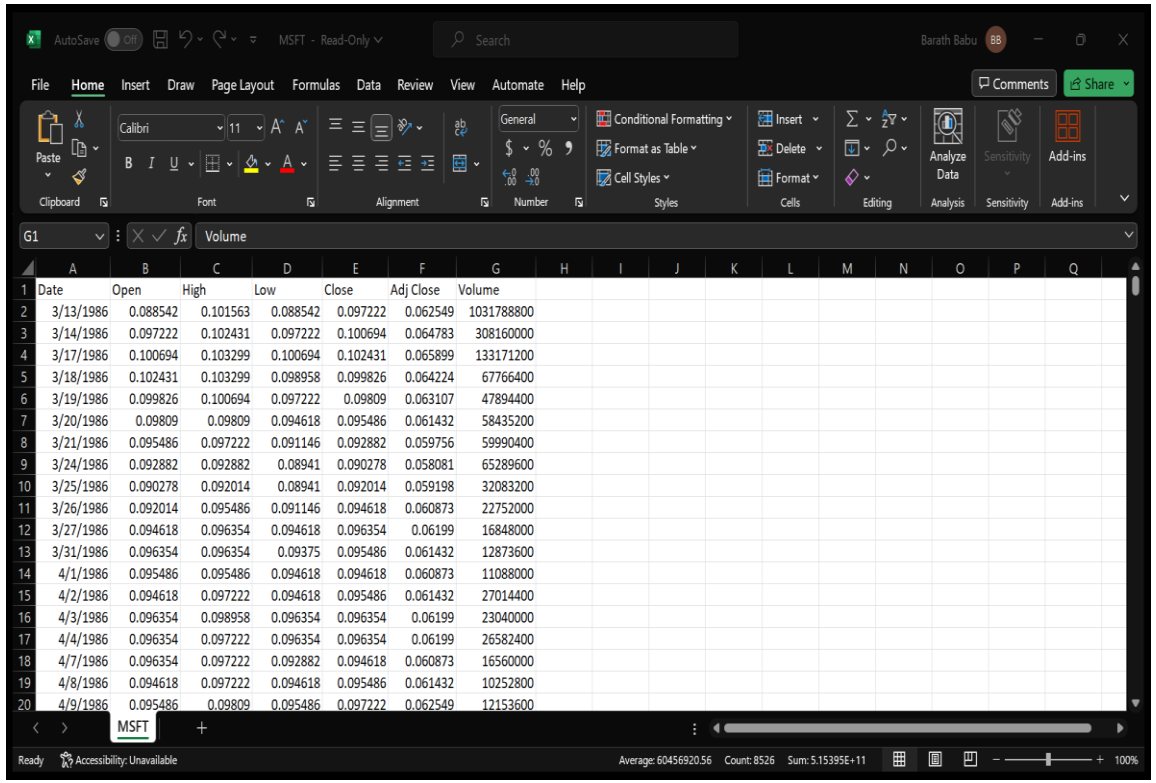
Project Title: Stock Price Prediction

Project Definition:

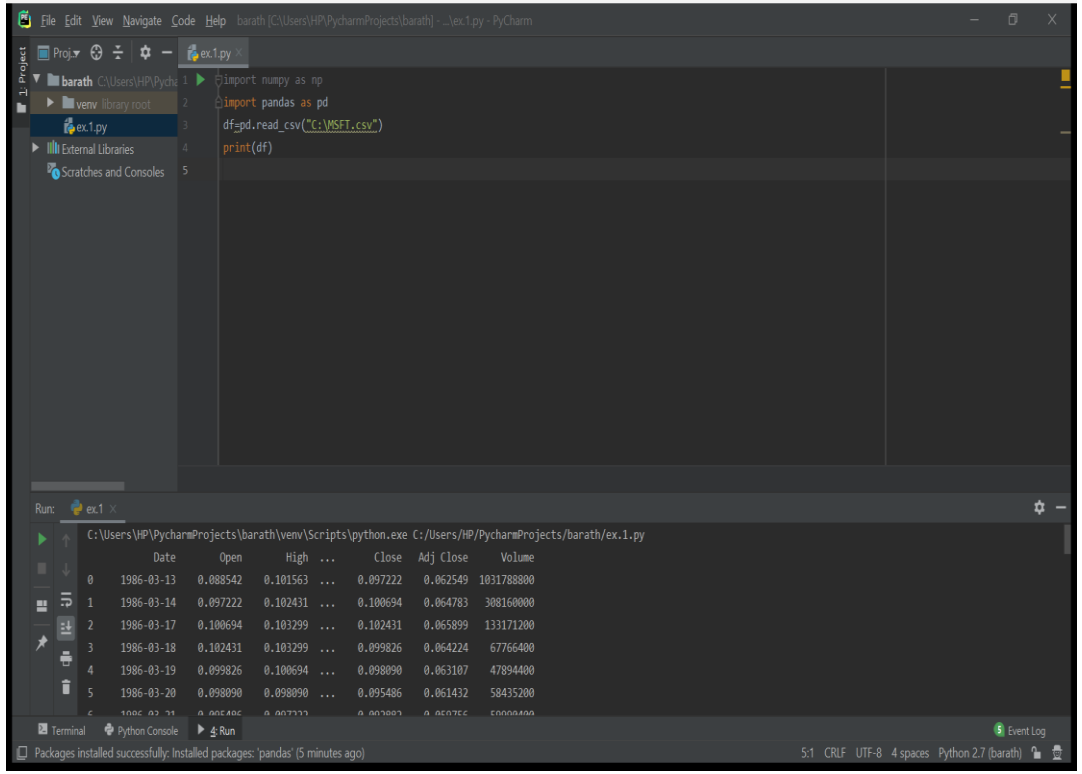
This PowerPoint contains an In-depth Analysis of Stock Price Prediction to build a predictive model that forecasts stock prices based on historical market data. This project involves data collection, data preprocessing, feature engineering, model selection, training, and evaluation.

Data Collection:

By making use of the link given below we can download the dataset for our project.
https://www.kaggle.com/datasets/prasoonkottarathil/microsoft-lifetime-stocks-dataset.

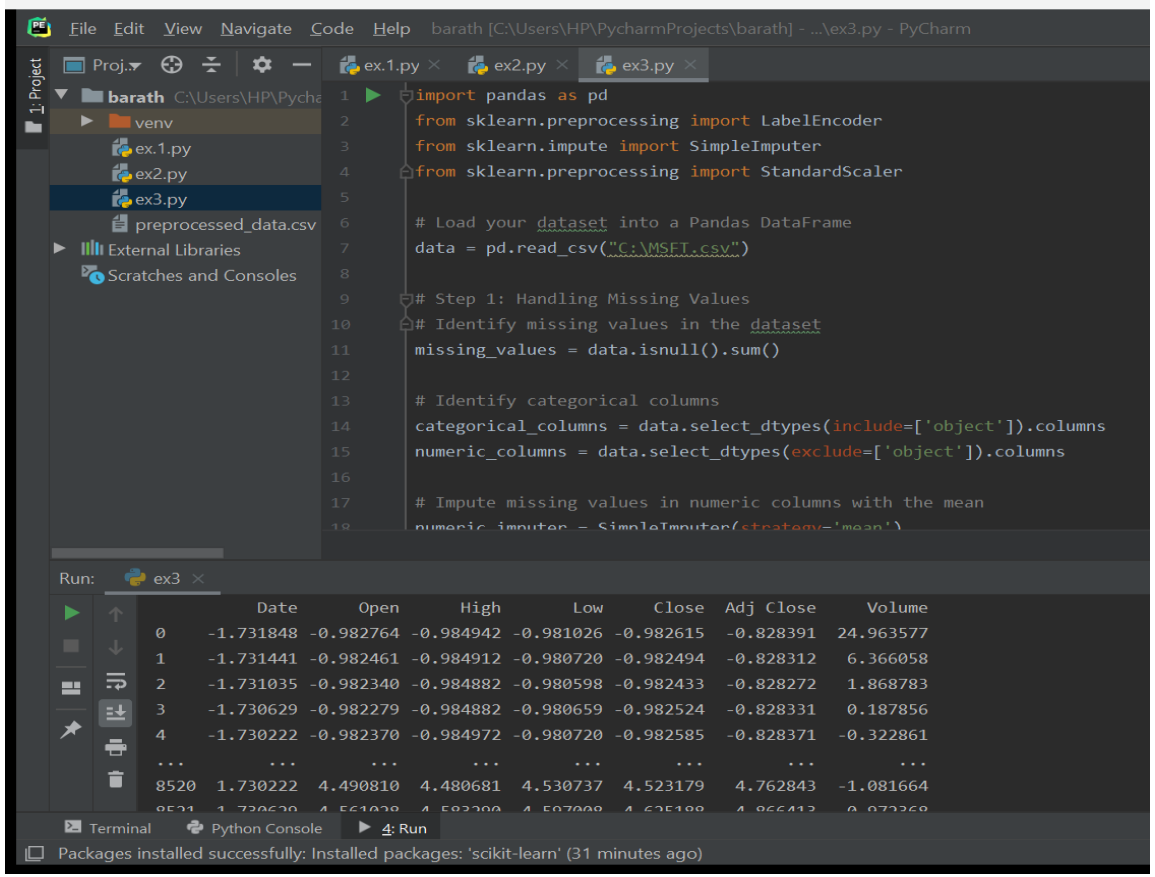


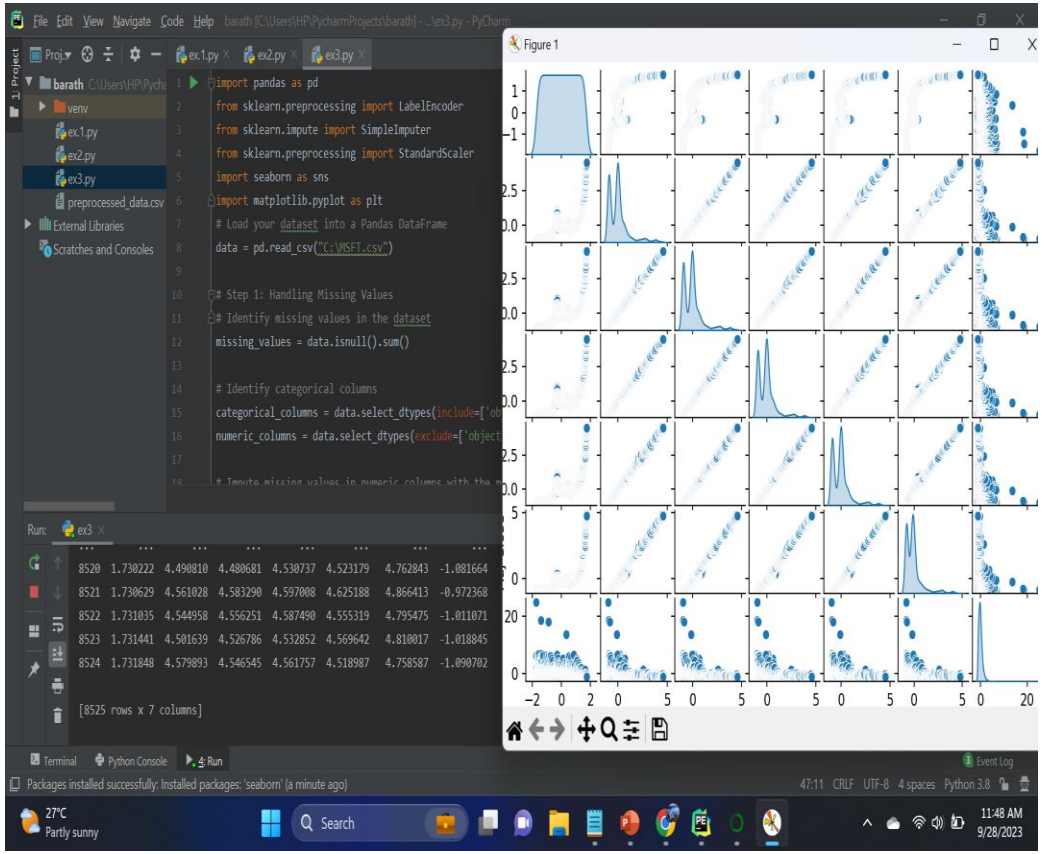
Date	Open	High	Low	Close	Adj Close	Volume
3/13/1986	0.088542	0.101563	0.088542	0.097222	0.062549	1031788800
3/14/1986	0.097222	0.102431	0.097222	0.100694	0.064783	308160000
3/17/1986	0.100694	0.103299	0.100694	0.102431	0.065899	133171200
3/18/1986	0.102431	0.103299	0.098958	0.099826	0.064224	67766400
3/19/1986	0.099826	0.100694	0.097222	0.09809	0.063107	47894400
3/20/1986	0.09809	0.09809	0.094618	0.095486	0.061432	58435200
3/21/1986	0.095486	0.097222	0.091146	0.092882	0.059756	59990400
3/24/1986	0.092882	0.092882	0.08941	0.090278	0.058081	65289600
3/25/1986	0.090278	0.092014	0.08941	0.092014	0.059198	32083200
3/26/1986	0.092014	0.095486	0.091146	0.094618	0.060873	22752000
3/27/1986	0.094618	0.096354	0.094618	0.096354	0.06199	16848000
3/31/1986	0.096354	0.096354	0.09375	0.095486	0.061432	12873600
4/1/1986	0.095486	0.095486	0.094618	0.094618	0.060873	11088000
4/2/1986	0.094618	0.097222	0.094618	0.095486	0.061432	27014400
4/3/1986	0.096354	0.098958	0.096354	0.096354	0.06199	23040000
4/4/1986	0.096354	0.097222	0.096354	0.096354	0.06199	26582400
4/7/1986	0.096354	0.097222	0.092882	0.094618	0.060873	16560000
4/8/1986	0.094618	0.097222	0.094618	0.095486	0.061432	10252800
4/9/1986	0.095486	0.09809	0.095486	0.097222	0.062549	12153600



Data Preprocessing:

In Case of this part we will be finding the Total number of missing values, Clean and preprocess the data in the given dataset and Handling the Categorical features into numerical representations.

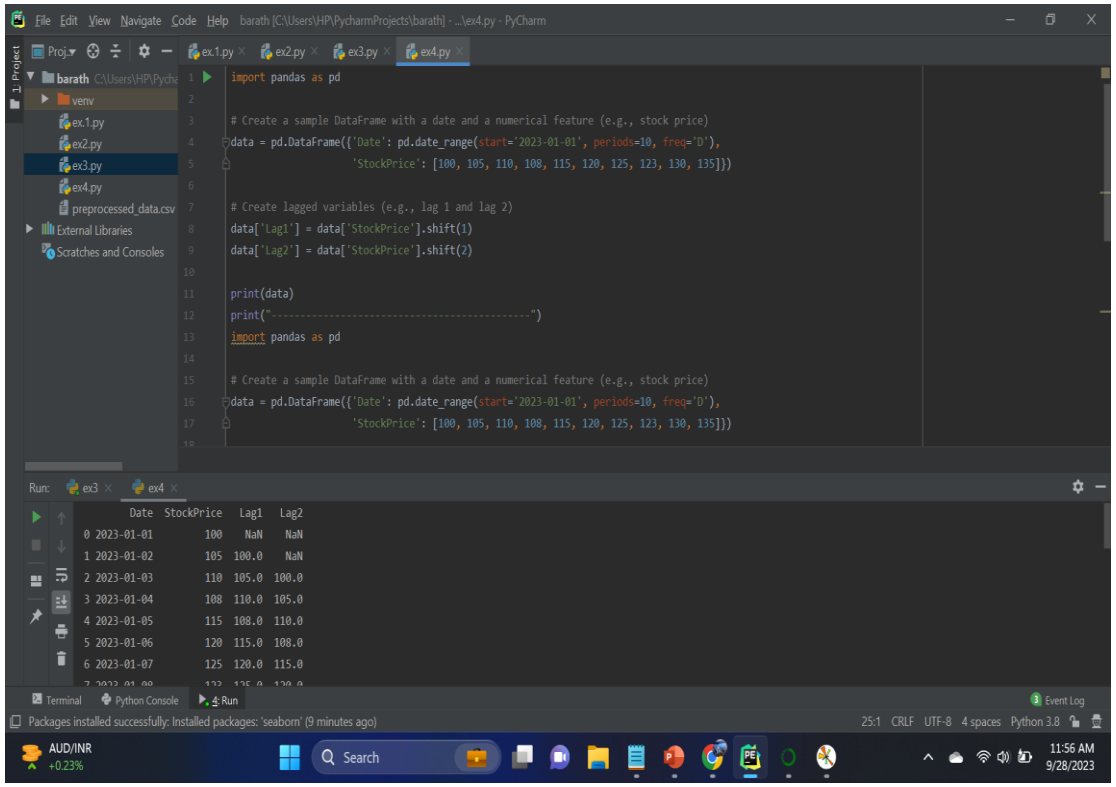




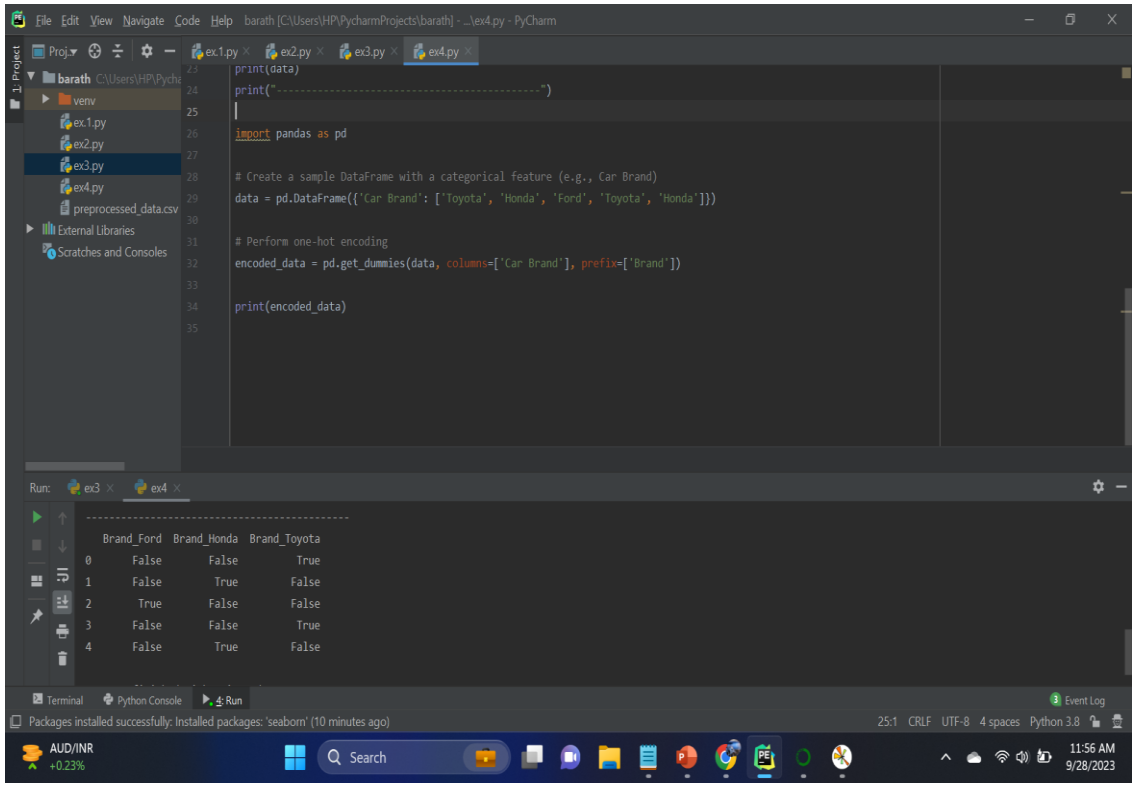
Feature Engineering:

Is the process of creating new features or modifying existing ones to improve the predictive power of a machine learning model. It involves extracting valuable information from the raw data or transforming features in a way that makes them more informative for the task at hand. Here are some common techniques

1. Creating Lagged Variable:



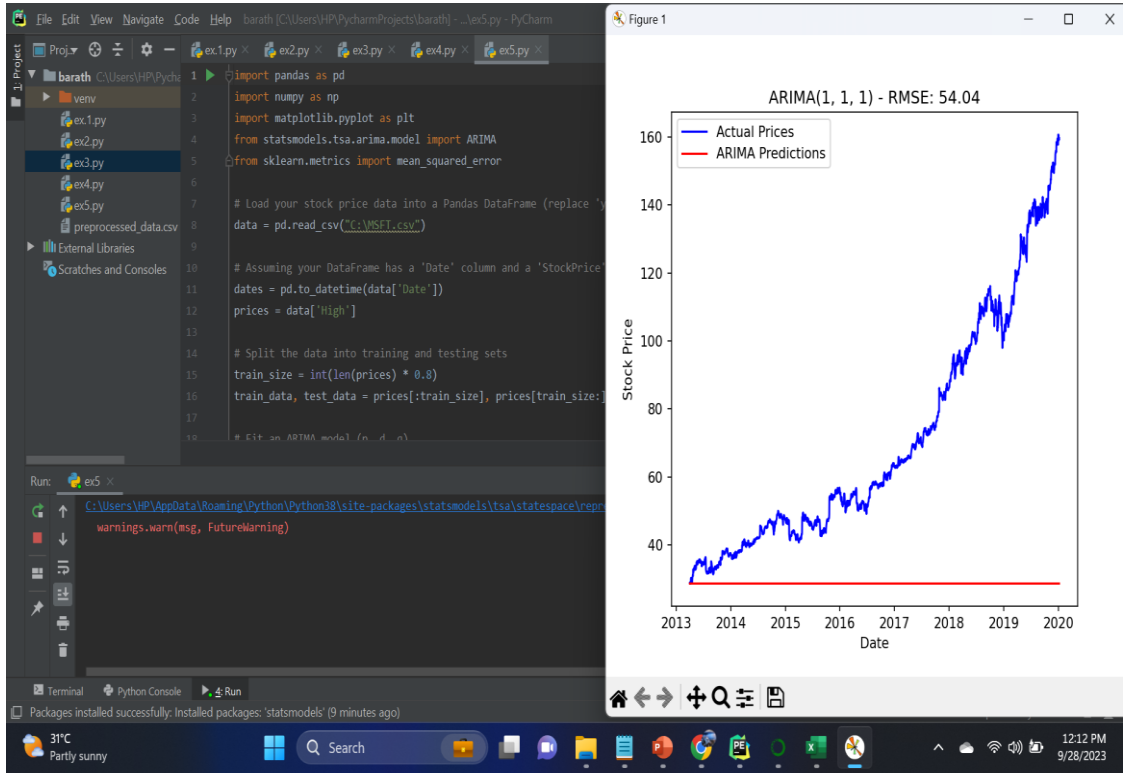
2.Moving Averages and One-Host Encoding



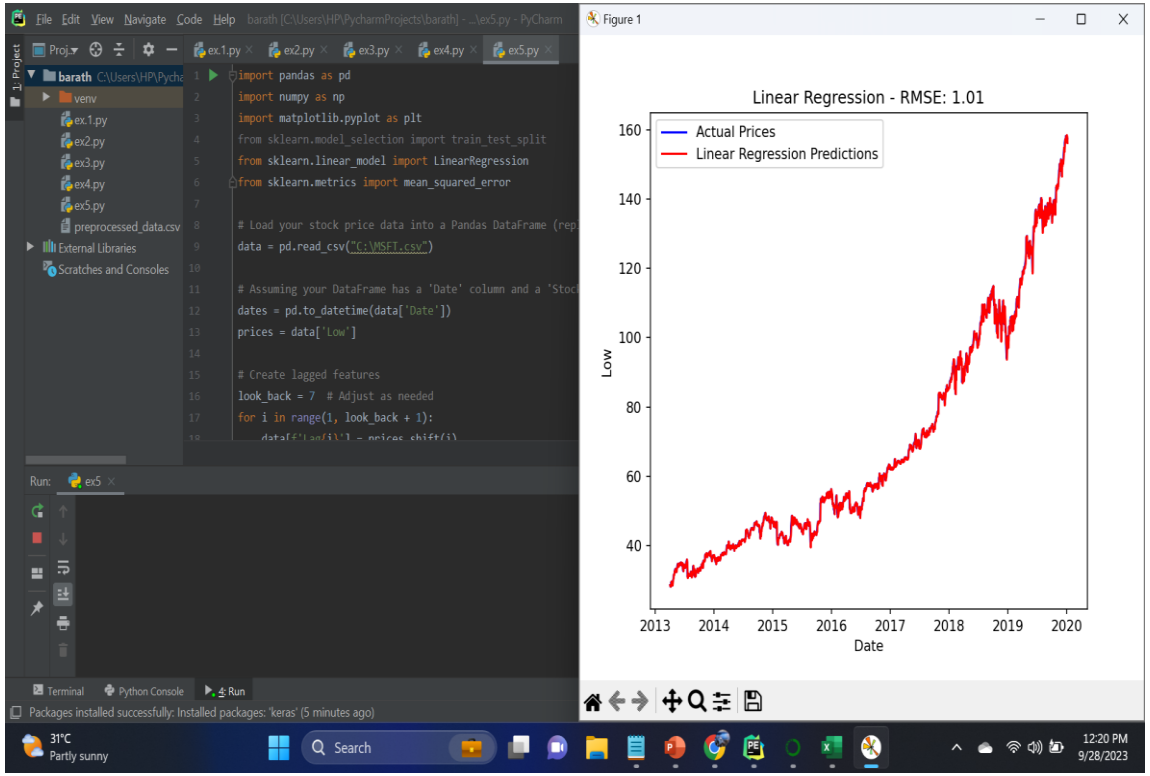
Model selection:

Is a critical step in time series forecasting, and it involves choosing the most suitable algorithm or model to make accurate predictions. When it comes to predicting stock prices, you have various options, including traditional statistical models like ARIMA (AutoRegressive Integrated Moving Average) and machine learning models like Long Short-Term Memory (LSTM) networks.

ARIMA Model Example:

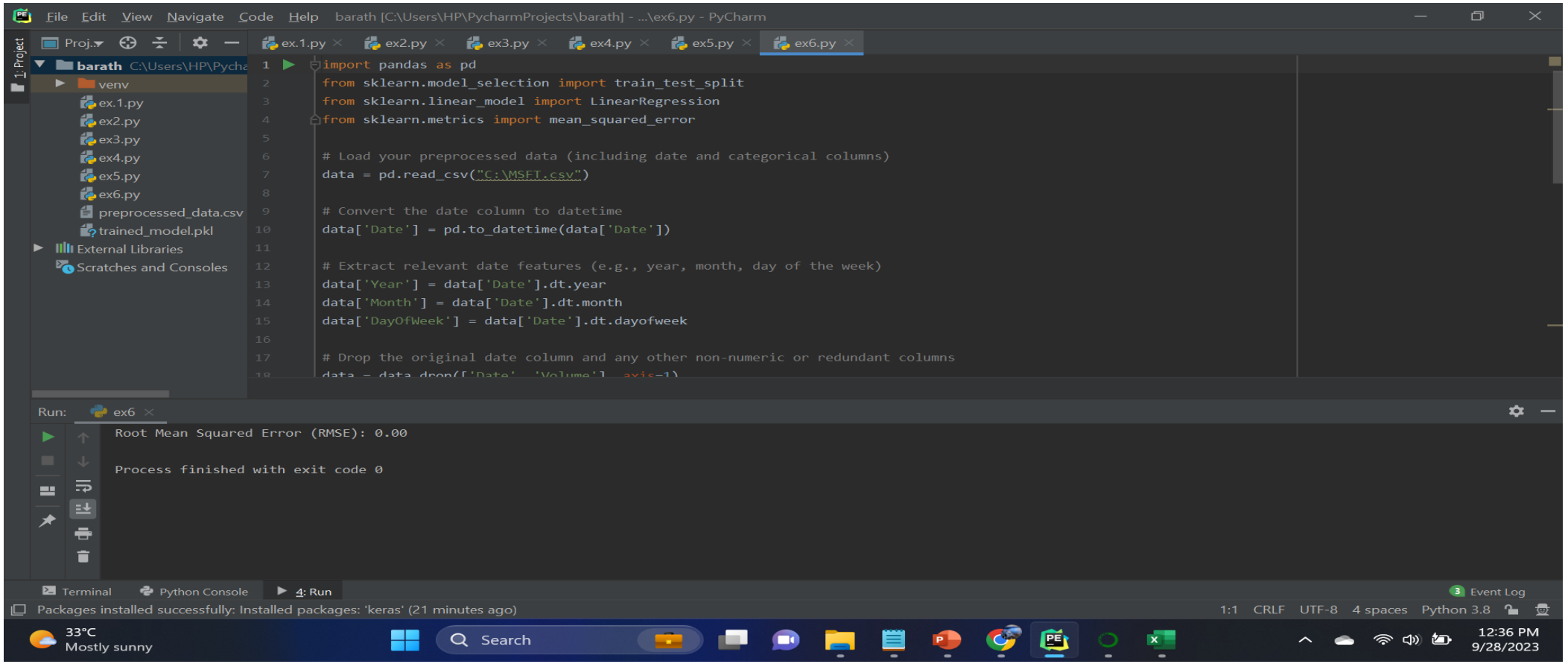


LSTM Model Example :



Model training :

Is the process of using a machine learning algorithm or model to learn patterns and relationships within your preprocessed data. During this phase, the model adjusts its internal parameters to minimize the difference between its predictions and the actual target values in your training dataset. Here's an overview of the model training process, along with Python code examples using scikit-learn for a simple linear regression model



Evaluation:

To evaluate the performance of your time series forecasting model, you can use various metrics, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and others. These metrics provide insights into how well your model's predictions align with the actual data. Here's how to compute these metrics using Python.