JULY 5, 2024

COFFE SHOP SALES
PROJECT REPORT

MUHAMMAD AWAIS
AWAISBARAK.77@GMAIL.COM
Domain: Data Science

Table of Contents

Overview:-

The Coffee Shop Sales Analysis aimed to analyze coffee shop sales data to understand their sales trends, find their profitable and loss-making products, predict future sales, and suggest strategies for mitigating the loss and enhancing the profit. To achieve these in the project I did data cleaning, analysis of the data, and visualization techniques to provide actionable insight.

Objectives:-

1. **Data Cleaning:** Remove null values and if there are any duplicates from the dataset to ensure the quality of the dataset.

2. **Sales Analysis:** Visualize and calculate total Sales by Product category.

3. **Profit/Loss Analysis:** Find those products that are in profit and as well as in loss-making and analyze their sales trends.

4. **Sales Prediction:** Used linear regression technique to predict future sales.

5. **Loss Mitigation:** Analyze factors contributing to loss and improve plans to minimize or remove them.

Tools Used:-

- **PyCharm**
- **Python**
- **Excel**
- **Pandas , matplotlib and Scikit-Learn**

In this project, I used PyCharm as an IDE to write and run Python code. Excel was used to store and explore the data initially. Python language was used to analyze and visualize data. Pandas were used to manipulate and analyze the data. I used matplotlib to create charts and plots. For linear regression and predictive modeling, I used Scikit-learn.

Methodology:-

Data Loading and Cleaning:

I loaded the dataset. Then Remove rows with null and duplicate values, especially from sheet 6. And at last, made sure the date column was in DateTime format to analyze and predict properly. And at last, after analysis, the code creates a new Excel file under the name **cleaned_coffee_Shop_Sale.xlsx** which is a cleaned dataset that does not contain null and duplicate values as mentioned above

Figures

```python
import pandas as pd

# Define the file path and loading the dataset
file_path = r'C:\Users\Zenbook\PycharmProjects\pythonProject\Coffee Shop Sales.xlsx'
xls = pd.ExcelFile(file_path)
sheets = {sheet_name: xls.parse(sheet_name) for sheet_name in xls.sheet_names}

# Remove rows with null values from sheet 6
sheets['Sheet6'].dropna(inplace=True)
data = sheets[xls.sheet_names[0]]
data['transaction_date'] = pd.to_datetime(data['transaction_date'])

data.dropna(inplace=True)

data.drop_duplicates(inplace=True)

data['total_price'] = data['transaction_qty'] * data['unit_price']
```

Figure 1 Data Loading Code

Sales Analysis:

Calculated total sales by product category of all categories. For that, I also created a clustered bar chart to visualize it

Figures

```python
import matplotlib.pyplot as plt

category_sales = data.groupby('product_category')['total_price'].sum().reset_index()

# Plot total sales by product category (Clustered Bar Chart)
plt.figure(figsize=(10, 6))
plt.bar(category_sales['product_category'], category_sales['total_price'], color='skyblue')
plt.title('Total Sales by Product Category')
plt.xlabel('Product Category')
plt.ylabel('Total Sales')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```
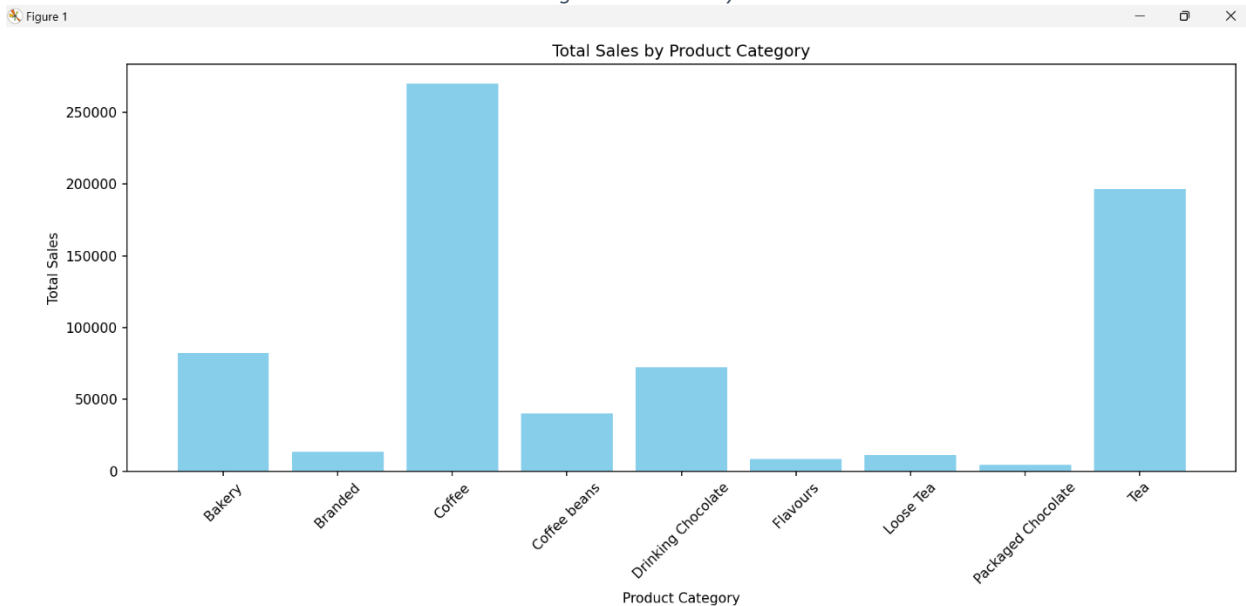
Figure 2 Sales Analysis code

Figure 3 Total Sales/Category

We can see that coffee and tea are the top two hot-selling categories of the Coffee shop.

Profit/Loss Analysis:

Total sales and profit are calculated using data grouped by product category. Using the median total sales, we identified profitable and loss-making products. To visualize that I created a line chart to show cumulative profit/loss over time.

Figures

```python
# Calculate daily profit and loss
daily_profit_loss = data.groupby('transaction_date')['total_price'].sum().reset_index()

# Calculate cumulative profit/loss
daily_profit_loss['cumulative_profit_loss'] = daily_profit_loss['total_price'].cumsum()

# Plot cumulative profit/loss over time (Line Chart)
plt.figure(figsize=(12, 6))
plt.plot(daily_profit_loss['transaction_date'], daily_profit_loss['cumulative_profit_loss'], label='Cumulative Profit/Loss')
plt.xlabel('Date')
plt.ylabel('Cumulative Profit/Loss')
plt.title('Cumulative Profit/Loss Over Time')
plt.legend()
plt.tight_layout()
plt.show()
```
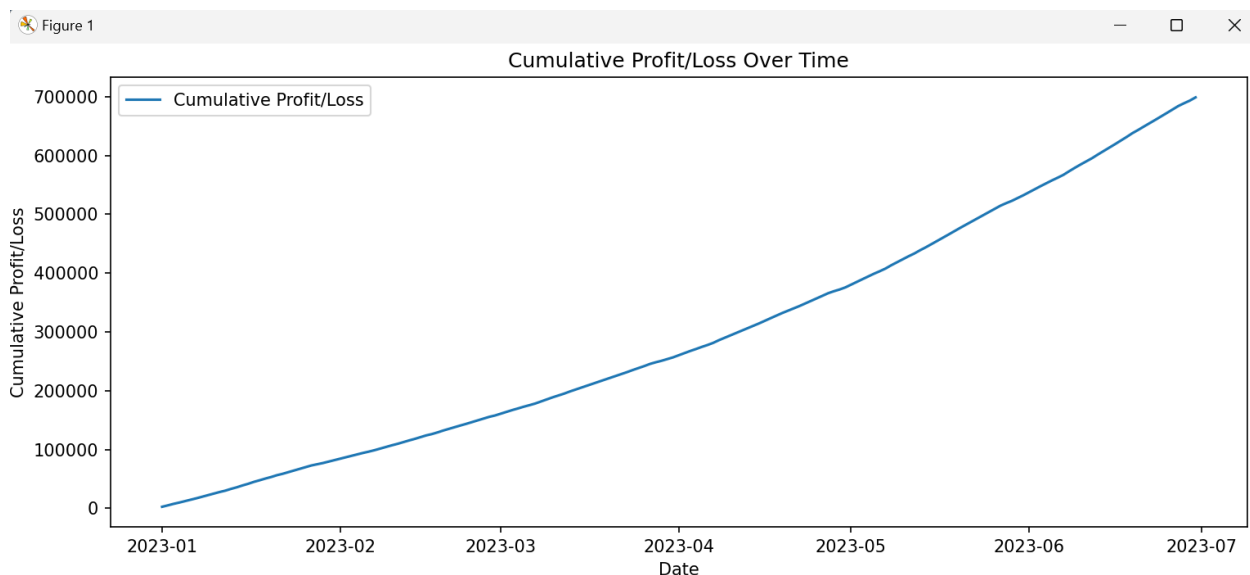
Figure 4 Profit/Loss Analysis code

Figure 5 Profit/Loss with Time

Sales Prediction:

To predict future sales based on monthly sales data by using linear regression technique. Visualize actual sales and predicted sales for future months by creating a line chart for that.

Figures

```python
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
data['month'] = data['transaction_date'].dt.month
monthly_sales = data.groupby('month')['transaction_qty'].sum().reset_index()
# Splitting data into training and testing sets
X = monthly_sales[['month']]
y = monthly_sales['transaction_qty']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
model = LinearRegression()
model.fit(X_train, y_train)
future_months = pd.DataFrame({'month': [7, 8, 9, 10, 11, 12]})
predictions = model.predict(future_months)


# Plot future sales predictions (Line Chart)
plt.figure(figsize=(10, 6))
plt.plot(monthly_sales['month'], monthly_sales['transaction_qty'], label='Actual Sales')
plt.plot(future_months['month'], predictions, label='Predicted Sales', linestyle='--')
plt.xlabel('Month')
plt.ylabel('Sales Quantity')
plt.title('Sales Predictions for Future Months')
plt.legend()
plt.tight_layout()
plt.show()
```
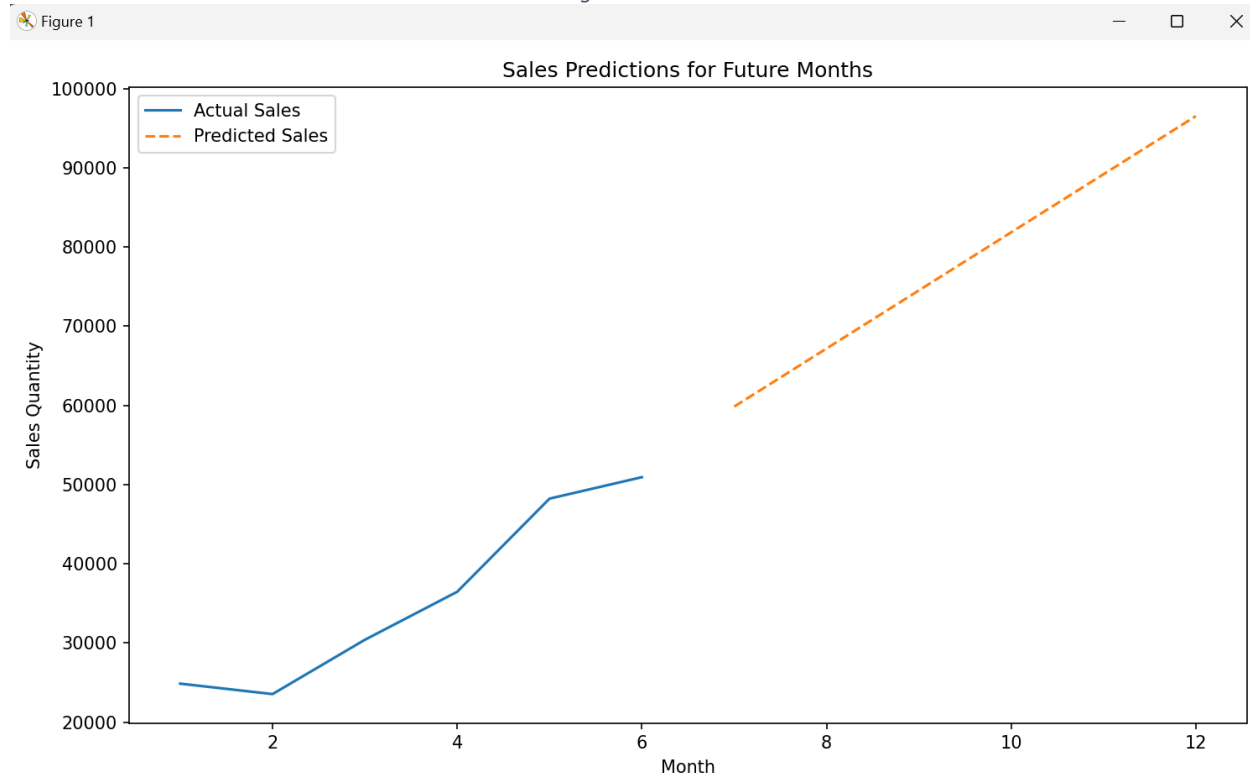
Figure 6 Prediction Code

Figure 7 Sales Prediction

Loss Mitigation:

Sales trends for loss-making products were analyzed over time. Pricing optimization, enhanced marketing efforts, promotional discounts, and product quality improvements are among the proposed tactics for turning loss-making items viable.

Figures

```python
# Identify profitable products
product_sales = data.groupby(['product_category', 'product_type']).agg({
    'transaction_qty': 'sum',
    'unit_price': 'mean'
}).reset_index()
product_sales['total_sales'] = product_sales['transaction_qty'] * product_sales['unit_price']
profitable_products = product_sales[product_sales['total_sales'] > product_sales['total_sales'].median()]
loss_making_products = product_sales[product_sales['total_sales'] <= product_sales['total_sales'].median()]

loss_making_analysis = data[data['product_type'].isin(loss_making_products['product_type'])]
plt.figure(figsize=(12, 6))
for product in loss_making_products['product_type']:
    product_data = loss_making_analysis[loss_making_analysis['product_type'] == product]
    product_sales_over_time = product_data.groupby('transaction_date')['total_price'].sum().reset_index()
    plt.plot(product_sales_over_time['transaction_date'], product_sales_over_time['total_price'], label=product)

plt.xlabel('Date')
plt.ylabel('Total Sales')
plt.title('Sales Trend of Loss-Making Products Over Time')
plt.legend()
plt.tight_layout()
plt.show()
```

Figure 8 Loss Mitigation Code

The following is the Sales trend of loss-making Products over time.
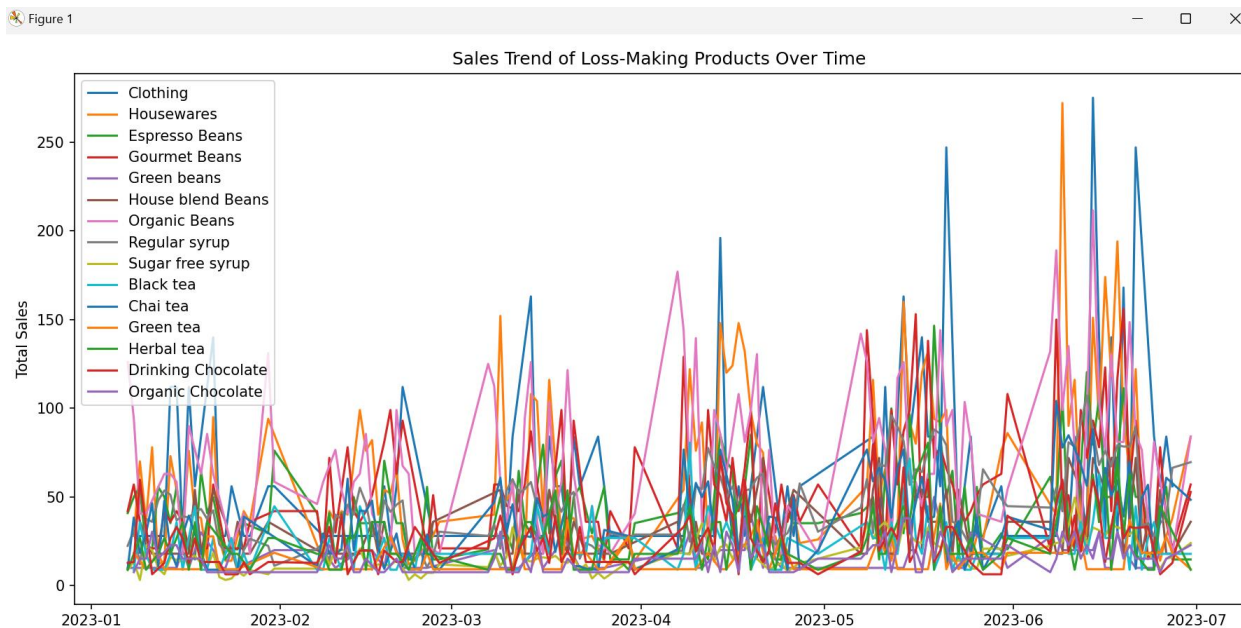

Figure 9 Sales Trend over time

Challenges:-

In The dataset **handling null values**, particularly in sheet 6. These were removed to maintain data integrity. Time-based analysis is required in this so for that the date column should be in the right **datetime format**. Another challenge was identifying products determining the threshold for **profitable** vs **loss-making** items required considerable analysis and a median-based approach.

Recommendations and Future Steps:-

To increase profit and reduce loss of the shop, the coffee shop should implement the offered strategies, which include improving pricing, targeting marketing, offering promotional discounts, and improving product quality based on consumer feedback. Regular data analysis is critical for following trends and making accurate decisions. Exploring modern predictive models, such as machine learning algorithms, can result in more accurate sales estimates.

Collecting and calculating consumer feedback will allow you to customize products and services to their preferences. Furthermore, using sales estimates to optimize inventory and supply chain management will ensure that popular products are available while reducing holding costs.

Conclusion:-

Used a planned strategy to provide full insights into the coffee shop's operations and financial performance. This included data cleaning, sales analysis, profit/loss analysis, sales prediction, and loss mitigation strategies. Key findings included the removal of null values and duplicates, the discovery of substantial sales differences by product category, the identification of profitable and loss-making products, and the use of linear regression techniques to predict future sales.

The study provided specific tactics for converting loss-making products into profitable ones, such as pricing optimization, greater marketing efforts, promotional discounts, and product quality improvements, as well as tangible recommendations to boost business growth and profitability.