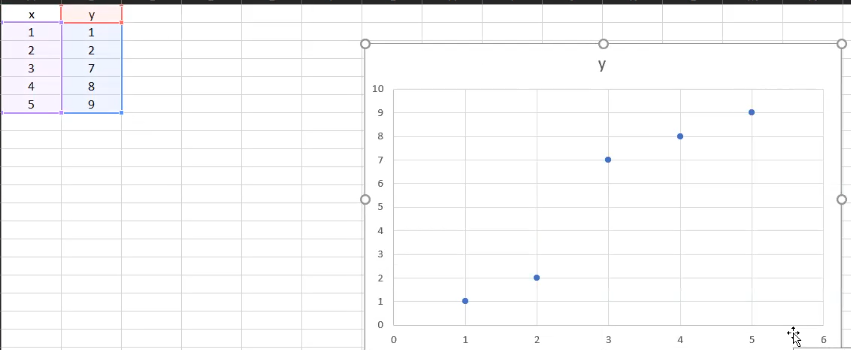
[Video Notes](https://zoom.us/rec/play/Pm6puKxkV6kIcRZlAO5iJIlYtM2IhfHhrvbKafZh1OjaZQVk5XHCFhP984sttiZvPccUJ8IdWKh-WV-j.-T8sDjBIfnq0tTXo)

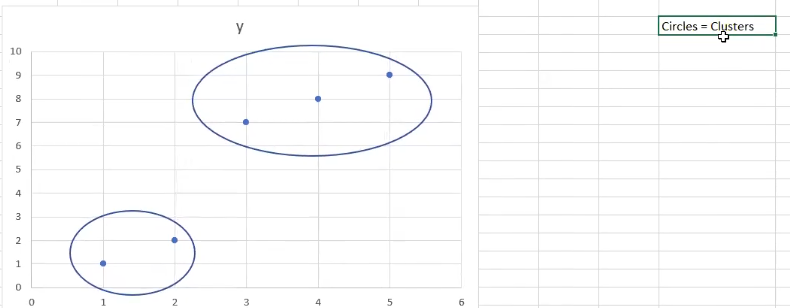
<https://docs.google.com/spreadsheets/d/1reyXq6Wb6QDGVB_vNqRwlHkUPK80Uqc57KAyjM8qv2Q/edit#gid=0>

<https://docs.google.com/document/d/1_seUca_lMrjpZeykhy5GjK4yDHE0xWspylkbMCeWg-4/edit>

[00:07:25]

* Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group are more similar to each other than to those in other groups.
* Here we have created a scatter plot based on the values of x and y as shown below:

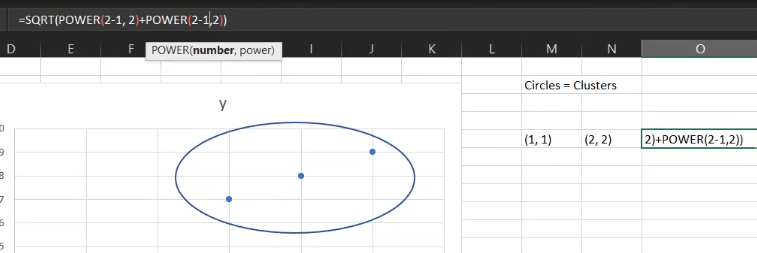




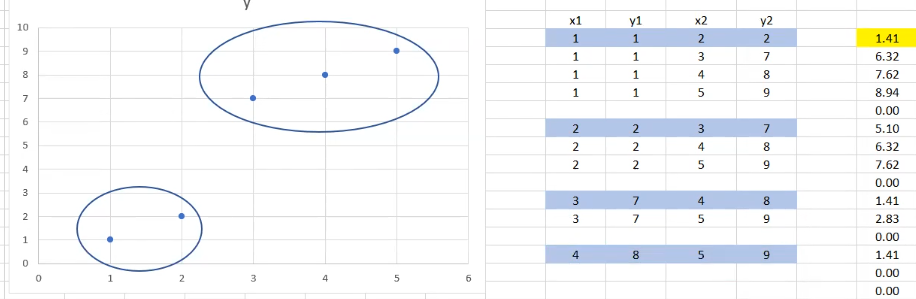
Cluster = Group of points that are “close” to each other.

[00:09:53]

Close? → Standard distance (sqrt((x2-x1)^2 + (y2-y1)^2)

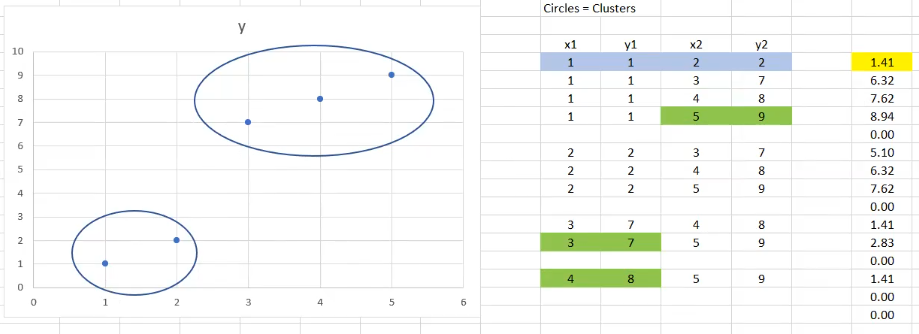






How many groups? →

Decree the number of groups = 2



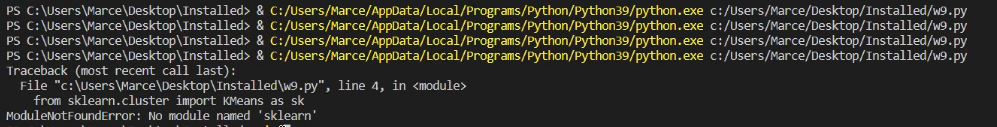
[00:19:12]

* Now we will create a python file in VS Code.

[00:21:06]

* Here we will import multiple python libraries.



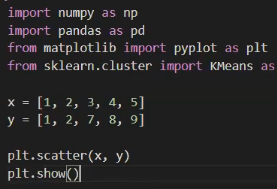


* The sklearn module is not installed so we will install it using pip install sklearn command:

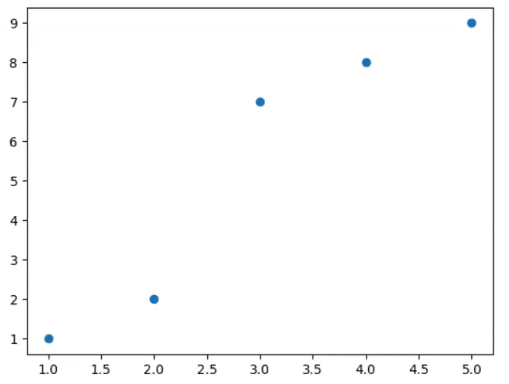


[00:23:10]

* The scatter() method in the matplotlib library is used to draw a scatter plot.
* The show() function in pyplot module of matplotlib library is used to display all figures.

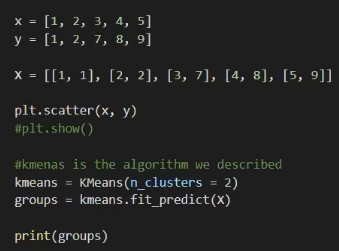


Output:



[00:25:03]

* k-means clustering is a method of vector quantization, originally from signal processing, that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster.

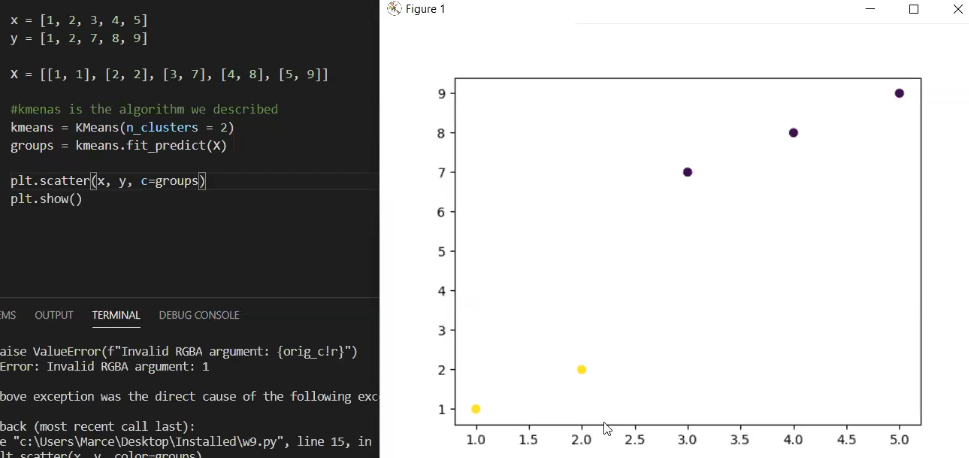


Output:



[00:29:08]

* Here we will give color to the clusters:



[00:33:34]

**Problem 1**:

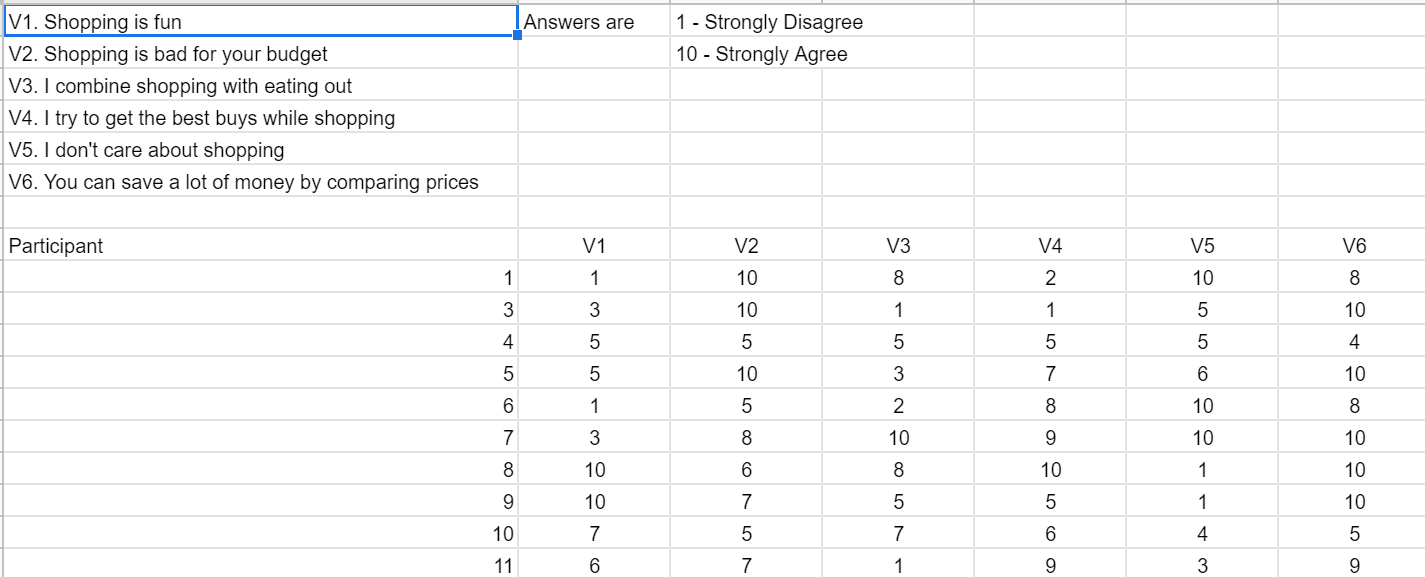
Create 3 clusters from these points.

|  |  |
| --- | --- |
| x | y |
| 1 | 8 |
| 2 | 9 |
| 3 | 8.5 |
| 4 | 1 |
| 5 | 2 |
| 6 | 1.5 |
| 7 | 12 |
| 8 | 13.5 |
| 9 | 12.5 |

Get a colored scatter plot with 3 clusters

[00:45:58]

**Problem 2**:



[00:51:36]

**Problem 3**:

Take the marketing.csv file and create a cluster that uses all the questions.

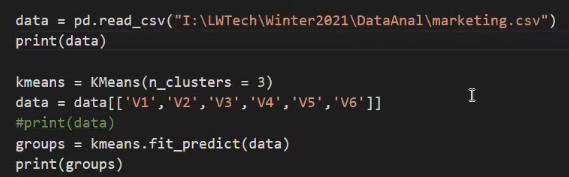
List the people in each of the three clusters

Cluster 1 - [1, 3, 4, 5…]

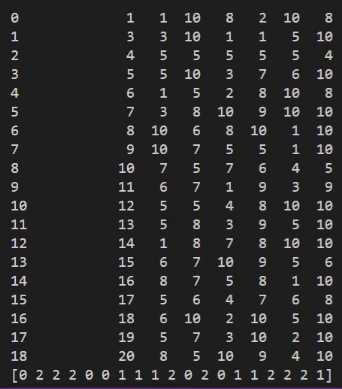
Cluster 2 - [2, 6, 9, …]

Cluster 3 - [7, 8, …]

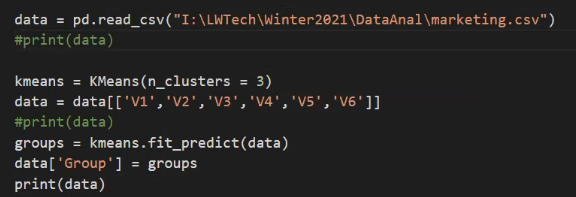
[01:24:27]



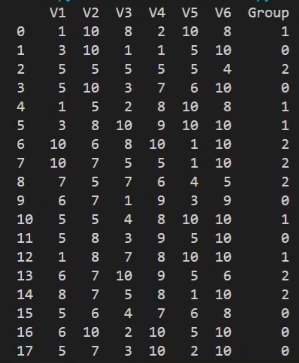
Output:



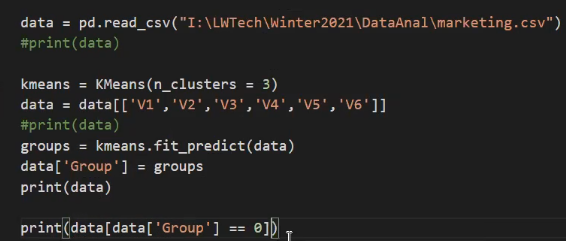
[01:25:48]



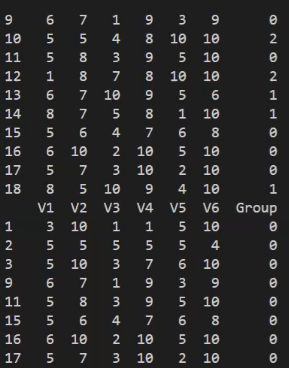
Output

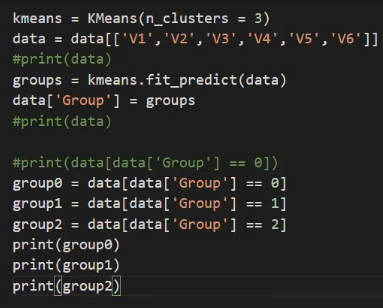


[01:30:07]



Output:



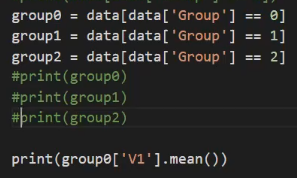


Output:



[01:32:55]

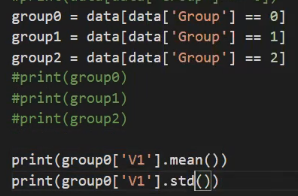
* Now we will try to figure out what these groups are.
* So we will calculate the means of the first question for the first group.



Output:



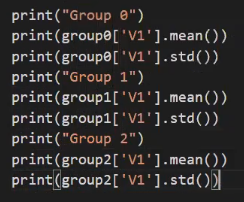
* Now we will calculate the standard deviation.



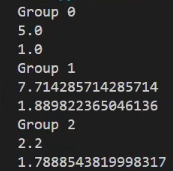
Output:



* Doing the same thing for other groups:



Output:



[01:49:38]

**Problem 4**:

* What kind of customer this is (i.e. frequent shopper, happy shopper, likely online shopper, etc.)

L Low - M Medium - H High

[02:09:05]

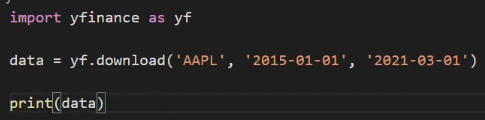
**Cluster Analysis**:

Group points together, understand the characteristics of the points, and then you try find commonality between those characteristics

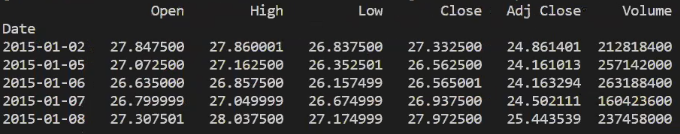
Marketing Research - Customer and Customer groups (cohorts)

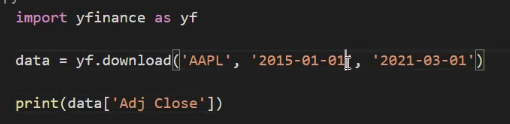
[02:14:38]

* Here we will import the yfinance package.
* If it is not installed, then use the pip install yfinance command to install it.
* YFinance was created to help the programs and users who were relying on the Yahoo Finance API.
* The following code will give the “AAPL” stock data within the mentioned range.

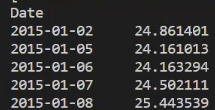


Output:

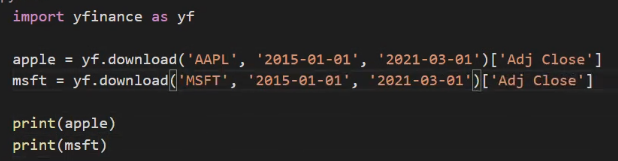




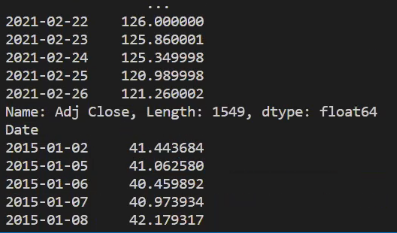
Output:



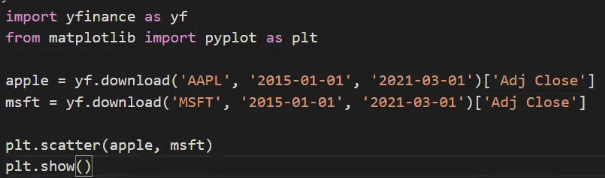
[02:16:56]



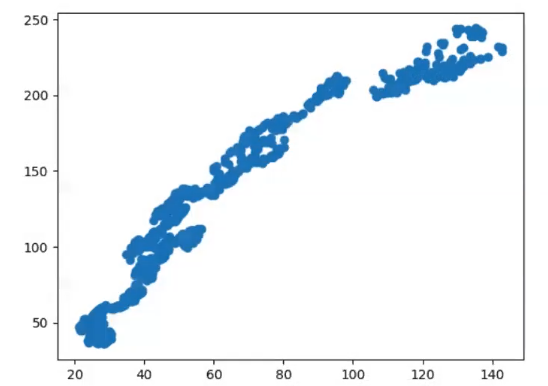
Output:



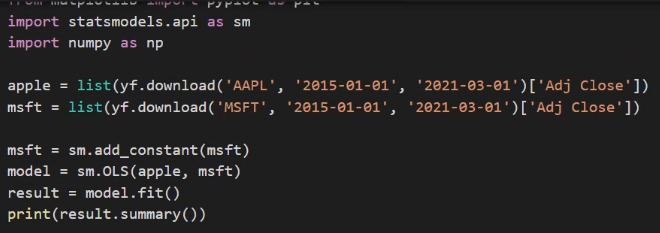
* So now we will plot a scatter plot for both the stocks together.



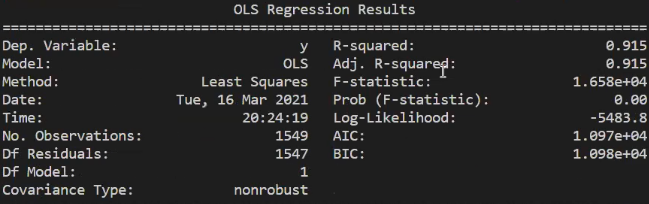
Output:



* Here we have created a model.



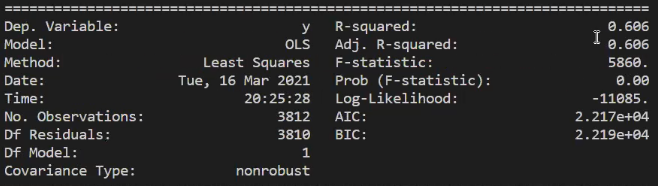
Output:



* Here the r-squared value is 91% which means the behavior of the stock of Apple and that of Microsoft is very co-related for the mentioned time frame.
* Changing the time frame:



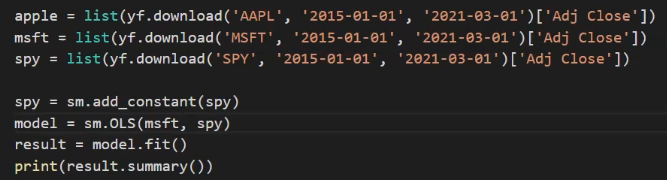
Output:



* Here the r-squared value is 60%.

[02:23:45]

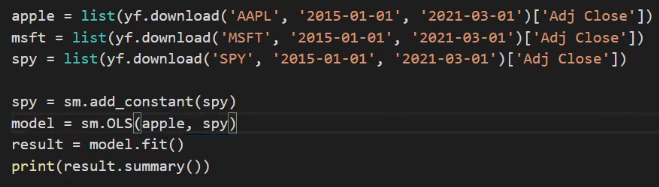
* Here we will try to predict the Microsoft stock price by the market.



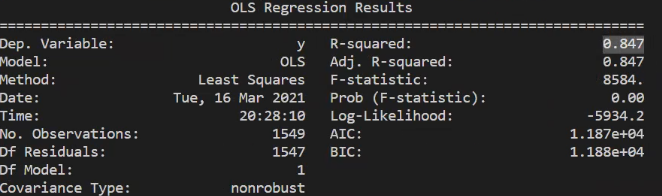
Output:



* For Apple stock:

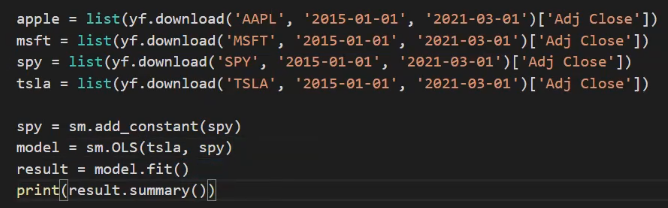


Output:

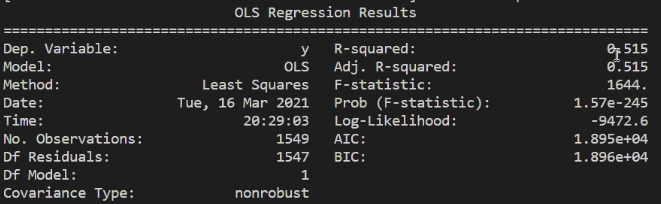


[02:26:14]

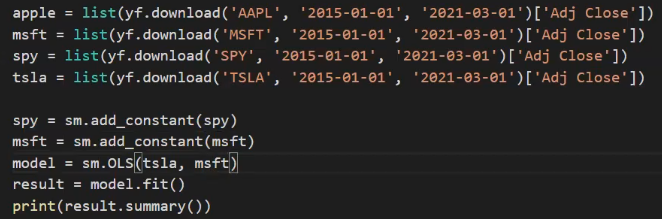
* Model the market vs Tesla stock:



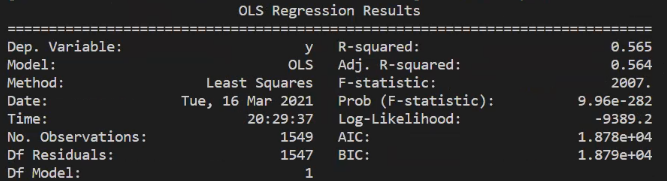
Output:



* Microsoft vs Tesla stocks:



Output:



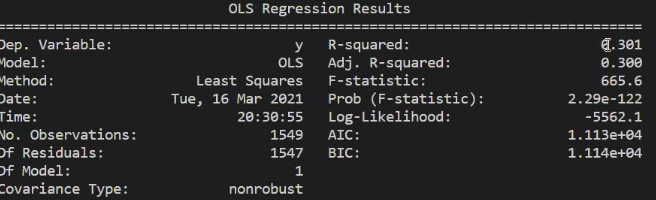
* They seem to be not co-related as the r-squared value is not high.

[02:27:30]

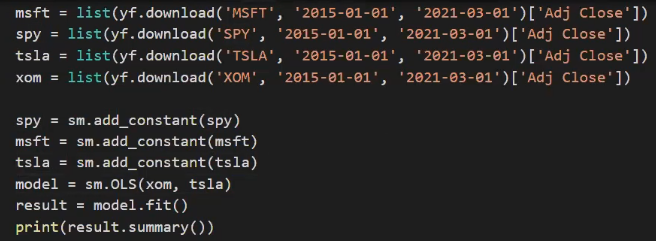
* Xom stock vs market:



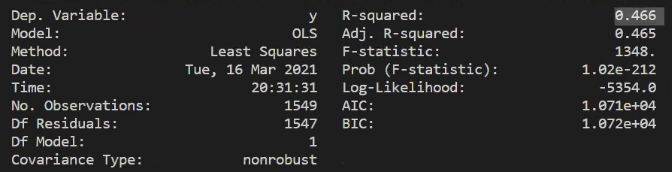
Output:



* Xom vs Tesla:



Output:



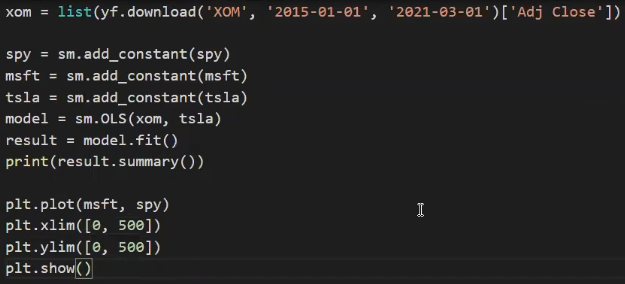
[02:29:51]

**Problem 5**:

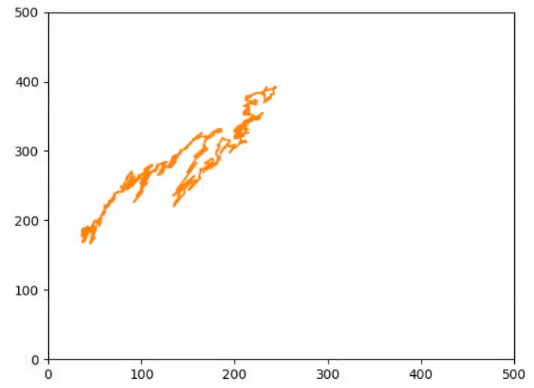
For these stocks. Group them with one of the previous ones or start a new group.

FB, CRM, CVX, AIG

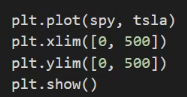
[02:43:27]



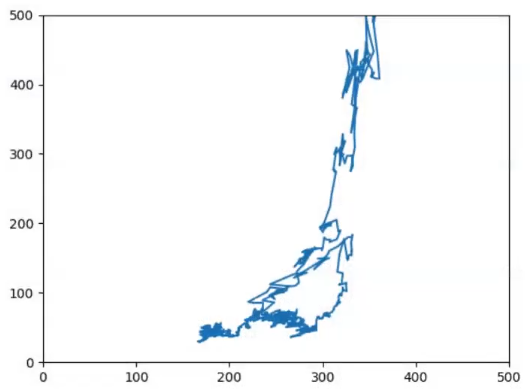
Output:



* Market vs Tesla:



Output:



* Market vs XOM:



Output:

