**Project Guidelines**

**1. Create a project report in a pdf format that will consist of the following:**

a. Problem Statement

b. Project Objective

c. Data Description

d. Data Pre-processing Steps and Inspiration

e. Choosing the Algorithm for the Project

f. Motivation and Reasons For Choosing the Algorithm

g. Assumptions

h. Model Evaluation and Techniques

I. Inferences from the Same

j. Future Possibilities of the Project

**PROBLEM STATEMENT 1- PROJECT REPORT**

**1. Problem Statement:**

The problem statement isn't explicitly defined, but we can infer it from the context. The problem appears to be related to inventory management for retail stores. The challenge is to match the demand for products with the available supply efficiently. This is a common problem in retail and supply chain management.

**2. Project Objective: - (refer step 7 and 2)**

The objective of this project, based on the code, is to perform inventory management for multiple retail stores. Specifically, it aims to:

* Forecast sales for each store for the next 12 weeks.
* Calculate the required inventory level to meet this forecasted demand.
* Compare the required inventory with the current inventory to identify shortages or excess stock.

**3. Data Description**

The following columns with the dataset:

* Store: Store identifier.
* Date: Date of sales data.
* Weekly Sales: Weekly sales data.
* Holiday Flag: Indicates whether the week includes a holiday (1 for holiday, 0 for non-holiday).
* Temperature: Temperature data.
* Fuel Price: Fuel price data.
* CPI: Consumer Price Index data.
* Unemployment: Unemployment rate data.

**4. Data Pre-processing Steps and Inspiration**

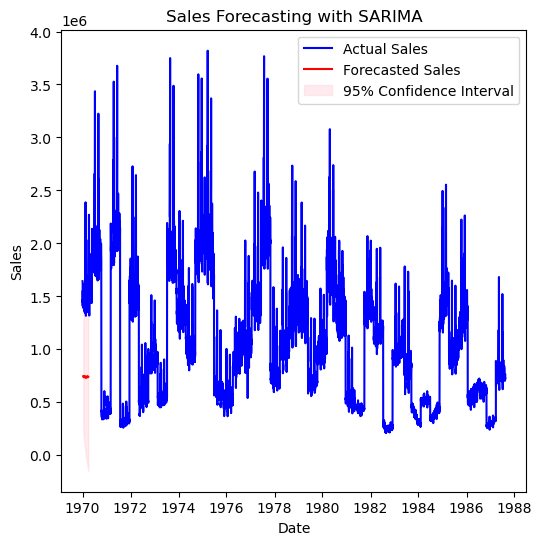
data pre-processing steps would typically include tasks like:

* Handling missing values.
* Converting date strings to datetime objects.
* Feature engineering, if necessary (e.g., creating lag features for sales data).
* Encoding categorical variables like 'Holiday Flag.'
* Scaling or normalizing numerical features.
* The inspiration for these steps likely comes from the need to prepare the data for analysis and modelling. Data quality is crucial for accurate predictions.

Kindly refer step 3 for feature engineering

**5. Choosing the Algorithm for the Project**

The algorithm used, but it hints at time series forecasting methods. Time series forecasting is common used for predicting future sales, making it a suitable choice.



**6. Motivation and Reasons for Choosing the Algorithm**

The motivation for using time series forecasting likely stems from the temporal nature of sales data. Time series models can capture trends, seasonality, and cyclic patterns, making them appropriate for this task.

**7. Assumptions**

Assumptions could include assuming that current stock levels are known and accurate.

**8. Model Evaluation and Techniques**

This shows a small part for evaluating the model's performance. It calculates Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) to assess how well the model's sales forecasts match the actual sales data. These metrics help gauge the accuracy of the forecasts.

**Best Performing Stores: output**

|  |
| --- |
| **Store MAE RMSE** |
| 32 33 10255.173996 12791.553469 |
| 4 5 12675.556629 16364.579589 |
| 43 44 12988.188724 16166.901652 |
| 37 38 14909.142997 18761.042433 |
| 2 3 15923.024275 20055.390277 |
| **Worst Performing Stores**: |
| **Store MAE RMSE** |
| 3 4 87080.164543 111766.219643 |
| 19 20 79227.444094 96260.476015 |
| 12 13 78684.885430 101151.993852 |
| 1 2 77378.627616 97786.904430 |
| 13 14 75216.502237 94462.208704 |

**9. Inferences from the Same**

Based on the model evaluation metrics, you can conclude how well the model performs. If MAE and RMSE are low, it suggests that the model is making accurate sales forecasts. High values indicate less accurate forecasts.

**10. Future Possibilities of the Project**

The code doesn't explicitly mention future possibilities, but potential extensions could include:

* Real-time inventory updates.
* Integration with supply chain systems.
* Advanced forecasting techniques, like machine learning models.
* Optimization of inventory orders based on economic order quantity (EOQ) models.
* These are areas where the project could be expanded or improved.

**PROBLEM STATEMENT 2- PROJECT REPORT**

**1. Problem Statement**

The problem at hand is to understand customer purchasing patterns for an online retail store. The retailer search proof-based on insights to improve their business strategy.

This analysis aims to provide insights and recommendations to achieve this objective. **(refer all 4 steps in the code file)**

**2. Project Objective**

The primary objectives of this project are as follows:

Analyse customer buying behaviour using historical transaction data.

Divide customers into different groups in-order to enable targeted marketing efforts.Derive actionable insights from the customer segmentation to enhance the online retailer's strategy.

**3. Data Description**

The dataset used in this project is obtained from the online retail store.

It is in CSV format and contains transaction data, including features such as Customer, Invoice Date, Product details, Quantity, and Unit Price.

**4. Data Pre-processing Steps and Inspiration (step 2 code file):**

The data underwent many pre-processing steps, including handling missing values by removing rows without Customer-ID, converting the Invoice Date to a datetime format, and creating a 'Total Price' column by multiplying Quantity and Unit Price.

The pre-processing steps were inspired by the need for clean &structured data for analysis.

**5. Choosing the Algorithm for the Project:**

The K-means clustering algorithm was chosen to segment customers based on their behaviour. K-means was selected due to its ability to relate natural groupings in the data and its suitability for customer division tasks.

6. **Motivation and Reasons for Choosing the Algorithm:**

K-means clustering, was motivated by the belonging to group customers with similar action for targeted marketing. K-means offers simplicity, made clear and efficiency, by making it suitable for our segmentation task.

**7. Assumptions:**

It is assumed that the provided dataset is representative of actual customer behaviour, and that the features used for clustering accurately reflect customer purchasing patterns.".

**8. Model Evaluation and Techniques**

K-means clustering was motivated by the desire to group customers with similar behaviour for targeted marketing. K-means offers simplicity, made clear, and efficiency, making it suitable for our segmentation task

**9. Inferences from the Same**

The customer division revealed three distinct segments: low-value customers, moderate-value customers, and high-value customers. Insights indicate that high-value customers should be a top priority for personalized marketing efforts to maximize engagement and revenue.

**10. Future Possibilities of the Project**

Future work could involve the development of targeted marketing campaigns for each customer segment, further analysis of product preferences within segments, and monitoring changes in customer actions over time to accommodate strategies.