

|  |
| --- |
| DATA VISUALISATION |
| *B9DA106*  *Continuous Assessment - Two* |
| April 28  DUBLIN BUSINESS SCHOOL  Module Guide: Basel Magableh |



TEAM\_SVM:

ROHAN VEDPATHAK (10392547)

ROHAN NAGALKAR (10403259)

ABHISHEK BERAD (10504842)

SHANTANU JADHAV (10393294)

Data Selection

For the exploration of interactive visualisation, we have selected the New York stock exchange dataset which can be downloaded by [this](https://www.kaggle.com/dgawlik/nyse) link.

As said by (Gawlik, 2017) this dataset consists of following files:

* **prices.csv:** raw, as-is daily prices. Most of data spans from 2010 to the end 2016, for companies new on stock market date range is shorter. There have been approx. 140 stock splits in that time, this set doesn't account for that.
* **prices-split-adjusted.csv:** same as prices, but there have been added adjustments for splits.
* **securities.csv:** general description of each company with division on sectors
* **fundamentals.csv:** metrics extracted from annual SEC 10K fillings (2012-2016), should be enough to derive most of popular fundamental indicators.

Why Stock Exchange dataset?

Everyone wants their life to be simple, and money should come to them quickly. To have a luxurious life one tends to find ways to find the quickest and shortest way to achieve it. Currently, a lot of development is done on stock markets. Though the nature of stock is unpredictable and unstable a lot of research has been done over the years. Many findings, statistical indicators, a lot of fundamental techniques have been suggested to predict stock. All the studies were previously just based on mere a guess. But now things have changed, a lot of efforts must be put by many quantitative analysts to derive predictions. In our assignment, we have not only done data visualisation but also predicted the stock price based on the given dataset.

The goal of this assignment is not only to gain familiarity implementing interaction techniques for visualisation but also show that by applying linear regression we can even predict the future price of the given stock.

The dataset consists of columns like date, symbol, open, close, low, high and volume

Prices were fetched from Yahoo Finance, fundamentals are from Nasdaq Financials, extended by some fields from EDGAR SEC databases.

Applications used

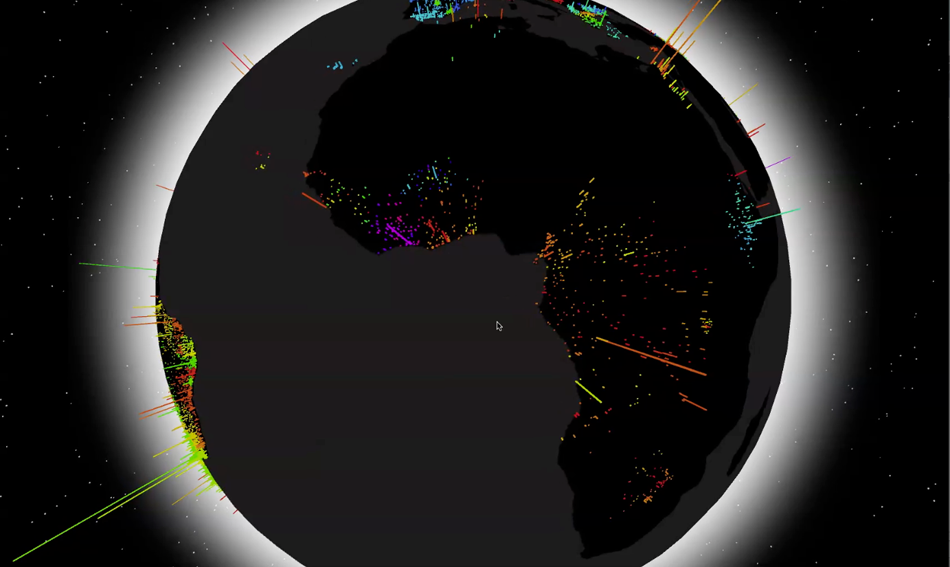
For this assignment we have used *Flask* which is a microservice python framework. Flask is considered more Pythonic than the Django or any other web framework in python because in common situations the equivalent Flask web application is more explicit.

We have used python’s dash library throughout our development. It is an open source Python library for creating reactive, Web-based applications.

*Dash* is a user interface library for creating analytical web applications. Those who use Python for data analysis, data exploration, visualisation, modelling, instrument control, and reporting will find Dash as an excellent option to start with.

We have also created an interactive dashboard using *Tableau*.

What is Interactive Data Visualisation?

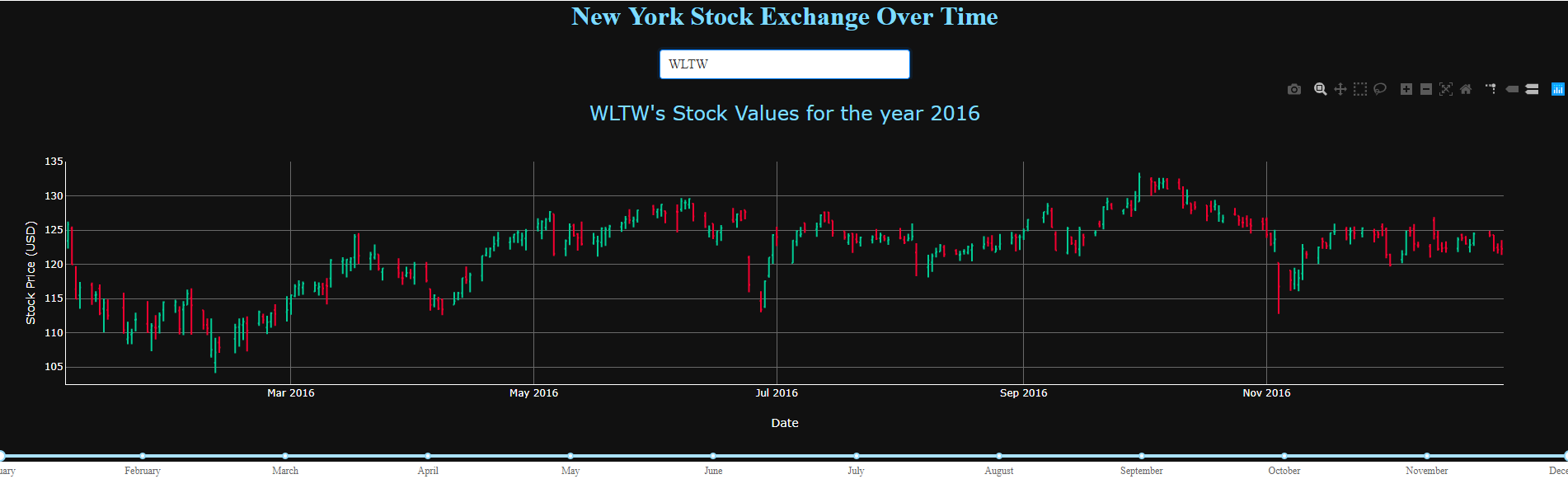


When it comes to business and big data, more isn’t always better. This is because raw data, when captured and stored in data silos, isn’t valuable unless it’s being broken down, evaluated, and acted upon. It’s been suggested that as little as 1% of the world’s data is effectively analyzed, with the rest either being captured and held indefinitely or never captured at all. And as the amount of data being created on a daily basis grows from year to year, ever more valuable information is slipping through the cracks. However, perhaps even more troublesome is when data is properly analyzed, and nonetheless still goes to waste. Bad data, which is to say data that is either inaccurate or incorrectly analyzed, has been shown to cost businesses as much as 25% of their revenue in certain situations. But it doesn’t have to be this way. Often, getting the most out of analyzed data is simply a matter of understanding data visualization.

Data Visualization

The first interactive visualisation is of candlestick plot, and it is a style of financial chart used to describe price movements of a security, derivative, or currency. It is like a combination of line-chart and a bar-chart: each bar represents all four important pieces of information for that day.

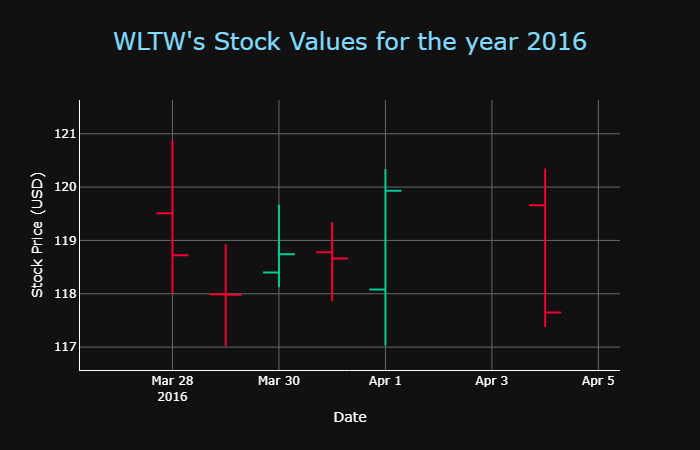
We have used candlestick plot because it will show us the high, low, open and close price of the stock. Also, by selecting the symbol from the dropdown, we will be displaying graph for that particular stock.



Features for this graph are:

* Zooming
* Paning
* Box Selection
* Lasso Selection
* Auto Scaling
* Toggle Spike Lines
* Taking a screenshot of graph.
* We have also provided RangeSlider which will help user to view the stock price for the range of dates he wants.

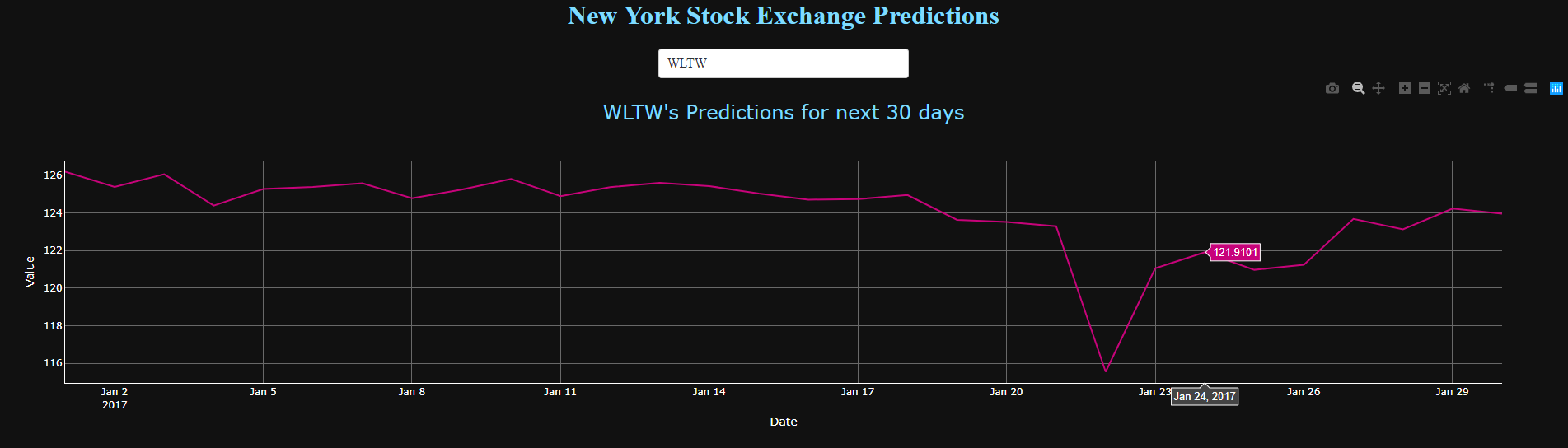
Below is the screenshot of above graph after using Box Selection.



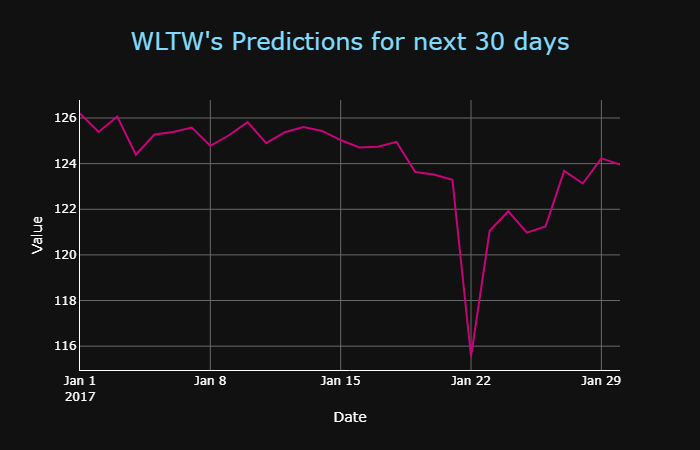
As we can see in many stock market applications, the charts are displayed in the black background and green-red combination of colors is used to depict increase and decrease of stock price respectively. We have maintained same color combinations here.

In the second graph, we have made the prediction of the stock price based on the data. We have done a prediction using *linear regression*.

By using the dropdown feature provided, user can select stock of which he wants to see the predicted rates for next 30 days.



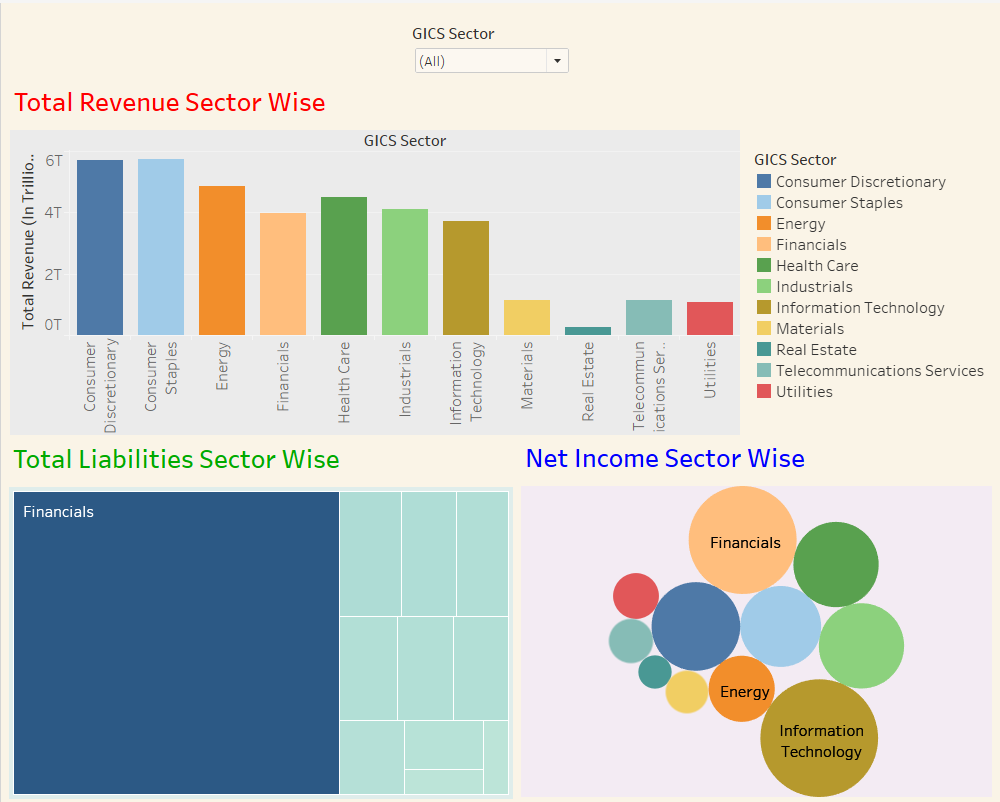
For the clearer picture of the graph, you can see the below image.



Features for this graph are:

Zooming, Paning, Box Selection, Lasso Selection, Auto Scaling, Toggle Spike Lines and Taking a screenshot of graph.

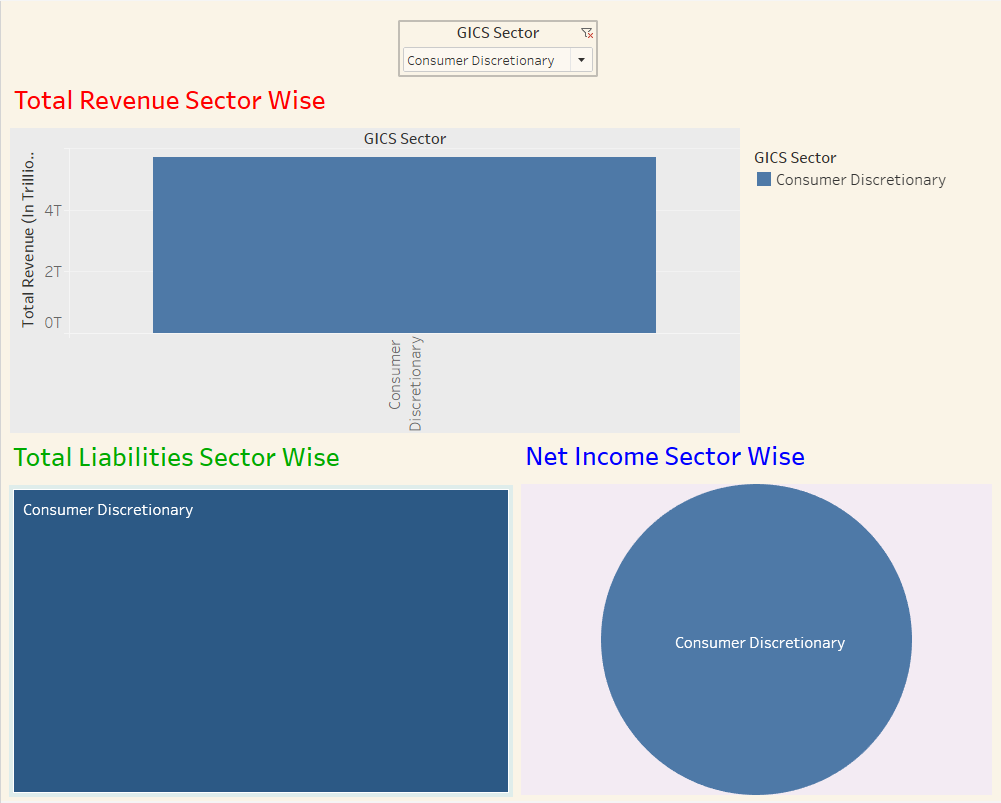
The third visualization is an interactive dashboard created on Tableau.



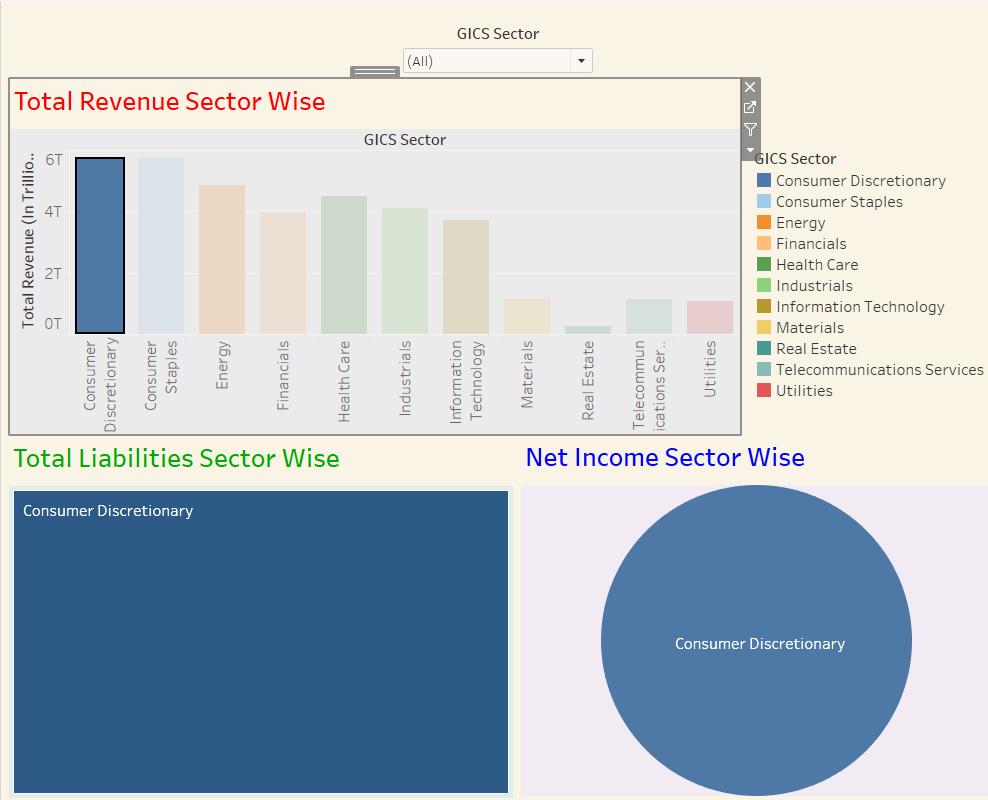
The above dashboard is made up of 3 reports, which explains the Total Revenue, Total Liabilities and Net Income of listed companies of New York Stock Market with respect to the sector they fall in.

What makes above dashboard interactive?

1. We have made it interactive my giving a ***dropdown*** feature which lets user to select a specific sector so that he can get all three information (Total Revenue, Total Liabilities and Net Income) provided for that selected sector.
2. ***Action***feature of Tableau for dashboard is used here, so that when ever user clicks on any of the sector from any of the 3 graphs provided in the dashboard, all the other graphs shows the information of the selected sector.



The above image shows the information of *Consumer Discretionary* sector, after selecting Consumer Discretionary from the dropdown provided.



The above image shows the information of *Consumer Discretionary* sector, after selecting we select Consumer Discretionary from any of the 3 reports of the dashboard.

Deployment

* We have added the source code on our [GitHub](https://github.com/rohanpn/data_visualisationasg2) profile.
* To deploy this application, we have used Microsoft Azure virtual machine, and that application will be accessible from any part of the world which have access to the URL of the application deployed on Azure machine. Our application is deployed on <http://137.117.66.227:8000/>

Contribution of Team Members

As it was a team assignment and not individual assignment, we all worked hard together right from selecting the dataset to deploying the application on Azure server, and there was an equal contribution from every single member of the team.

With using advanced development and deployment techniques like flask and azure we no only got to learn data visualisation but also how to work in a team where we can get a lot of disagreement and overcome over those with your team’s support.