Project Proposal
Group-11

1. Team members:

Name	Email	GitHub
Shrivats Agrawal	shriv9@seas.upenn.edu	ShrivatsAgrawal
Abhinav Atrishi	abhinav1@seas.upenn.edu	AbhinavAtrishi
Samuel Kamien	skamien@seas.upenn.edu	skamien
Bushra Kidwai	bkidwai@seas.upenn.edu	bushrakidwai

2. Description:

As undergrad or grad students, we are in throes of finding a summer internship or a post-college job search. The process can be exhausting and time taking considering the research that goes behind finding the perfect role and more importantly the right company. Hence our application is aimed at making the 'research' simpler for people looking for jobs.
The application will enable the user to view a range of stats of a company and filter out companies that match their interests. Eventually, the user can use it to find jobs at these companies and view recommendations of similar job opportunities.

3. Datasets:

- **Dataset 1- Financial Modelling Prep (FMP)**

Description- provides one of the most comprehensive financial data API. With each update being audited and standardized in real time to ensure consistency in the data.

We choose to work with FMP API as it is freely and easily available, while also it is guaranteed to be consistent and accurate. In addition to the raw financial data teh FMP API also provides access to a host of other interesting data about companies like what are a company's stock peers (essentially mapping how companies are related to each other) and what are the social sentiment of a company (with data from a host of social media sites like twitter, reddit, etc.).

Link to FMP- https://financialmodelingprep.com/developer/docs#Company-Profile

Attributes of the Data:

The following attributes will be queried for all the companies available through the FMP API:

S.No	Attribute	Type	Primary Key?
1	Symbol	String	Yes
2	companyName	String	No
3	price	Float	No
4	exchange	String	No
5	website	String	Yes
6	description	String	No
7	sector	String	No
8	country	String	No
9	image	String (url)	No

Stats of the FMP Dataset:

- Total Number of Rows (Including all companies): 27807
- Total Number of Companies traded in a US based exchange:
    - NASDAQ: 6185
    - NYSE: 4831

Data Collection:

We will use the Requests package in python to query for the information corresponding to all the companies available through the FMP API and collect the attributes specified above to fill the table. This table then can be joined with the table from the Indeed.com scrapping to perform various queries as desired.

- **Dataset 2- Indeed (Scrapping the data)**

Description - Indeed.com is one of the top websites job seekers prefer for finding work opportunities. Due to the availability of up-to-date information displayed in an organized and structured fashion, Indeed will serve as a befitting source of job/internships data for our project.
Indeed came out as the top choice amongst Linkedin and Glassdoor due to a well defined html structure and free availability of information without a paywall or requirements for login.

Link to Indeed- https://www.indeed.com

Attributes of the Data:

The following attributes for the top 5-10 search results will be scraped (provisional list):

S.No	Attribute	Type
1	Role Title	String
2	Company	String
3	Role Location	String
4	Role Description	String
5	Indeed URL – for Application	String
6	Company Rating	Float
7	Job Posting Date (In Days Ago Format)	String

In addition the following search parameters will be stored as attributes with the scraped data in the table:

S.No	Attribute	Type
1	Job Type	String
2	Search Location	String

How:
Technologies to be used:

a. Python Requests Library
b. Selenium - Python
c. Beautiful Soup

Methodology:
Generic URL Format of Indeed.com:
https://www.indeed.com/jobs?q=<Company_Name>&l=<Location> &jt=<Job_Type>
Eg. https://www.indeed.com/jobs?q=facebook&l=United%20States&jt=internship

➔ Get Requests to the general URL of Indeed.com will be made by filling in the variables of Company_Name [with Company Name from the primary dataset], Location [= United States] and Job_Type [= Internship | Full Time].

➔ Top 5-10 results from each search will be saved to a cloud database daily (tentatively) for every combination of the input variables as specified above.

4. Queries

1. Compare the stocks of companies in the same market
2. Job opportunities at tech companies that have an average review above a given rating
3. List companies by stock price and average employee happiness
4. List all companies whose stock rose in value in the past 6 months and have job openings
5. Order companies by number of job openings with starting pay above $15/hour