

Data Visualisation Project 2021-2022

Lorcan Turner : A00264376

Date: 5th January 2022

**STUDENT PLAGIARISM DISCLAIMER FORM**

**PLAGIARISM DISCLAIMER**

STUDENT NAME: \_\_LorcanTurner\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

STUDENT NUMBER: \_A00264376\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

PROGRAMME: \_\_BSC software design (game development)\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

YEAR: \_\_2022\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

MODULE: \_data Visualisation\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

LECTURER: \_\_\_David Scott\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

ASSIGNMENT TITLE: \_data visualization project \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

DUE DATE: \_\_\_5th jan\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

DATE SUBMITTED: \_5th Jan\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

ADDITIONAL INFORMATION: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

I understand that plagiarism is a serious academic offence, and that AIT deals with it according to the AIT Policy on Plagiarism.

I have read and understand the AIT Policy on Plagiarism and I agree to the requirements set out therein in relation to plagiarism and referencing. I confirm that I have referenced and acknowledged properly all sources used in preparation of this assignment. I understand that if I plagiarise, or if I assist others in doing so, that I will be subject to investigation as outlined in the AIT Policy on Plagiarism.

I understand and agree that plagiarism detection software may be used on my assignment. I declare that, except where appropriately referenced, this assignment is entirely my own work based on my personal study/or research. I further declare that I have not engaged the services of another to either assist in, or complete this assignment.



**Signed: \_\_\_lorcanturner \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Dated: \_\_\_\_\_5th jan\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

Data visualisation report

Within this report I will discuss my experience with gathering mass amounts of tweets from twitter to analyse how positive or negative a person is based on people tweeting about them. Gathering large amounts of data allow for more accurate statistics so to see an average of what people are saying. This will paint a more accurate depiction of the person they portray to the public.

The first thing I had to do was determine what packages I had to use. I had the option of text blob sentiment analysis and Vader sentiment analysis to determine if a string had positive meaning or a negative meaning. I sent in multiple items and determined that Vader had a more accurate output. For example, text blob didn’t think being left out in the rain was that big of a deal whereas Vader sentiment seen this as a catastrophe. I created a function with Vader to output how negative the result is. This was to add additional data to the web framework. This function took a list of tweets and returns two lists, one with the rating score and the other with the score converted to words ie. “negative”, “super-positive” ect

Text

Description automatically generated

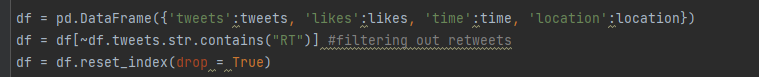
Graphical user interface, text

Description automatically generated

Graphical user interface, text

Description automatically generatedText

Description automatically generated

For this assignment I’ve used the tweepy package to gather 40 tweets about Elon musk. I then put the tweets into a data frame using pandas and filter out the retweets. This was to clean up the data for the web framework. I didn’t use these data frames in the end as I was having trouble keeping a consistent number of tweets for the two datasets, I was comparing. It was good learning but since I needed the same number of tweets between the first data set and the comparative data set I left the filtered data sets alone..

I then send each of the tweets into a csv document. An observation I had made when printing the tweets is that some tweeters put commas in their tweets, so I though it would be better putting a multi char delimiter such as \*\_\* into the document at the end of every tweet making the end of each tweet more identifiable. I then send in the rated data. Each tweet, score and rating have the same indexes. I never used the csv data but sorted it so that it could have been processed in a web framework using high chart or other packages with java script.

Chart, histogram

Description automatically generated

I then used matplotlib to draw a line graph of the tweets. Each tweet represented as one unit on the x axis and its score on the y axis. I then calculate an average and draw the line through the results in dotted format. This is useful as it allows us to take a large dataset and see if people think Elon is positive or negative. It’s a clever way to see what a mass amount of people think.

I then used the function I created for this graph to create another “comparegraph” function. This is used to draw two data sets to see which one is more positive or negative

I added a text area and submit button to the charts html page. This allows a user input. This user input goes to two places. The first place is to the graphic.py file as the user input so it will lable the legand in the graph. The second place is into the tweepy cursor query. This is to gather 40 more tweets about a person of your choice. I used trump as an example.

A picture containing graphical user interface

Description automatically generated

Graphical user interface, application

Description automatically generated

In hindsight I find the primitive line graph to look a little messy so perhaps I should have gone with a different chart but overall, I am happy to see the result. I’m glad I calculated an average as looking at the data we can now see on average people talk more negatively about Donald trump and people talk more positively about Elon musk.

Before I began this project I knew very little about python. Unlike pervious project where the code was just given to me, this project had me stuck to figure out many things on my own. I’ve learned a lot about how the language works and different functions, variables, packages, and keywords to use under different circumstances. For example

the “dir()” function came in handy for finding what functions an object/variable had within it. This allowed me to find the Vader sentiments score to determine how positive or negative a segment of text is “”

score = analyser.polarity\_scores(i).get('compound')

Text

Description automatically generatedI also had immense trouble finding what was causing the sentiment analysing function to produce two scores for one tweet when writing out to the csv. It turned out I was saving the sentiment data to global variables and continued appending data to it. This problem just required fresh eyes to solve. Another aspect of global variables in python that I was having trouble with was to intentionally overwrite global variables within a function. Unlike other languages that implicitly assume you are referring to a global variable, In python the key word global is used to explicitly state this.

in this case I was setting the user Input for the Graphs legend based on the user input.

Overall, I have gained knowledge of python that I am satisfied with. I enjoyed seeing multiple packages come together to create a useful piece of software that processes mas amounts of data to gain more accurate statistics. however, I wish I had invested more time into this to get it to a web framework. It would have been very useful as I am not very oriented in web and it would be useful to gain more skills in that area. However, processing statistics is a useful skill to use in game development (the area I am orientated in) processing user data is a very important aspect. Eg leader boards, time spent in games, available storage ect.