Project Report

# GitHub URL

(insert URL here)

# Abstract

I will be analysing financial data – specifically stock prices and volumes over time. I will look to include multiple data sets and/or sources and compare. I will look to manipulate these sets to look for trends.

# Introduction

(Explain why you chose this project use case)

I used my own company **Flutter Entertainment** and will analyse the company’s stock price over a historical period. The historical data was retrieved from the Yahoo finance website.

The stock price will be analysed against two of the worlds largest stock market index’s – the FTSE 100 (UK) and the S&P 500 (US).

What happens when the volume of this stock trading increases/decreases in a short and long period of time? What happens when there is a discrepancy between the adjusted close and the next day's opening price?

# Dataset

(Provide a description of your dataset and source. Also justify why you chose this source)

I wanted non-complex, large dataset with as few null or incorrect outliers as possible – the dataset/sources needed to show stock prices and volumes of shares (sold) over time.

There would have been a several financial websites I could have retrieved this open-source information from but went with **yahoo finance** for my first FLTR dataset as that was a popular and reliable site from my reading on Kaggle.

# Implementation Process

(Describe your entire process in detail)

I set up my version control of this project by setting up a new repository on git.

I first checked (on Mac terminal) git was installed and what version

I needed to read in and manipulate my initial dataset FLTR.L.csv.

Text

Description automatically generated\

I wanted to insure I had the latest version so I checkd on the Git website and needed to install **Homebrew** and

The data could have been read in with either pandas or numpy librarys. The dataset contained a ‘Date’ column and then the remaining columns all featured numerical data - integers or floats.

I decided to initially read the data in with pandas that deals best with different data types and I could easily set the ‘Date’ column to be the index in the dataframe.

I used my own company **Fluter Entertainment** and wanted to get all the historical price data of the London stock listing – symbol FLTR.

To do this I went to Yahoo finance and typed in the ticker symbol FLTR.L (.l = London listing). Then the historical data tab and did the time period from 01-01-2018 to 01-03-2023.

The frequency was *daily* and this.*csv* file was downloaded.

Because the data is primarily all numbers, I will use the **NumPy** package.

Graphical user interface

Description automatically generated with low confidence

I will also collate this data for the FTSE 100 and S&P 500 for the same time period.

**Anaconda/Jupiter**

Graphical user interface, text, application

Description automatically generated

Before using NumPy the *csv* file needs to be read into the

# Results

(Include the charts and describe them)

# Insights

(Point out at least 5 insights in bullet points)

# References

(Include any references if required)

<https://uk.finance.yahoo.com/quote/FLTR.L?p=FLTR.L&.tsrc=fin-srch>

1. Create histograms to visualize the distribution of data.
2. Calculate descriptive statistics such as mean, median and standard deviation.
3. Compute correlation coefficients to determine the degree of relationship between variables.
4. Use linear regression to model the relationship between variables.
5. Calculate the Fourier transform of signals to analyze periodic patterns.
6. Create scatter plots to visualize the relationship between two variables.
7. Perform hypothesis testing to determine if two samples are statistically different.
8. Use gradient descent to optimize a function.
9. Use principal component analysis to reduce the dimensionality of data.
10. Apply machine learning algorithms such as k-means clustering or decision trees to classify or predict data.