**Final Project Proposal**

**DATA620 - Web Analytics - Summer 2021**

**Layla Quinones, Karim Hammoud, Maliat Islam**

**Overview**

The S&P 500 stock market index, maintained by S&P Dow Jones Indices, comprises 505 common stocks issued by 500 large-cap companies and traded on American stock exchanges and covers about 80 percent of the American equity market by capitalization. In this project we are going to create network analysis using top nodes in terms of headquarters. We are also going to implement sentiment analysis on news articles, customers’ reviews, and try find relationship between customers’ satisfaction, opinion and financial performance.

**Data Source:**

[**https://en.wikipedia.org/wiki/List\_of\_S%26P\_500\_companies**](https://en.wikipedia.org/wiki/List_of_S%26P_500_companies)

News paper articles, customer reviews on five highest and five lowest performing companies.

**Hypothesis**

Through this project, we intend to examine data pertaining to S&P 500 companies and their relationship between stakeholders’ sentiments and financial performance using NLTK library. The influence of headquarters will be analyzed using tools from NetworkX. Our team also intend to project sector based on company’s name using classification model. In terms of visualization, we intend to create animated map based on the year and headquarter of each company.

* Collect the data from web scraping.
* Create network analysis.
* Focus on the top nodes in terms of headquarters, implementation of island method.
* Create sector projection, to predict the sector for each name.
* Implement sentimental analysis for the news articles, reviews for the positive, negative and, neutral sentiments
* Visualize animated map based on the year and headquarter of each company.

**Deliverables**

We intend to deliver our findings in the form of. ipynb notebook and video presentation.

**Workflow**

1) Data preparation

2) NLP / Network Analysis

3) Evaluation

4) Tuning and Optimization

5) Finalize analysis

**Work Division**:

Project work will be distributed among team members evenly. Every team member’s analysis, research will be uploaded in the GitHub before creating the finalized version of notebook.

**Resources:**

**NLTK, NetworkX.**