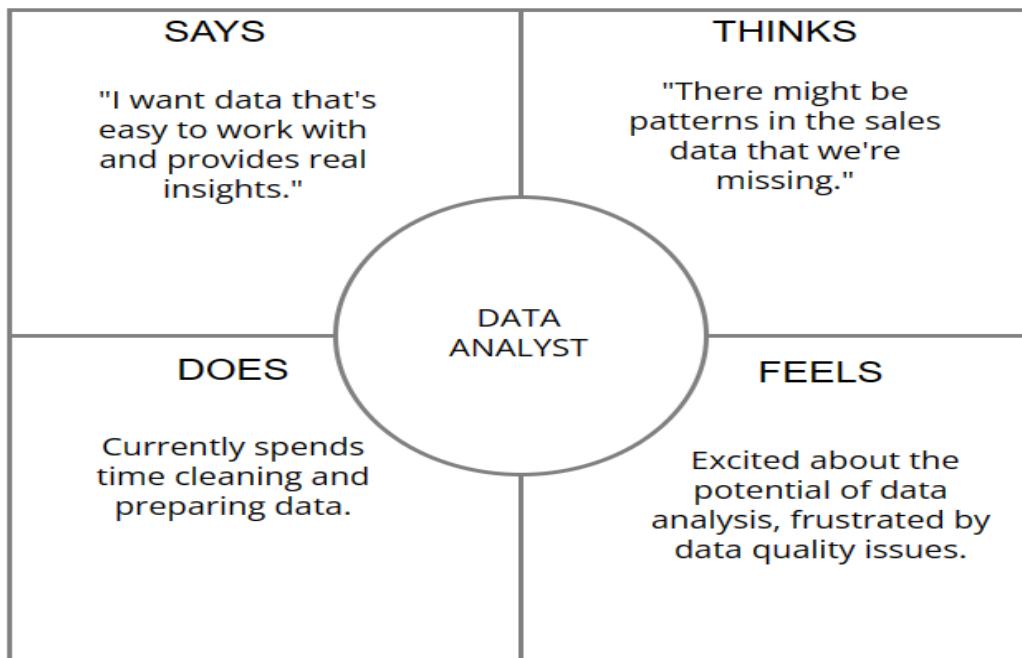
## Empathy map as a Business Owner:



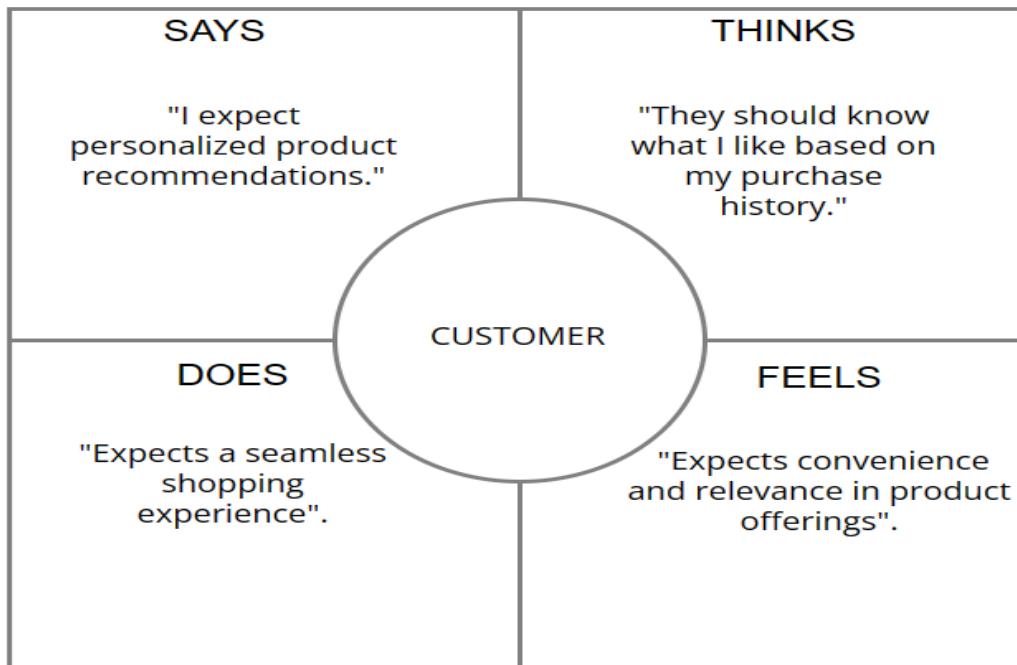
## Empathy map as a Marketing Manager:



## Empathy map as a Data Analyst:



## Empathy map as a Customer:



## STEPS FOR DOING THE PROJECT:

### Step 1: Given the Dataset

We start with a data set with complete product information. We make sure we have collected all the important data points including item ID, sale date, sales quantity, prices and number of customers.

### Step 2: Data Preprocessing

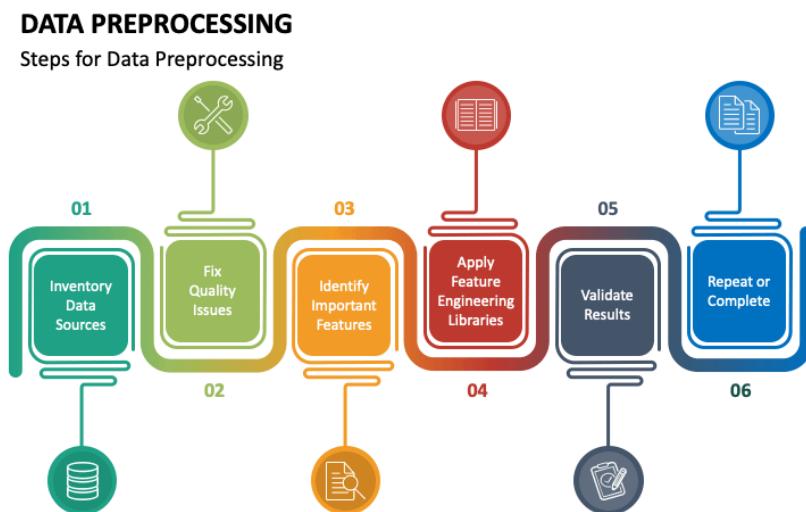
To prepare a dataset for analysis, we perform the following operations.

**We replace missing values:** we carefully address any missing data points, either by creating appropriate statistics or by removing incomplete records.

**Remove duplicates:** We check the dataset for duplicate entries and remove them to retain the data.

**Data format:** We ensure that data are correct; Let's change them if necessary. For example, we make sure the dates are always set.

**Data consistency:** We check for data set inconsistencies or data entry errors and make corrections. Data consistency is critical for meaningful analysis



### Step 3: Data Analysis

Before diving into the rigorous analysis, we conduct an initial exploratory data analysis (EDA) to gain insight:

**We calculate descriptive statistics:** To understand statistical variables such as sales volume, we calculate basic statistics such as average, median, standard deviation and quartiles.

**Visualize data:** Create data visualizations, such as histograms, box plots, scatter plots, and density plots, to visualize data distribution and identify outliers or trends.

**Summarize the data:** Prepare summary reports to state key features and trends in the data structure. These reports will serve as valuable references throughout the investigation.



### Step 4: Define the purpose of the evaluation

We clearly stated our main objectives and research questions. This accuracy ensures that our research efforts remain focused and results-driven. immediately:

We identify the highest sales in a given period of time.

We see seasonal trends in sales.

We segment customers based on buying behaviour.

## Step 5: Select Analysis Techniques

In order to meet our objectives, we selected the most appropriate research methods according to our specific requirements:

**Descriptive Statistics:** We use these to summarize and understand data.

**Time series analysis:** To identify temporal patterns and trends in sales data.

**Customer Classification:** We use clustering algorithms such as K-means or hierarchical clustering to group customers with similar purchase behaviour.

**Regression analysis:** We use this to predict future sales based on historical data.

## Step 6: Data Visualization

Visual aids play an important role in effectively communicating our research results:

**We use graphs:** we show sales trends over time, so that seasons or growth patterns can be seen.

**We use Bar Charts or Pie Charts:** we show the distribution of sales.

**We create Heatmaps:** We highlight connections and patterns in a data structure.



## Step 7: Data analysis

We perform the selected analytical methods on the dataset:

For example, when looking to find the best-selling products, we calculate and rank the products based on sales volume.

When analysing sales trends, we use time series techniques such as moving averages or exponential smoothing to identify underlying trends.

## Step 8: Derive Insights

We interpret the results of our study in broad terms:

We analyse data and images for meaningful insights. For example, we may find that a product consistently performs better than others, or that sales peak during particular months.

## Step 9: Recommendations

We formulate actionable recommendations based on the insights we have gathered:

We provide specific guidance on how the project can use research findings. For example, we suggest increasing the volume of high-demand

products or launching targeted marketing campaigns during peak sales periods.

## Step 10: Documentation

Produce a well-structured document or report that accompanies our entire research journey:

Includes professional summaries, data visualizations, interpretations of findings, recommendations, and any technical issues or methods used in the research.

## Step 11: Test and validate

To ensure the integrity of our research:

Analyse calculations, data changes, and any assumptions made during analysis.

We believe that the evaluation process adheres to best practices and is consistent with the defined objectives.

## Step 12: Peer Review

- Ask for feedback and validation from colleagues or peers:
- Collaborate with others to contribute to our analysis, gather new insights, and identify potential blind spots or errors.



### Step 13: Presentation

- If necessary, prepare and provide stakeholders with an explanation of our research findings:
- Use illustrations and short descriptions to effectively communicate results and engage the audience.



## Step 14: Implementation

Helps implement recommended changes or strategies:

Collaborates with relevant teams to ensure implementation of actionable recommendations.



## Step 15: Analysis

Continuously monitoring and evaluating the impact of changes implemented:

Examine whether recommended strategies lead to improvements in inventory and marketing strategies over time.

# Continuous Monitoring



### 3. Development Part 1 (Phase3)

#### Data Collection Process:

It involves collecting and sourcing data to produce data that can be used for analysis, modeling, and decision making.

For this project we have given a dataset

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

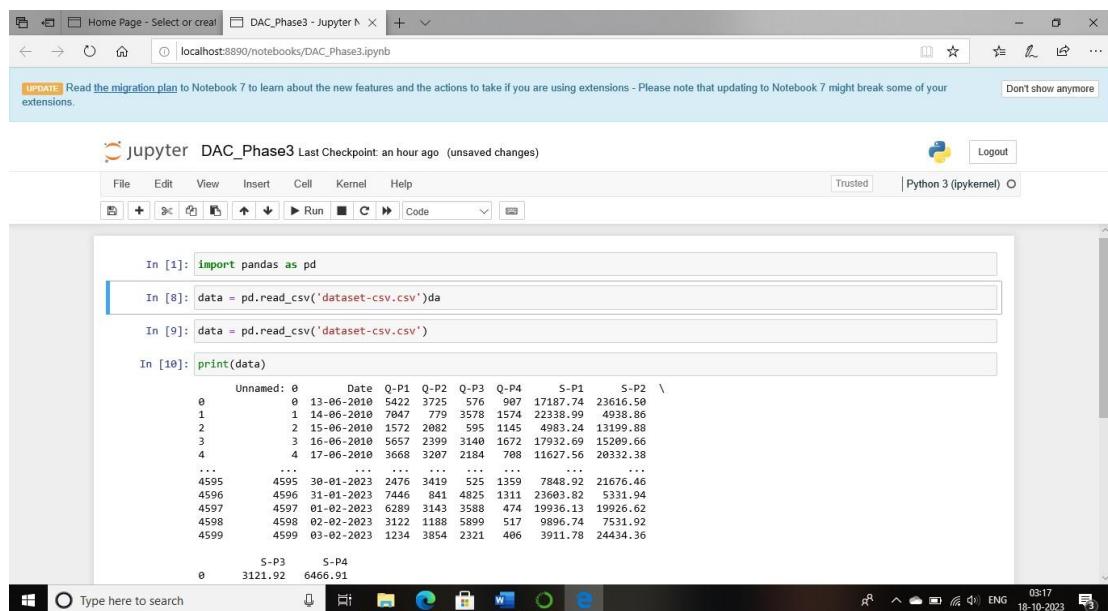
#### Data Loading and Preprocessing:

#### Data Loading:

I've successfully loaded the dataset we were provided with into Our analysis environment. It contains essential product sales information.

## Data preprocessing:

Preprocessing is necessary to make the data quality and ready for analysis. I am actively working to address missing values, eliminate duplicates, refine the data structure, and ensure data consistency.

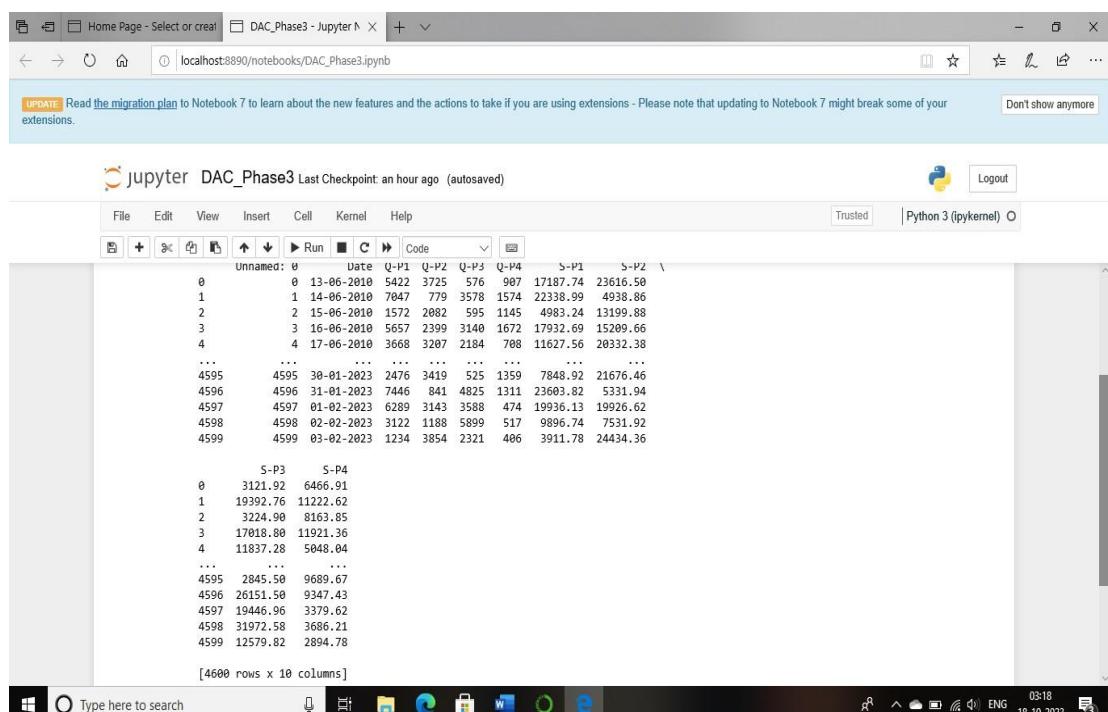


The screenshot shows a Jupyter Notebook interface running on a Windows desktop. The browser tab is titled "DAC\_Phase3 - Jupyter" and the URL is "localhost:8890/notebooks/DAC\_Phase3.ipynb". The notebook contains the following code and its output:

```
In [1]: import pandas as pd
In [8]: data = pd.read_csv('dataset-csv.csv')
In [9]: data = pd.read_csv('dataset-csv.csv')
In [10]: print(data)
```

	Unnamed:	Date	Q-P1	Q-P2	Q-P3	Q-P4	S-P1	S-P2
0	0	13-06-2010	5422	3725	576	907	17187.74	23616.50
1	1	14-06-2010	7847	779	3578	1574	22338.99	4938.86
2	2	15-06-2010	1572	2882	595	1145	4983.24	13199.88
3	3	16-06-2010	5657	2399	3148	1672	17932.69	15269.66
4	4	17-06-2010	3668	3207	2184	708	11627.56	20332.38
...	...	...	...	...	...	...	...	...
4595	4595	30-01-2023	2476	3419	525	1359	7848.92	21676.46
4596	4596	31-01-2023	7446	841	4825	1311	23603.82	5331.94
4597	4597	01-02-2023	6289	3143	3588	474	19936.13	19926.62
4598	4598	02-02-2023	3122	1188	5899	517	9896.74	7531.92
4599	4599	03-02-2023	1234	3854	2321	406	3911.78	24434.36

S-P3 S-P4  
0 3121.92 6466.91



The screenshot shows a Jupyter Notebook interface running on a Windows desktop. The browser tab is titled "DAC\_Phase3 - Jupyter" and the URL is "localhost:8890/notebooks/DAC\_Phase3.ipynb". The notebook contains the following code and its output:

```
In [1]: import pandas as pd
In [8]: data = pd.read_csv('dataset-csv.csv')
In [9]: data = pd.read_csv('dataset-csv.csv')
In [10]: print(data)
```

	Unnamed:	Date	Q-P1	Q-P2	Q-P3	Q-P4	S-P1	S-P2
0	0	13-06-2010	5422	3725	576	907	17187.74	23616.50
1	1	14-06-2010	7847	779	3578	1574	22338.99	4938.86
2	2	15-06-2010	1572	2882	595	1145	4983.24	13199.88
3	3	16-06-2010	5657	2399	3148	1672	17932.69	15269.66
4	4	17-06-2010	3668	3207	2184	708	11627.56	20332.38
...	...	...	...	...	...	...	...	...
4595	4595	30-01-2023	2476	3419	525	1359	7848.92	21676.46
4596	4596	31-01-2023	7446	841	4825	1311	23603.82	5331.94
4597	4597	01-02-2023	6289	3143	3588	474	19936.13	19926.62
4598	4598	02-02-2023	3122	1188	5899	517	9896.74	7531.92
4599	4599	03-02-2023	1234	3854	2321	406	3911.78	24434.36

S-P3 S-P4  
0 3121.92 6466.91  
1 19392.76 11222.62  
2 3224.90 8163.85  
3 17818.80 11921.36  
4 11837.28 5048.04  
...

4595 2845.50 9689.67  
4596 26151.50 9347.43  
4597 19446.96 3379.62  
4598 31972.50 3686.21  
4599 12579.82 2894.78

[4600 rows x 10 columns]

# MISSING VALUES:

A screenshot of a Jupyter Notebook interface. The browser tab is titled "DAC\_Phase3 - Jupyter". The notebook content shows the following code execution:

```
In [14]: data = data.drop(columns=['Unnamed: 0'])

In [15]: 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: 323.6+ KB

In [16]: data.isnull().sum()

Out[16]: 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
```

The status bar at the bottom indicates the date as 18-10-2023 and the time as 03:24.

A screenshot of a Jupyter Notebook interface. The browser tab is titled "DAC\_Phase3 - Jupyter". The notebook content shows the following code execution:

```
In [14]: data = data.drop(columns=['Unnamed: 0'])

In [15]: data.info()

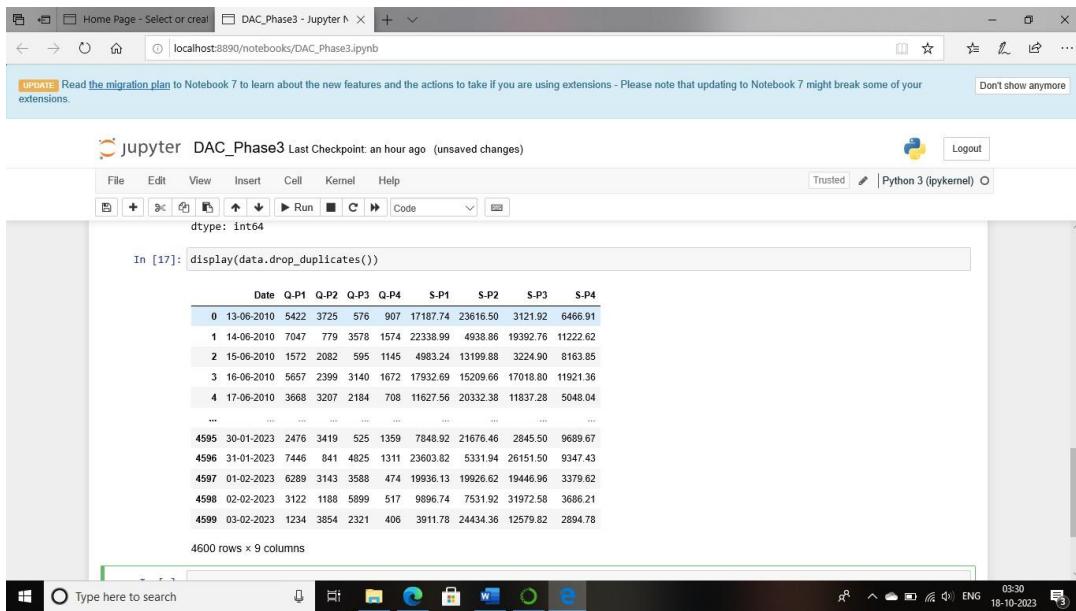
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4600 entries, 0 to 4599
Data columns (total 9 columns):
 #   Column   Non-Null Count  Dtype  
--- 
 0   Q-P2    4600 non-null    int64  
 1   Q-P3    4600 non-null    int64  
 2   Q-P4    4600 non-null    int64  
 3   S-P1    4600 non-null    float64 
 4   S-P2    4600 non-null    float64  
 5   S-P3    4600 non-null    float64  
 6   S-P4    4600 non-null    float64  
 7   S-P1    4600 non-null    float64  
 8   S-P2    4600 non-null    float64  
 9   S-P3    4600 non-null    float64  
 10  S-P4    4600 non-null    float64  
dtypes: float64(4), int64(4), object(1)
memory usage: 323.6+ KB

In [16]: data.isnull().sum()

Out[16]: 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
```

The status bar at the bottom indicates the date as 18-10-2023 and the time as 03:25.

# CHECKING FOR DUPLICATES:



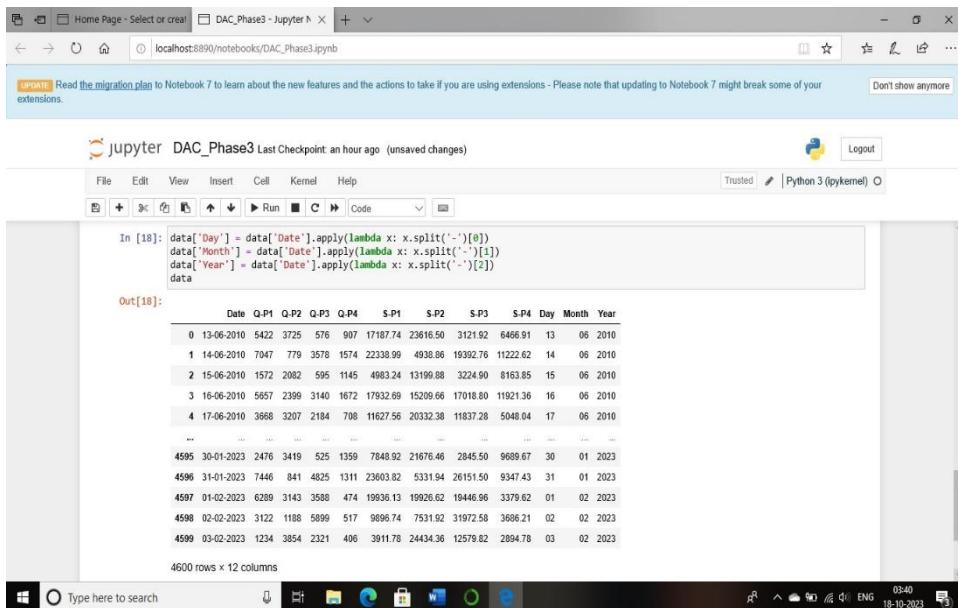
Jupyter DAC\_Phase3 Last Checkpoint: an hour ago (unsaved changes)

In [17]: `display(data.drop_duplicates())`

	Date	Q-P1	Q-P2	Q-P3	Q-P4	S-P1	S-P2	S-P3	S-P4
0	13-06-2010	5422	3725	576	907	17187.74	23616.50	3121.92	6466.91
1	14-06-2010	7047	779	3578	1574	22338.99	4938.86	19392.76	11222.62
2	15-06-2010	1572	2082	595	1145	4983.24	13199.88	3224.90	8163.85
3	16-06-2010	5657	2399	3140	1672	17932.69	15209.66	17018.80	11921.36
4	17-06-2010	3668	3207	2184	708	11627.56	20332.38	11837.28	5048.04
...	...	...	...	...	...	...	...	...	...
4595	30-01-2023	2476	3419	525	1359	7848.92	21676.46	2845.50	9689.67
4596	31-01-2023	7446	841	4825	1311	23603.82	5331.94	26151.50	9347.43
4597	01-02-2023	6289	3143	3588	474	19936.13	19926.62	19446.96	3379.62
4598	02-02-2023	3122	1188	5899	517	9896.74	7531.92	31972.58	3686.21
4599	03-02-2023	1234	3854	2321	406	3911.78	24434.36	12579.82	2894.78

4600 rows × 9 columns

Splitting each date string into a list of substrings using - as a separator and assign the resulting values to the new columns.



Jupyter DAC\_Phase3 Last Checkpoint: an hour ago (unsaved changes)

In [18]: `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`

Out[18]:

	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

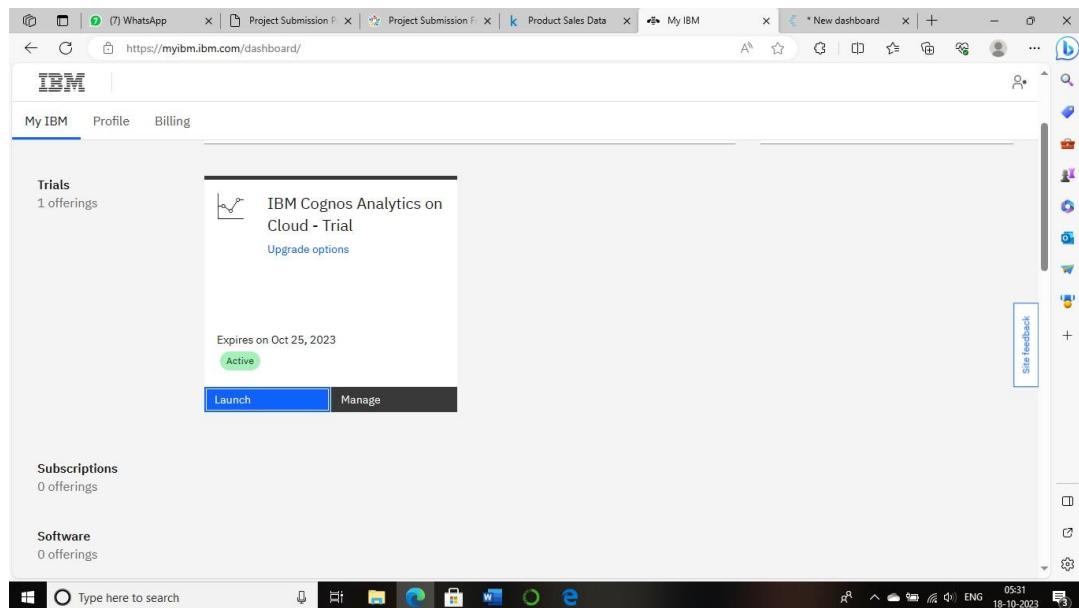
## Defining Analysis Objectives:

I've laid out our analysis objectives to guide our project. We're keen on identifying top-selling products, exploring sales trends over time, and gaining insights into customer purchasing patterns.

## Data Visualization in IBM COGNOS:

### Visualization Setup (IBM Cognos)

Setting Up IBM Cognos: I've set up our visualization environment using IBM Cognos, a powerful tool for creating effective data visualizations.

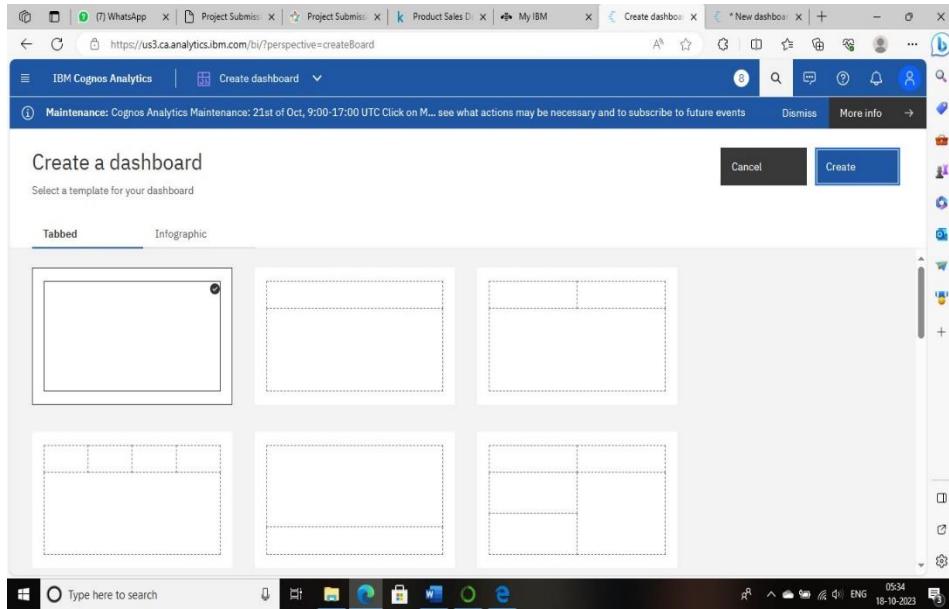


The screenshot shows the 'Profile' tab selected in the IBM My IBM interface. The left sidebar lists 'ID and password', 'Contact information', 'Company or organization', and 'Addresses'. The main content area displays contact information: Name (Jeswin A A), Display name (empty), Email address (oodo5210@gmail.com, Primary), and Phone number (empty). A security section indicates two-step verification is off. The bottom navigation bar includes a search bar, pinned icons for Home, Project Submission, Product Sales Data, and My IBM, and system status indicators.

The screenshot shows the 'Upload data and stats' page in IBM Cognos Analytics. A progress bar at the top indicates 'Initializing file upload...'. Below it, a message says 'You can upload supported file types that are stored in any location to which your computer has local or LAN access.' A large dashed rectangular area is provided for dragging and dropping files. At the bottom, there are 'Cancel' and 'Next' buttons, along with a link to 'Learn more' about uploading data.

## Layout Design:

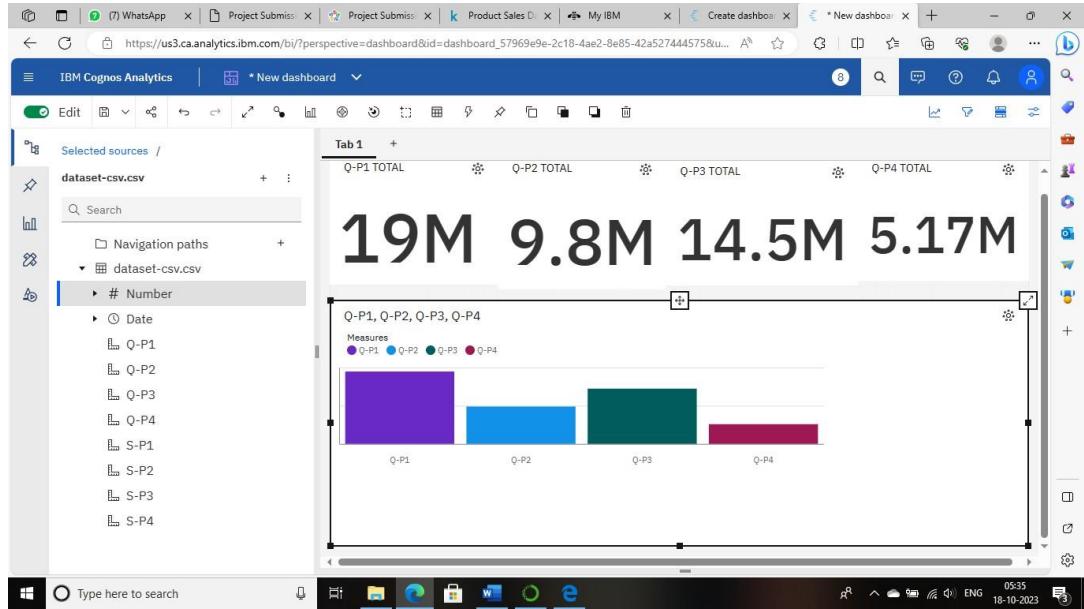
We're actively planning the layout and structure for our dashboard or visualizations. This design will be instrumental in presenting our findings effectively.



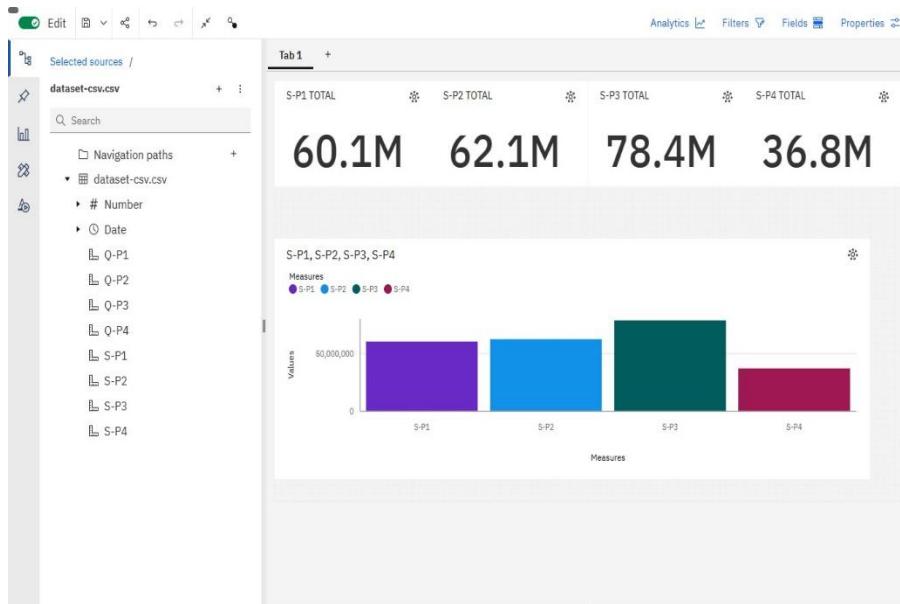
## Initial Visualizations:

We're in the process of creating initial visualizations that align with our analysis objectives. These include a range of chart types, such as line charts, bar charts, and pie charts, to effectively represent our data.

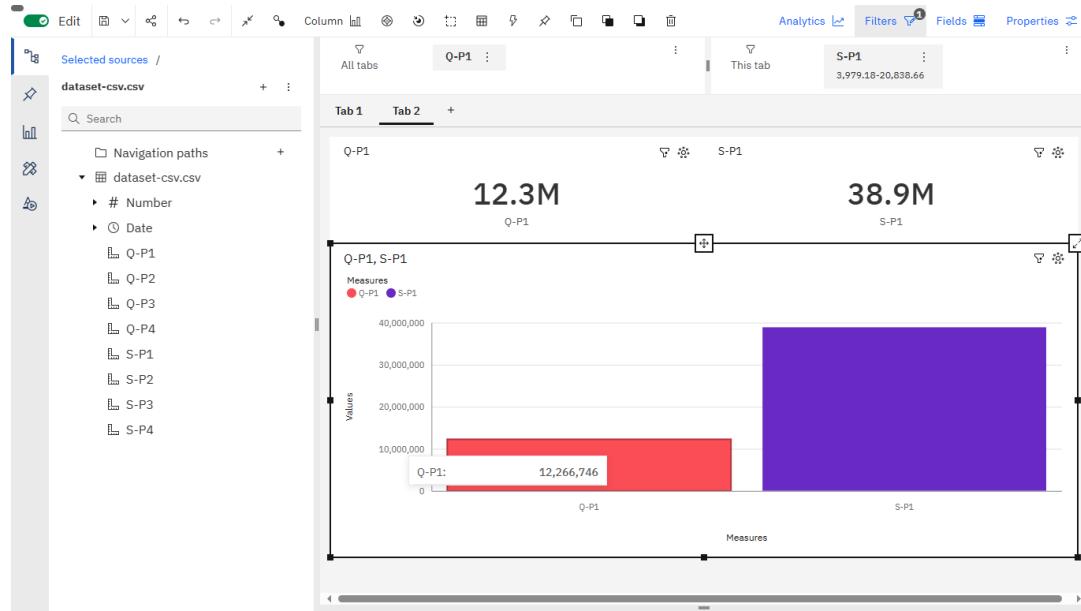
## Total Unit sold in all the years:



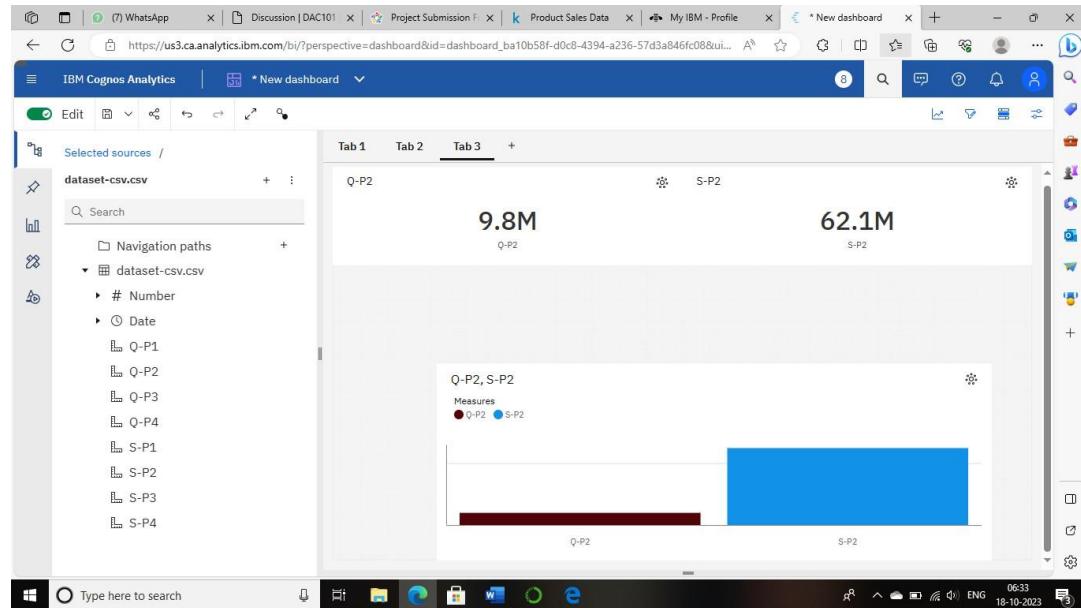
## Total Revenue generated by Each Unit:



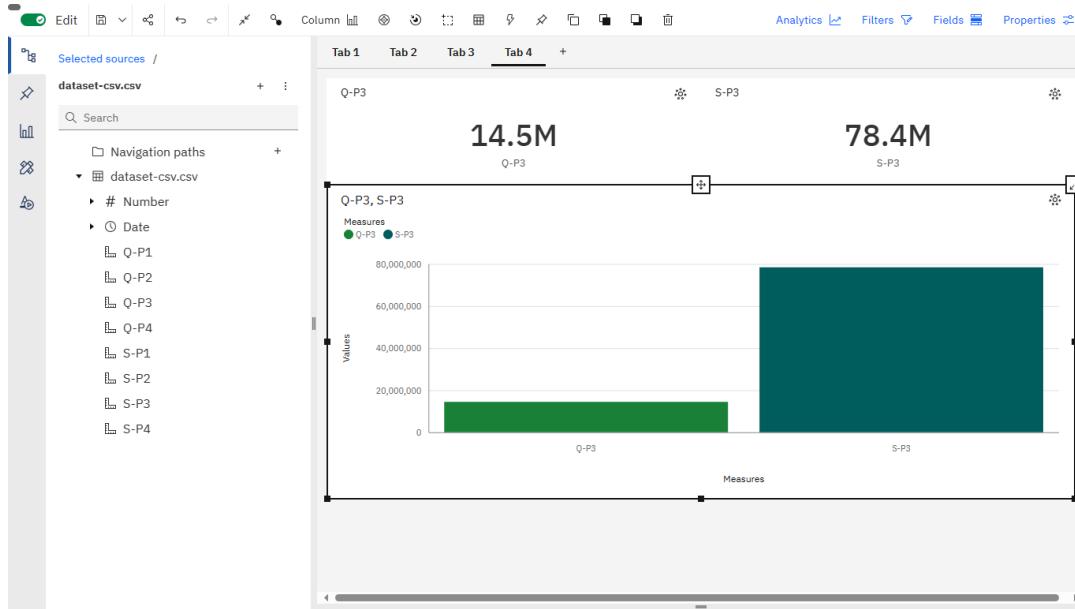
## Total Unit Sold and Total Revenue Generated by Product1



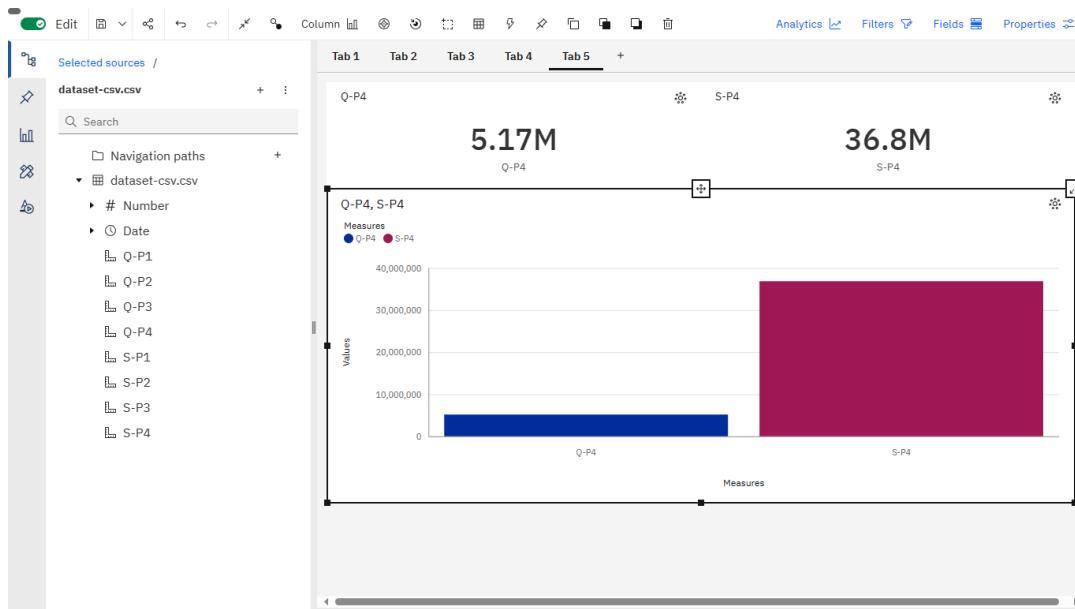
## Total Unit Sold and Total Revenue Generated by Product 2



## Total Unit Sold and Total Revenue Generated by Product 3



## Total Unit Sold and Total Revenue Generated by Product 4



By these visualizations we can say that P1 is the highest sold product followed by P3, P2 and P4

And P4 could be considered the best product in terms of revenue per unit sold.

## Defining the objectives of analysis:

### 1. Identification of top selling products:

The objective of the project is to identify the products that consistently generate the highest sales. This knowledge is critical for companies to prioritize their inventory, and to ensure that popular brands are adequately stocked to meet customer needs.

### 2. Analysis of peak sales periods:

Understanding when sales peak is important for efficient product distribution. By analysing the timing of peak sales, companies can make informed decisions about staffing, promotions and restocking during these critical periods.

### 3. Understanding customer preferences:

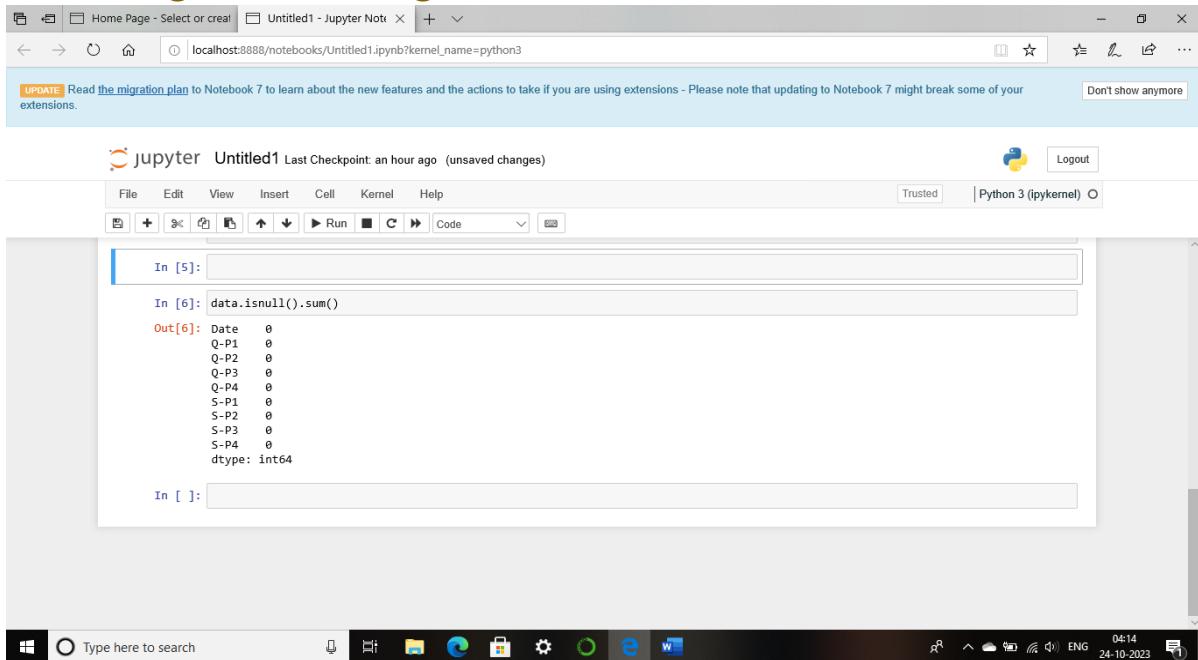
Depth on customer preferences and behaviours is important to tailor marketing strategies. This includes distinguishing between consumer preferences, reactions to marketing campaigns, and factors that influence their purchase decisions.

## Predict Future Sales Trends:

Utilize historical sales data and predictive analytics to forecast future sales patterns. This forecasting is essential for demand planning and resource allocation.

## 4. Development Part 2 (Phase4)

### Checking For Missing Values:



The screenshot shows a Jupyter Notebook window titled "Untitled1 - Jupyter Notebooks". The URL in the address bar is "localhost:8888/notebooks/Untitled1.ipynb?kernel\_name=python3". A message at the top says "UPDATE Read the migration plan to Notebook 7 to learn about the new features and the actions to take if you are using extensions - Please note that updating to Notebook 7 might break some of your extensions." Below this, there is a "Don't show anymore" button. The notebook interface includes a toolbar with File, Edit, View, Insert, Cell, Kernel, Help, and a Python 3 (ipykernel) button. The code cell In [6] contains the command `data.isnull().sum()`. The output cell Out[6] displays the following data:

	Date	Q-P1	Q-P2	Q-P3	Q-P4	S-P1	S-P2	S-P3	S-P4
dtype:	int64	0	0	0	0	0	0	0	0

😊 From this we can see that our dataset has no missing values.

### EDA Exploratory Data Analysis:

EDA is normally carried out as a preliminary step before undertaking extra formal statistical analyses or modelling.

## Let us extract the Year, Month and day

The screenshot shows a Jupyter Notebook interface. In the code cell (In [7]), the following Python code is written:

```
In [7]: 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
```

The output cell (Out[7]) displays a DataFrame with 4600 rows and 12 columns. The columns are Date, Q-P1, Q-P2, Q-P3, Q-P4, S-P1, S-P2, S-P3, S-P4, Day, Month, and Year. The data includes various numerical values and dates like 13-06-2010, 14-06-2010, etc.

## Let us Drop Rows for Years 2010 and 2023

The screenshot shows a Jupyter Notebook interface. In the code cell (In [8]), the following Python code is written:

```
In [8]: data_reduced = data.query("Year != '2010' and Year != '2023'")
```

The output cell (In [ ]:) is currently empty.

# Graph Representation of TOTAL and MEAN of Each Product's Unit Sold

Home Page - Select or creat Untitled1 - Jupyter Note + localhost:8888/notebooks/Untitled1.ipynb?kernel\_name=python3

UPDATE Read the migration plan to Notebook 7 to learn about the new features and the actions to take if you are using extensions - Please note that updating to Notebook 7 might break some of your extensions. Don't show anymore

jupyter Untitled1 Last Checkpoint: an hour ago (unsaved changes)

File Edit View Insert Cell Kernel Help Trusted Python 3 (ipykernel) ●

In [ \*]:

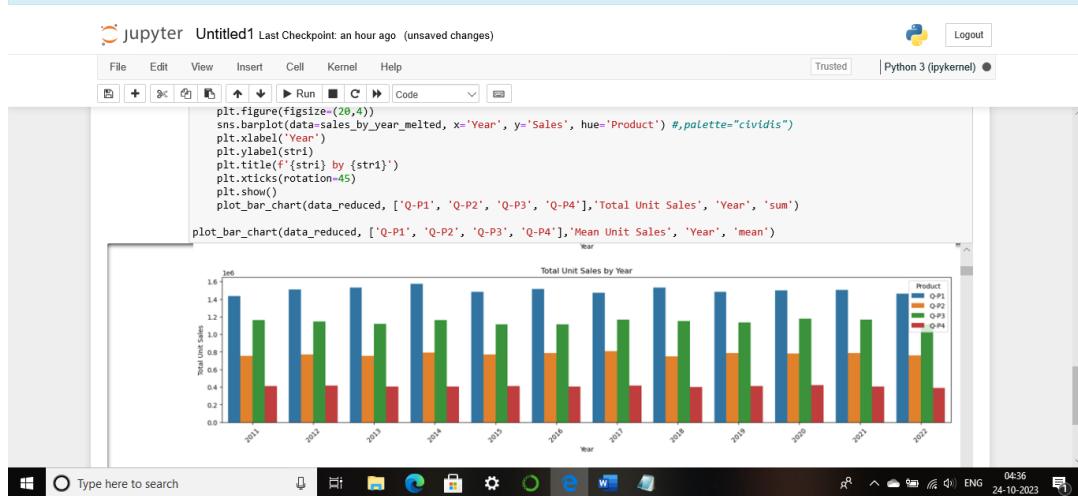
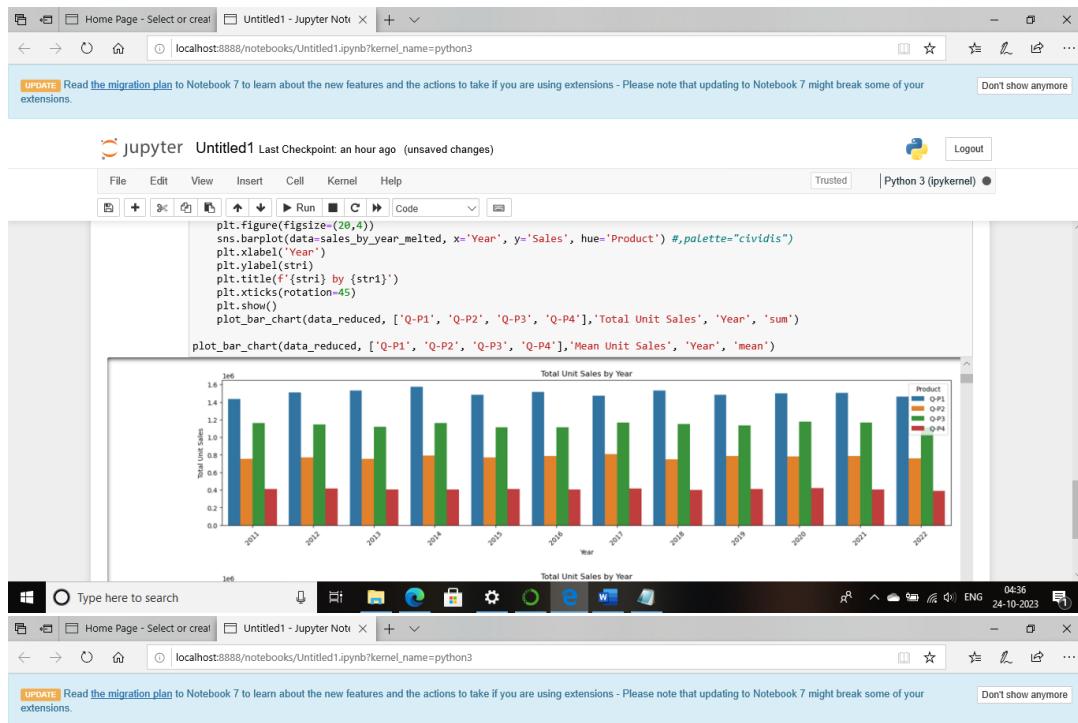
```
def plot_bar_chart(df, columns, str1, str2, val):
    if val == 'sum':
        sales_by_year = df.groupby('Year')[columns].sum().reset_index()
    elif val == 'mean':
        sales_by_year = df.groupby('Year')[columns].mean().reset_index()

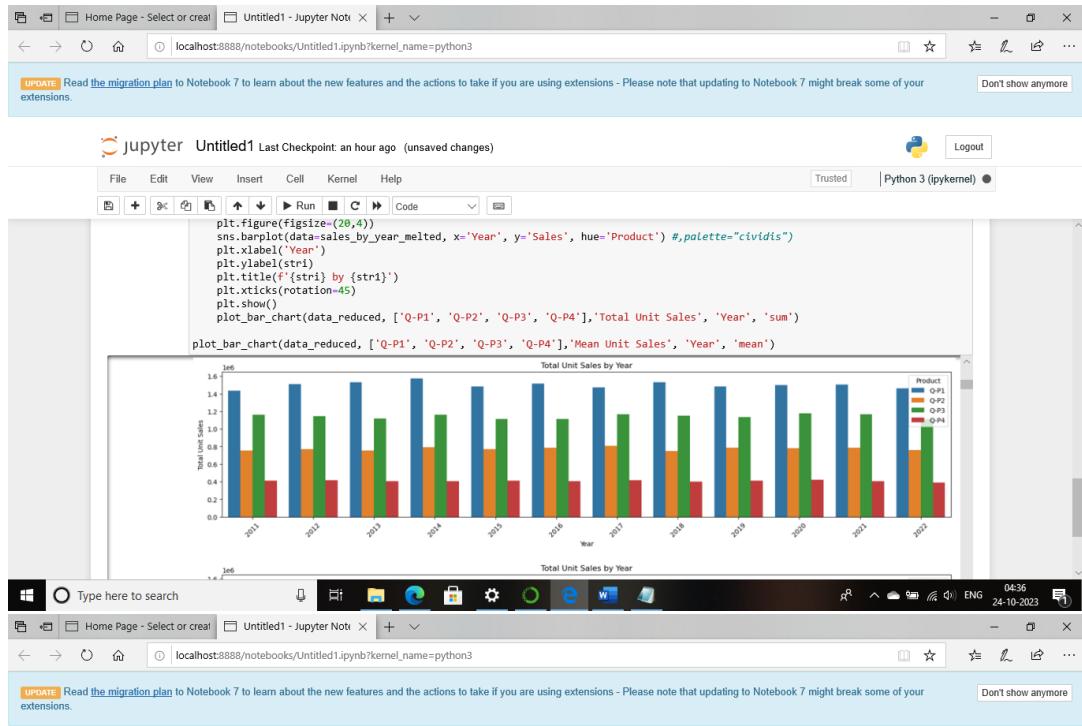
    sales_by_year_melted = pd.melt(sales_by_year, id_vars='Year', value_vars=columns, var_name='Product', value_name='Sales')

    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 {str2}')
    plt.xticks(rotation=45)
    plt.show()
    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')
```

Mean Unit Sales by Year

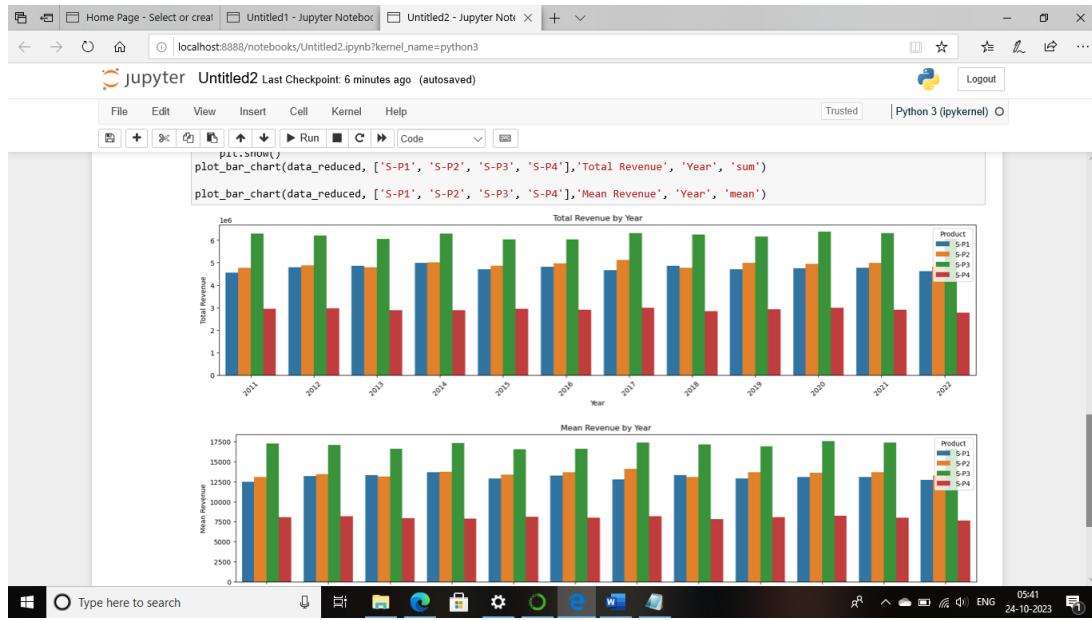
Year	Q-P1	Q-P2	Q-P3	Q-P4
2011	3500	3200	3100	3000
2012	3800	3500	3400	3300
2013	3600	3300	3200	3100
2014	3900	3600	3500	3400
2015	3700	3400	3300	3200
2016	3800	3500	3400	3300
2017	3600	3300	3200	3100
2018	3900	3600	3500	3400
2019	3700	3400	3300	3200
2020	3800	3500	3400	3300
2021	3600	3300	3200	3100
2022	3900	3600	3500	3400
2023	3700	3400	3300	3200





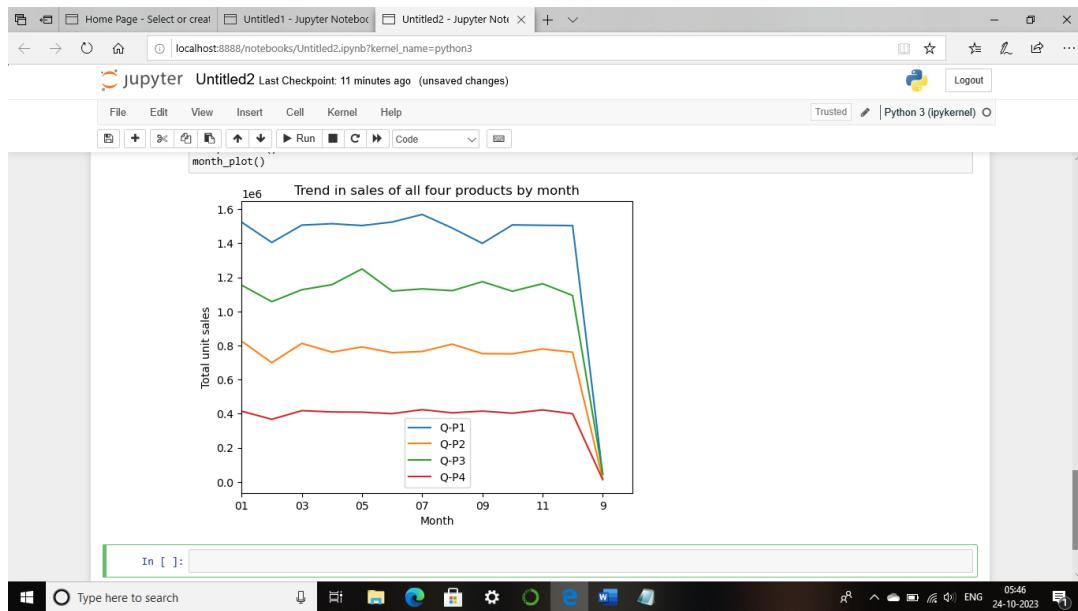
- 😊 We can see that the product P1 is the Highest Unit Sold for Each Year.
- 😊 P1 Sold Highest in the Year 2014
- 😊 P4 is the Lowest sold Product of all the Years.

# Graph Representation of TOTAL and MEAN of Revenue of Each Product

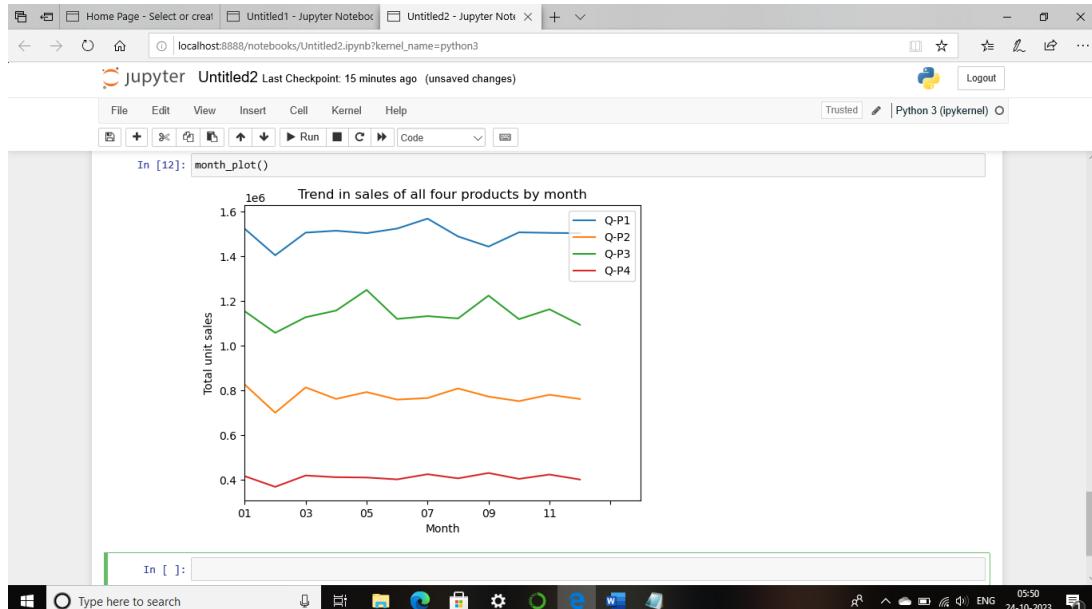


- 😊 We can see that P3 brought us the most revenue.
- 😊 We can observe that P1 and P2 brought in similar revenues for each year. With P2 bringing in slightly more.
- 😊 P1 despite having the most unit sold, brought in the second lowest revenue each year.

## Trend in Sales of all the Four Products:



😊 We can see that all the product's sales decrease massively in February.



😊 We can see that the Months February and December is the lowest rate of product sold for all the Four.

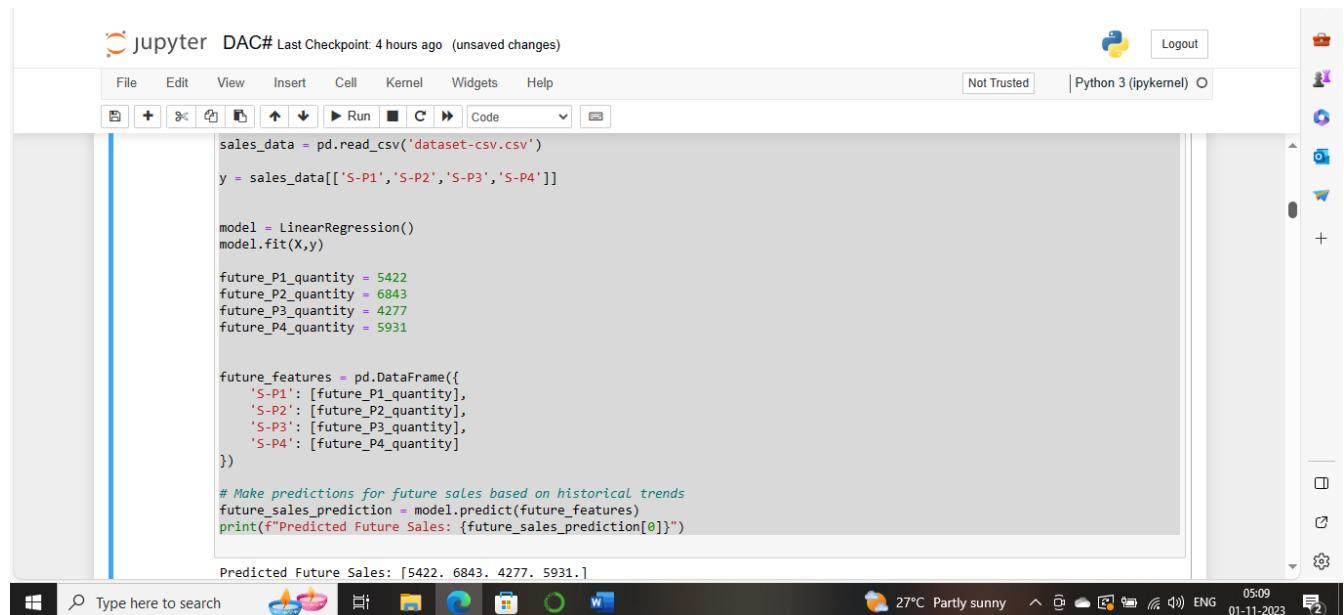
- 😊 The months March and July is the highest sold month for the Product P1
- 😊 The months January, March and August is the highest sold month for the Product P2

😊 The months May and September is the highest sold month for the Product P3

😊 From the month January to December the Product P4 has Uniform Sales.

## To Predict Future Sales:

Predicting future sales with machine learning is a game changer for industry. By analysing past sales data, these smart algorithms reveal trends, helping with inventory, demand planning and pricing. It sounds like you have a crystal ball for optimizing performance and staying ahead in a dynamic market.



The screenshot shows a Jupyter Notebook interface running on a Windows operating system. The notebook displays a single code cell containing Python code for a linear regression model to predict future sales. The code reads a CSV dataset, defines features and labels, fits a LinearRegression model, and makes predictions for four products (P1-P4) based on their current quantities. The output of the code cell shows the predicted future sales values for each product.

```
jupyter DAC# Last Checkpoint: 4 hours ago (unsaved changes)
File Edit View Insert Cell Kernel Widgets Help
Not Trusted Python 3 (ipykernel) ○
sales_data = pd.read_csv('dataset-csv.csv')
y = sales_data[['S-P1','S-P2','S-P3','S-P4']]

model = LinearRegression()
model.fit(X,y)

future_P1_quantity = 5422
future_P2_quantity = 6843
future_P3_quantity = 4277
future_P4_quantity = 5931

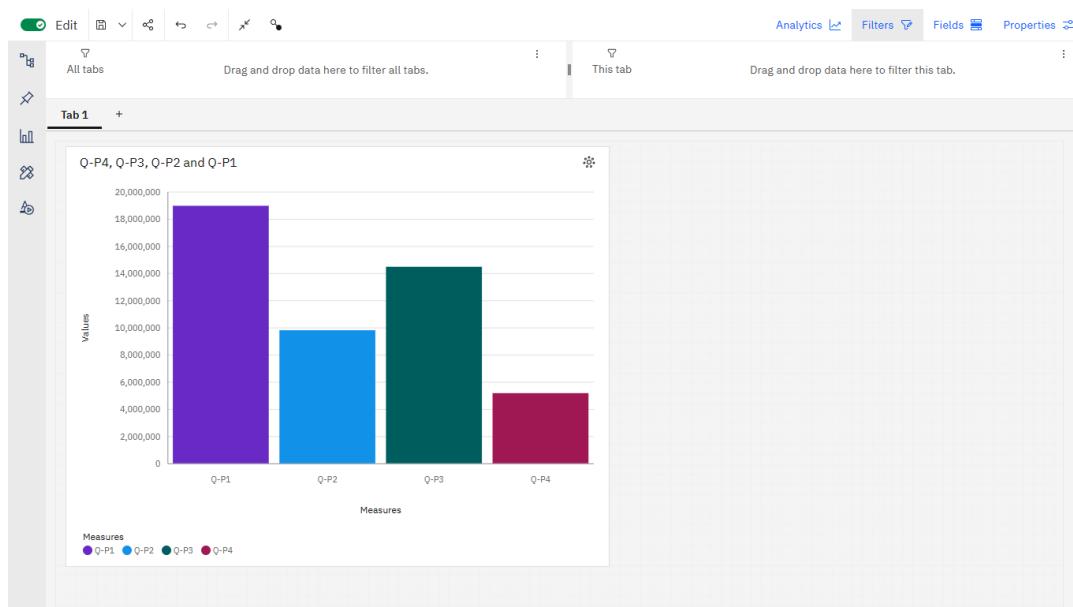
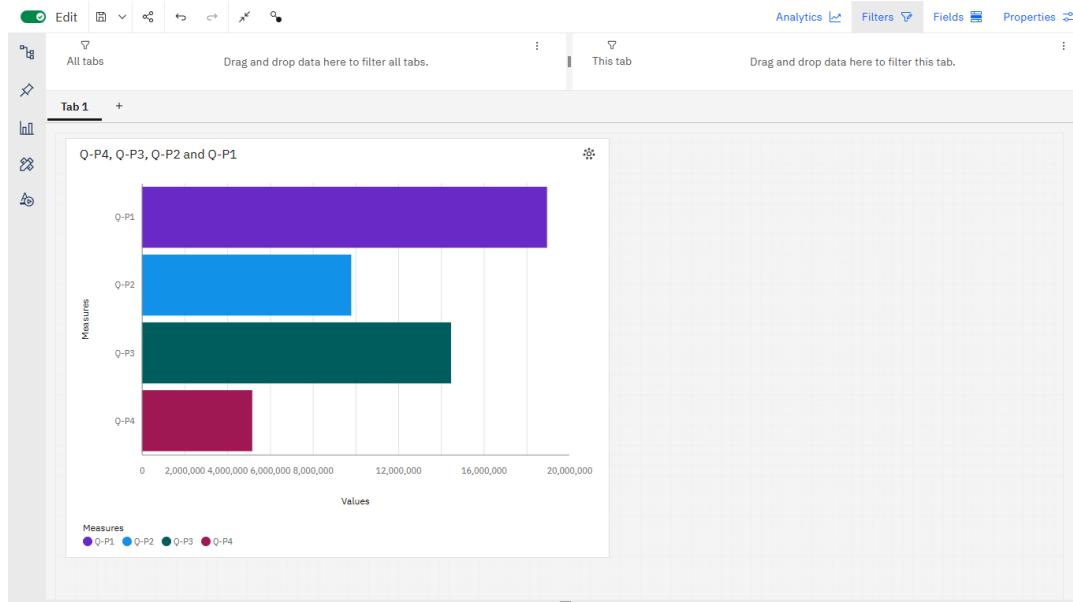
future_features = pd.DataFrame({
    'S-P1': [future_P1_quantity],
    'S-P2': [future_P2_quantity],
    'S-P3': [future_P3_quantity],
    'S-P4': [future_P4_quantity]
})

# Make predictions for future sales based on historical trends
future_sales_prediction = model.predict(future_features)
print(f"Predicted Future Sales: {future_sales_prediction[0]}")
```

Predicted Future Sales: [5422. 6843. 4277. 5931.]

# Data Visualization in IBM Cognos Analytics:

## Representation on highest sold products over the years



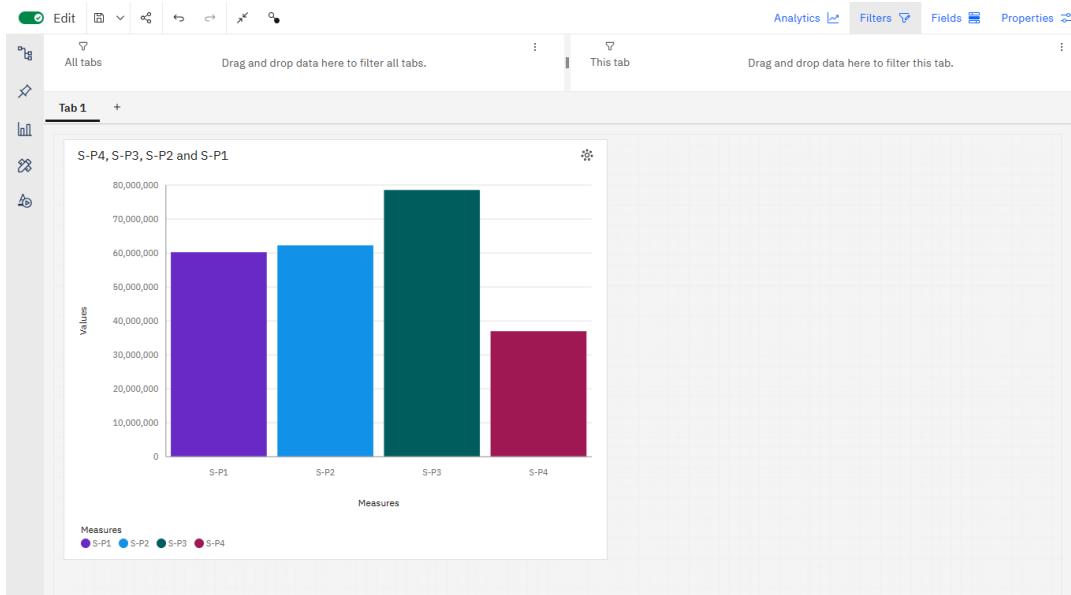
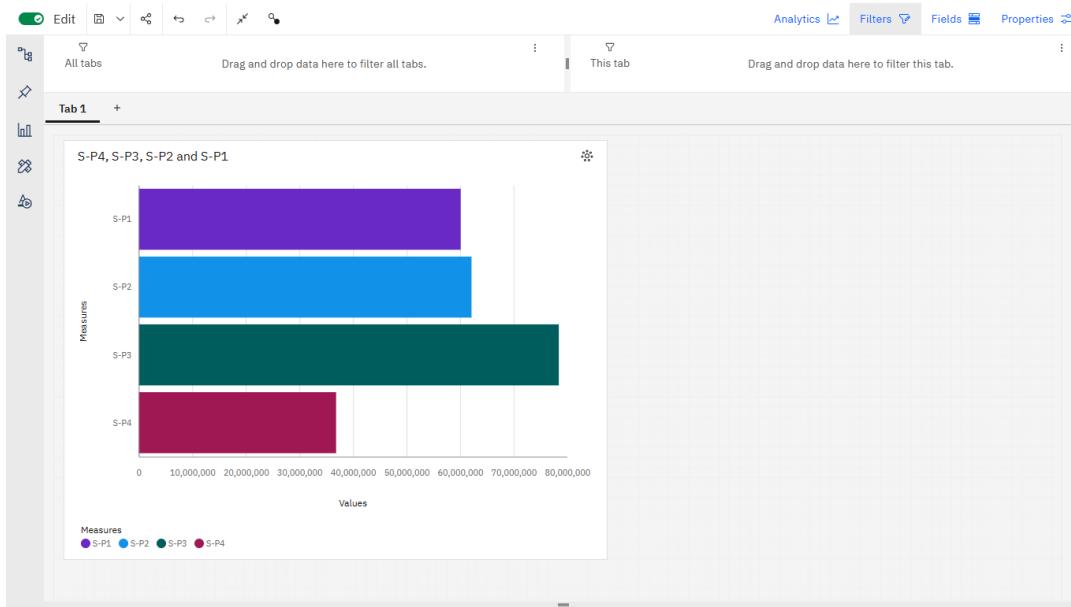
## Total sum of units sold of each product Q-P1, Q-P2, Q-P3, Q-P4

Q-P1, Q-P2, Q-P3 and Q-P4			
Q-P1	Q-P2	Q-P3	Q-P4
18,960,506	9,799,295	14,470,404	5,168,100

😊 From these we can say that the P1 is the highest sold product

FIRST HIGHEST SOLD PRODUCT	P1
SECOND HIGHEST SOLD PRODUCT	P3
THIRD HIGHEST SOLD PRODUCT	P2
FOURTH HIGHEST SOLD PRODUCT	P4

# Representation on highest Revenue generated products over the years



## Total sum of revenue generated by each product

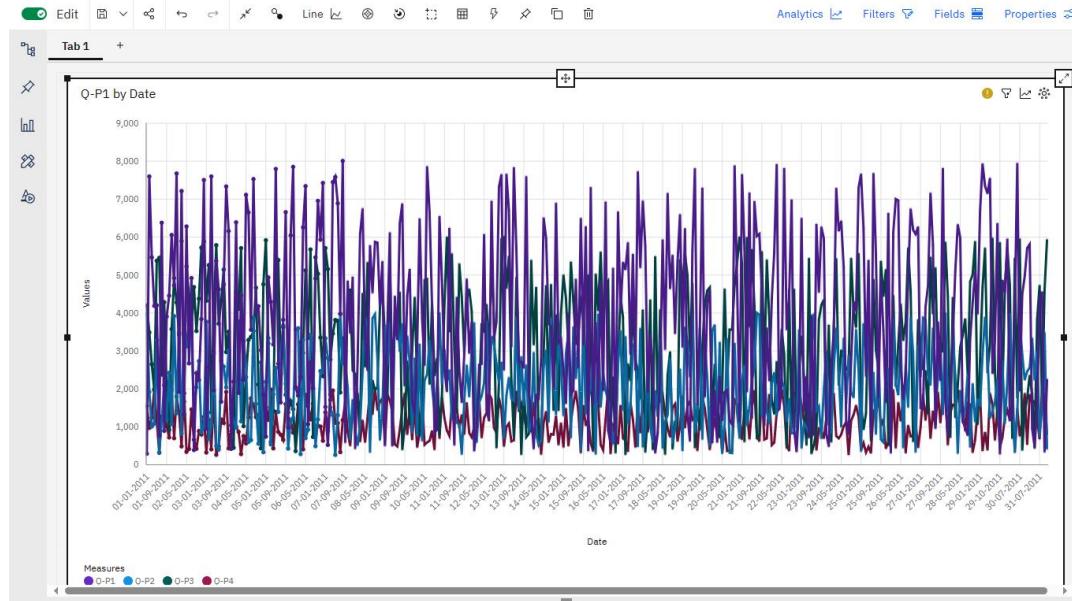
The screenshot shows a data visualization interface with a toolbar at the top. Below the toolbar, there are two tabs: "All tabs" and "This tab". A message "Drag and drop data here to filter all tabs." is displayed above the "All tabs" tab, and "Drag and drop data here to filter this tab." is displayed above the "This tab" tab. The main area contains a table titled "Tab 1" with four columns labeled "S-P1", "S-P2", "S-P3", and "S-P4". The data rows show the following values:

S-P1	S-P2	S-P3	S-P4
60,104,804.02	62,127,530.3	78,429,589.68	36,848,553

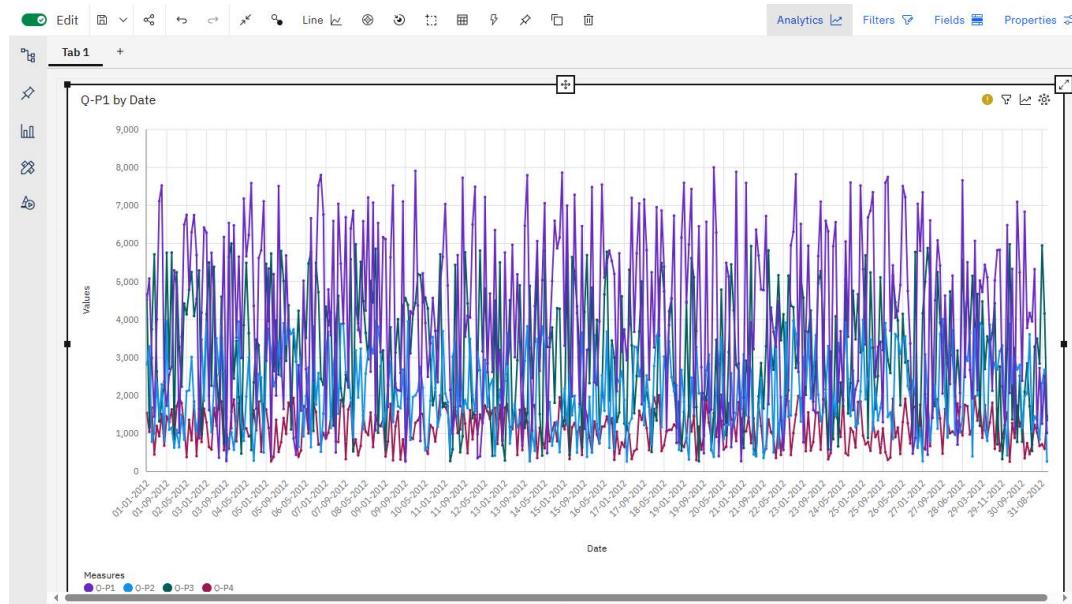
FIRST HIGHEST REVENUE GENERATED PRODUCT	P3
SECOND HIGHEST REVENUE GENERATED PRODUCT	P2
THIRD HIGHEST REVENUE GENERATED PRODUCT	P1
FOURTH HIGHEST REVENUE GENERATED PRODUCT	P4

## Line Chart Over Time to find Peak Sales

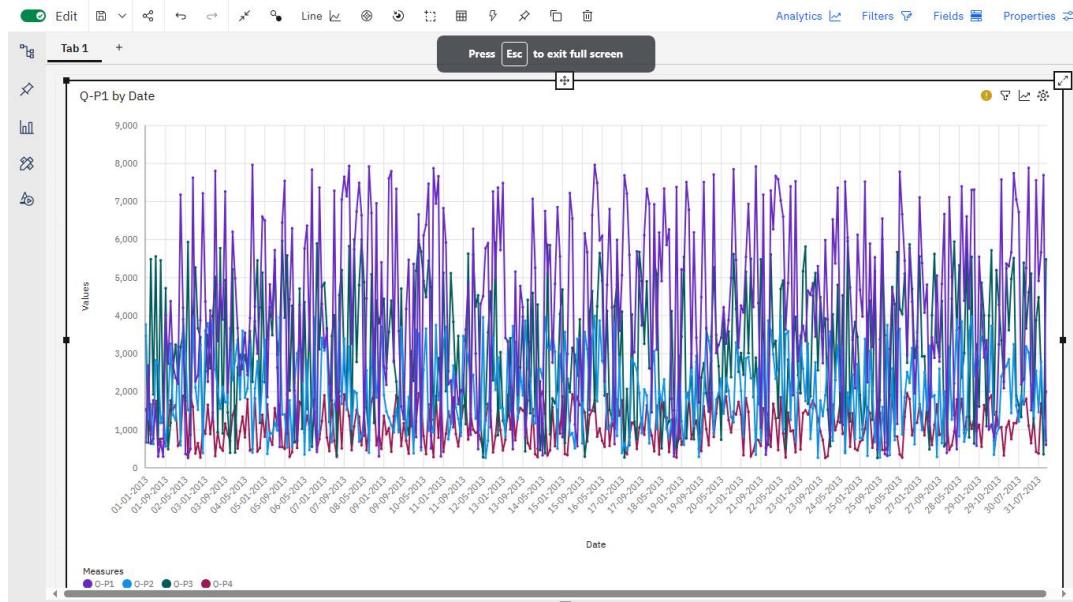
In 2011 for Q-P1, Q-P2, Q-P3, Q-P4



In 2012 for Q-P1, Q-P2, Q-P3, Q-P4



In 2013 for Q-P1, Q-P2, Q-P3, Q-P4



## SALES TREND:

By analysing all the years sales of all products, we can say that,

- 😊 All Products unit sale drop in February month.
- 😊 In all the years the month January, March, August has peak sale for the product P2.
- 😊 In all the years the month May, September has peak sale for the product P3.
- 😊 The Product P4 has a uniform sales all over the Years.

## CUSTOMER PREFERENCES:

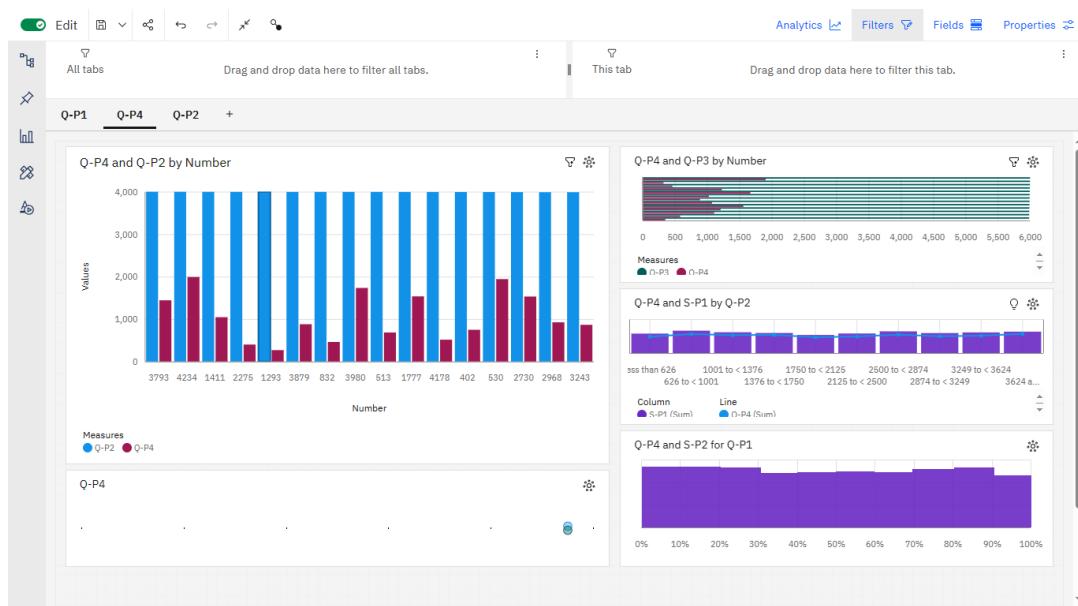
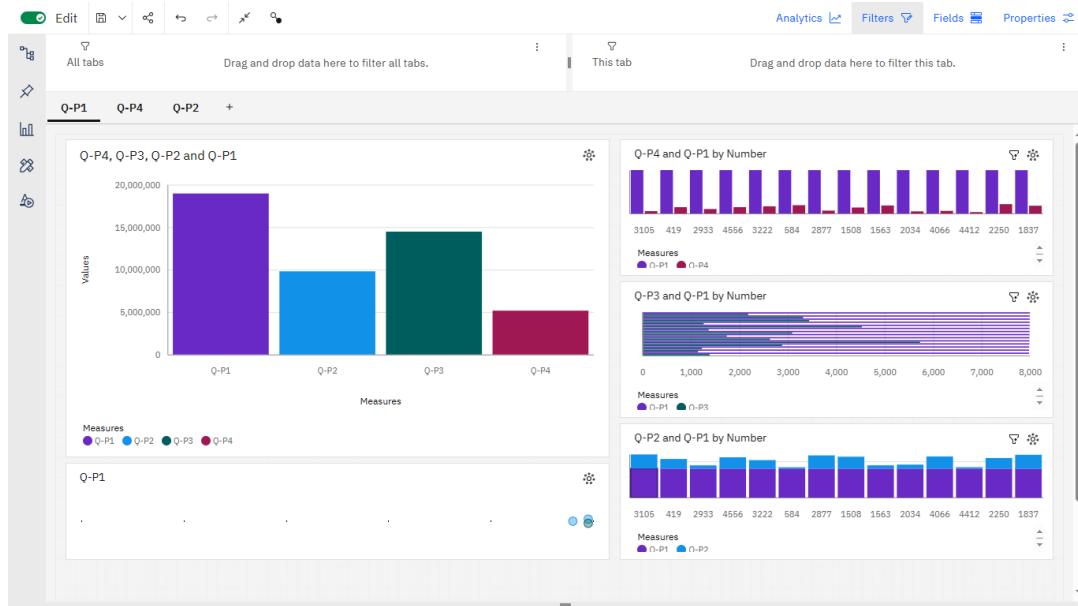
By these analysis and the visualizations we can see the customers prefers the,

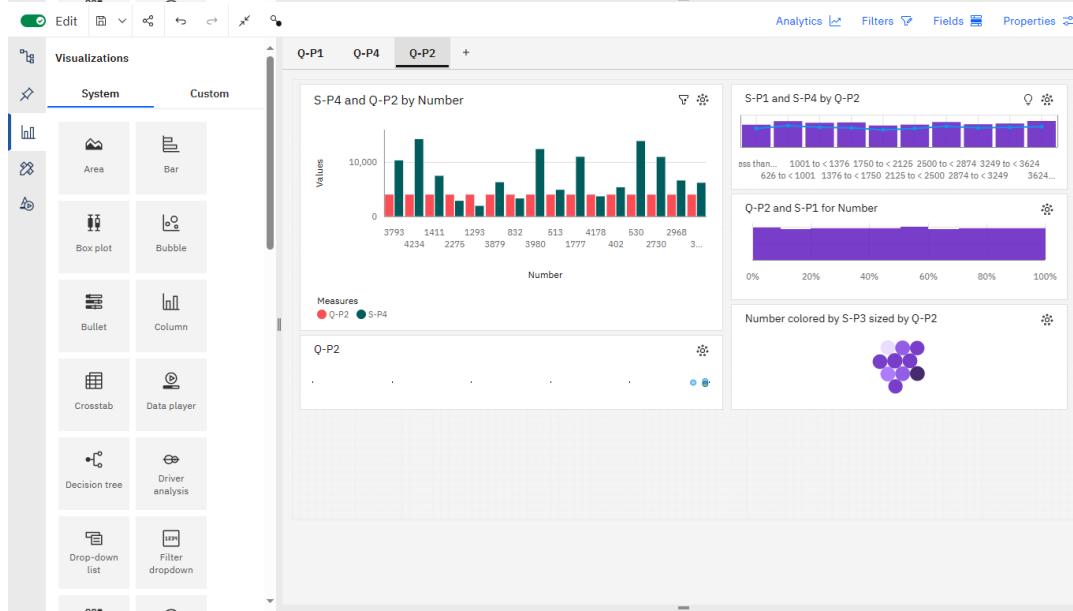
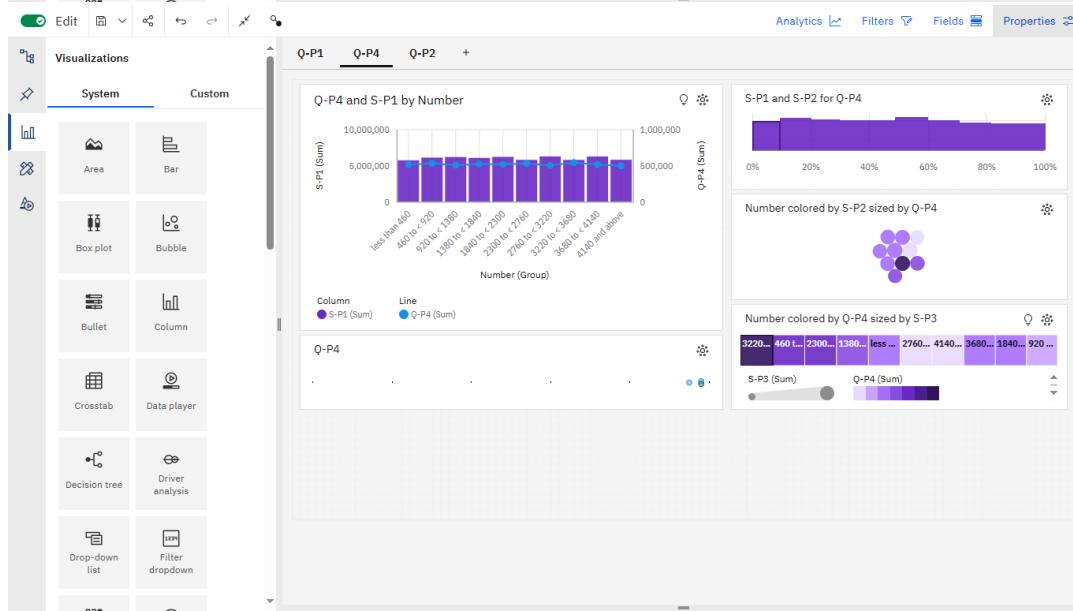
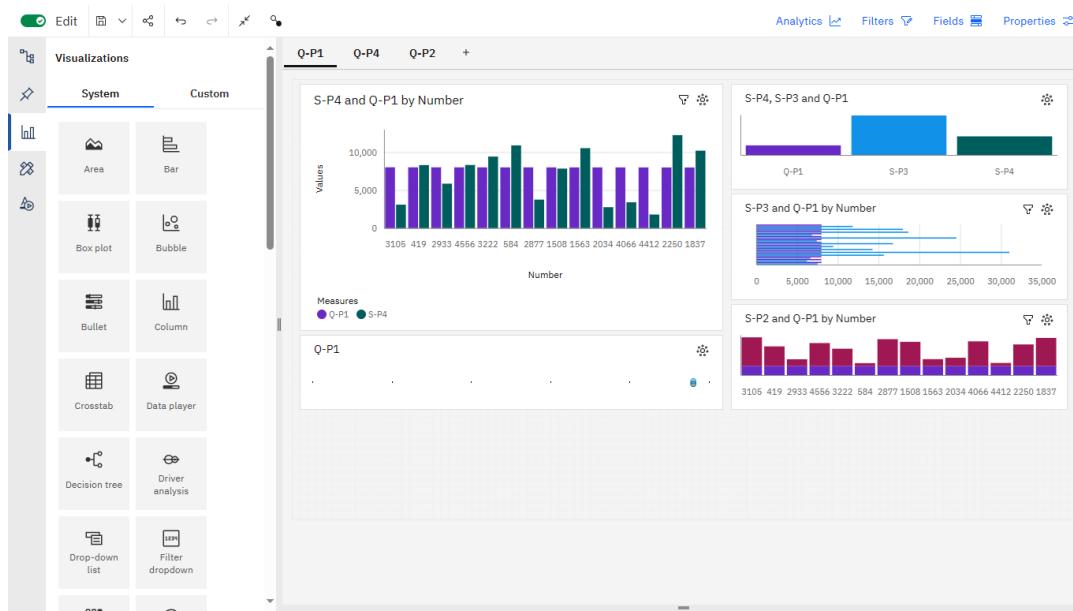
Product P1 the most and

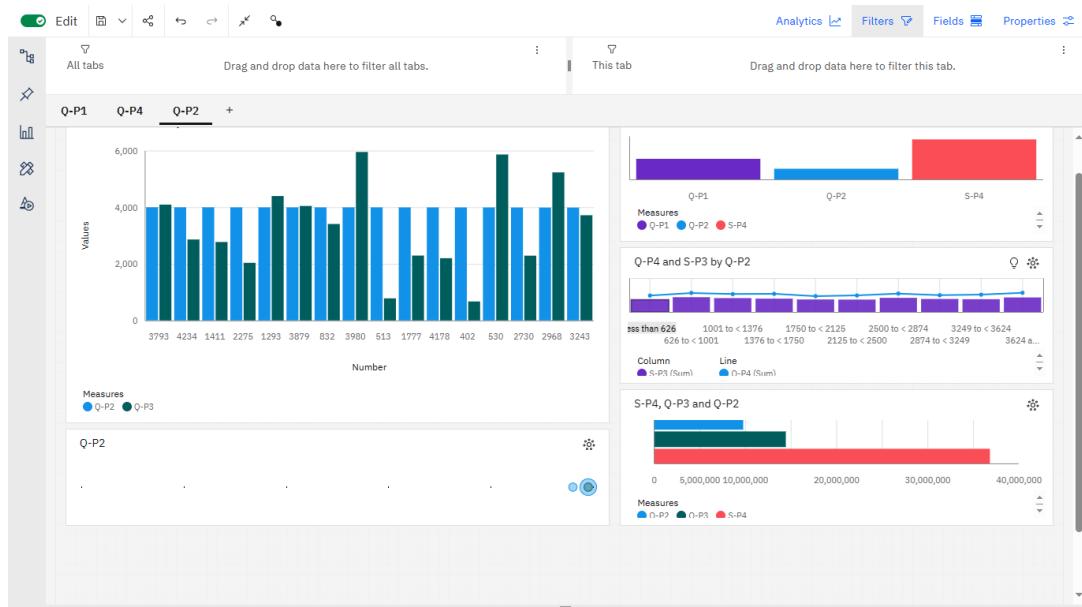
Product P3 is the next preferred, and

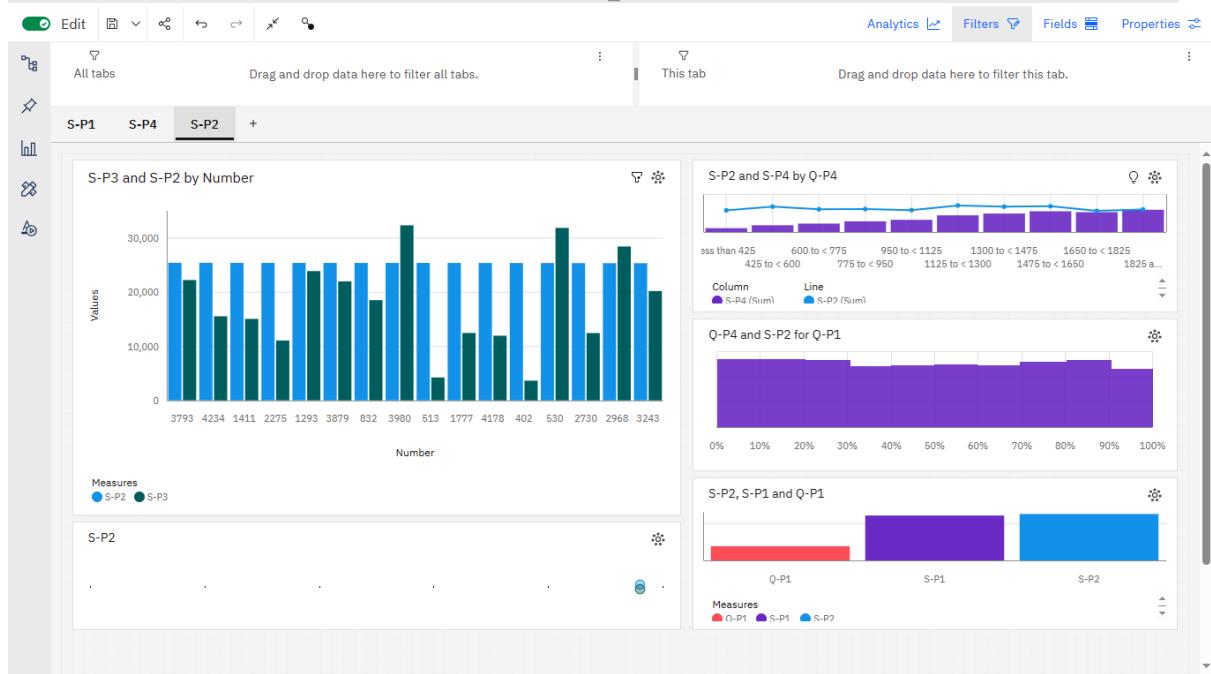
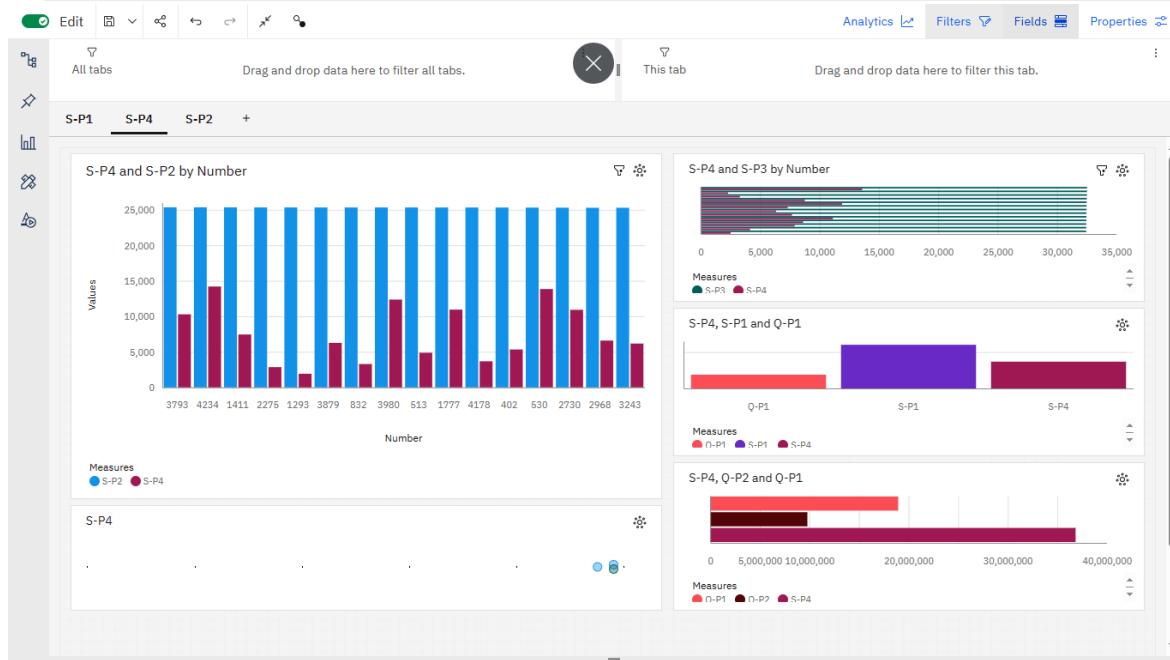
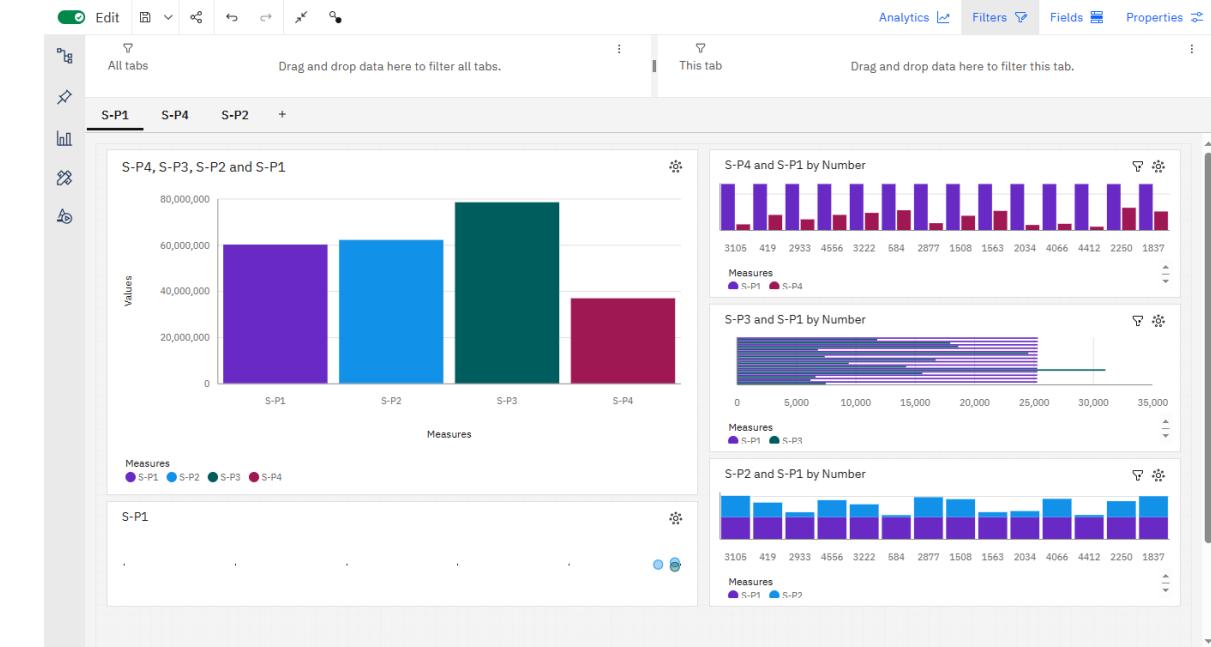
P2 and P4 as next.

## DASHBOARDS:









How my insights from the analysis can guide inventory management and marketing strategies:

## Inventory Management:

### Optimizing stock levels:

By identifying the best-selling products, companies can strategically diversify their inventory. This insight allows for optimal stocking of popular brands, ensuring that customer's needs are met with minimal inventory. This practice prevents underutilisation, which can lead to missed sales opportunities, and overutilisation, which ties up capital and storage space.

### Adjustment in Time:

Identifying sales trends over time is invaluable for seasonal inventory management. When products experience an increase in sales during certain holidays or seasons, companies can actively adjust their inventory to meet anticipated demand. This reduces the risk of dropping inventory during peak periods and reduces overloading during slow periods.

### Reduction in Dead Stock:

Your research may also have highlighted a lot of sales. By identifying these inefficiencies, companies can make data-driven decisions to phase out such features. This reduces dead stock and frees up storage space and capital for more profitable construction.

## Marketing Strategies:

### Target Marketing:

Understanding consumer preferences and behaviours allows companies to create more targeted marketing campaigns. Consumers can be segmented based on their preferences and past purchase behaviours. By tailoring promotions and advertising to specific customer segments, companies can maximize the relevance of their marketing efforts, increasing conversion rates and optimizing marketing budgets.

### Product Recommendations:

Gaining insight into customer preferences enables successful product recommendation programs to be implemented. Companies can enhance user experience by providing personalized offerings. This not only facilitates new purchases but also increases customer loyalty.

### Optimized pricing:

Sales data and consumer behaviour data can provide valuable insights into pricing strategies. For example, if a product performs consistently well during peak sales periods, companies may choose to offer discounts or promotions during those periods to gain capital Pricing options based on data can increase sales volume without sacrificing profitability.

## Product Combining:

Identifying common purchases together can guide bundling strategies. Companies can create bundles to meet typical customer preferences, resulting in higher average retail prices. Bundling also encourages customers to look for extras they might not otherwise have considered.

## Content Creation:

Sales information data can impact product development and marketing efforts. For example, if certain products are consistently popular, companies can create products and advertising campaigns around those products to further increase sales. This approach leverages existing customer interest and maximizes return on marketing investment.

By successfully integrating these insights into inventory and marketing strategies, businesses can achieve efficiencies, increased revenue and an enhanced user experience. This ensures that products offerings are better aligned with customer demand, improving overall customer satisfaction and performance.

## Conclusion:

In conclusion, this project was a strong incentive to demonstrate valuable creativity, hard work and insight. It all started with understanding what stakeholders needed and the challenges they faced managing their products and sales. We brainstormed alternative solutions and developed an action plan.

Using IBM Cognos, we dug deeper into the data, uncovering useful insights about best-sellers, products, and customer preferences. The impact of our research on how businesses manage their inventory and market their products

is critical. We are at the crossroads of data-driven decision-making and operational efficiency.

Our findings provide businesses with a clear path to effectively manage their inventories, adjust to seasonal changes, and reduce unsold inventory. They also help with targeted marketing efforts and creating user-friendly experiences by identifying products that customers are likely to appreciate. What's more, our ability to forecast future sales attests to the success of our business in making proactive decisions.

This project is not just the end; This is also the beginning. We learned from the challenges we faced and are excited about the impact our findings will have on decision-making and business development.