

DHAANISH AHMED COLLEGE OF ENGINEERING

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Domain Name: Data Analytics with Cognos

Project Title: product sales data analysis

phase 3: Development Part 1

Name of the Student : Imran Sheriff

machine learning to predict future sales trends or customer behaviors, you can follow these steps:

1. Title and Introduction:

- Title: "Machine Learning for Predicting Sales Trends and Customer Behaviors"

- Introduction: Provide an overview of the document's purpose, which is to explore how machine learning can be leveraged to address the challenge of predicting sales trends and understanding customer behaviors.

2. Problem Statement:

- Define the problem: Explain the challenge of predicting future sales trends and understanding customer behaviors in the context of your business or scenario.

3. Data Preparation and Feature Engineering:

- Describe the data used for machine learning, including its source, structure, and relevant features.

- Explain any data preprocessing steps, such as handling missing values, data scaling, and encoding categorical variables.

- Discuss feature engineering, where you can create new features from the existing data that might be relevant for predicting sales trends or customer behaviors.

4. Machine Learning Model Selection:

- Explain the machine learning algorithms or models chosen for this task. It could include regression models for sales prediction or clustering/classification models for customer behavior analysis.

- Mention why these models were selected and their suitability for the problem.

5. Training and Evaluation:

- Describe how the dataset is split into training and testing sets.

- Explain the evaluation metrics you plan to use to assess the performance of the machine learning models. For instance, mean squared error (MSE) for regression models or accuracy for classification models.

6. Model Training and Results:

- Present the results of model training, including model performance metrics. Discuss the accuracy, precision, recall, or any other relevant metrics.

- Share visualizations of the model's predictions, such as time series forecasts for sales trends or clustering results for customer behavior.

7. Feature Importance:

- If applicable, discuss the importance of different features in making predictions. You can use techniques like feature importance scores from tree-based models or correlation analysis.

8. Insights and Business Recommendations:

- Share insights gained from the machine learning models. What trends, patterns, or customer segments have been identified?

- Provide actionable recommendations for your business based on these insights. For example, marketing strategies or inventory management improvements.

9. Future Work:

- Suggest potential areas for further research or improvements in the machine learning approach. This could include exploring different algorithms, acquiring more data, or refining feature engineering.

10. Conclusion:

- Summarize the key findings and the value of incorporating machine learning for predicting sales trends and customer behaviors.

11. References:

- If you've used any external sources, research papers, or libraries, provide proper citations.

12. Appendix:

- Include any supplementary information, code snippets, or detailed data descriptions that support your analysis.

Once you've created this document, you can share it for assessment as per the instructions provided to you. If you have any specific questions or need further assistance with any section of the document, please let me know

Algorithm:

1. Import necessary Python libraries, including pandas, numpy, matplotlib, and seaborn.
2. Read a CSV file ('statsfinal.csv') into a DataFrame named 'data.'
3. Perform data preprocessing, including removing the 'Unnamed: 0' column and extracting day, month, and year from the 'Date' column.
4. Filter the data to exclude records from the year 2010 and 2023.
5. Define a function called 'plot_bar_chart' to create bar charts for unit sales and revenue.
6. Generate bar charts to analyze unit sales and revenue trends by year.

7. Create visualizations to analyze sales trends for each product by month.
8. Extract data for the 31st day of each month for selected months and visualize it.
9. Calculate the average unit sales and revenue for all 31st days across all years for each product.

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

Data Analysis program:

```
# Import necessary libraries
```

```
import pandas as pd
```

```
import numpy as np
```

```
import matplotlib.pyplot as plt
```

```
import seaborn as sns
```

```
import warnings
```

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

```
# Load dataset
```

```
data = pd.read_csv(r'C:\ibm project\statsfinal.csv') # Use 'r'  
before the file path to interpret it as a raw string
```

```
# Preprocess dataset
```

```
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])
```

```
# Define analysis objectives
```

```
data_reduced = data.query("Year != '2010' and Year != '2023'")
```

```
# Create a function that allows us to plot a pie chart for the 4 products
```

```
def plot_pie_chart(df, columns, stri, str1, val):  
    # Aggregate sales for each product by year, by sum or mean  
    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()
```

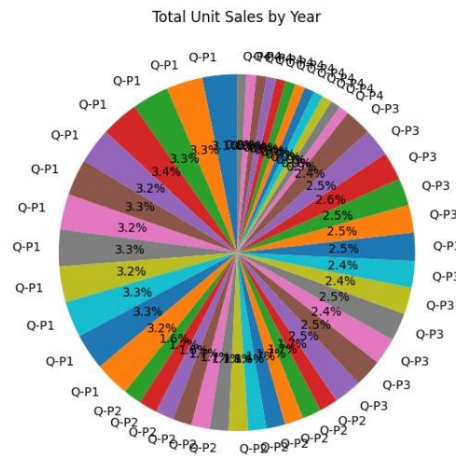
```
    # Melt the data to make it easier to plot  
    sales_by_year_melted = pd.melt(sales_by_year,  
id_vars='Year', value_vars=columns, var_name='Product',  
value_name='Sales')
```

```
    # Create a pie chart  
    plt.figure(figsize=(20, 4))  
    plt.pie(sales_by_year_melted['Sales'],  
labels=sales_by_year_melted['Product'], autopct='%1.1f%%',  
startangle=90)  
    plt.title(f'{stri} by {str1}')  
    plt.show()
```

```
plot_pie_chart(data_reduced, ['Q-P1', 'Q-P2', 'Q-P3', 'Q-P4'],  
'Total Unit Sales', 'Year', 'sum')  
plot_pie_chart(data_reduced, ['Q-P1', 'Q-P2', 'Q-P3', 'Q-P4'],  
'Mean Unit Sales', 'Year', 'mean')  
plot_pie_chart(data_reduced, ['S-P1', 'S-P2', 'S-P3', 'S-P4'],  
'Total Revenue', 'Year', 'sum')  
plot_pie_chart(data_reduced, ['S-P1', 'S-P2', 'S-P3', 'S-P4'],  
'Mean Revenue', 'Year', 'mean')
```

1 output:

Figure 3



2 ouput:

Figure 3

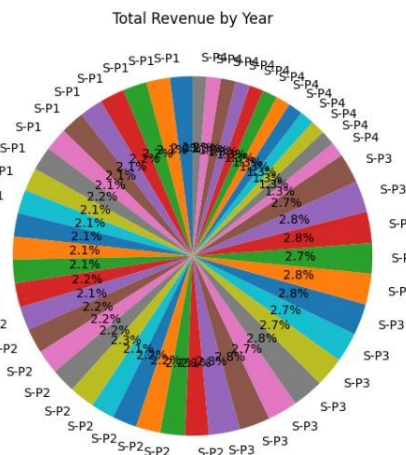


Figure 3



Figure 3



CONCLUSION:

Here by I am concluding my knowledge at phase 3: Development part of Product sales analysis.

