README File

Predicting Walmart Sales and Performing Exploratory Data Analysis

Project Description

This project aims to address inventory management issues at Walmart by providing actionable insights and accurate sales forecasts. Using historical sales data and external factors, the project performs exploratory data analysis (EDA) and develops predictive models to forecast sales for each store over the next 12 weeks.

Objectives

1. **Data Analysis**: Perform EDA to derive meaningful insights from the dataset.
2. **Sales Forecasting**: Develop robust time series forecasting models to predict weekly sales for each Walmart store.

Dataset Information

The dataset walmart.csv contains 6435 rows and 8 columns:

- **Store**: Store number
- **Date**: Week of sales
- **Weekly_Sales**: Sales for the given store in that week
- **Holiday_Flag**: Indicates if the week includes a holiday
- **Temperature**: Temperature on the day of sale
- **Fuel_Price**: Cost of fuel in the region
- **CPI**: Consumer Price Index
- **Unemployment**: Unemployment rate

Installation Instructions

Ensure you have the following libraries installed. You can install them using pip:

```
pip install pandas numpy matplotlib seaborn scikit-learn statsmodels pmdarima fbprophet tbats
```

Installation Links for Required Packages

Below are the links for details and commands (if required) to install the necessary Python packages:

- **pandas**: Go to https://pypi.org/project/pandas/ or use command: pip install pandas
- **numpy**: Go to https://pypi.org/project/numpy/ or use command: pip install numpy
- **matplotlib**: Go to https://pypi.org/project/matplotlib/ or use command: pip install matplotlib
- **seaborn**: Go to https://pypi.org/project/seaborn/ or use command: pip install seaborn
- **scikit-learn**: Go to https://pypi.org/project/scikit-learn/ or use command: pip install scikit-learn
- **statsmodels**: Go to https://pypi.org/project/statsmodels/ or use command: pip install statsmodels
- **pmdarima**: Go to https://pypi.org/project/pmdarima/ or use command: pip install pmdarima
- **fbprophet**: Go to https://pypi.org/project/fbprophet/ or use command: pip install fbprophet
- **tbats**: Go to https://pypi.org/project/tbats/ or use command: pip install tbats

Usage

Clone the repository and navigate to the project directory:

```
git clone <repository_url>
cd <repository_directory>
```

Run the Jupyter Notebook to perform the analysis:

```
jupyter notebook walmart_project.ipynb
```

Insights and Analysis

The project involves the following steps:

1. **Data Preprocessing**
    - Cleaning the data
    - Handling missing values and duplicates
    - Transforming data types and creating new features
2. **Exploratory Data Analysis (EDA)**
    - Visualizing data distributions and trends
    - Analyzing the impact of external factors (e.g., unemployment rate, CPI, temperature) on sales
3. **Model Selection**
    - Evaluating time series forecasting models: ARIMA, SARIMAX, AutoARIMA, Prophet, TBATS
    - Selecting the best-performing model based on evaluation metrics
4. **Model Evaluation**
    - Using metrics like RMSE, MAE, and MAPE to assess model performance
5. **Business Applications**
    - Providing actionable insights for inventory management and strategic decision-making

Conclusion

The TBATS model was identified as the best-performing model for sales forecasting, providing Walmart with accurate forecasts and valuable insights to optimize inventory management and improve store operations. The project's outcomes emphasize the importance of precise time series forecasting in enhancing operational efficiency and maintaining a competitive edge in the retail industry.

Contact

For any questions or suggestions, please contact:

Tharakh George Chacko
Email: tharakhgeorge@yahoo.co.in
LinkedIn: www.linkedin.com/in/tharakh-george-chacko-27a50051