

DATA ANALYTICS WITH COGNOS

PRODUCT SALES ANALYSIS

PHASE:3

INTRODUCTION:-

The project involves analyzing Product sales data using IBM Cognos. The objective is to compare and contrast the mean values and standard deviations of sales and associated profits per day and by a company. This project encompasses defining analysis objectives, collecting Product sales data, designing relevant visualizations in IBM Cognos, and deriving insights from the data.

DATASET:-

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

DATA PREPROCESSING AND DATA COLLECTION:-

Data preprocessing is a crucial step in any data analysis project. Here are the general steps you might follow for the Product sales analysis:

1. Load the Data:

First, you need to load the data into a suitable data structure. If you're using Python, pandas Data Frame is a good option for this.

CODE:

```
import pandas as pd  
data=pd.read_csv('productsales_data.csv')
```

2. Inspect the Data:

Take a look at the data to understand its structure and contents. Check the number of rows and columns, column names, data types, etc.

CODE:

```
data.info()
```

3. Handle Missing Values:

Check for missing or null values in the data. You might fill them with appropriate values (like mean or median), or drop the rows or columns containing them, based on the situation.

CODE:

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

4. MANIPULATION OF DATA:

Data manipulation refers to the process of adjusting data to make it organised and easier to read. Data manipulation language, or DML, is a programming language that adjusts data by inserting, deleting and modifying data in a database such as to cleanse or map the data.

CODING:-

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
data=pd.read_csv("/kaggle/input/product-sales/statsfinal.csv")
data.describe()
data.info()
x=data.drop("S-P4",axis=1)
y=data['S-P4']
```

OUTPUT:-

[4]:

```
data.describe()
data.info()
x=data.drop("S-P4",axis=1)
y=data['S-P4']
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4600 entries, 0 to 4599
Data columns (total 10 columns):
#   Column      Non-Null Count  Dtype
---  -
0   Unnamed: 0   4600 non-null   int64
1   Date         4600 non-null   object
2   Q-P1         4600 non-null   int64
3   Q-P2         4600 non-null   int64
4   Q-P3         4600 non-null   int64
5   Q-P4         4600 non-null   int64
6   S-P1         4600 non-null   float64
7   S-P2         4600 non-null   float64
8   S-P3         4600 non-null   float64
9   S-P4         4600 non-null   float64
dtypes: float64(4), int64(5), object(1)
memory usage: 359.5+ KB
```

CODING:-

```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.5,random_state=0)
x_train
x_train.shape
x_test.shape
y_train.shape
y_test.shape
y_test
x_test
x
display(data.drop_duplicates())
plt.plot(data.Q-P1,data.S-P1)
```

```
plt.xlabel('data.Q-P1')
plt.ylabel('data.S-P1')
plt.title('Product sales graph')
plt.show()
```

OUTPUT:-

```
display(data.drop_duplicates())

plt.plot(data.Q-P1,data.S-P1)
plt.xlabel('data.Q-P1')
plt.ylabel('data.S-P1')
plt.title('Product sales graph')
plt.show()
```

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

CONCLUSION:-

Product Sales Analysis insights aid decision-makers in understanding current scenarios, predicting future trends, and making informed choices. These insights guide healthcare professionals in allocating resources and implementing them.