Readme file for Oren Shmuel Capstone project:

The project includes the following files:

1) data4.csv: include raw tweets scraped from twitter when using the scrape-data2.py file.

2) scrape-data2.py: python code to scrape tweets, from twitter, that include the following key words (economic slowdown, expensive, cost of living).

3) Jupyter notebook capstone12.ipynb: include two separate parts

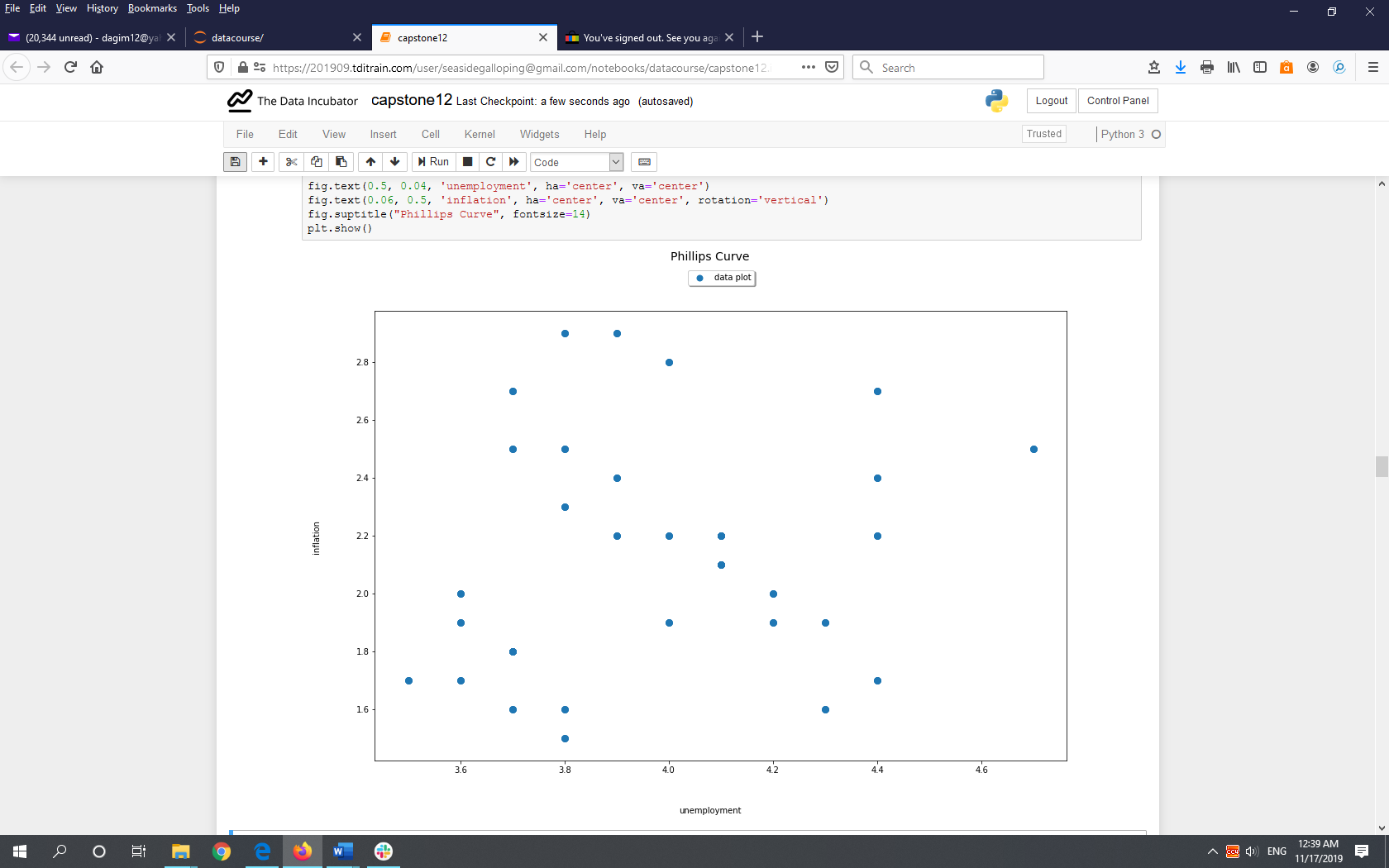
Part1: reads as input the data4.csv file, calculates the sentiment scores for each tweet, and output the sentiment scores in the file sentiments4.csv.

Part2: reads as input the file alldata41.csv file, which includes the tweets sentiments scores, unemployment rates, and inflation rates. Then, Support Vector Regressions with four different kernels (linear, RBF, polynomial, and sigmoid) are performed. Cross Validation on the training set is used to choose the SVR parameter values for each kernel. To compare the performance between the four SVRs the R^2 and MSE are calculated. A graphical comparison of the performance of the four SVRs concludes the notebook.

4) sentiments4.csv: file contains the sentiment scores of each tweet. This file generated in part 1 of the notebook code.

5) alldata41.csv: file contains the sentiment scores of each tweet, unemployment rates, and inflation rates.

The X matrix (features matrix) includes the tweets sentiment scores and unemployment rates. The target (y) is the public expectation of inflation which is used in the Philips curve, as part of the linear line intercept, to model actual inflation rates.



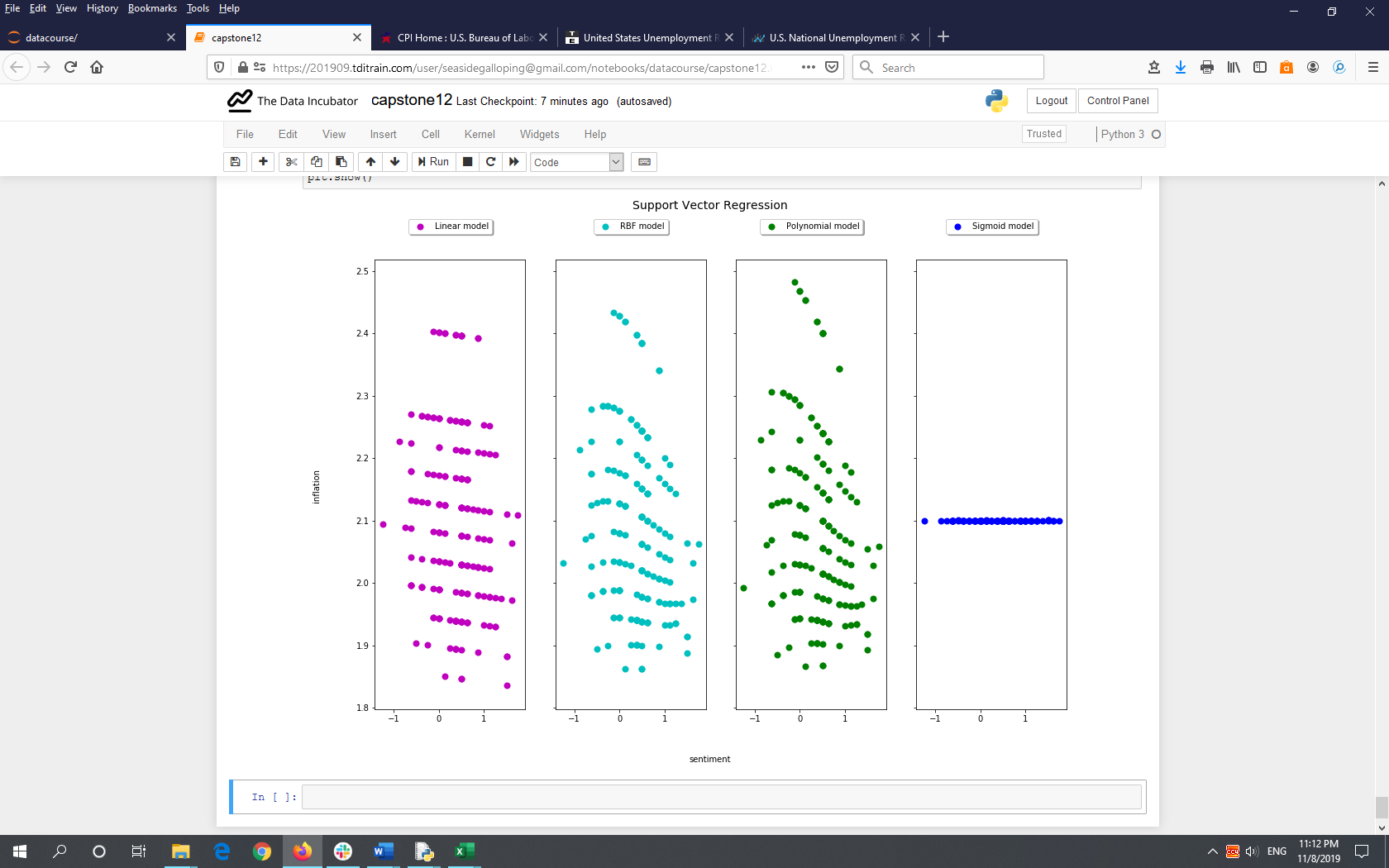
Plotting the data (above) represents a typical Phillips Curve. The two main groups of points represent two parallel, downward sloping linear lines, each represents a Phillips Curve before and after an economic policy change. The shift from the lower Phillips Curve to the upper one was caused due to a change in economic policy that caused the intercept of the Phillips Curve to increase. Hence, the new Phillips Curve is parallel to the old curve but with a higher intercept with respect to the Y axis.

The table below summaries the performance of the four SVRs:



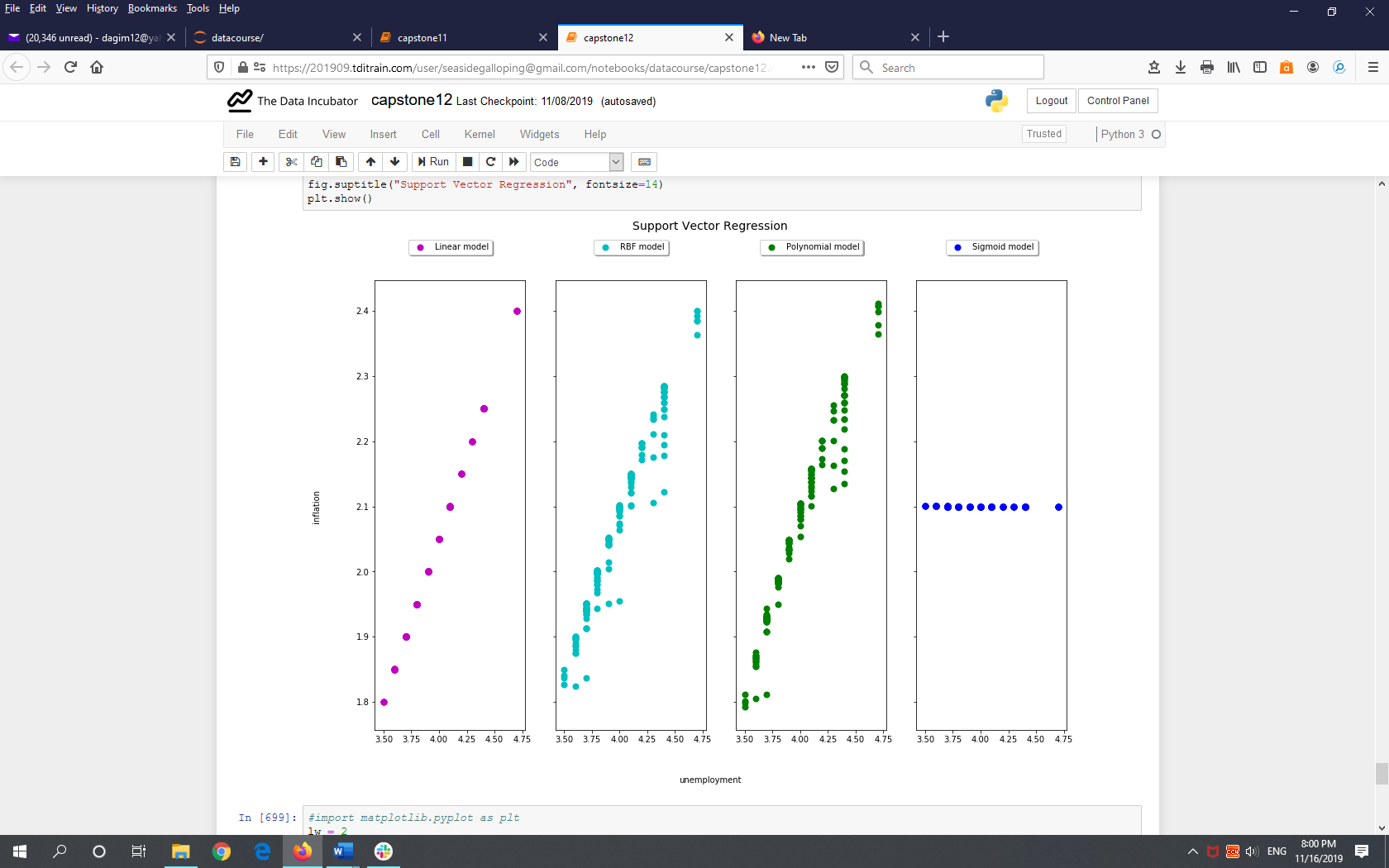
From the table the SVR with the linear kernel performs the best.

A comparison graph of the relation between tweet sentiment score and the public’s expectation of inflation is below:



For the SVR with the linear kernel one can see a downward sloping linear line that represents the relation between tweet sentiment scores and the public’s expectation of inflation. The economic logic here is that when the tweet sentiments have a more positive score, public confidence in the industry is high which causes people to believe that inflation rates will decrease, hence the negative slope.

A comparison graph of the relation between unemployment rates and the public’s expectation of inflation is below:



For the SVR with the linear kernel one can see a linear line that represents the relation between unemployment rates and public’s expectation of inflation. The economic logic here is that when unemployment rates increase, public confidence in the industry is decreases which causes people to believe that inflation rates will increase, hence the positive slope.