DEVELOPMENT PHASE PART 2

PRODUCT SALES ANALYSIS

Date	24-10-2023
Team ID	1289
Project Name	Product Sales Analysis

Table Of Content:

1. Introduction:

In the development part 2 the project is Continued building the analysis by creating visualizations and building a predictive model. Using visualization libraries like Matplotlib, Seaborn the histograms were created. A predictive model is built to determine the top sales product and peak seal product and customer preference.

2. Data Preprocessing:

In the previous phase the Data processing which is essential in data analysis to increase data quality. Data processing is described as “the collection and manipulation of data components to produce meaningful information.” Through meticulous handling of missing values, dynamic feature scaling, and real-time outlier detection, the dataset attained a level of precision essential for accurate predictions. The data preprossing is done by using Jupyter notebook .

3. Data visualization:

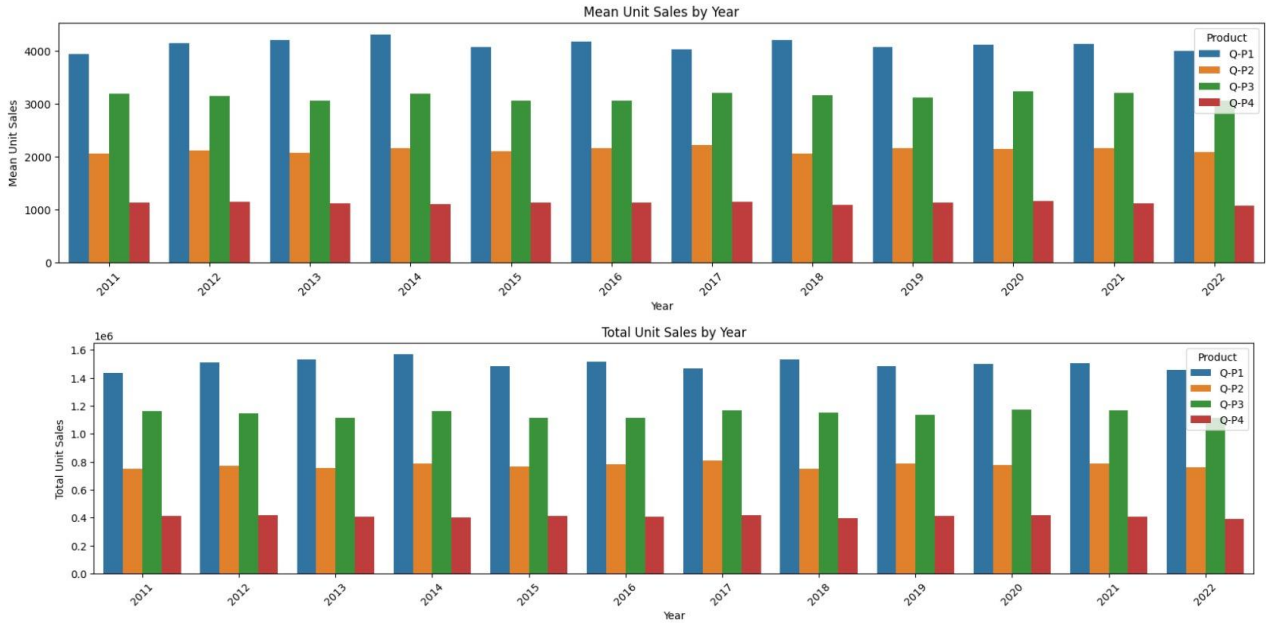
3.1 HISTOGRAM AND DISTRIBUTION.

Graph our TOTAL & MEAN unit sold for each product using a histogram.

```
    def plot_bar_chart(df, columns, stri, str1, val):
        if val == 'sum':
        sales_by_year = df.groupby('Year')[columns].sum().reset_index()
    elif val == 'mean':
        sales_by_year = df.groupby('Year')[columns].mean().reset_index()
    sales_by_year_melted = pd.melt(sales_by_year, id_vars='Year', value_vars=col
umns, 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(stri)
    plt.title(f'{stri} by {str1}')
    plt.xticks(rotation=45)
    plt.show()
```

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

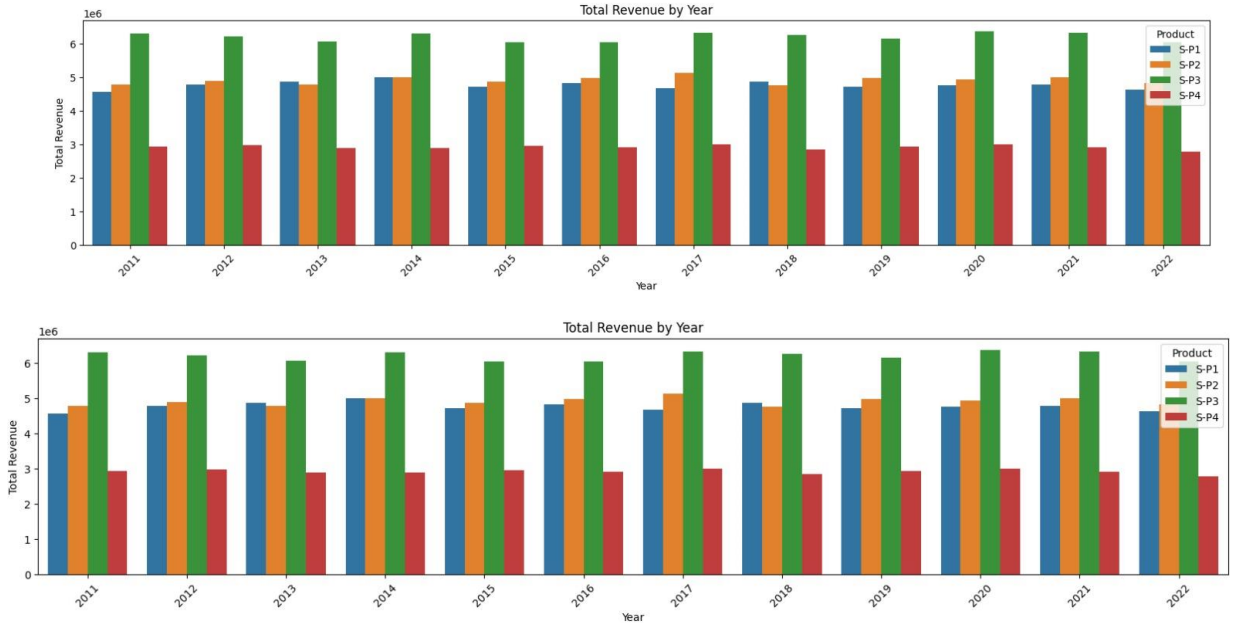
```
plot_bar_chart(data_reduced, ['Q-P1', 'Q-P2', 'Q-P3', 'Q-P4'],'M
ean Unit Sales', 'Year', 'mean')
```



Observation

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

Graph our TOTAL & MEAN revenue of sales for each product using historgram.



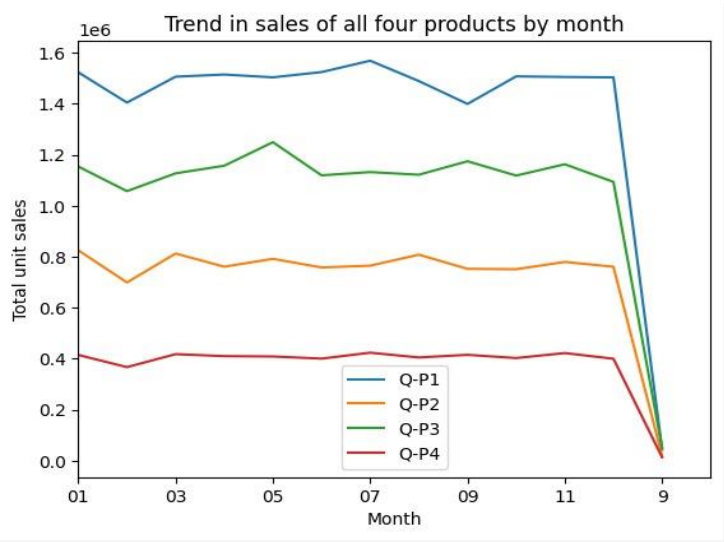
Observation

- We have 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 have 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.

3.1. LINE CHART.

Trend in sales of all four products during certain months

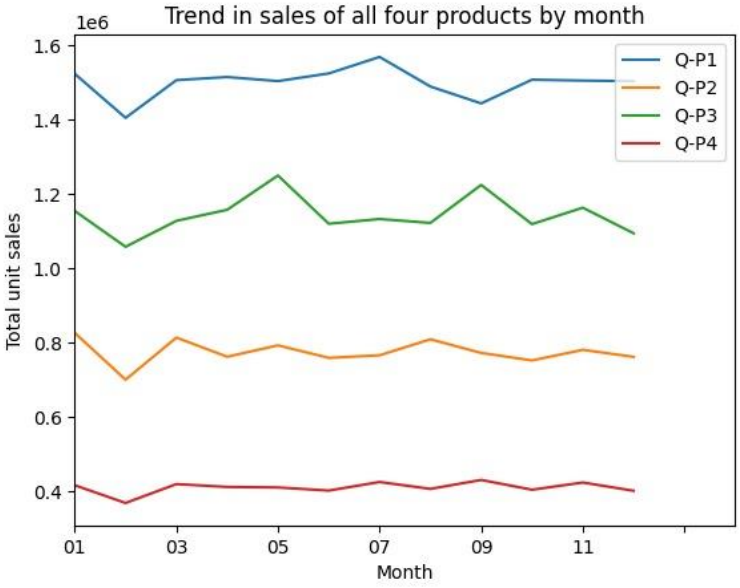
```
    def month_plot():
    fig, ax = plt.subplots()
    data_reduced.groupby('Month')[['Q-P1', 'Q-P2', 'Q-P3', 'Q-P4']].sum().plot(a
x=ax)

    ax.set_xlim(left=0, right=13)
    ax.set_xlabel('Month')
    ax.set_ylabel('Total unit sales')
    ax.set_title('Trend in sales of all four products by month')
    plt.show()
    month_plot()
```



Observation

- We have 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.

```
data_reduced['Month'] = data['Month'].replace('9', '09')
In [15]:
linkcode
month_plot()
```



Observation

- We have merged the sales for months 9 and 09.
- We have observe that Feb and Dec have the lowest sales for each product
- For P1 We have observe Mar - Jul having the highest unit sales
- For P2 We have observe Jan, Mar - Aug having the highest unit sales
- For P3 We have observe May & Sep having the highest unit sales
- For P4 We have observe uniform sales from Jan - Dec

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

Question:

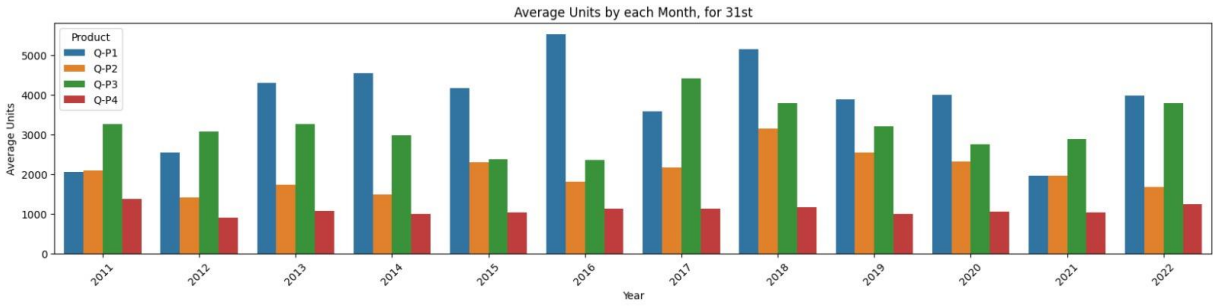
The company has all it's retail centers closed on the 31st of December every year. Mr: Hariharan , the CEO , would love to get an estimate on no: of units of each product that could be sold on 31st of Dec , every year , if all their retail centers were kept open.

```
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', '0
7', '08', '09', '10', '11', '12'])
_31_months
```

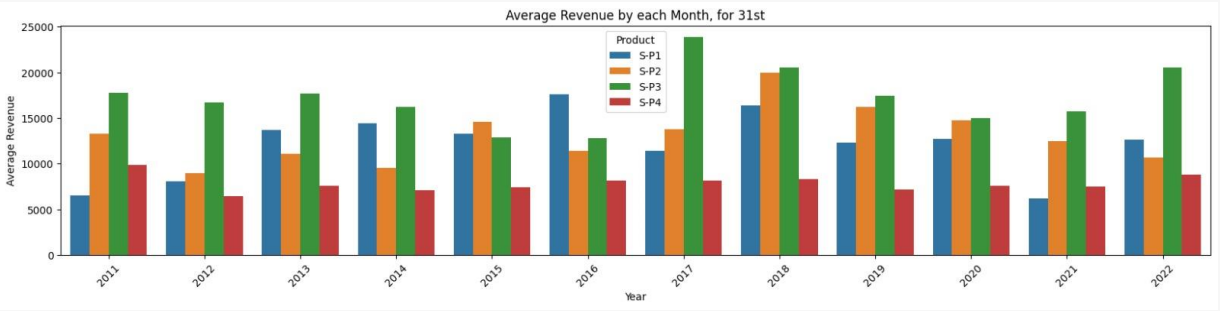
	Date	Q-P1	Q-P2	Q-P3	Q-P4	S-P1	S-P2	S-P3	S-P4	Day	Month	Year
231	31-01-2011	939	3325	1863	1612	2976.63	21080.50	10097.46	11493.56	31	01	2011
290	31-03-2011	464	2220	421	1663	1470.88	14074.80	2281.82	11857.19	31	03	2011
351	31-05-2011	1507	2980	3816	1202	4777.19	18893.20	20682.72	8570.26	31	05	2011
412	31-07-2011	4336	744	4717	667	13745.12	4716.96	25566.14	4755.71	31	07	2011

	Date	Q-P1	Q-P2	Q-P3	Q-P4	S-P1	S-P2	S-P3	S-P4	Day	Month	Year
442	31-08-2011	4548	1484	1596	1974	14417.16	9408.56	8650.32	14074.62	31	08	2011
...	...	...	...	...	...	...	...	...	...	...	...	...
4352	31-05-2022	3669	2710	3067	1593	11630.73	17181.40	16623.14	11358.09	31	05	2022
4413	31-07-2022	1437	833	1867	1270	4555.29	5281.22	10119.14	9055.10	31	07	2022
4443	31-08-2022	1035	1639	3658	841	3280.95	10391.26	19826.36	5996.33	31	08	2022
4474	31-9-2022	6964	1873	5481	1336	22075.88	11874.82	29707.02	9525.68	31	09	2022
4535	31-11-2022	4600	2006	3796	1426	14582.00	12718.04	20574.32	10167.38	31	11	2022

```
plot_bar_chart(_31_months, ['Q-P1', 'Q-P2', 'Q-P3', 'Q-P4'], 'Average Units', 'each Month, for 31st', 'mean')
```




```
plot_bar_chart(_31_months, ['S-P1', 'S-P2', 'S-P3', 'S-P4'], 'Average Revenue', 'each Month, for 31st', 'mean')
```

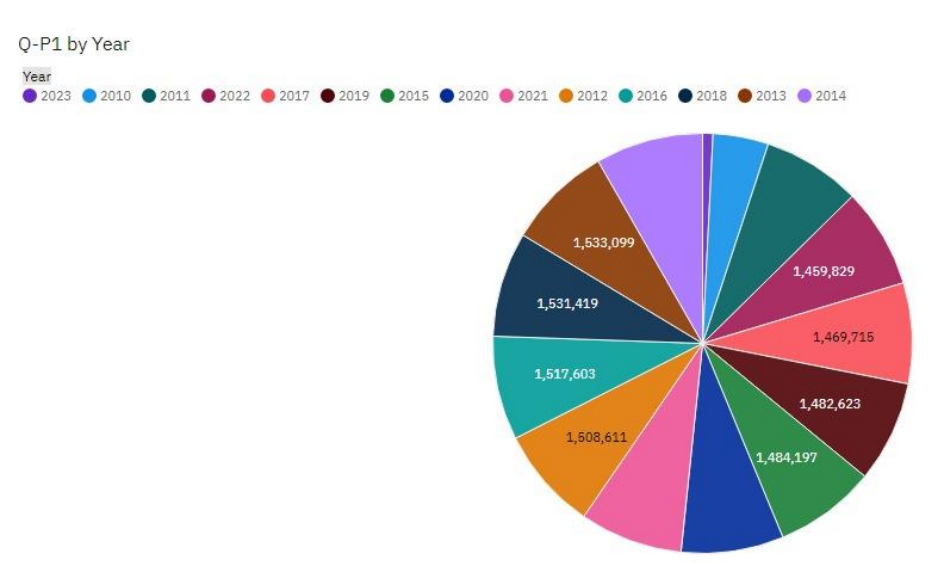


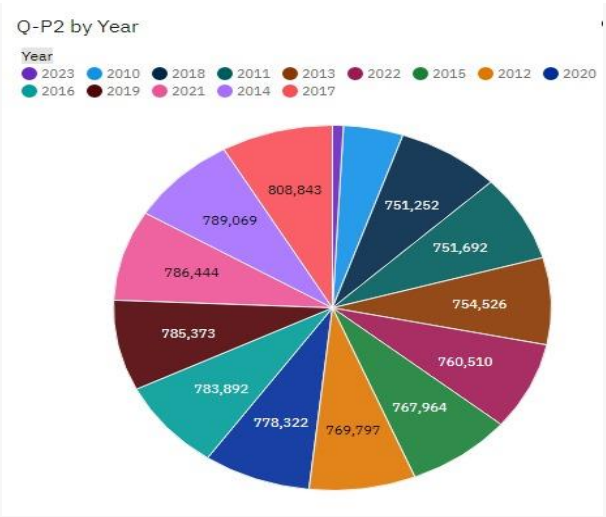
Observation

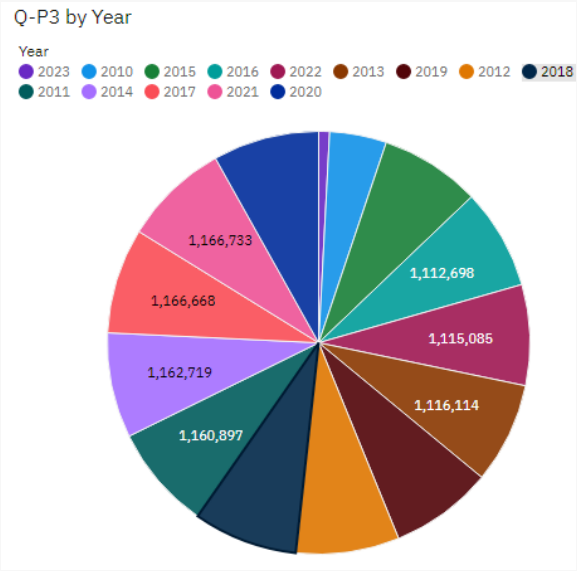
- Overall we can see that P1 has the highest unit sales on the 31st for each year, except for 2021 and 2022. (These could be as a result to Covid and other economy issues.)
- P3 has the second highest unit sales for all the 31st in each year.

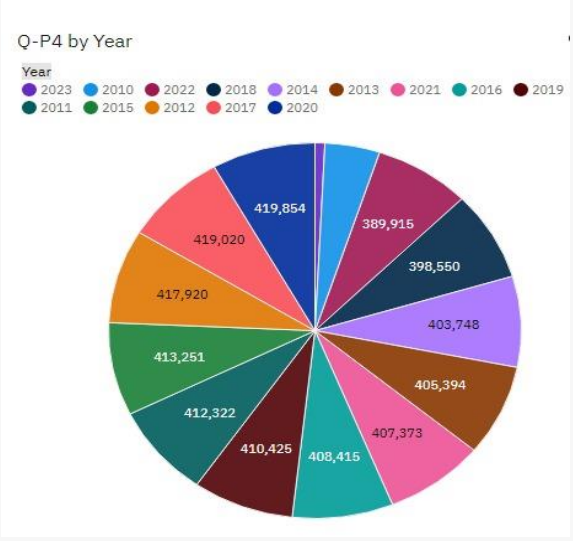
COGNOS VISUALIZATION:

Visualization between products and year:

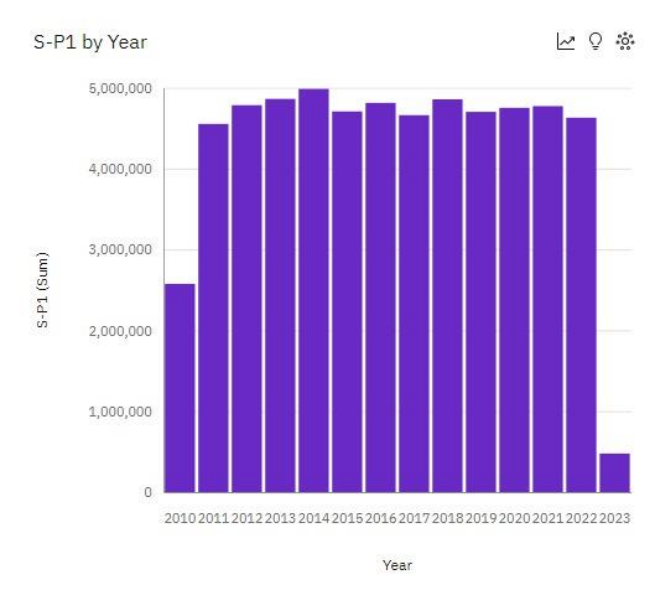


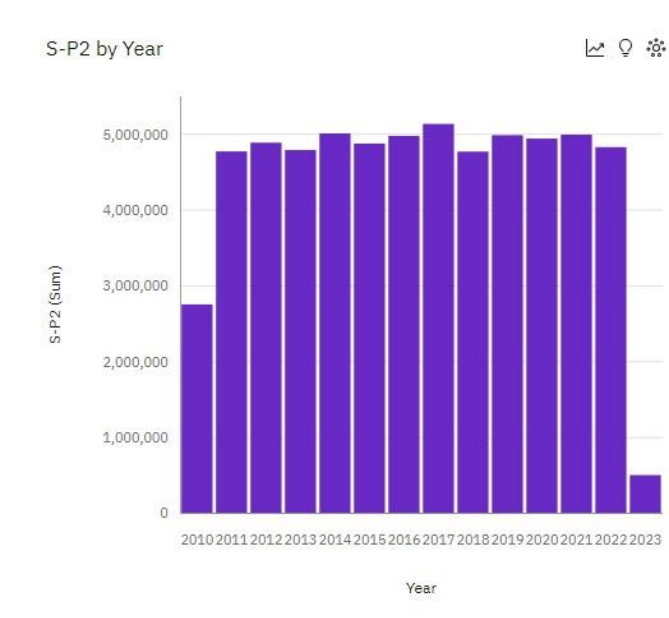

Q-P2 by Year
Year
2023 2010 2018 2011 2013 2022 2015 2012 2020
2016 2019 2021 2014 2017
808,843
751,252
751,692
754,526
760,510
767,964
769,797
778,322
783,892
785,373
786,444
789,069

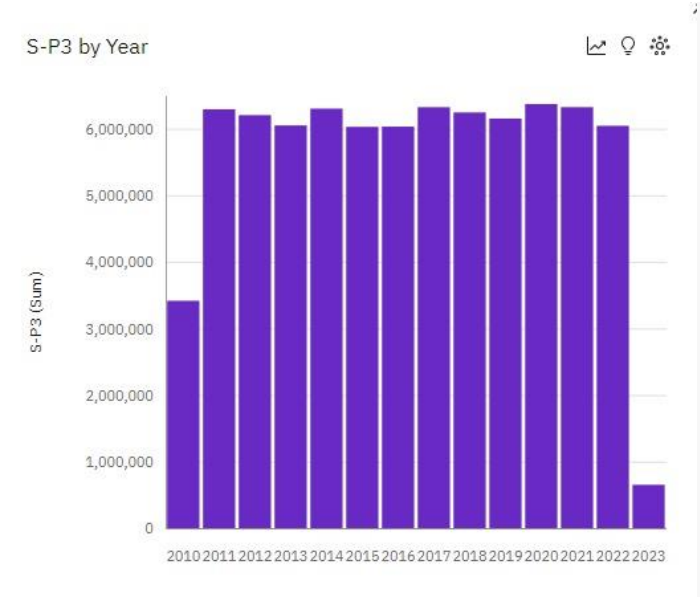

Q-P3 by Year
Year
2023 2010 2015 2016 2022 2013 2019 2012 2018
2011 2014 2017 2021 2020
1,166,733
1,112,698
1,166,668
1,115,085
1,162,719
1,116,114
1,160,897

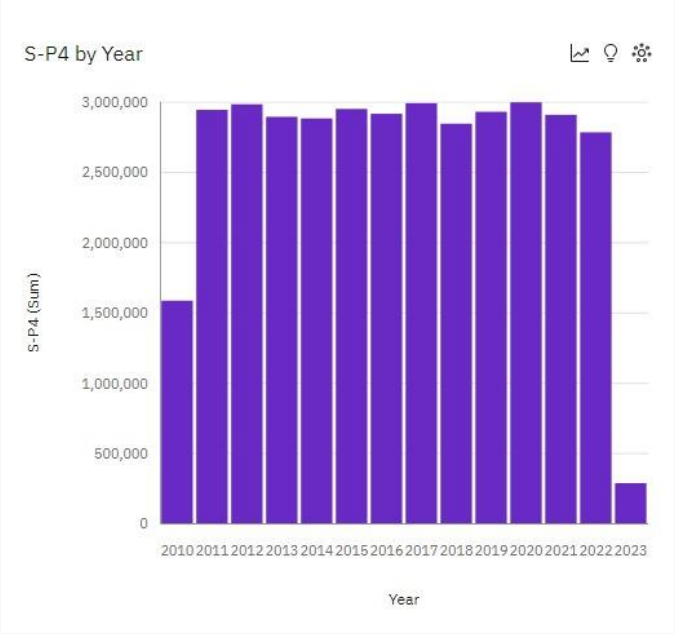

Q-P4 by Year
Year
2023 2010 2022 2018 2014 2013 2021 2016 2019
2011 2015 2012 2017 2020
419,854
389,915
419,020
398,550
417,920
403,748
413,251
405,394
412,322
407,373
410,425
408,415

Visualization between sales and year:









PREDICTION:

```
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

```
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

- We have see that our previous observation correlate as Q-P1 has the higest estimate, follwed by Q-P3
- We have approxiamte that the company will make:
  - Q-P1: 3813.74
  - Q-P2: 2058.80
  - Q-P3: 3183.88
  - Q-P4: 1098.61

Conclusion

Unit Sales 2011 - 2022

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

Revenues 2011 - 2022

- We have 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 have 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 have observe that all Products unit sales drop in Feb.
- We have 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 can not 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 have 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.
- We have see that our previous observation correlate as Q-P1 has the higest estimate, followed by Q-P3
- We have approxiamte that the company will make:
    - Q-P1: 3813.74
    - Q-P2: 2058.80
    - Q-P3: 3183.88
    - Q-P4: 1098.61