Analysis and linear regression on Microsoft stock Dataset
Project report

2

Ghadeer Alsandi & Zahra Sameer

Abstract :

Our second project in the Data Science Course with SDAIA Academy. The project was about using web scraping methods to collect more than 4229 data records from any website we choose. We decided to choose the Microsoft Stock dataset because we need to think about less and more prices and analyze it more.

Data :

The dataset used in this project was extracted from Microsoft Stock website, we used data from 2005 to 2021 and to know the least and the most in stock.

Date: Date of stock information.
Open: Opening price.
High: Highest price.
Low: Lowest price.
Close: Closing price.
Adj Close: Closing price after adjustments for all applicable splits and

dividend adjustments.
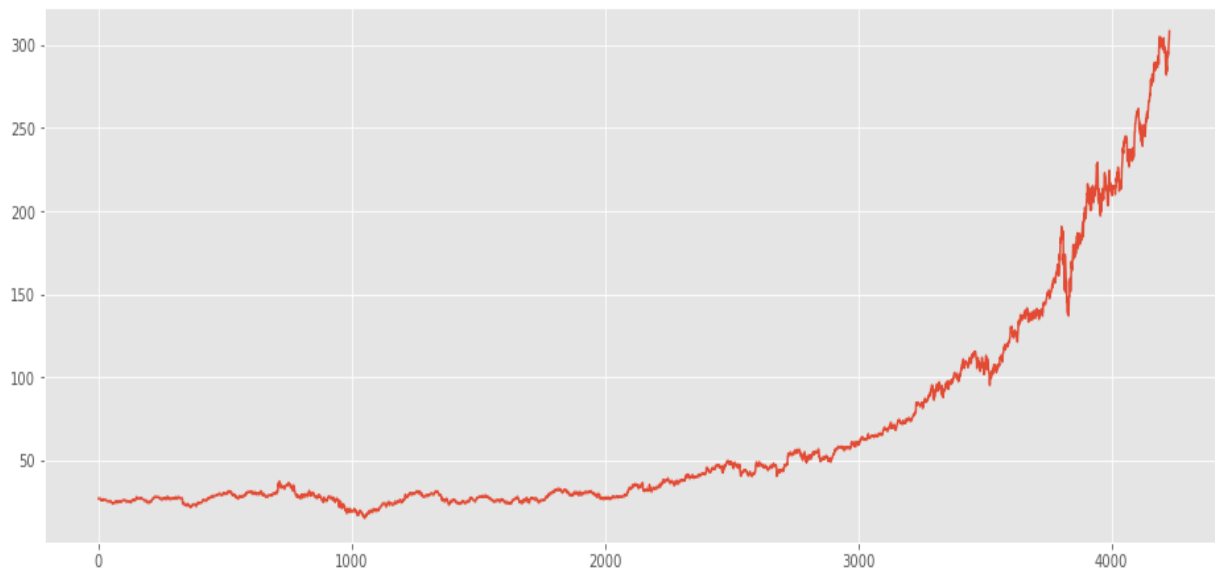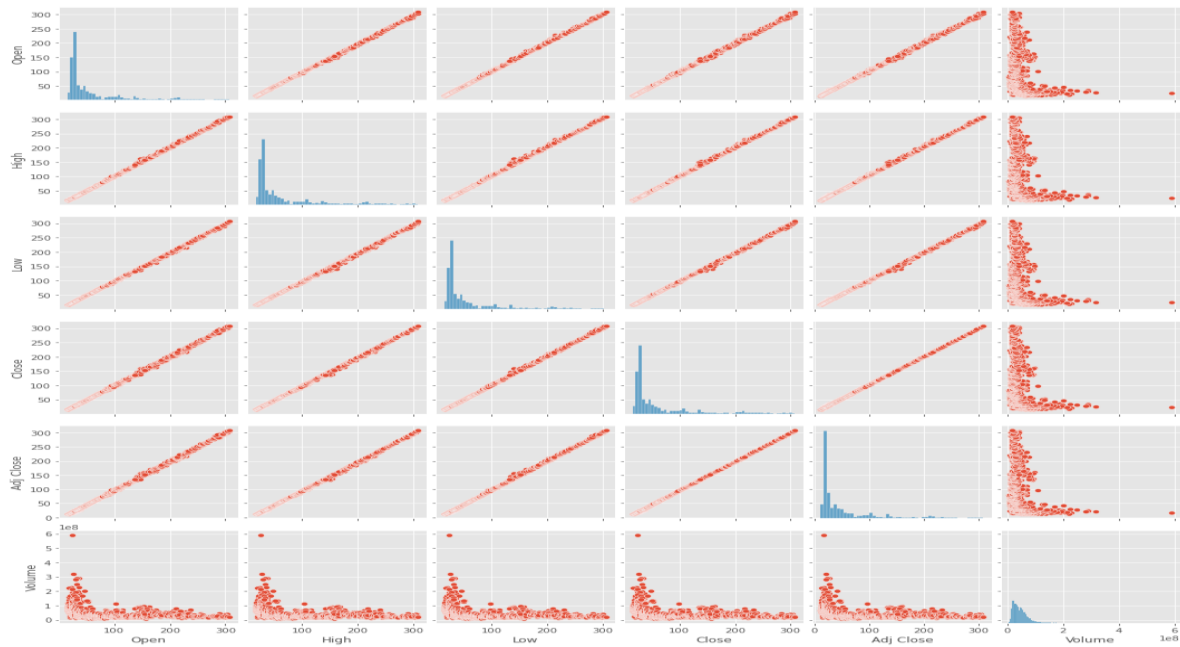**Volume:** Amount of stock.

Design :

Exploratory data Analysis (EDA) will be performed on the data we gathered
. Finally, preparing data to be used in different regression models.

Approche and methodology :

We start by, exploring, and checking for duplicated data and whit space,
using selection feature to select some data from the dataset, then exploratory
Data Analysis for all features. Finally, we used linear regression methods to
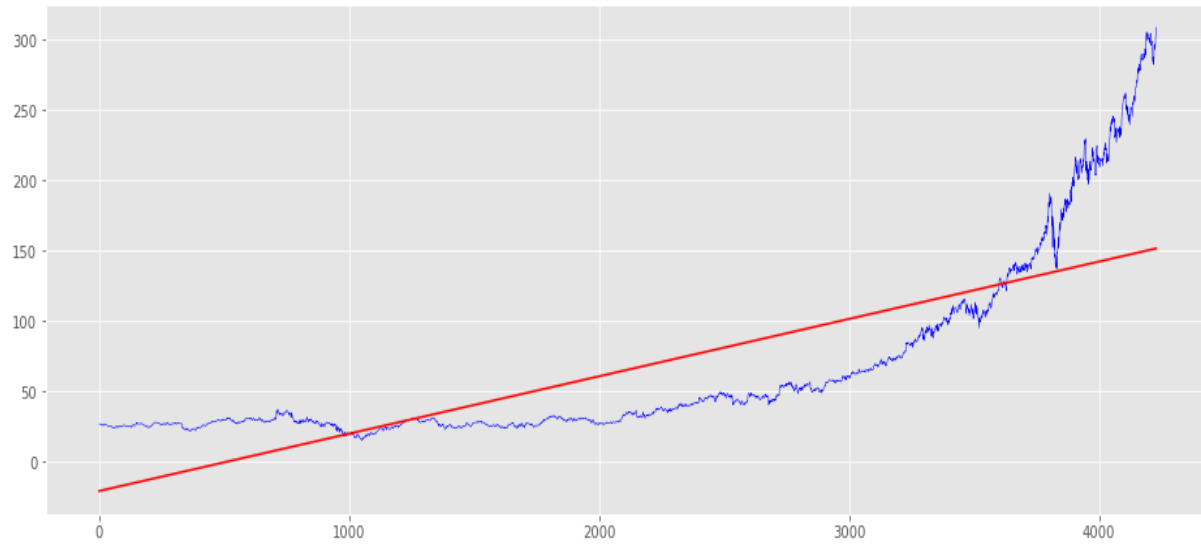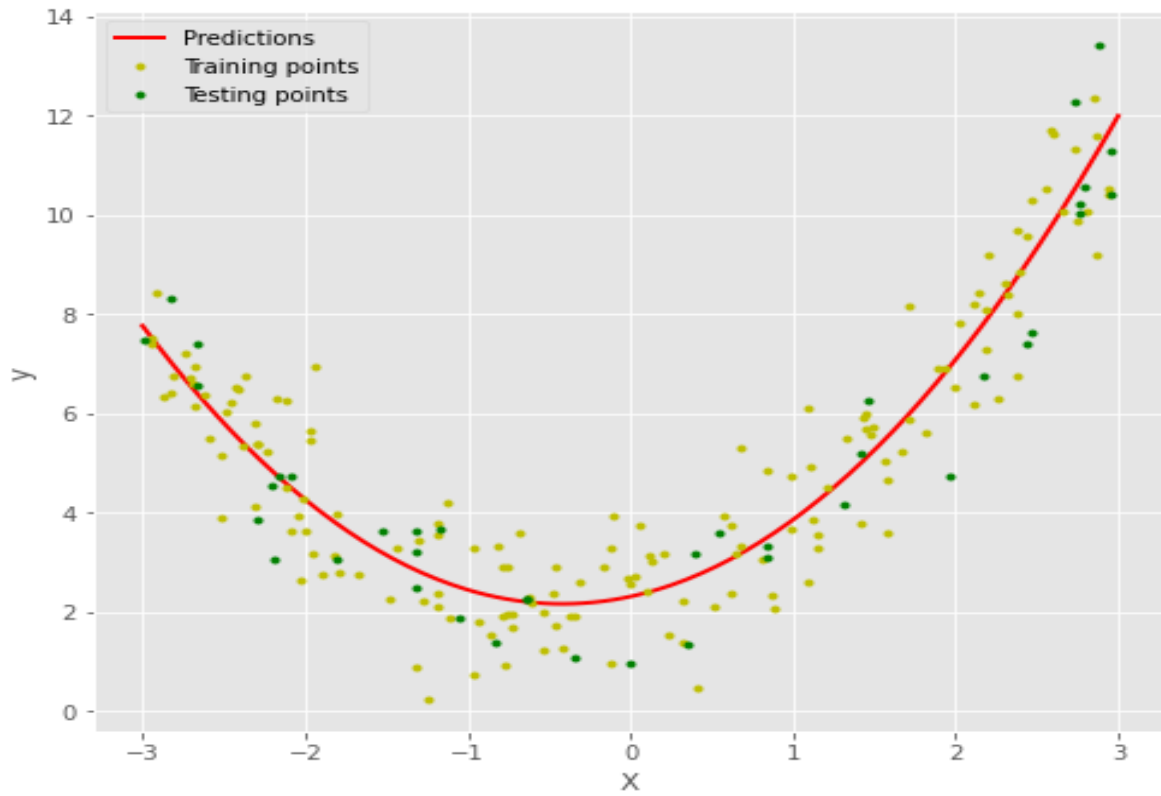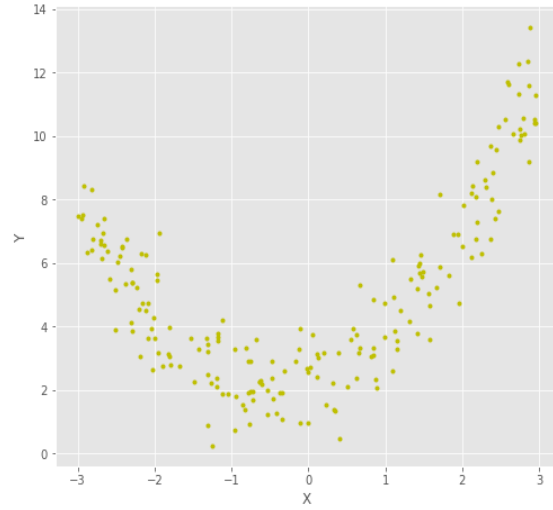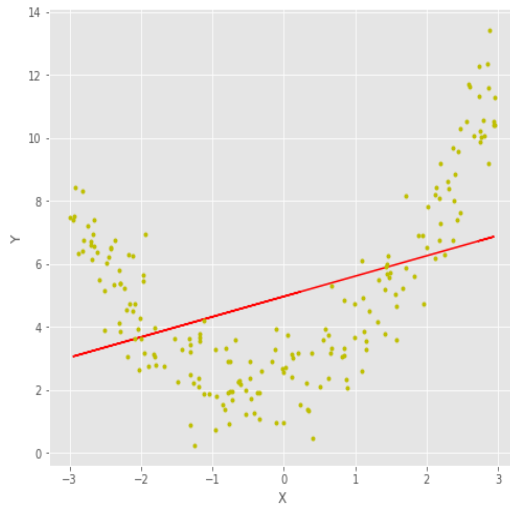do more analysis for that data.

Result :

At the end of our research we found the result in the year 2005, the stock increased and 2021, it will decrease.

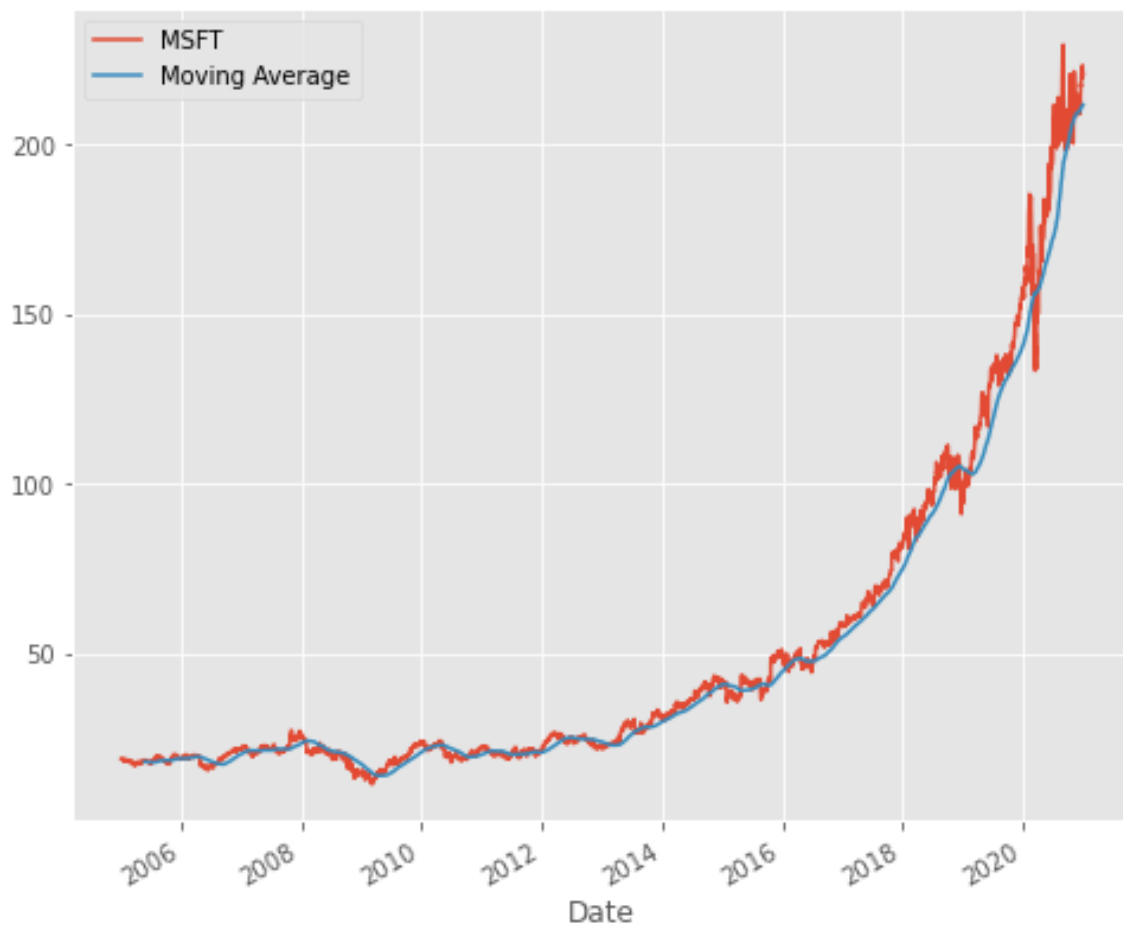$$a = \overline{y} - b\overline{x} \qquad\qquad b = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}$$

Predictions
Training points
Testing points

Tools :

Technologies : Jupyter Notebook, Python

Libraries : pandas, Numpy, Matplotib, seaborn, Request, and Sklearn