**Case study**: Nifty-50 stocks

**About Dataset**

The National Stock Exchange of India Limited (NSE) is the leading stock exchange of India, located in Mumbai. The NIFTY 50 index is National Stock Exchange of India's benchmark broad based stock market index for the Indian equity market.

Apart from NIFTY 50 index, there are also other indices like NIFTY Next 50, Nifty Midcap 150 etc. Exploring these indices may help in taking investment decisions.

**Content**

This dataset has day level information on major NIFTY indices starting from 01 January 2000.

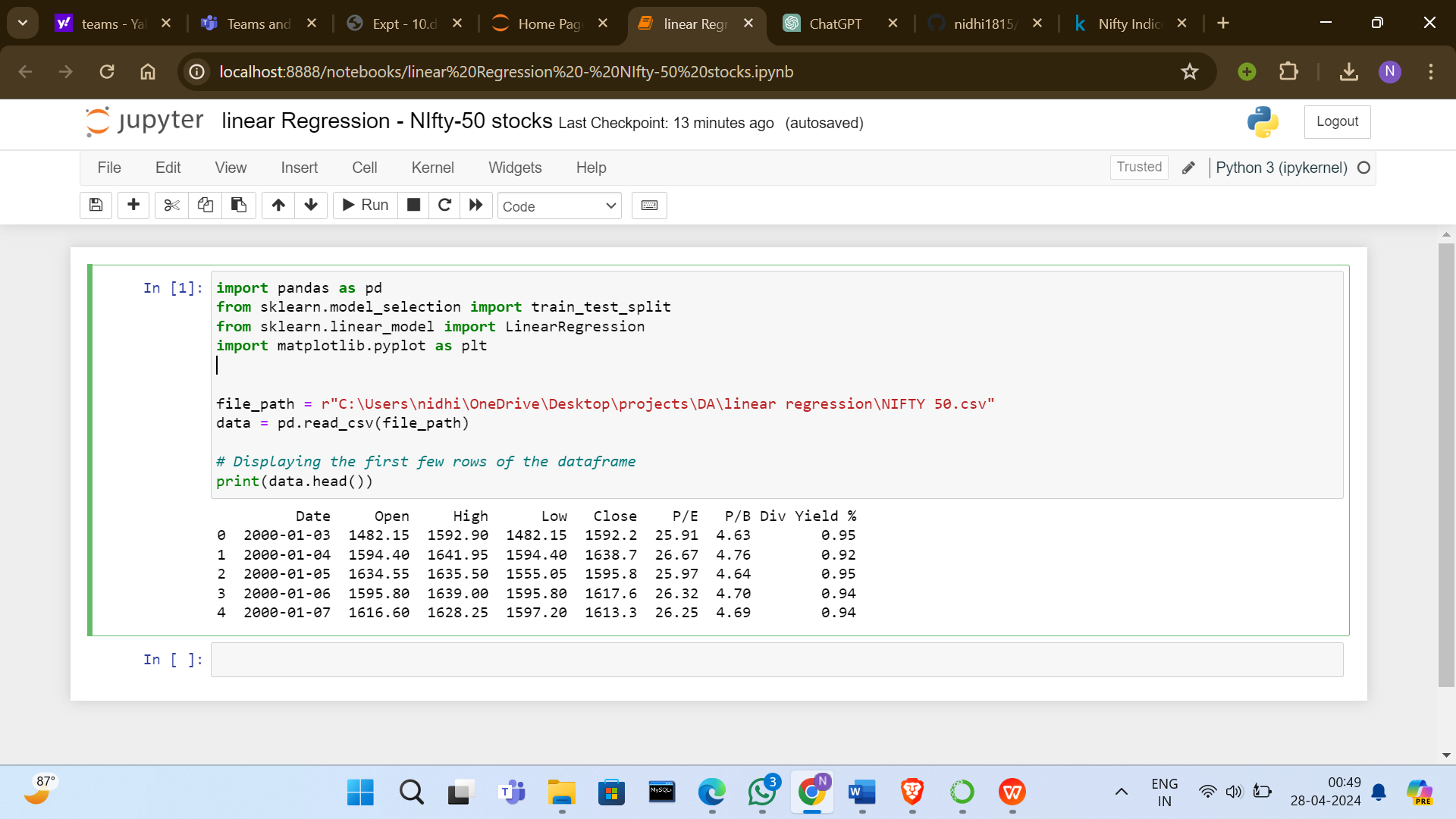
Each file represents an index and has the following columns

* Date - date of observation
* Open - open value of the index on that day
* High - highest value of the index on that day
* Low - lowest value of the index on that day
* Close - closing value of the index on that day
* Volume - volume of transaction
* Turnover - turn over
* P/E - price to earnings ratio
* P/B - price to book value
* Div Yield - dividend yield

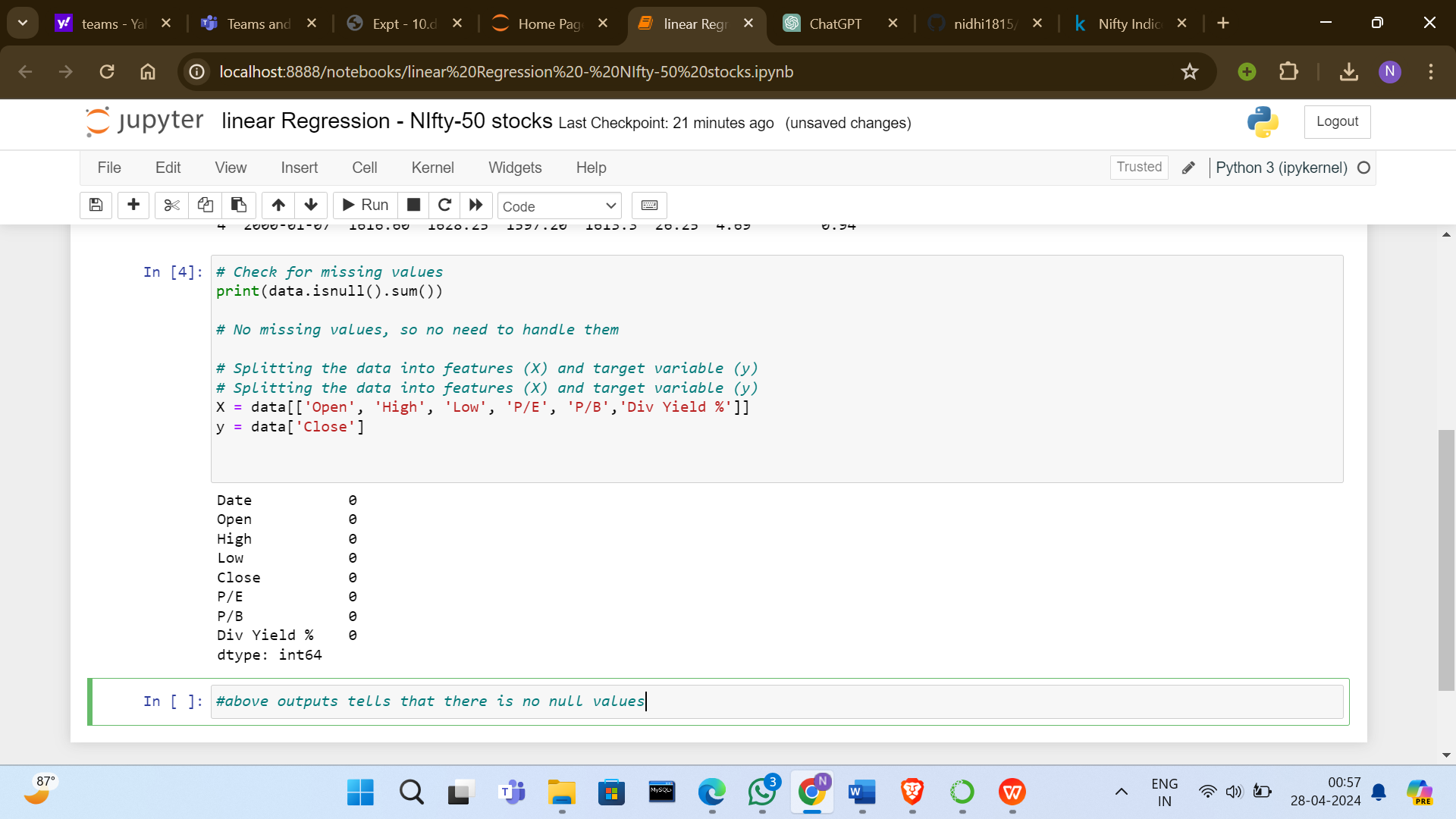
**Acknowledgements**

The data is obtained from [NSE](https://www1.nseindia.com/products/content/equities/indices/historical_index_data.htm) website with the help of nsepy python package.

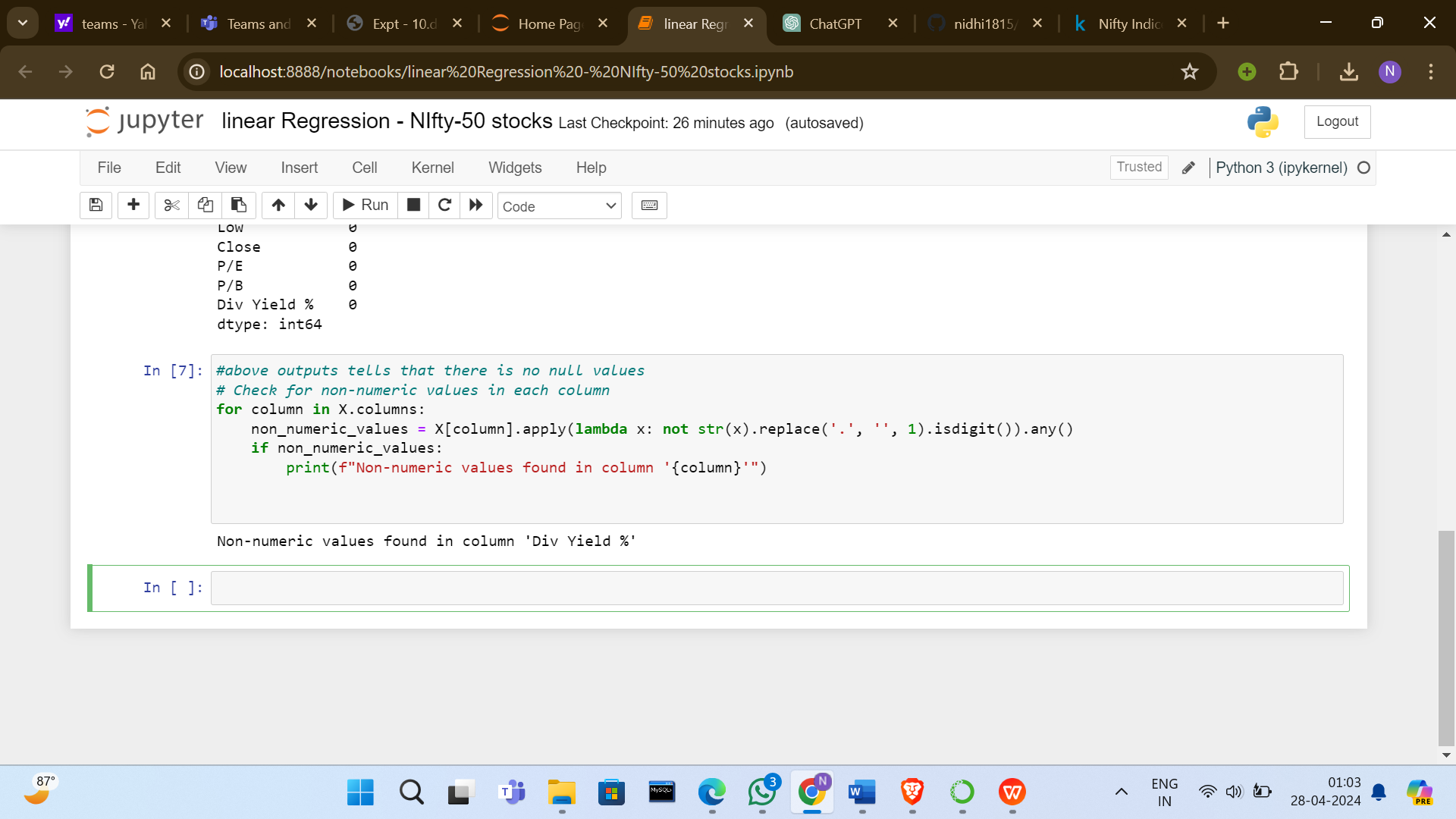
**Step 1**: Import Libraries and loading the data



Step 2: Data Preprocessing



Before moving to next step we need to check if data set contains non-numeric values

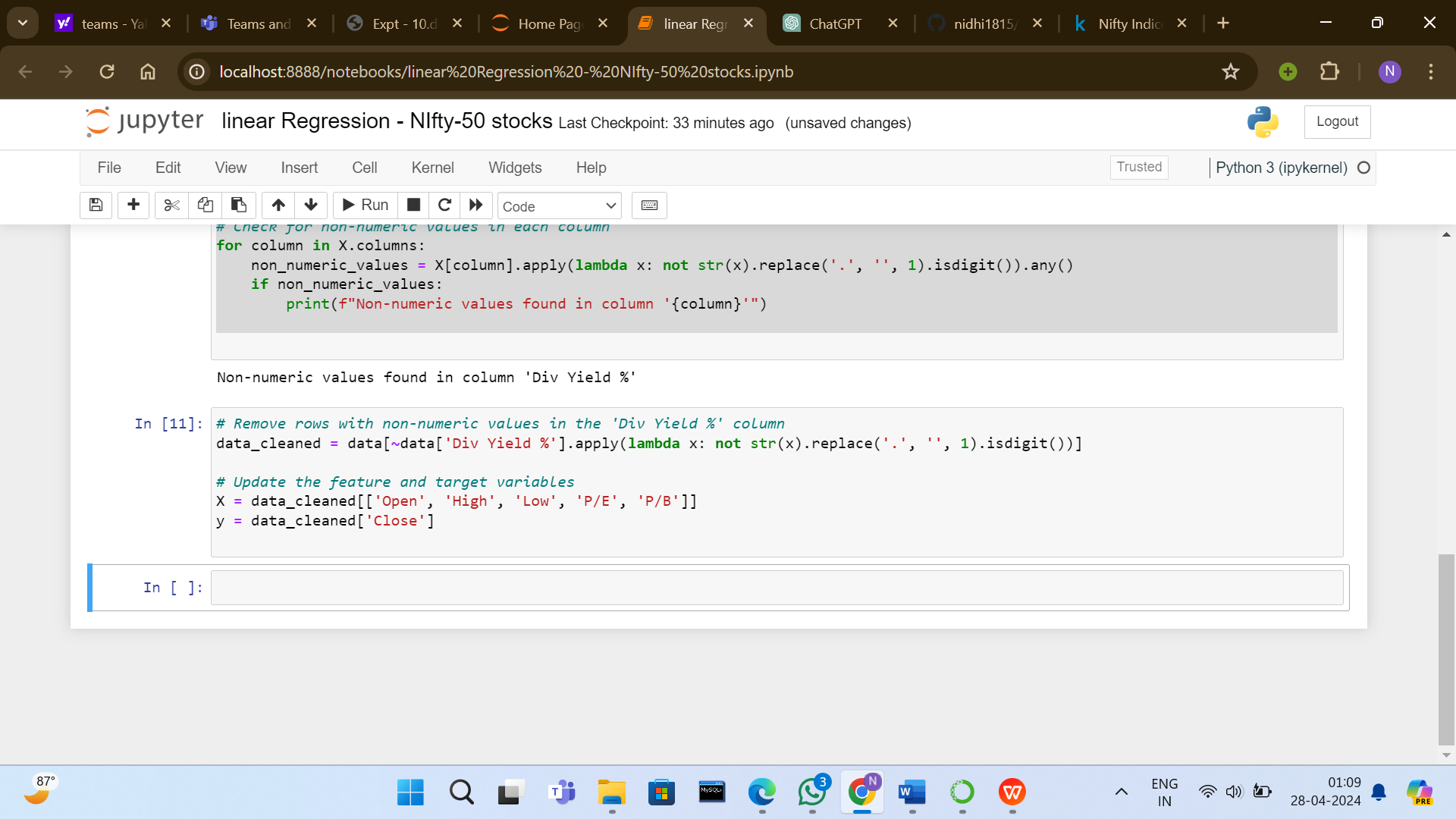


Therefore , Div yield column contains non numeric values, so we need to do something about it.

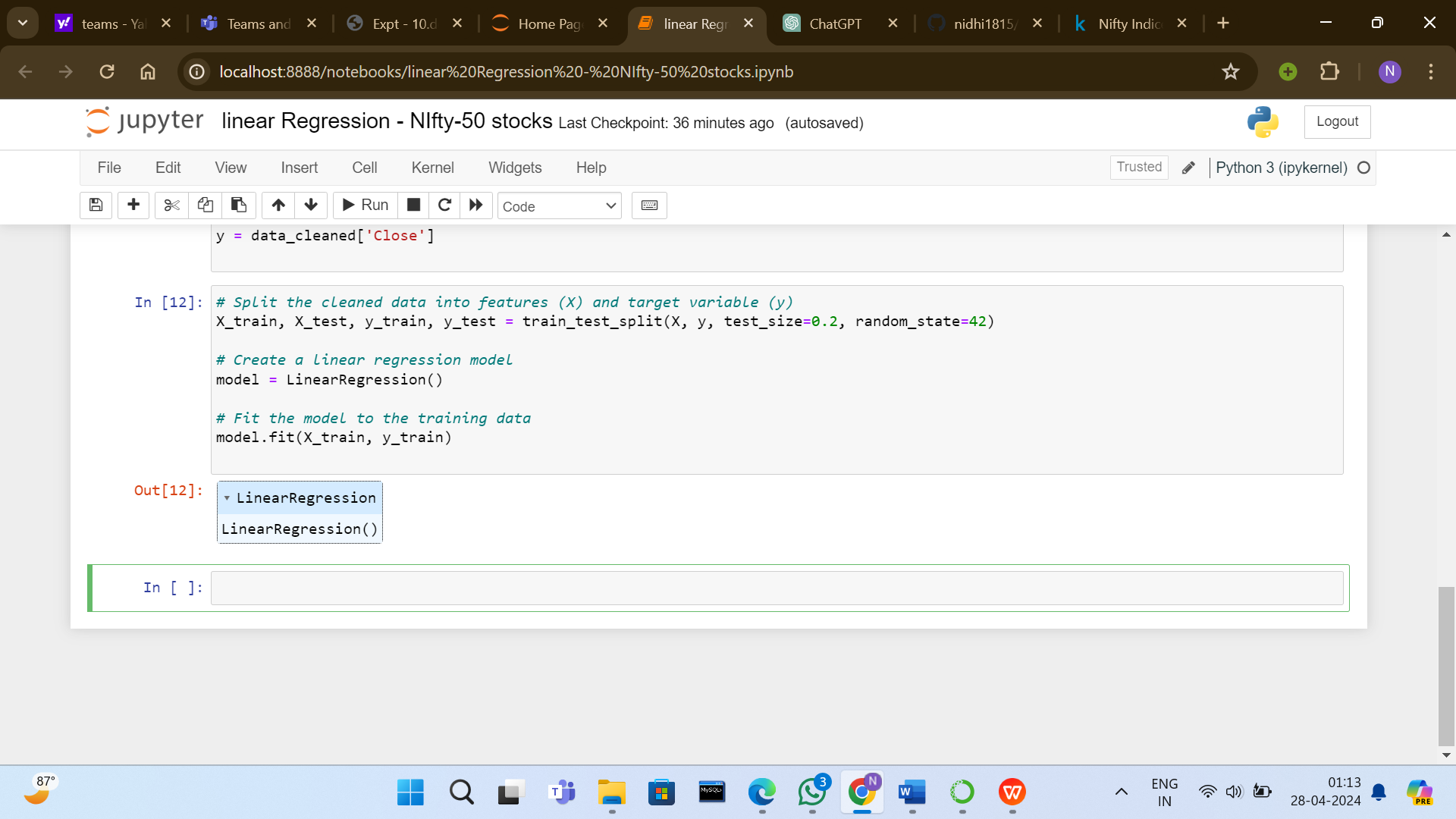
Option 1: Remove rows with non-numeric values

Option 2: Impute non-numeric values with appropriate values

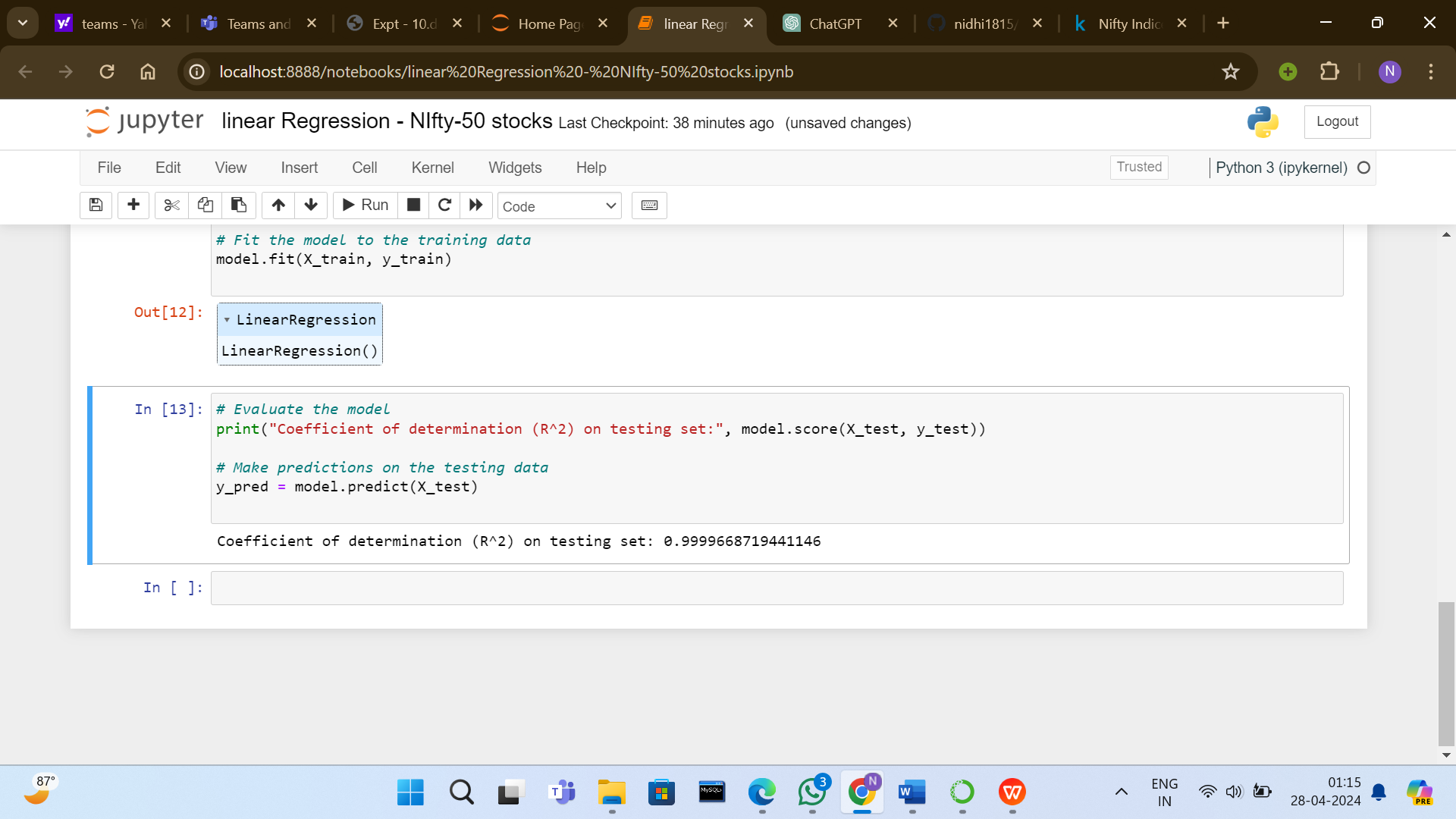
For out analysis we don’t need these non-numeric values , so we will remove them.



Step 3: Model Training



Step 4: Evaluation and Prediction

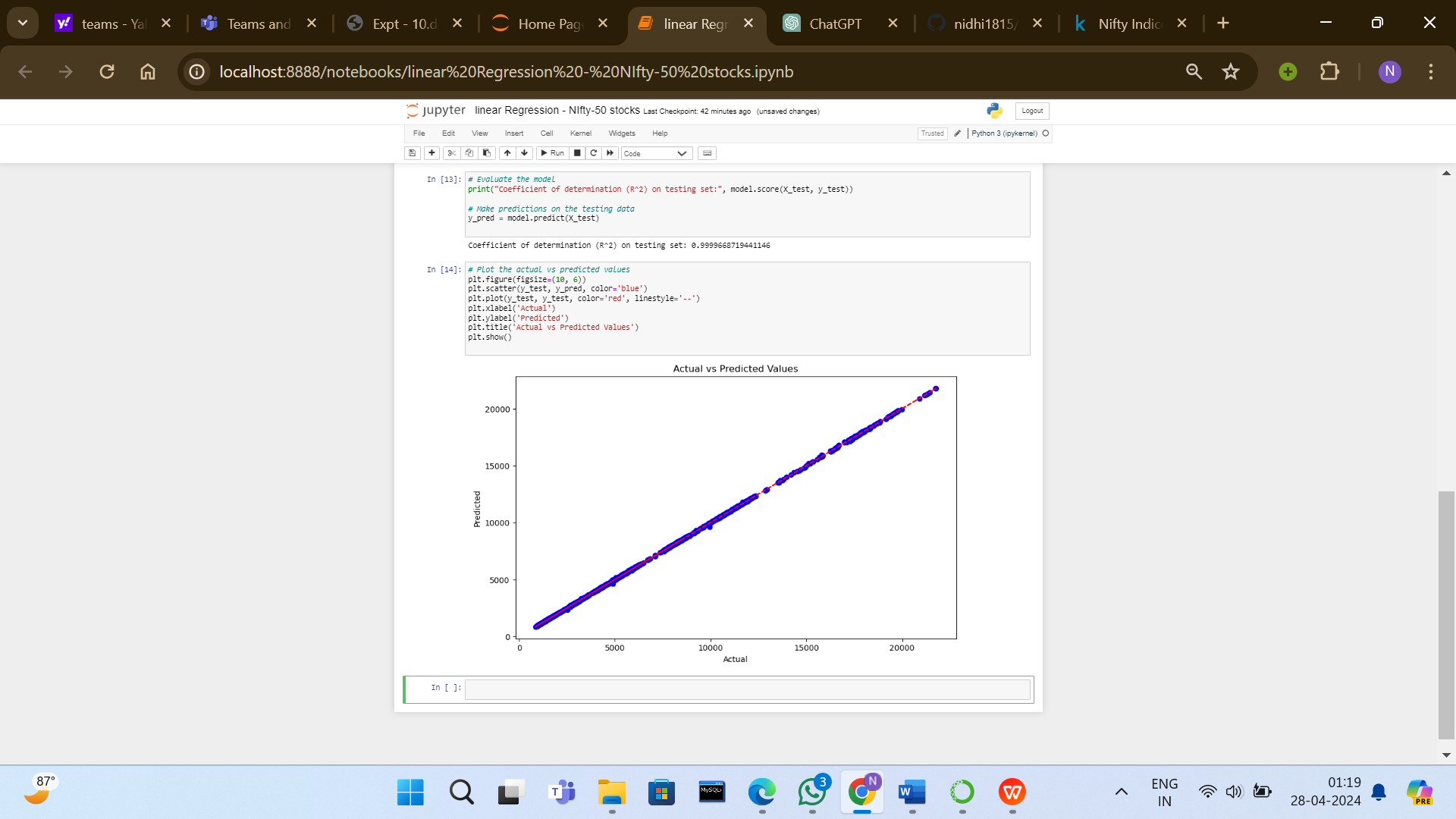


Explanation of o/p:

An R^2 value close to 1 indicates that the model explains a high percentage of the variance in the target variable (in this case, the NIFTY 50 index's stock rate). A value of 0.9999668719441146 suggests that the model fits the testing data extremely well.

With such a high R^2 value, it appears that the linear regression model performs exceptionally well on the testing data. This indicates that the features (OHLC prices, P/E ratio, P/B ratio) are highly predictive of the NIFTY 50 index's stock rate.

Step 5: Conclusion



Now , we need to check if our model can predict correct values or not , so we train our model and want to check stock price for 26-04-2024 , based on available data which is upto 02-01-2024

NOTE:-   
While linear regression can provide valuable insights into stock price movements, it's important to recognize its limitations in predicting stock prices accurately. However, stock price movements often exhibit complex patterns and nonlinear relationships that may not be captured effectively by linear regression models alone

Stock prices are influenced by a multitude of factors, including but not limited to, market sentiment, economic indicators, company performance, geopolitical events, and investor behaviour. Linear regression assumes a linear relationship between the predictor variables (such as historical prices, volume, or financial ratios) and the target variable (stock price).

