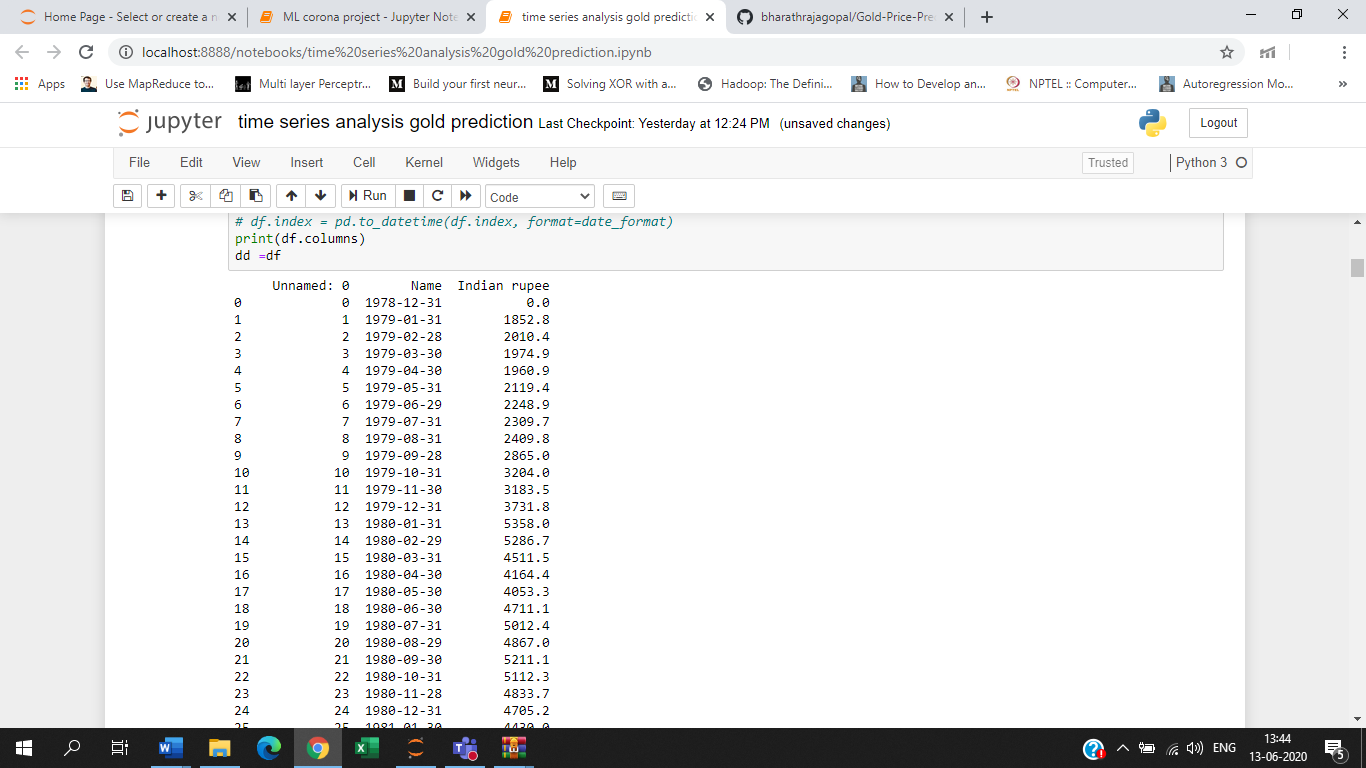
GOLD PRICE ANALYSIS AND PREDICTION USING

TIME SERIES ANALYSIS

**AIM OF THE PROJECT-**

* The aim of this project is to forecast the daily gold price in India and help the common people and the producers to choose the right date to buy gold and save the money in a more accurate way than traditional mind calculations.
* Predicting the gold price in India using the time series methods multiple linear regression, SARIMA. From these models, the best fitting model is selected, and prediction is done.
* After this project, any people who are looking to buy or sell gold in India can depend on the final model selected and can sell or buy this precious metal at the right time and for the right money.

**DATASET-**

**PREPROCESSING-**

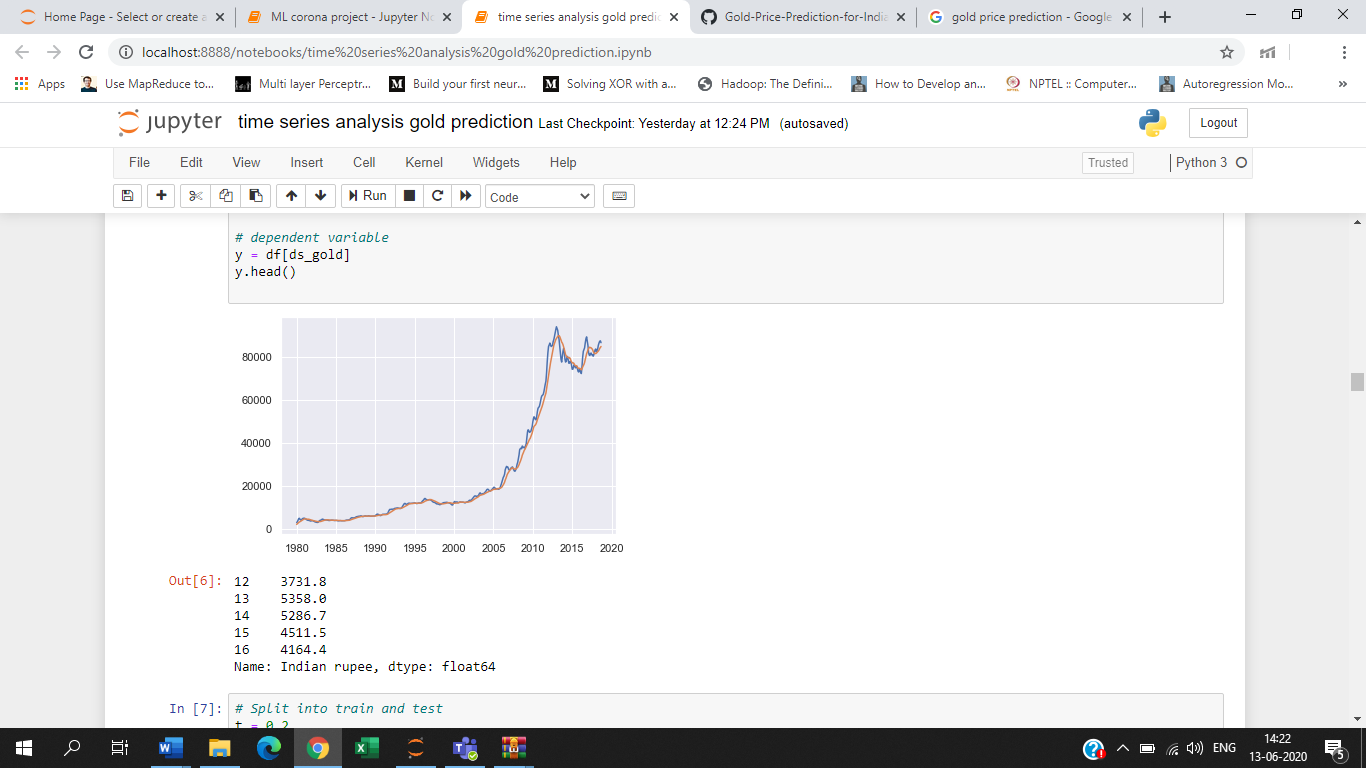
* Preprocess all the data by using basic techniques like dropping all the rows with missing values.
* The most preprocessing was done to the dates, since the data was collected from different sources.
* The dates were in different formats and they were formatted to a common format which would be understood by matplotlib for plotting it appropriately.



**CREATING A MODEL:**

**Multiple Regression**

* Multiple regression model with input parameters as the moving average of the past 1 month and the past 2 months.
* We can clearly observe overfitting in this model.
* This overfitting can be attributed to the data being non-stationary.



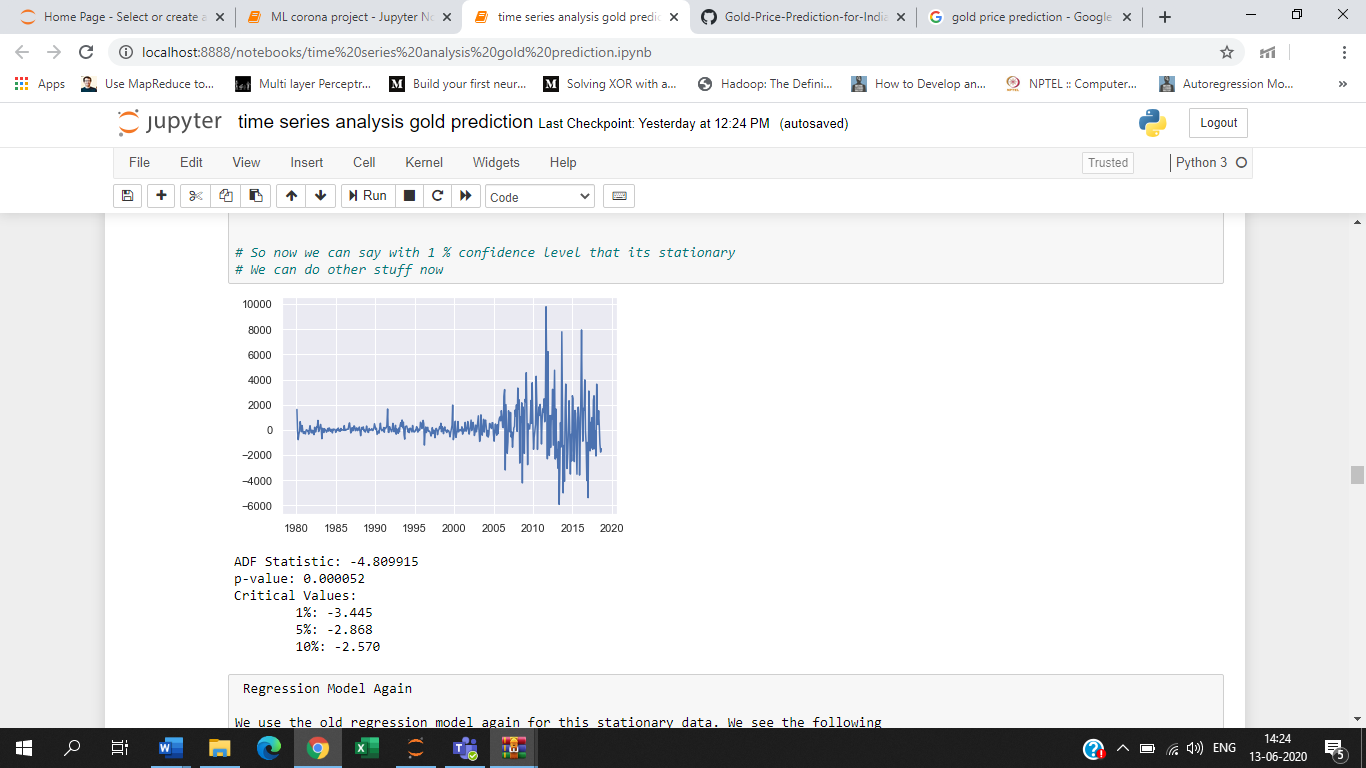
**Check Stationarity**

* To check the stationarity of the data we plot the data along with the dates.
* Perform the dickey-fuller test to confirm the stationarity.
* ADF statistic is higher than any of the critical values, and the p value is much greater than 0.05, so we cannot reject the null hypothesis that the data is non-stationary.
* To make the data stationary, we use simplest technique of taking a log

transform.

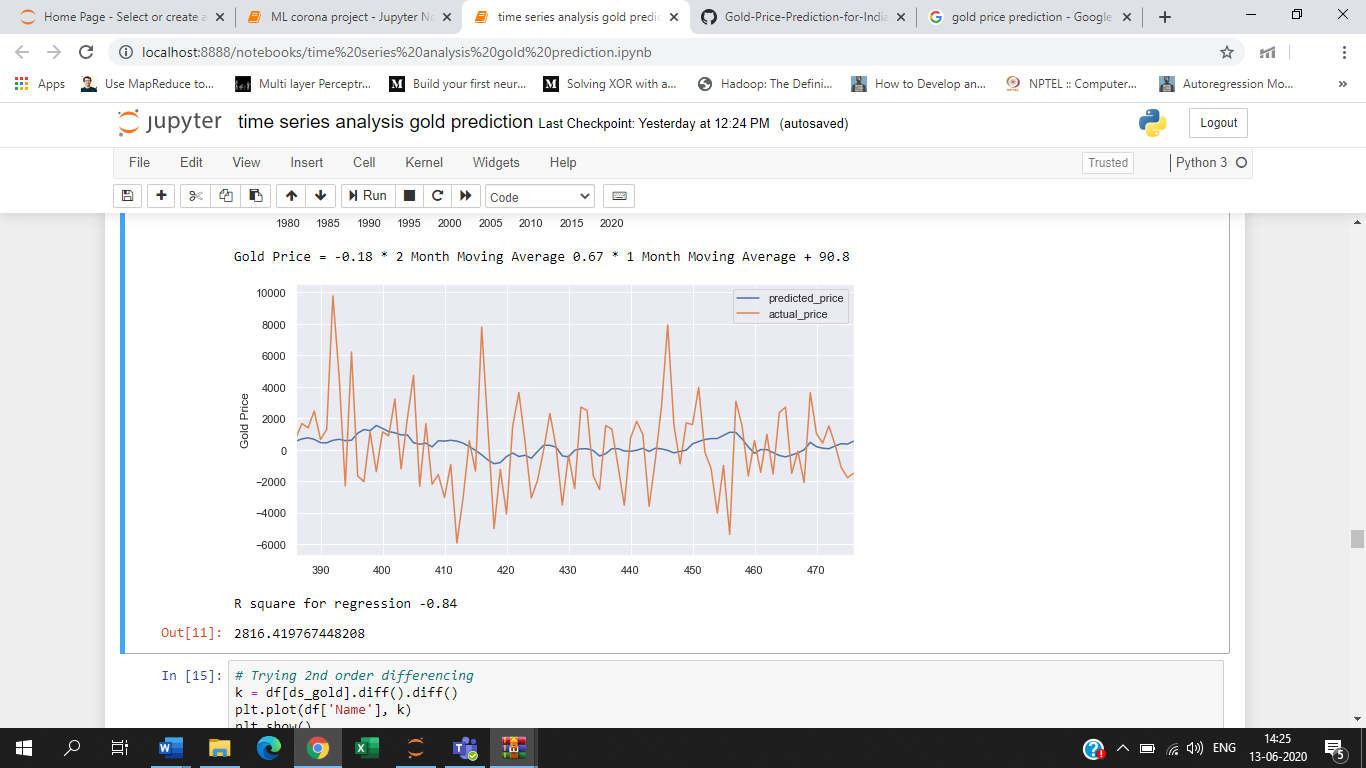
**DIFFERENCING**

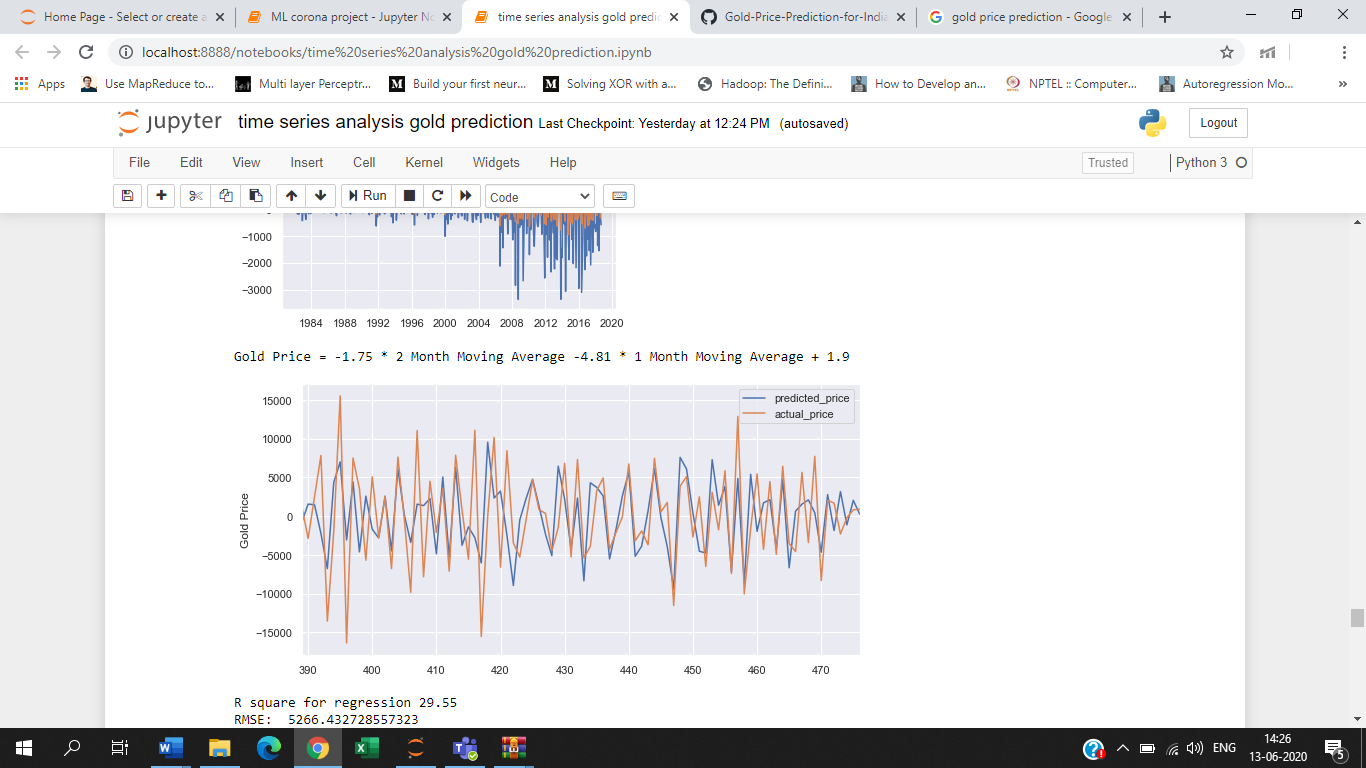
* We perform differencing of order 2 and observe the following results.
* We can see that the ADF statistic is less than 1% critical value hence we can reject the null hypothesis and conclude with a confidence level of 99% that the data is stationary.
* We can now use this data for further modelling.



**Regression Model Again**

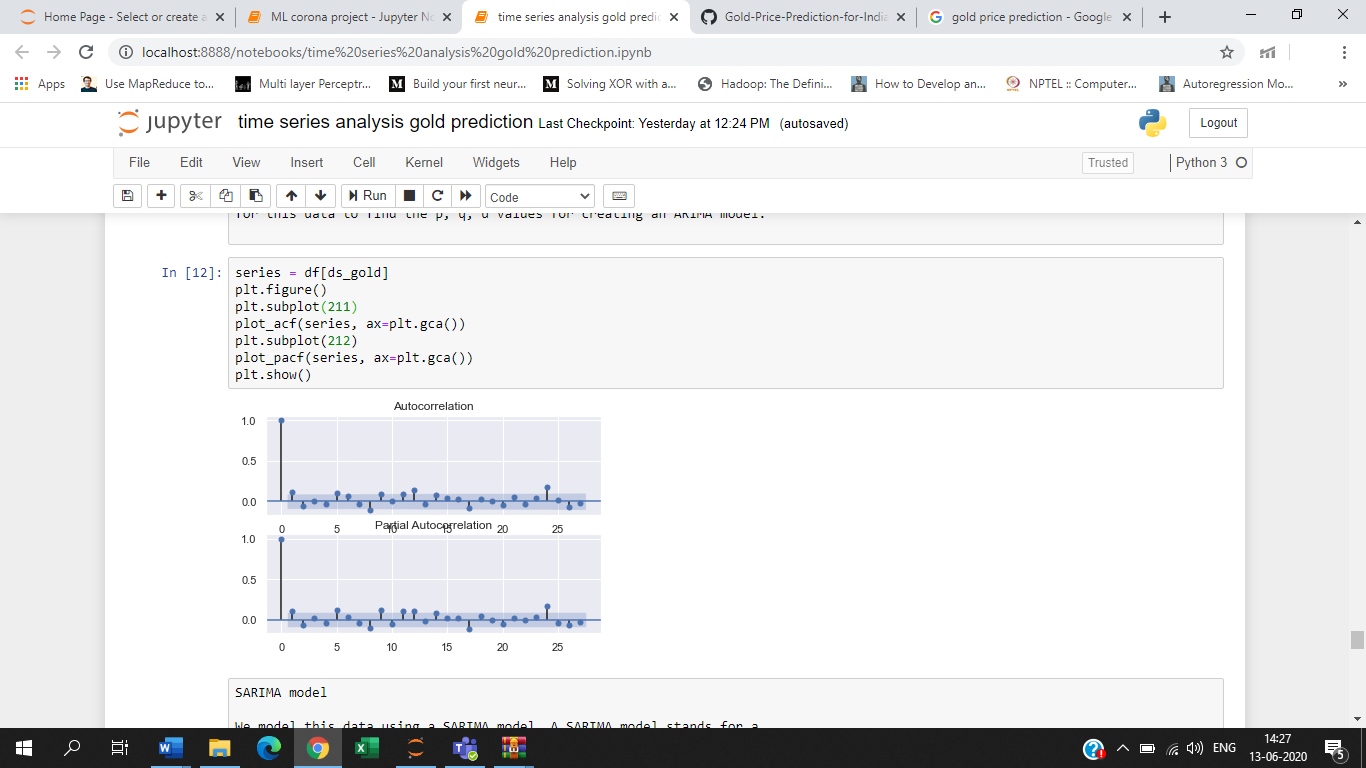
* We use the old regression model again for this stationary data. We see the following results. We obtain R square value of 30% which is below par. And the root mean square error is also very high.
* Though the RMSE is an absolute statistic and it cannot be used to judge the goodness of fit, we will use this value for further comparison with other models





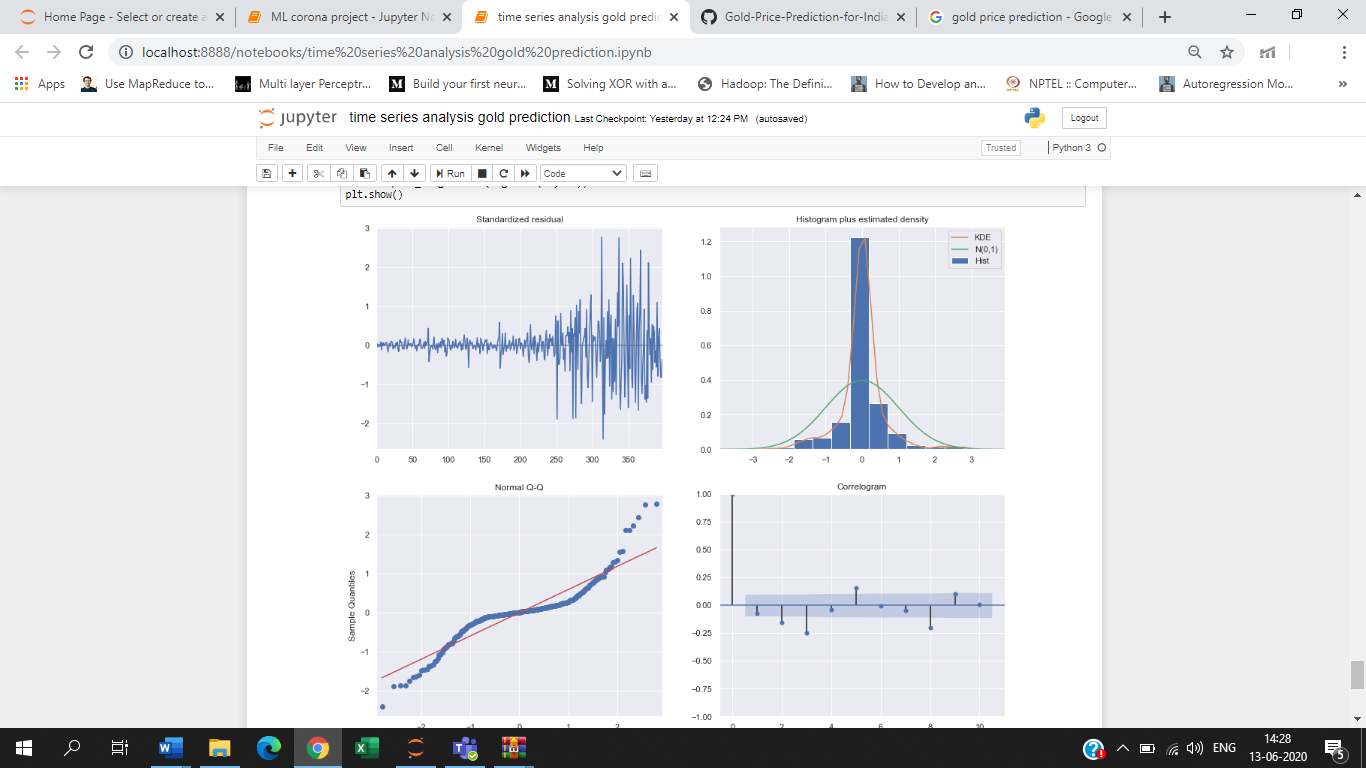
**ACF and PACF Plots**

* We now try to plot ACF i.e. Autocorrelation and PACF i.e. Partial Autocorrelation plots for this data to find the p, q, d values for creating an ARIMA model.



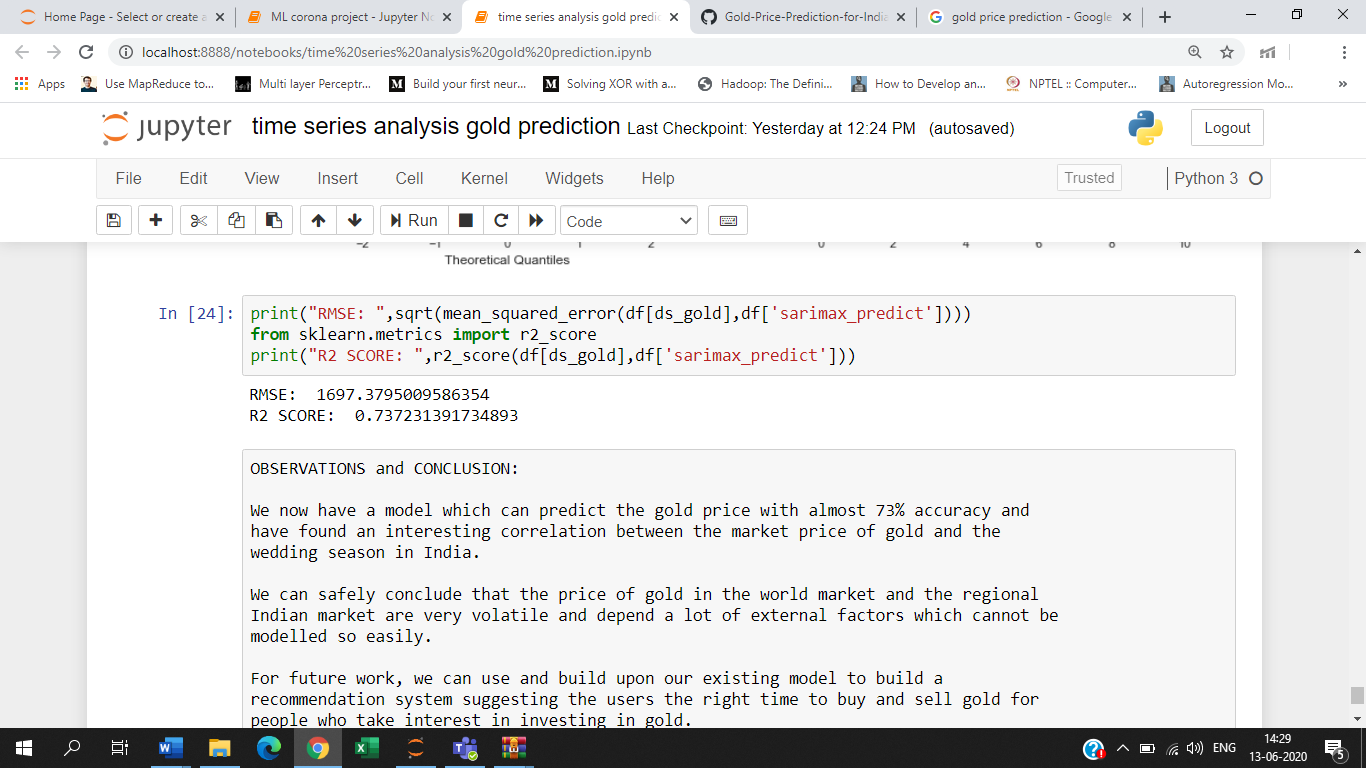
**SARIMA model**

* We model this data using a SARIMA model. A SARIMA model stands for a Seasonal ARIMA model. SARIMA Model is better over a simple ARIMA model when there is seasonal data. I.e. the timeseries data has repeating cycles.
* We observe that the model fits much better than any of the previous models.
* R square value is 73% which is acceptable and the RMS Error has reduced to 1715 from 5000, which is a good sign.



**OBSERVATIONS and CONCLUSION:**

* We now have a model which can predict the gold price with almost 73% accuracy and have found an interesting correlation between the market price of gold and the wedding season in India.
* We can safely conclude that the price of gold in the world market and the regional Indian market are very volatile and depend a lot of external factors which cannot be modelled so easily.



**FUTURE WORK**

We can use and build upon our existing model to build a recommendation system suggesting the users the right time to buy and sell gold for people who take interest in investing in gold.