INTRODUCTION

Product sales analysis is an indispensable tool for modern businesses striving for success in an increasingly competitive market. By delving insights into customer preferences, market trends, and the performance of their products. This project explores the significance of product sales analysis, its key components, and its vital role in shaping business strategies.

OBJECTIVE

The objective of product sales analytics is to analyse data and gain insights into product performance, identifying top-selling products, peak sales periods, and customer preferences. This aids in effective inventory management, marketing strategies, and decision-making for optimizing pricing and market segmentation. Businesses aim to enhance efficiency, customer satisfaction, and overall revenue growth through informed analytics-driven strategies.

DEVELOPMENT PHASE

Phase1: In phase1, the basic idea was understood and we obtained clear vision in our objectives.

Phase2: Our understanding was presented as flowchart to enhance our understanding and cleared our working path.

Phase3: In this phase, our major topics were covered analysing, out of 5,3 were completed and we visualized the 3 major topics using IBM cognos. Also, we explained about the data.

Phase4: The remaining 2 topics were completed and visualized using IBM cognos and the python program for modifying and cleaning the dataset was provided. And the main objective of the Project was accomplished.

VISUALIZATION ON IBM COGNOS TOOL

The IBM Cognos tool is used for analysing the files such as csv files and other files to visualize data from them.

TOP SELLING PRODUCTS

1. Average of product P1 is nearly 4000
2. Average of product P2 is nearly 2129
3. Average of product P3 is nearly 3143
4. Average of product P4 is over 1123
5. Comparing the 4 products, product P1 has sold more than other products.
6. So, product P1 can be considered as Top selling product.
7. The below graph represents the total sales units for 4 products on every year(2011-2023) .
8. P1>P3>P2>P4 is the order of top selling products.

PEAK SALES PERIOD

1. From the Below graph we can assume that the period 2020-2022 has the highest peak sales period for the four products .
2. Next to that 2011-2013 has the highest sales period.
3. Year 2022 has the highest sales year and 2011 is more or less has similar sales as year 2022.
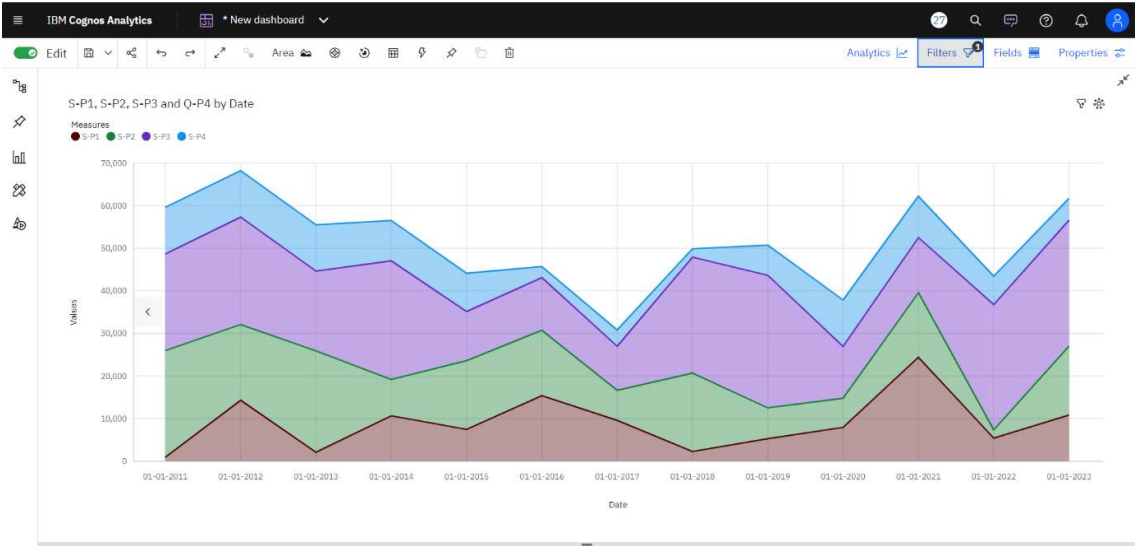
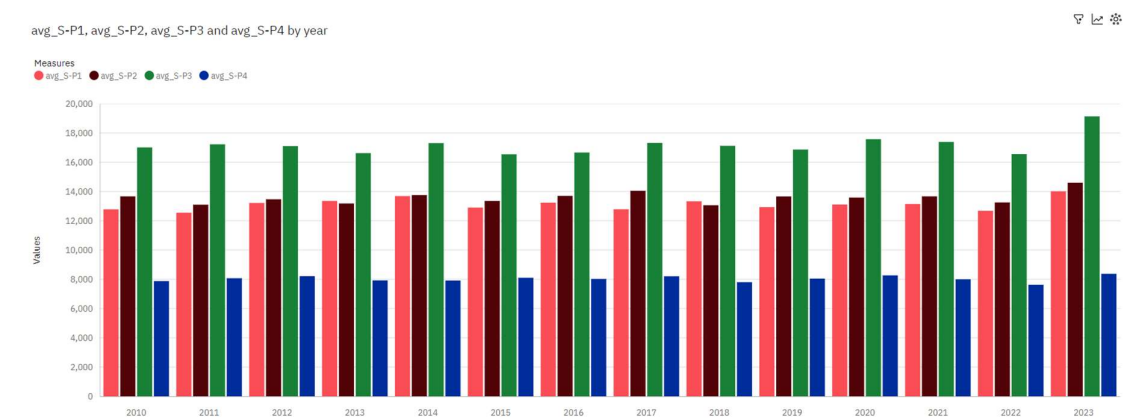## CUSTOMER PREFENCES

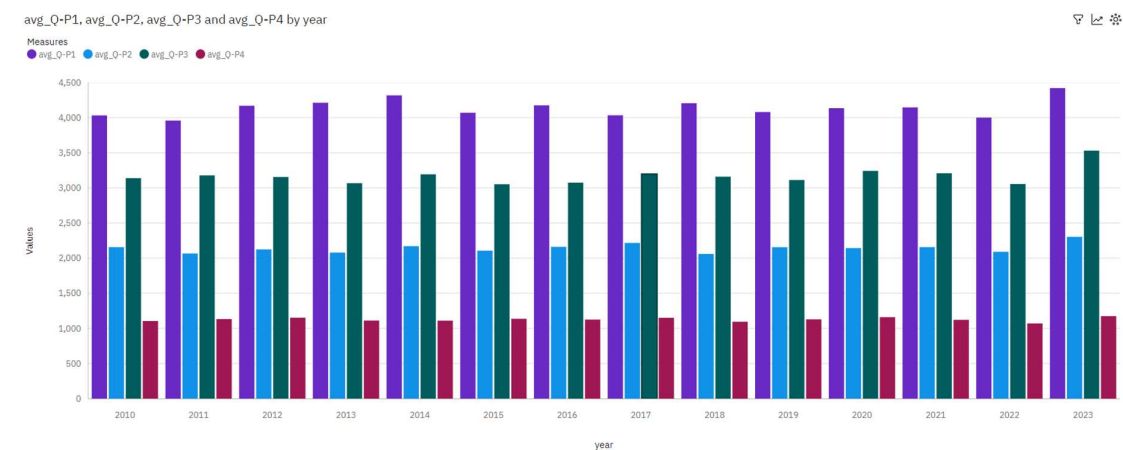Product P1 has more customer preferences.

TOTAL SELLING UNIT

TOTAL REVENUE

REVENUE COLUMN BAR GRAPH

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

UNIT SOLD COLUMN BAR GRAPH

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

AIDING INVENTORY MANAGEMENT

Inventory is one of the most crucial aspects of any business model. A close tab on the movement of inventory can make or break your business and that's why entrepreneurs always emphasise on effective inventory management. While a few business owners do understand the significance and cruciality of tracking inventory on a regular basis, some fail to realise its importance making their business fall through the unseen cracks.

From the graph we can come into conclusion that the product P1 has more sold units and it has to manufacture more units to sell in the future, because as every year the demand and unit produce is increasing.

year, avg_Q-P1, avg_Q-P2, avg_Q-P3 and avg_Q-P4

year	avg_Q-P1	avg_Q-P2	avg_Q-P3	avg_Q-P4
2010	4,031.92	2,155.97	3,138.81	1,104.77
2011	3,958.75	2,065.79	3,177.85	1,131.94
2012	4,169.25	2,124.31	3,155.4	1,152.02
2013	4,212.21	2,078.96	3,067.12	1,111.36
2014	4,317.52	2,169.77	3,193.31	1,110.06
2015	4,069.33	2,106.16	3,051.92	1,136.44
2016	4,175.22	2,160.68	3,074.47	1,125.11
2017	4,033.45	2,216.01	3,195.43	1,150.67
2018	4,205.34	2,059.76	3,159.22	1,094.28
2019	4,079.47	2,155.18	3,112.19	1,128.42
2020	4,135.11	2,143.02	3,242.14	1,159.56
2021	4,146.35	2,156.08	3,207.99	1,121.4
2022	4,000.73	2,090.14	3,054.72	1,069.48
2023	4,420.88	2,302.97	3,530.29	1,174.26
Summary	4,123.34	2,129.71	3,143.77	1,123.74

MARKETING STRATEGIES

1. As comparing the sales unit column graph and revenue column graph many things seems odd.
2. Because P1 sold more than P3 in every year, but P3 has more revenue than P1.
3. Other products like P2,P4 has similar hike in the units sold and revenue graph.
4. P1 has to change it's Sales strategy, Since it has more popularity among customers compared to other products.
5. P3 has to increase its customers, since P1 has more customers than P3.

DATASET ANALYSIS

The collection of data is done from the Products P1,P2,P3,P4.And, the dataset provided the number of units sold every day from the year 2010 till 2023.Also, they provided the revenue collected from each product from every same day.

Using the python program the dataset is cleaned and modified for visualization.

The visualization is done on IBM cognos tool, which was easy to use and we obtained expected outputs and any type visualization types can utilized from IBM cognos tool.

DATA CLEANING AND PROCESSING

The data cleaning and data processing process done in python using the below code.

Import necessary libraries

```python
import pandas as pd
from sklearn.preprocessing import StandardScaler
from sklearn.impute import SimpleImputer
```

Load the dataset

```python
df = pd.read_csv('statsfinal.csv')
```

Step 1: Handle Missing Data ,Identify and handle missing values

```python
df.dropna(inplace=True)
```

Drop rows with missing values

#Alternatively, use imputation for missing values

```python
imputer = SimpleImputer(strategy='mean')
df['column_with_missing_values'] =
imputer.fit_transform(df[['column_with_missing_values']])
```

Step 2: Remove Duplicate Records

```python
df.drop_duplicates(inplace=True)
```

Step 3: Standardize Data

Standardize numerical data using StandardScaler

```python
scaler = StandardScaler()
df[['numeric_column1', 'numeric_column2']] =
scaler.fit_transform(df[['numeric_column1', 'numeric_column2']])
```

Step 4: Encoding Categorical Variables

Convert categorical variables to numerical using one-hot encoding

```
Sdf = pd.get_dummies(df, columns=['categorical_column'])
```

Previously we cleaned and processed the data's in dataset.Now we're modifying the dataset by adding averages of every products sales unit and revenue.

CODE

```python
import pandas as pd
```

Load your dataset into a Pandas DataFrame

Assuming you have a CSV file with 'Date' and 'value' columns

```python
df = pd.read_csv('statsfinalsample.csv')
```

Assuming 'Date' is in a datetime format, if not, you may need to convert it

```python
df['Date'] = pd.to_datetime(df['Date'], format='%d-%m-%Y', errors='coerce')
```

Extract the year from the 'Date' column

```python
df['year'] = df['Date'].dt.year
```

Specify the columns for which you want to calculate averages

```python
columns_to_average = ['Q-P1', 'Q-P2', 'Q-P3', 'Q-P4']
```

Calculate average for each specified column

```python
for column in columns_to_average:
    avg_column_name = f'avg_{column}'
    df[avg_column_name] =
df.groupby('year')[column].transform('mean')
```

Drop the 'year' column if you don't want it in the final output

```python
df = df.drop('year', axis=1)
```

Save the modified DataFrame back to the CSV file

```
df.to_csv("D:\CSE\PROGRAM\PYTHON\statsfinalsample.csv",
index=False)
print(df.head(5))
```