DEVELOPMENT PHASE

PRODUCT SALES ANALYSIS PROJECT

Date	**17-10-2023**
Team ID	**1289**
Project Name	**Product Sales Analysis**

Table of Contents

1. Project Introduction :

The aim of this project to analyze sales data and extract insights about top selling products, peak sales periods, and customer preferences by using IBM Cogons. The objective is to help businesses improve inventory management and marketing strategies by understanding sales trends and customer behaviour.

2. Problem Statement :

The Problem Statement od Product Sales Analysis that involves using IBM Cognos to analyze sales data and extract insights about top selling products, peak sales periods, and customer preferences. The objective is to help businesses improve inventory management and marketing strategies by understanding sales trends and customer behaviour. This project includes defining analysis objectives, collecting sales data, designing relevant visualizations in IBM Cognos, and deriving actionable insights.

3. Data Pre-Processing :

REC corp LTD. is small-scaled business venture established in India. They have been selling FOUR PRODUCTS for OVER TEN YEARS .

The products are:

- P1
- P2
- P3

- P4

They have collected data from their retail centers and organized it into a small csv file , which has been given to you.

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

Step 1: Import libraries

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

%matplotlib inline


import warnings
warnings.filterwarnings("ignore")
```

Step 2: Loading the datasets

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

Unnamed 0	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	579.82	94.78

[4599 rows x 10 columns]

```
data.head(-1)
```

Unnamed 0	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	579.82	94.78

```
[4599 rows x 10 columns]
```

Observations:

- There is a column called 'Unnamed: 0' which we can drop as it is a repeat of our ID.
- The data contains date.
    - And for each date the total unit of sales for P1, P2, P3 & P4.
    - Also the total revenue from sales for P1, P2, P3 & P4.
- We can observe the first entry in the data, starts at 13-06-2010. This means the data for year 2010 is not complete.
- We can observe the last entry in the data, ends at 02-02-2023. This means the data for year 2023 is also not complete.
    - it will be best to drop year 2010 and year 2023.

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

Step 3: Checking the info of the training data

```
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
```

Observations:

- The train dataset has 4600 entries(rows) and 9 columns. (we dropped one column)
- Date is an object data type. the rest of numerical in nature.

Step 4: Check for missing values

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

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

	Date	Q-P1	Q-P2	Q-P3	Q-P4	S-P1	S-P2	S-P3	S-P4	Date	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

```
dtype: int64
```

Step 5: EDA

```
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
```

```
[4599 rows x 13 columns]
```

4. Data Visualization:

CHART-1

Chart Insights were not computed because this visualization is based on clipped data. Consider applying a filter to reduce the number of records, and to prevent the data from being clipped, before creating the visualization.

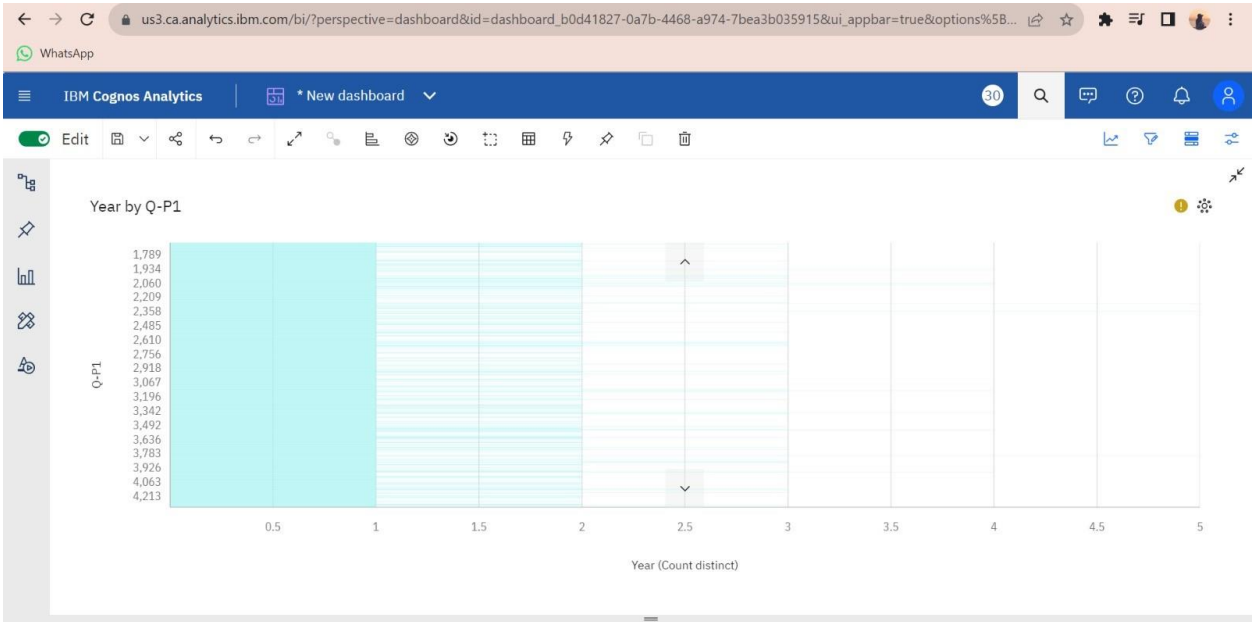


CHART-2

Chart Insights were not computed because this visualization is based on clipped data. Consider applying a filter to reduce the number of records, and to prevent the data from being clipped, before creating the visualization.

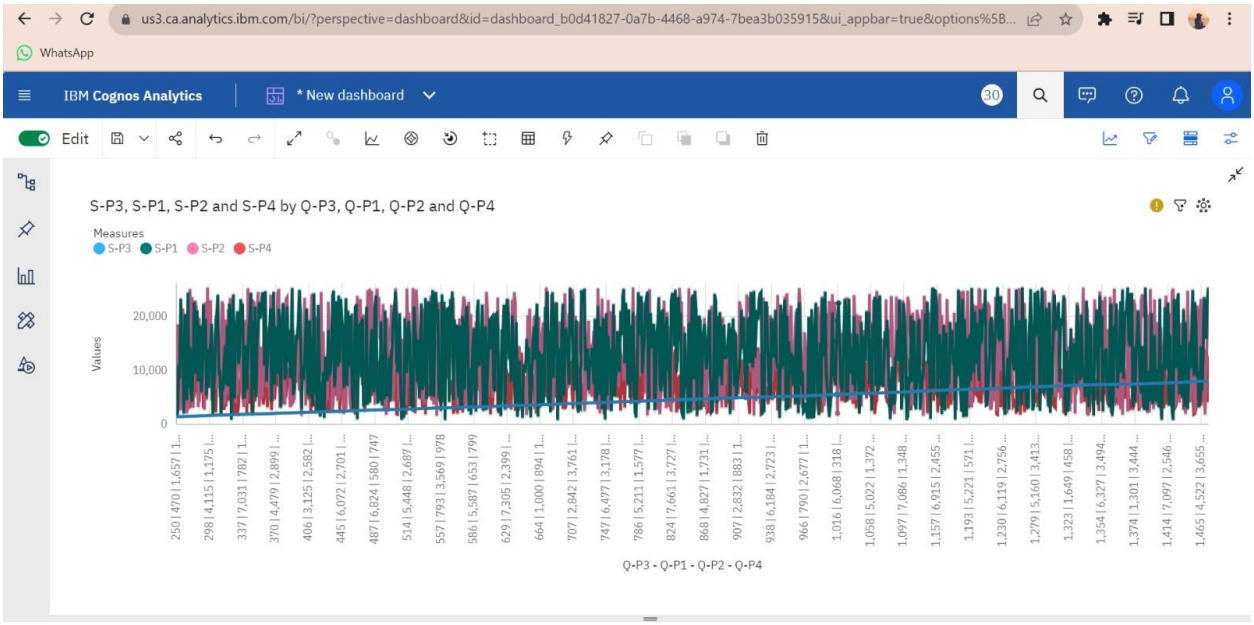
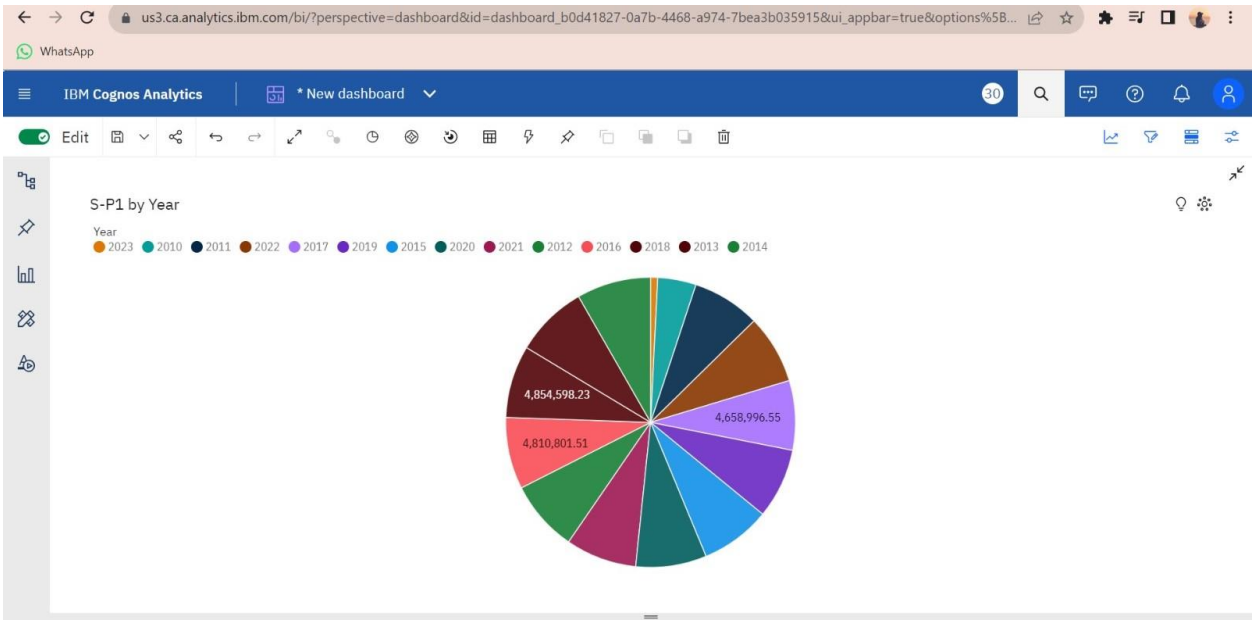

us3.ca.analytics.ibm.com/bi/?perspective=dashboard&id=dashboard_b0d41827-0a7b-4468-a974-7bea3b035915&ui_appbar=true&options%5B...
WhatsApp
IBM Cognos Analytics
* New dashboard
Edit
S-P3, S-P1, S-P2 and S-P4 by Q-P3, Q-P1, Q-P2 and Q-P4
Measures
S-P3
S-P1
S-P2
S-P4
Values
20,000
10,000
0
Q-P3 - Q-P1 - Q-P2 - Q-P4

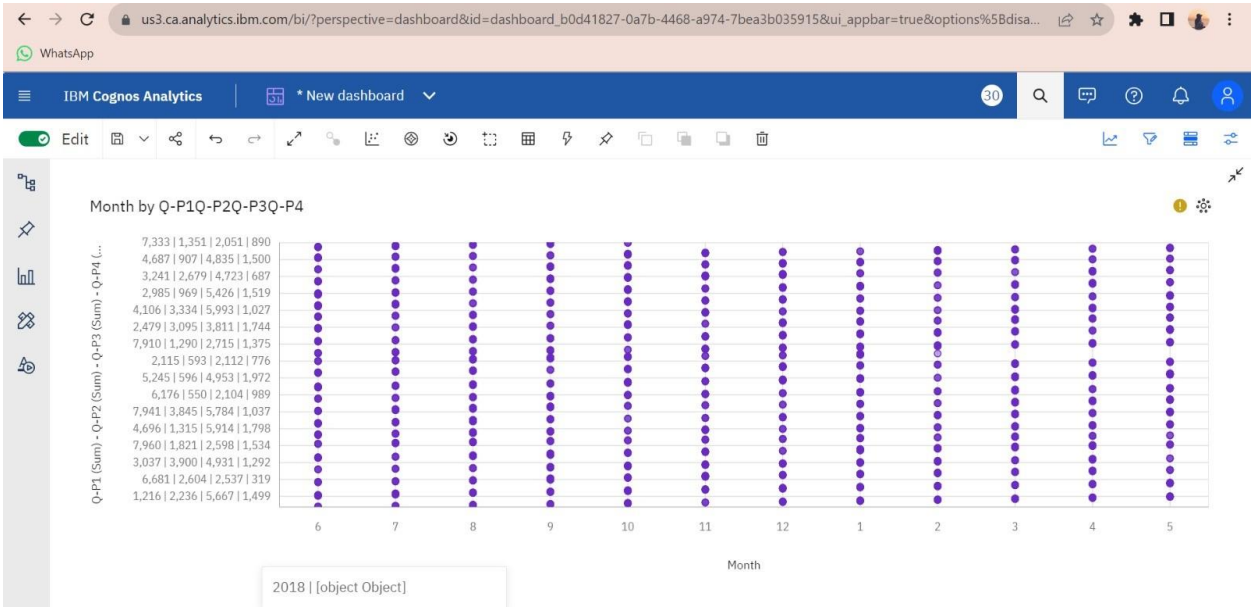
CHART-3


us3.ca.analytics.ibm.com/bi/?perspective=dashboard&id=dashboard_b0d41827-0a7b-4468-a974-7bea3b035915&ui_appbar=true&options%5B...
WhatsApp
IBM Cognos Analytics
* New dashboard
Edit
S-P1 by Year
Year
2023
2010
2011
2022
2017
2019
2015
2020
2021
2012
2016
2018
2013
2014
4,854,598.23
4,658,996.55
4,810,801.51

    - S-P1 has a moderate upward trend.
    - S-P1 is unusually low in 2023 and 2010.
    - Based on the current forecasting, S-P1 may reach over 6.2 million by Year 2026.
    - 2011 (7.9 %), 2022 (7.9 %), 2021 (7.9 %), 2019 (7.9 %), and 2018 (7.9 %) are the most frequently occurring categories of Year with a combined count of 1820 items with S-P1 values (39.6 % of the total) .
- 2011 (7.9 %), 2022 (7.9 %), 2021 (7.9 %), 2019 (7.9 %), and 2018 (7.9 %) are the most frequently occurring categories of Year with a combined count of 1820 items with S-P2 values (39.6 % of the total) .
- Across all years, the average of S-P1 is over thirteen thousand.
- Over all years, the average of S-P2 is nearly fourteen thousand.
- The total number of results for S-P1, across all years, is over 4500.
- The total number of results for S-P2, across all years, is over 4500.
- S-P1 ranges from over 476 thousand, in 2023, to almost 5.0 million, in 2014.
- S-P2 ranges from over 496 thousand, in 2023, to over 5.1 million, in 2017.

CHART-4

Chart Insights were not computed because this visualization is based on clipped data. Consider applying a filter to reduce the number of records, and to prevent the data from being clipped, before creating the visualization.



4. Overall Observation:

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