

DATA ANALYTICS WITH COGNOS-GROUP 2

PROJECT 5: Product Sales Analysis

Phase3: Development Part 1

Introduction:

In Phase 3, we are entering the development phase of our product marketing research project. This phase marks the beginning of building the foundation of the project, including loading and preprocessing the data set, defining the objectives of the analysis, and beginning to use IBM Cognos for visualization.

Objectives:

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.

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 DAC_Phase3 Last Checkpoint: an hour ago (unsaved changes) Logout

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

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

	Unnamed: 0	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	7047	779	3578	1574	22338.99	4938.86
2	2	15-06-2010	1572	2082	595	1145	4983.24	13199.88
3	3	16-06-2010	5657	2399	3140	1672	17932.69	15209.66
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					

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 DAC_Phase3 Last Checkpoint: an hour ago (autosaved) Logout

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

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

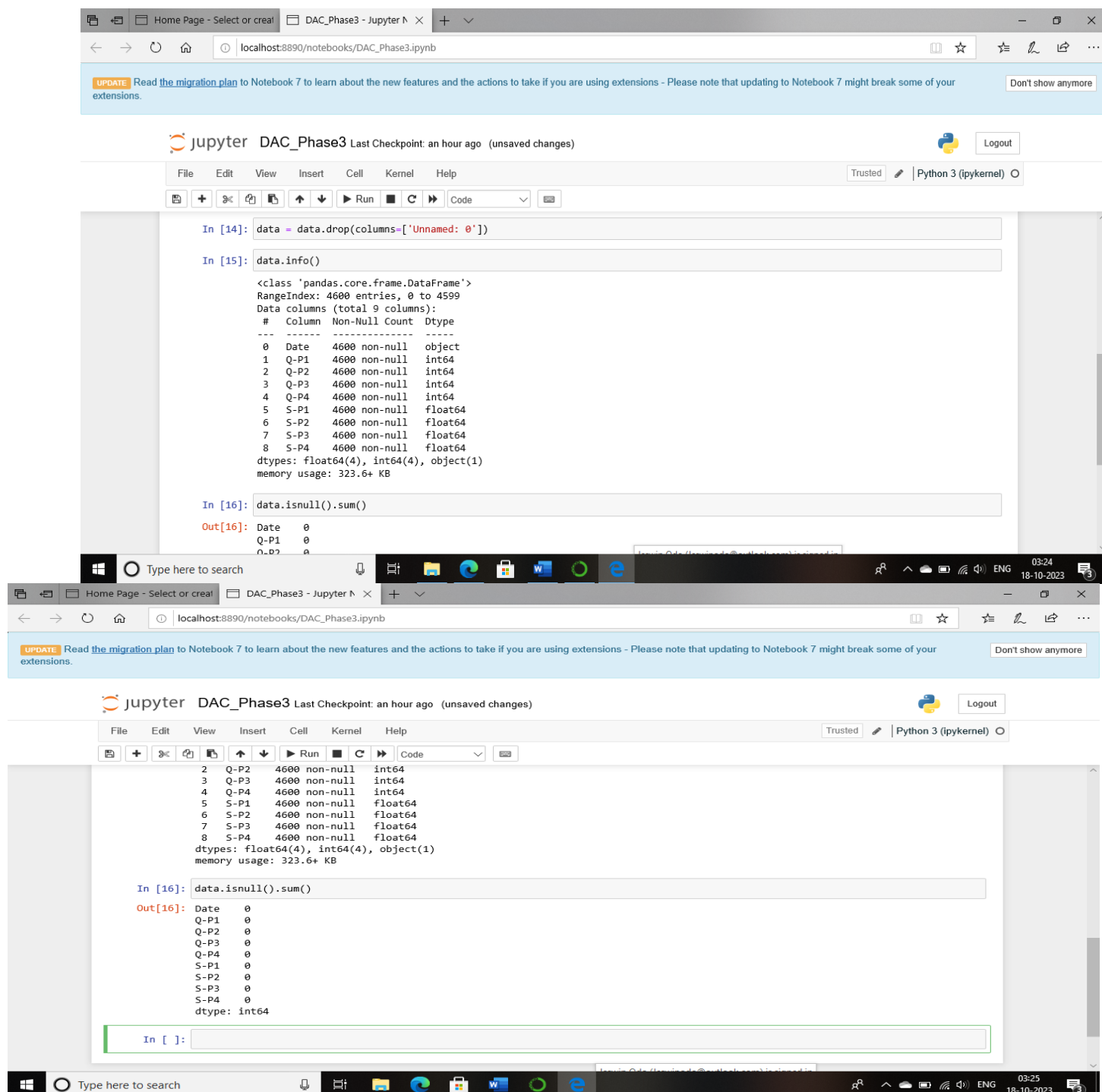
	Unnamed: 0	Date	Q-P1	Q-P2	Q-P3	Q-P4	S-P1	S-P2
0	0	13-06-2010	5422	3725	576	907	17187.74	23616.50
1	1	14-06-2010	7047	779	3578	1574	22338.99	4938.86
2	2	15-06-2010	1572	2082	595	1145	4983.24	13199.88
3	3	16-06-2010	5657	2399	3140	1672	17932.69	15209.66
4	4	17-06-2010	3668	3207	2184	708	11627.56	20332.38
...
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		17018.80	11921.36					
4		11837.28	5048.04					
...						
4595		2845.50	9689.67					
4596		26151.50	9347.43					
4597		19446.96	3379.62					
4598		31972.58	3686.21					
4599		12579.82	2894.78					

[4600 rows x 10 columns]

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.

MISSING VALUES:



The screenshot displays a Jupyter Notebook interface with the following content:

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

In [15]: data.info()

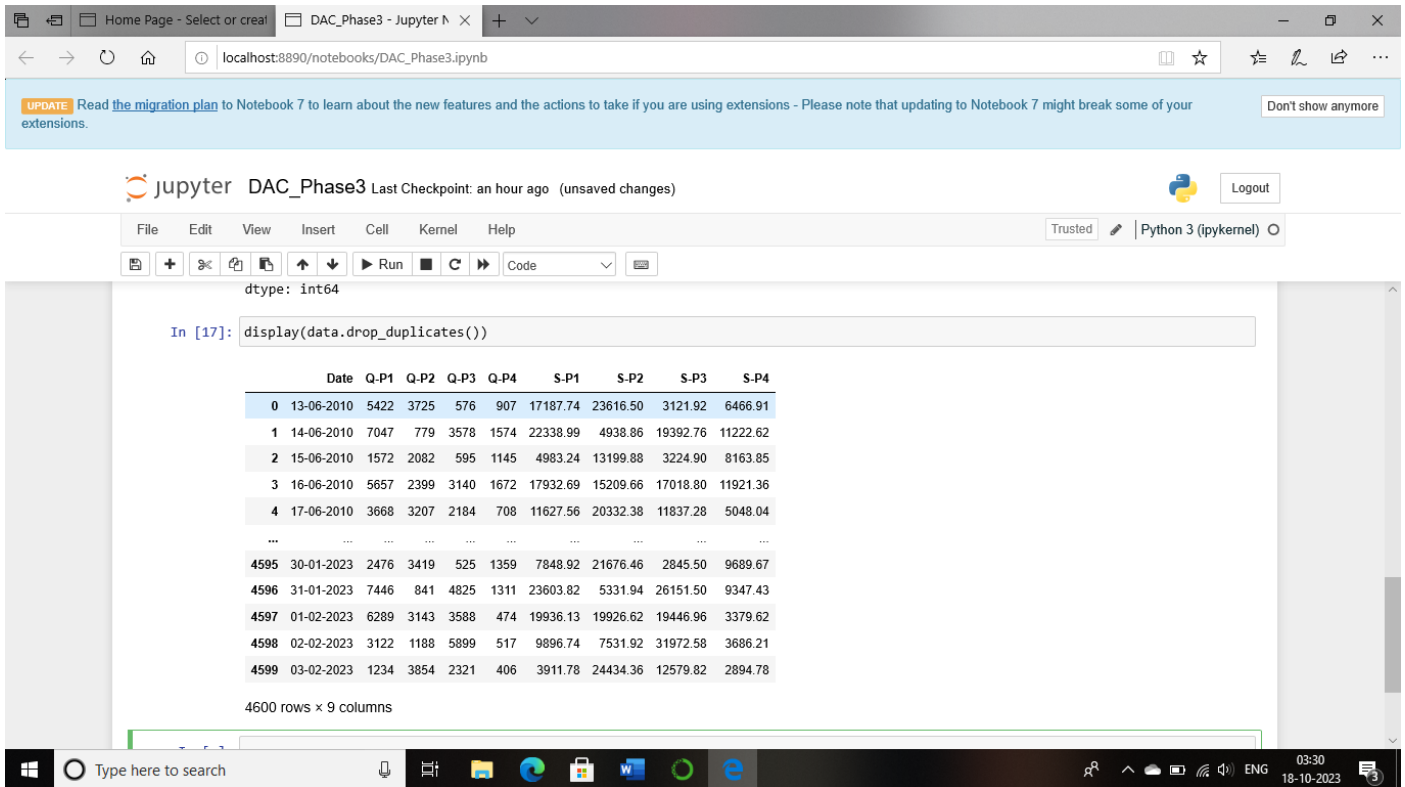
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4600 entries, 0 to 4599
Data columns (total 9 columns):
#   Column      Non-Null Count  Dtype  
---  -
0    Date        4600 non-null   object  
1    Q-P1        4600 non-null   int64   
2    Q-P2        4600 non-null   int64   
3    Q-P3        4600 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 notebook interface includes a top bar with the title "Jupyter DAC_Phase3" and a "Last Checkpoint: an hour ago (unsaved changes)" message. The bottom bar shows the file explorer, search bar, and system tray with the date "18-10-2023" and time "03:24".

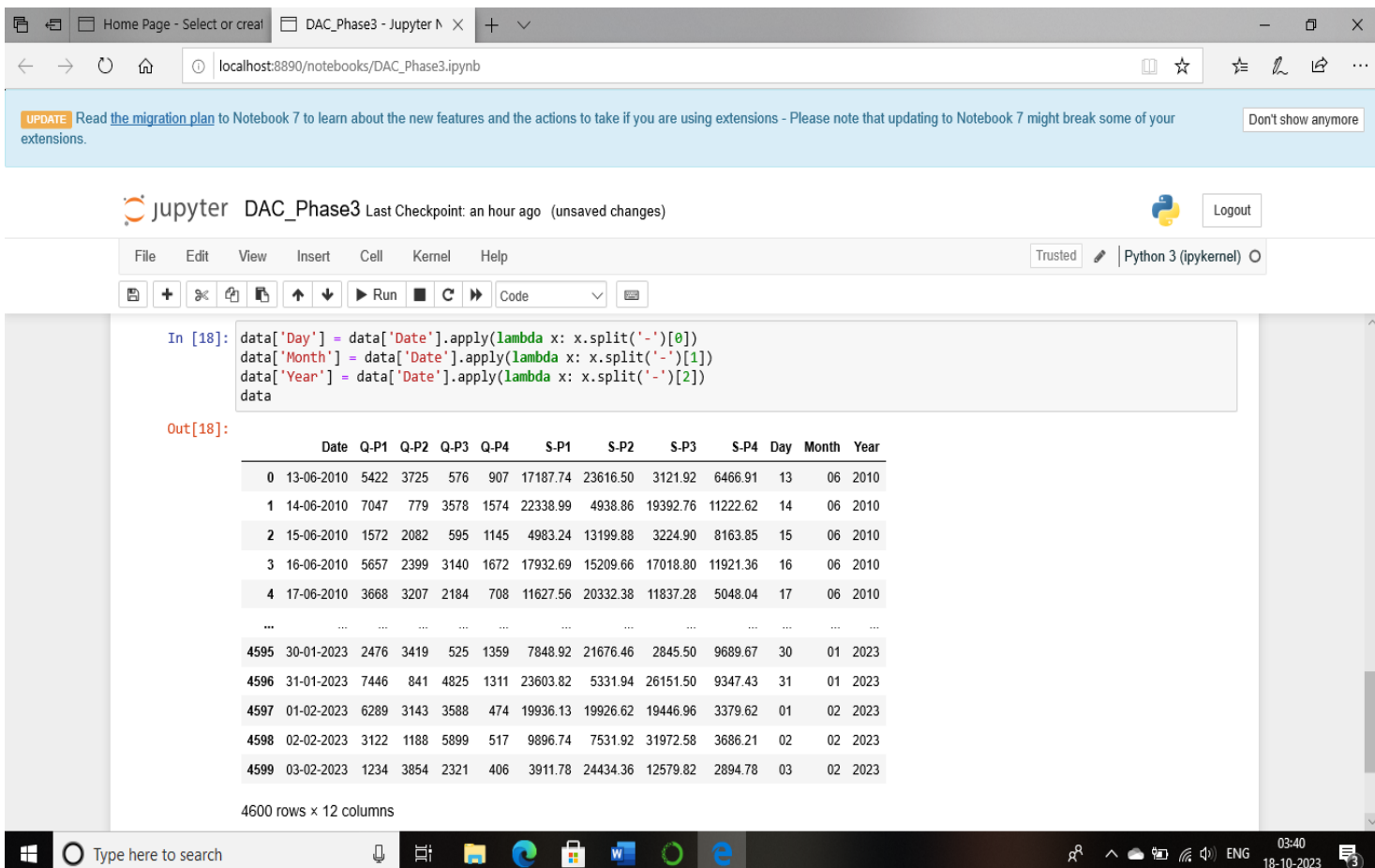
CHECKING FOR DUPLICATES:



The screenshot shows a Jupyter Notebook interface with the following components:

- Browser Address Bar:** localhost:8890/notebooks/DAC_Phase3.ipynb
- Update Banner:** 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.
- Header:** jupyter DAC_Phase3 Last Checkpoint: an hour ago (unsaved changes) Logout
- Menu Bar:** File Edit View Insert Cell Kernel Help
- Toolbar:** Includes icons for saving, running, and other notebook functions. The selected mode is 'Code'.
- Code Cell:** Contains the command `display(data.drop_duplicates())`.
- Output:** A table showing the result of dropping duplicates. The first column is an index (0 to 4599). The subsequent columns are Date, Q-P1, Q-P2, Q-P3, Q-P4, S-P1, S-P2, S-P3, and S-P4. The table shows data for various dates from 2010 to 2023. Below the table, it states '4600 rows x 9 columns'.

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



The screenshot shows a Jupyter Notebook interface with the following components:

- Browser Address Bar:** localhost:8890/notebooks/DAC_Phase3.ipynb
- Update Banner:** 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.
- Header:** jupyter DAC_Phase3 Last Checkpoint: an hour ago (unsaved changes) Logout
- Menu Bar:** File Edit View Insert Cell Kernel Help
- Toolbar:** Includes icons for saving, running, and other notebook functions. The selected mode is 'Code'.
- Code Cell:** Contains the following code:

```
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
```
- Output:** A table showing the result of splitting the date strings. The first column is an index (0 to 4599). The subsequent columns are Date, Q-P1, Q-P2, Q-P3, Q-P4, S-P1, S-P2, S-P3, S-P4, Day, Month, and Year. The table shows data for various dates from 2010 to 2023. Below the table, it states '4600 rows x 12 columns'.

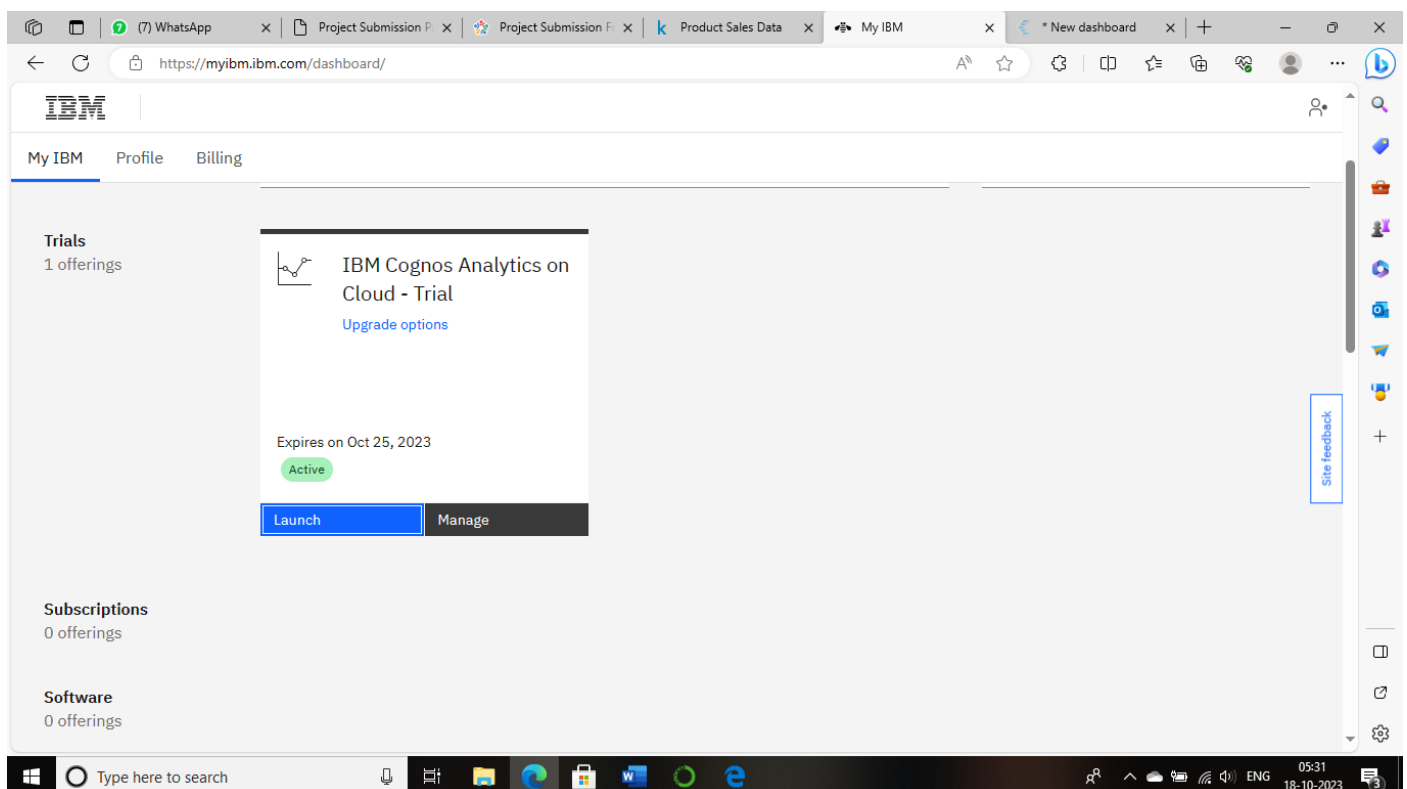
Defining Analysis Objectives:

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 Visualisation 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.



IBM | My IBM | Profile | Billing

https://www.ibm.com/ibmweb/myibm/us-en/profile

ID and password

Contact information

Company or organization

Addresses

Password

.....

Last updated September 25, 2023

Edit

Security

You can enable/disable your two-step verification options using the toggle button.

Off

Contact information

Name

Jeswin A A

Edit

Display name

You haven't added display name.

Edit

Email address

oodo5210@gmail.com

Primary

Edit

Phone number

Business

Edit

Site feedback

05:31 18-10-2023

IBM Cognos Analytics

https://us3.ca.analytics.ibm.com/bi/7perspective=home

Upload data and start analysis

Initializing file upload...

Cancel Details

You can upload supported file types that are stored in any location to which your computer has local or LAN access.

Drag and drop file here or click to upload

Want to know more about uploading data? [Learn more](#)

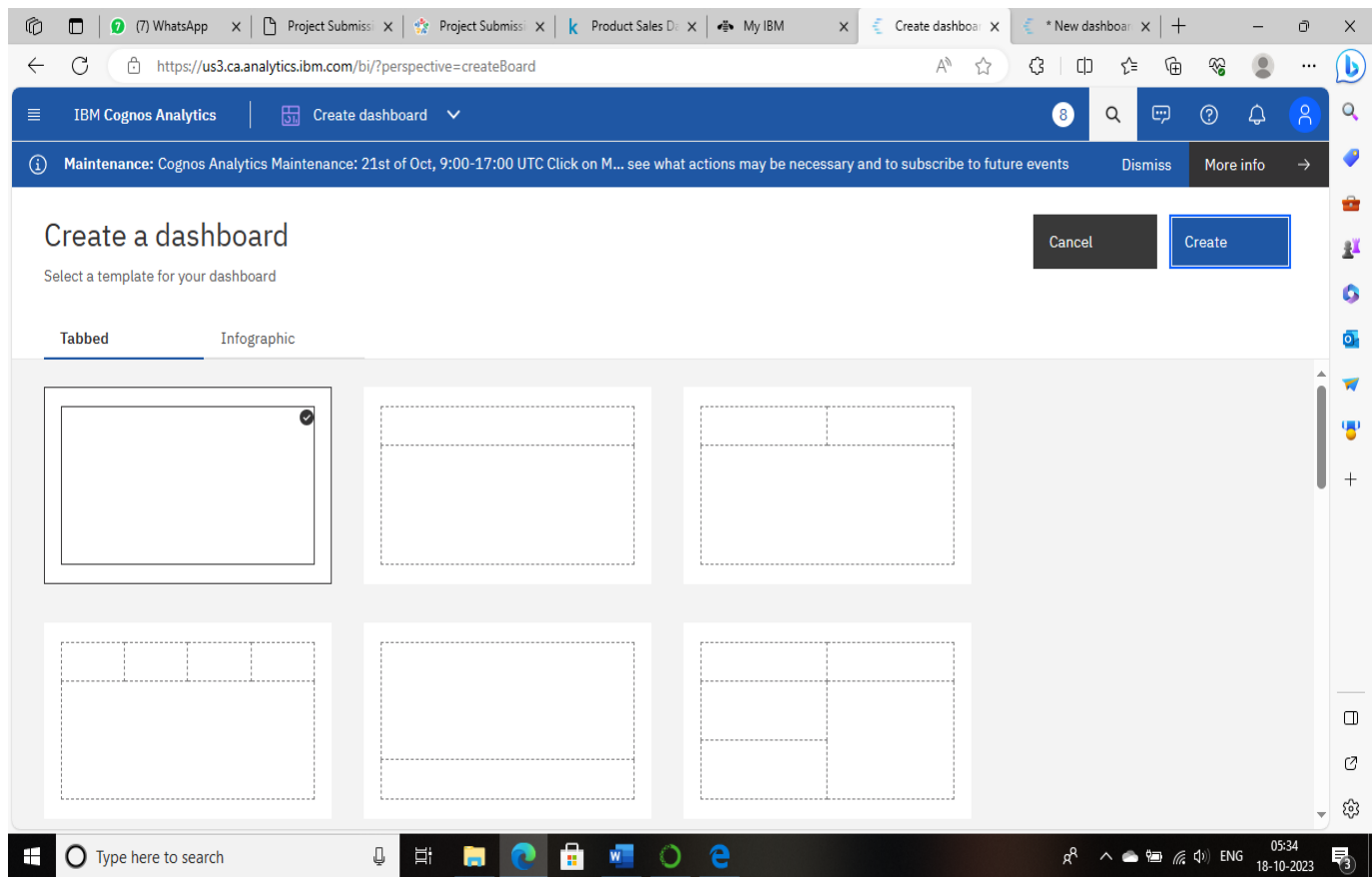
Cancel

Back

Next

Layout Design:

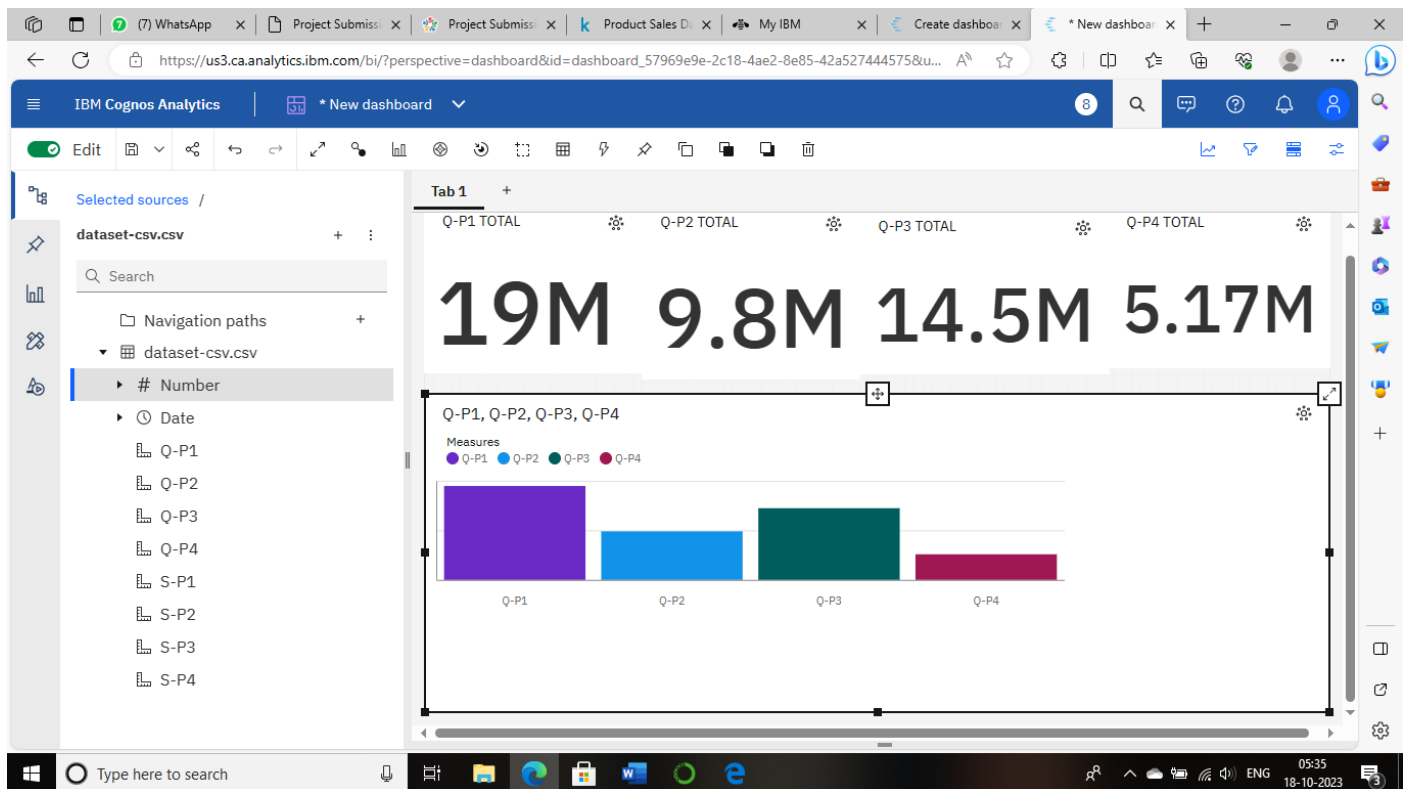
We're actively planning the layout and structure for our dashboard or visualizations. This design will be instrumental in presenting our findings effectively.



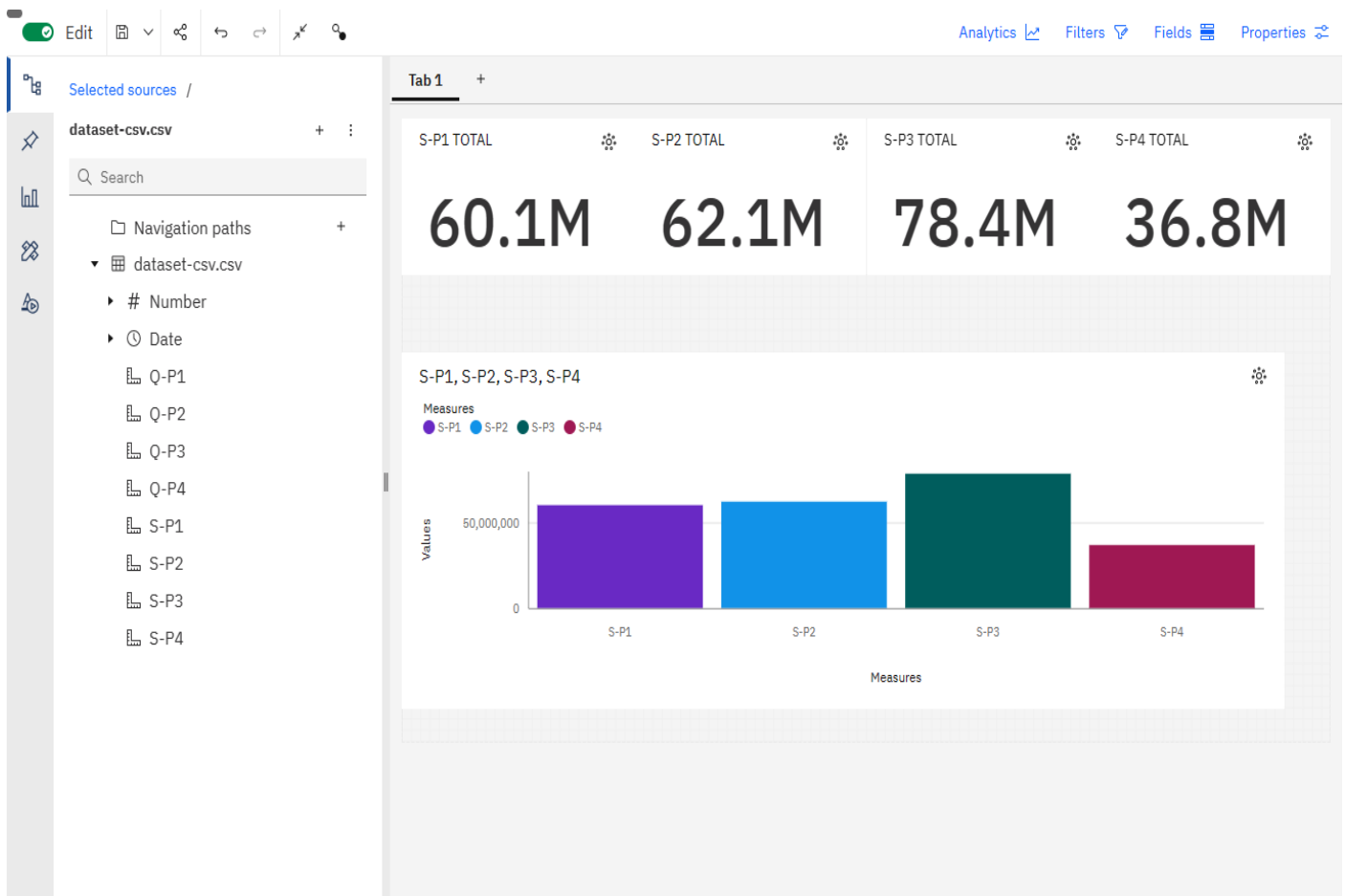
Initial Visualizations:

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.

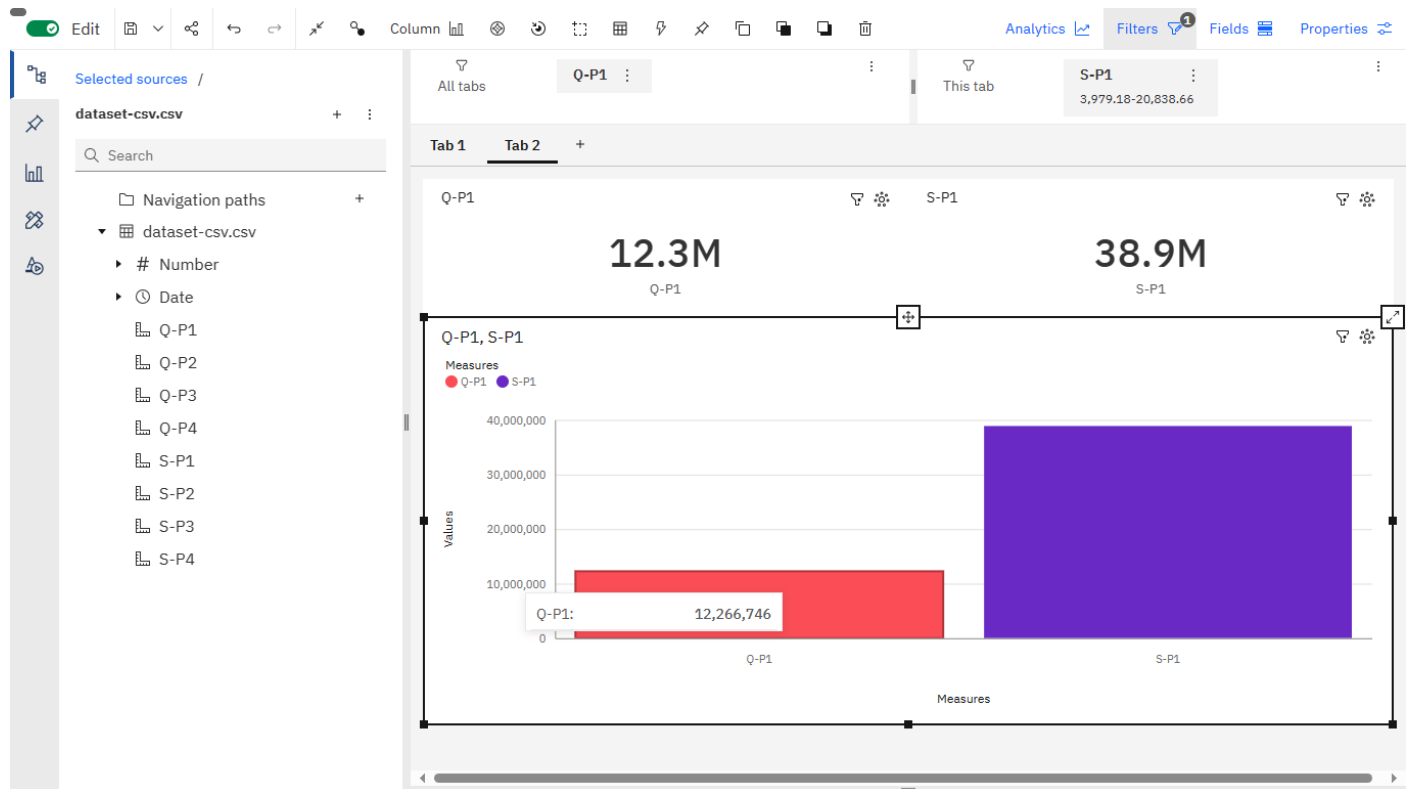
Total Unit sold in all the years:



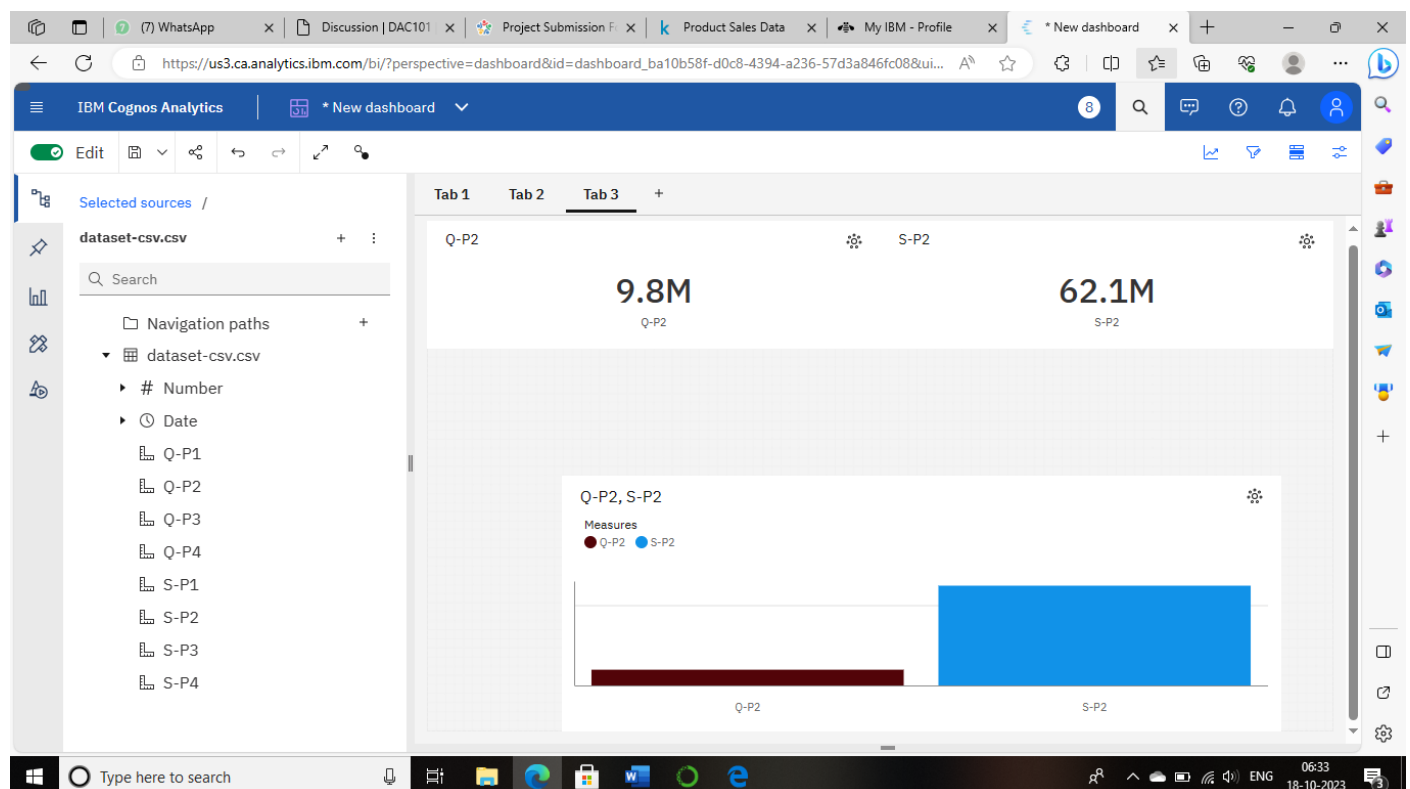
Total Revenue generated by Each Unit:



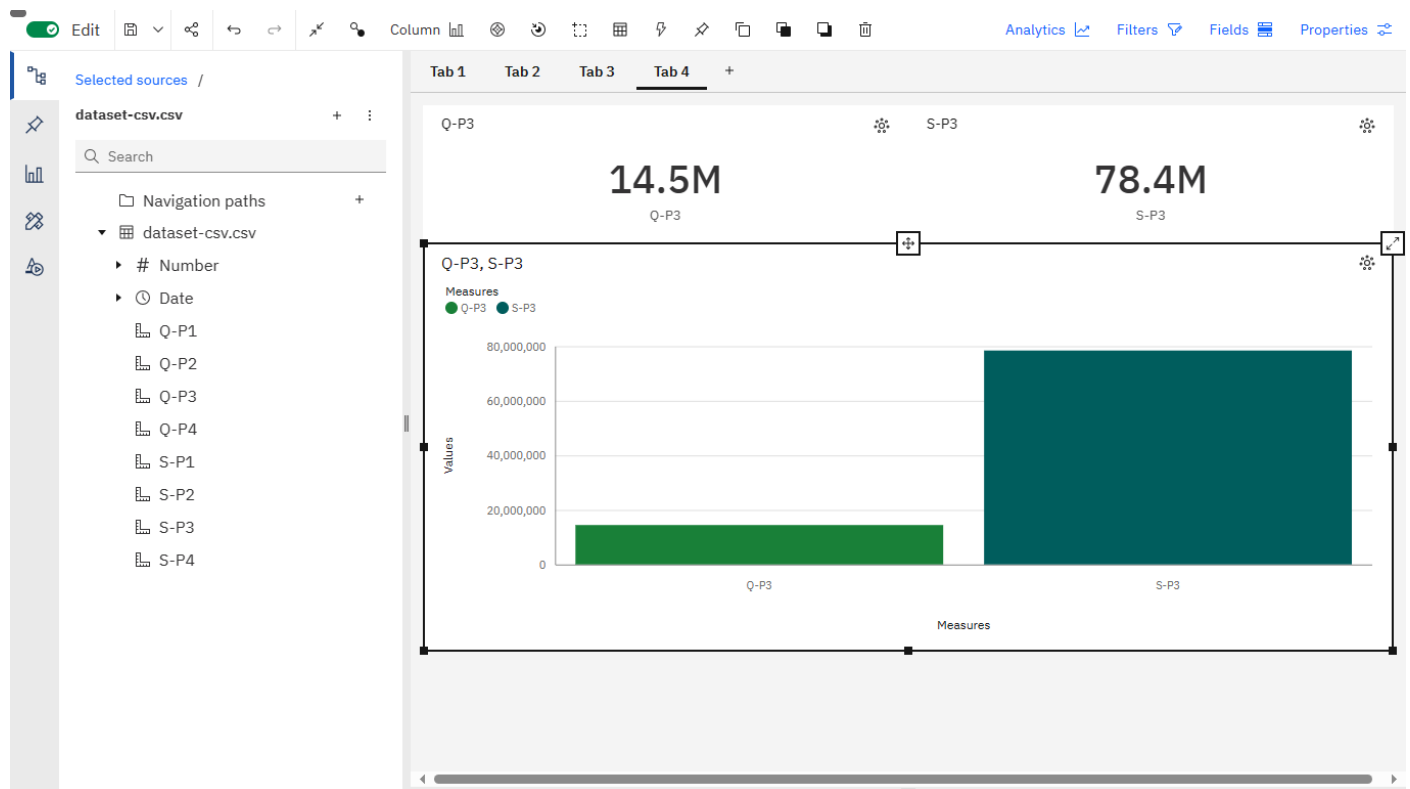
Total Unit Sold and Total Revenue Generated by Product1



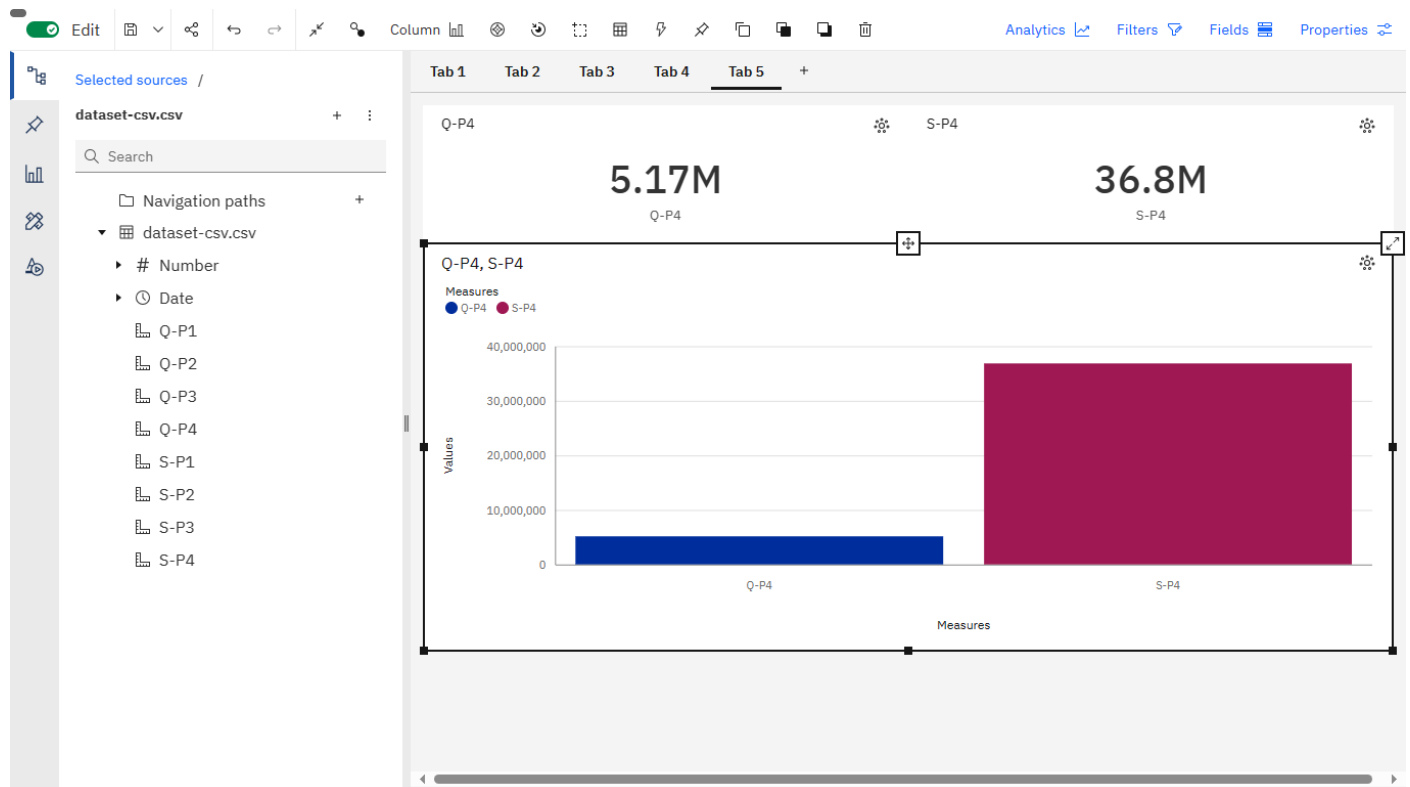
Total Unit Sold and Total Revenue Generated by Product 2



Total Unit Sold and Total Revenue Generated by Product 3



Total Unit Sold and Total Revenue Generated by Product 4



By these visualisations 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 analyzing 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.