**STORE ITEM FORECAST**

1. **Project Overview:**

This project uses time series forecasting techniques to predict future sales for various items across multiple stores. The primary objective is to explore seasonal trends, understand the impact of seasonality, and forecast sales for three months using SARIMA and SARIMAX models.

1. **Table of Contents:**

1. Tools and Libraries

2. Dataset

3. Key Steps and Analysis

4. Results

5. How to Run

6. Project Structure

7. Future Work

* 1. **Tools and Libraries:**

The project uses the following tools and libraries:

* Python: For data processing and analysis.
* Pandas: Data manipulation and preprocessing.
* Matplotlib and Seaborn: Data visualization.
* Statsmodels: Time series analysis and forecasting.
* Google Colab: For running the code in a cloud environment.
* Numpy: For numerical operations.
  1. **Dataset:**

The dataset includes sales data for 50 items across 10 stores over five years. Key fields include:

* **Date:** Date of the sale.
* **Store:** Store identifier.
* **Item:** Item identifier.
* **Sales:** Number of items sold on the corresponding date.

**2.3. Key Steps and Analysis:**

**1. Data Loading and Preprocessing:**

* + Load training and testing datasets.
  + Handle missing values and reduce memory usage by converting columns to appropriate data types.
  + Remove outliers by filtering sales above the 98th percentile.

**2. Feature Engineering:**

* + Convert the `date` column into a datetime object for time series operations.
  + Set the `date` column as the index for easier manipulation.
  + Generate new features like:
  1. Rolling averages
  2. First-order and seasonal differencing

**3. Visualization:**

* + Distribution of sales through histograms and boxplots.
  + Seasonal trends using autocorrelation and partial autocorrelation plots.
  + Decompose the time series into trend, seasonal, and residual components.

**4. Stationarity Tests:**

* + Perform ADF (Augmented Dickey-Fuller) and KPSS tests to check stationarity and decide if differencing is needed.

**5. Time Series Modeling:**

* + Build and evaluate a SARIMA model with parameters optimized based on ACF and PACF plots.
  + Extend SARIMAX by including additional external factors if necessary.

**6. Forecasting:**

* + Predict three-month sales using the trained SARIMA model for each store-item combination.
  + Save forecasts in a structured format for submission.

**2.4. Results:**

**1.** **Stationarity Tests:**

* + ADF and KPSS results indicated the need for first-order and seasonal differencing.
  + Forecasting Performance:
  + The SARIMA model captured seasonal patterns effectively and provided sales forecasts for each store-item combination.

**2.** **Visualization Insights:**

* + Clear seasonality patterns, with weekly sales trends and peak seasons identified.

**2.5. How to Run:**

**1. Setup Environment:**

If using Google Colab, mount your Google Drive to access the dataset:

“python

from google.colab import drive

drive.mount '/content/drive' “

**2. Install the required libraries:**

“bash

pip install pandas numpy matplotlib seaborn statsmodels “

**3. Run the Code:**

3.1. Load the dataset:

“python

train = pd.read\_csv '/content/train.csv', parse\_dates= 'date'

test = pd.read\_csv '/content/test.csv', parse\_dates= 'date'

“

Follow the steps in the notebook to preprocess, analyze, and forecast sales.

**4. Save predictions:**

“python

final\_forecasts.to\_csv 'store\_item\_forecasts.csv', index=False

“

**5. Download Results:**

“python

from google.colab import files

files.download 'store\_item\_forecasts.csv'

“

**2.6. Project Structure:**

* + train.csv: The training dataset with historical sales data.
  + test.csv: The testing dataset for which predictions are generated.
  + Store\_Item\_Forecast.ipynb: Jupyter Notebook with the entire workflow for preprocessing, analysis, and forecasting.

**2.7. Future Work:**

* + Experiment with deep learning models like LSTMs for time series forecasting.
  + Incorporate additional external variables such as holidays, promotions, or weather data into SARIMAX for enhanced prediction accuracy.