

Data Visualisation Project 2021

Roisin Brady, A00263209

05/01/2022

**STUDENT PLAGIARISM DISCLAIMER FORM**

**PLAGIARISM DISCLAIMER**

STUDENT NAME: \_Roisin Brady\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

STUDENT NUMBER: \_A00263209\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

PROGRAMME: \_BSc (Hons) in Software Design\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

YEAR: \_4\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

MODULE: \_Data Visualisation 4\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

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

ASSIGNMENT TITLE: \_Data Visualisation Project 2021 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

DUE DATE: \_05/01/2022\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

DATE SUBMITTED: \_05/01/2022\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

ADDITIONAL INFORMATION: \_Group project done alongside Tara Da Silva \_\_\_\_\_\_\_\_\_\_\_\_\_\_

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: \_\_Roisin Brady\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Dated: \_\_05/01/2022\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

Table of Contents

[Introduction 4](#_Toc92315420)

[System Overview 4](#_Toc92315421)

[Architecture 4](#_Toc92315422)

[Design Decisions 4](#_Toc92315423)

[Topic Selection and Presupposition 5](#_Toc92315424)

[Post-process Analysis 6](#_Toc92315425)

[Conclusion 6](#_Toc92315426)

[What you would do differently 6](#_Toc92315427)

[Future work 6](#_Toc92315428)

# **Introduction**

This report will be covering the specifications and work process of our semester 1 *Data Visualisation Project*. This will include a system overview covering the project architecture and thought process behind the system’s design. We’ll then discuss the process behind the chosen topics for this project. Following that will be a post-process analysis and finally a conclusion detailing the authors plans going forward.

# **System Overview**

For this project, we were tasked with creating an application that obtained a critical mass of tweets, applying sentiment analysis when finished. This sentiment analysis would be redone with a different function (results being compared to initial analysis) and the results would be visualised using a third-party extension. We would then need to convert this application to a web framework, providing not only a drill-down of user demographic information, but also allow users to two live topics to compare data on based on their own chosen topics.

### Architecture

This project started as a basic python project. Over time we implemented various third-party extensions to implement the various functions this project required. The first extension we implement was *tweepy*, which was used to gather the tweets needed for analysis by connecting to Twitter’s API through a developer account we created. These gathered tweets would then be run through a *textblob* function that applied sentiment analysis to each obtained tweet, judging whether each tweet was positive, neutral or negative in tone. These would then be printed to a .csv file which will be used for further analysis.

We would then take this .csv file and run it through another form of sentiment analysis, this time using an extension called VADER which works for more casual language and the scope of this project as we will discuss a bit later. We would also use this to do a drill-down on information on the users we gathered tweets from, such as their location and age.

Finally we had to visualise the data and convert our app to run on a web framework. For the data visualisation, we started out using matplotlib to convert our data to a visual format. This format was then displayed on our webpage using JavaScript and two other extensions called Highcharts and Flask. Highcharts is used to display information charts in a visual format while Flask allows python programs to run on a webpage. This functionality would be furthered extended with JavaScript by allowing users to do their own searches to gather tweets for sentiment analysis, the results of which would then be displayed on the page.

### Design Decisions

Much of the functionality driving this project was split into separate python files (one for gathering tweets, one for sentiment analysis and so on). This was done to extend functionality so that different functions could be used for different parts of the program and to keep the code simple by keeping it in distinct, easy to read files.

For this project two different types of sentiment analysis were used: textblob and VADER. The reason for this was to ensure that our gathered tweets were properly analysed and assigned a sentiment value accordingly. While textblob by itself is usually quite accurate, it’s best suit for instances of formal language and does not handle idiosyncrasies such as slang or use of emojis. VADER on the other hand is designed to take these quirks into account, and so by using both we ensure that our gathered tweets are given a proper and thorough sentiment analysis.

Interestingly, the results from our sentiment analysis showed that most of our gathered tweets were neutral to positive in tone. Granted, this could be the result of many things, such as a small data pool, context missed by one/both of the analysers or the nature of the tweets themselves. This result is interesting to note mainly because of the controversial nature surrounding the two figures we chose for analysis, though given Johnny Depp’s status as a Hollywood classic with a still-dedicated following, this may not be as surprising as expected.

The data obtained from the process of sentiment analysis was represented using bar charts created using an extension called *matplotlib*. This extension was used because it was easy to use and a quick way of visualising information. Pandas, another python extension, was used to manipulate our data and convert it to chart format. Bar charts were used to represent the data because it is a simple visual guide for comparing statistics.

For this project the website visualisation was done with a combination of python code and JavaScript. This was done to ensure that the charts would display on the webpage when the user clicked into the page. The JavaScript also allowed for user inputs, which was needed so that users could enter a topic for either single or dual-topic analysis. The python end was handled using an extension called Flask which allowed us to easily convert our code to run on a web framework.

# **Topic Selection and Presupposition**

Before beginning the project, there were a few elements that needed to be decided on in advance. One such factor that needed to be planned was the selected topic for the project. For this project we could pick any topic to collect and analyse tweets from. For this project we chose Johnny Depp and Amber Heard as our topics of analysis because we thought the divided public opinion on them based on their split and the controversies surrounding them made both people interesting figures to analyse tweets from.

Another element of the project that had to be decided on beforehand was what packages were to be used for this project. We would need at least two packages that included sentiment analysis, one for visualising raw data, and one for converting python code to a web framework. For our project, we used VADER and textblob for sentiment analysis at different stages of the project, matplotlib and Highcharts for visualising our gathered data, and finally Flask, Highcharts and JavaScript to convert our application to a web framework.

# **Post-process Analysis**

On paper, this process of creating a project seemed simple and easy to follow, with plenty of information and resources available. In practice, this project turned out to be a bit more complicated than expected, mostly due to the process of visualising that data and then implementing the application into a website. Compounding this was the fact that this was my first major project created in python.

As a result, there were some instances where issues occurred, primarily in the form of syntax/programming errors late in the project’s development during the process of visualisation and web conversion. Thankfully these issues were eventually corrected through debugging, testing and re-checking of our code to gain basic functionality on all project elements. There were some minor hitches at the start of the project when we were first figuring out how to set up our project. Thankfully we had enough time and resources to develop a plan on what work needed to be done and when.

# **Conclusion**

*Overall, I think this was an interesting and fun project to work on.* This project was a nice opportunity to work on a group project again, something that I hadn’t done in a while. As I will discuss in a bit, I also feel like I’ve learned a lot from this project and will take what I’ve learned with me going forward. That being said, there are some things that I think I’d approach differently if I were to do this type of project again.

### What you would do differently

One thing that I would’ve done differently if I were to redo this project was to include a translation component to support multiple languages. The final project only gathers tweets written in English, which limits our pool of data and adds unnecessary bias by way of excluding tweets from non-English speakers. Adding a method for translating tweets in different languages to English for sentiment analysis would eliminate these issues and allow for a more objective data analysis along with a wider data pool.

Another thing I would do differently would to put more research/focus into the visual and web framework sections of the project. As this was my first project working with Highcharts and Flask, I think in retrospect it would’ve been ideal if I’d done more in-depth research into how these applications work before trying to implement them into our project, which may have also prevented certain issues with the final product.

### Future work

I’d like to take everything I’ve learned from working on this project and apply it to future projects. It was worth doing this project as it gave me some experience and practice coding in *python*, which will certainly be useful given the popularity of python as a coding language. This project has shown me how easy programming in python can be and has not only provided me a chance to broaden my programming skills but has also given me another array of resources I can use for projects in the future.

I think especially the web framework aspect of this project has given me what is potentially the most useful information that I can apply for future projects. Framework tools like Flask or Highcharts allow for python programs to be extended to a web framework, giving python programs extra reach and functionality. This is useful to know for programming, especially in the field of web design and app development, and something I’ll keep in mind as a way of executing/developing future web-based projects.

I’ve also learned a lot about implementing extensions and interpreters into python projects. This is useful know-how because these interpreters provide and extend useful functionality for python projects, such as data visualisation and the ability to implement python programs into HTML webpages. This broadens how python can be used and is very useful for software engineers to know and have under their belt.