DS 210 Final Project

By: Sophia Shiu

What this project does

This project uses the dataset of Technology Layoffs Since COVID-19. The dataset had originally had around 3,500 vertices, however, many lacked important information that was necessary for the project. I cleaned the set and ended up with a revised version with around 1,600 vertices. I am interested to see the trends and potentially impacts the pandemic caused on the people working in the technology field of various industries. The dataset specifies the location, date, and industries the layoffs took place in. This information was used to explain the trends and similarities in each industry. I placed the company as the nodes and the edges as the shared similarities between the companies, such as industry. I created clusters based on industry to develop the trends and insights about layoffs after the pandemic.

The questions I wanted to ask is, is there a correlation between a certain industry and technology layoffs? Can we identify clusters or communities of companies that have experienced larger layoffs than other industries? I am interested to see if layoffs were heavily prevalent in specific industries or if it rather the influence of the pandemic as a whole.

I answered these concerns through clustering companies based on their industries and finding the similarities in each group. This way, I am able to analyze and discover impacts and influences on certain industries and layoffs within the technology field.

Modules:

In my project, I have three modules called: main, graph, and utils. Each is connected to one another.

Main module

The main module imports the libraries and modules (graph and utils) it needs. In the main function, it initializes a ‘LayoffsGraph’ instance to represent a graph data structure for analyzing layoffs data. It opens the “layoffs(1).csv” file, reads the data, and parses the data to add information about companies, industries, layoffs, and years to ‘LayoffsGraph’. After, it generates summaries of the data, calculates the statistics, and prints out the average number of layoffs and degree centrality for each industry.

The graph operations in the main module manipulates the graph data structure and calculates other statistics such as median, mode, and standard deviation of layoffs in the industry. The output is the program printing out the calculated statistics and summaries.

Graph module

It defines a struct called ‘LayoffsGraph’ that contains two fields, industries and company_layoffs, and implements methods to manipulate and analyze data related to layoffs in different industries. The ‘to_adjacency_list()’ function converts the graph into an adjacency list representation and each industry is mapped to a list of tuples that has their connected company name and layoff count. The ‘calculate_degree_centrality()’ function analyzes the graph by calculating the degree centrality for each industry. It measures the importance of a node (industry) in a graph and calculates the number of connections (number of companies) it has. It provides insights on which industries have more companies, therefore could be more affected by the pandemic and layoffs. The ‘get_industry_summary()’ function creates a summary of layoffs data for each industry. The rest of the functions calculate certain statistics needed for a comprehensive understanding and conclusion of the relationship between layoffs and industries.

There are three tests created to check accuracy. The first one checks the test_add_company function. It makes sure that when a company is added using the add_company method that the related data structures are updated and produces accurate, expected results.

Utility module

This is called the utility module as it provides more detail and code that helps with insights and cleans data to be easier to use and read for statistical calculation. The ‘calculate_median’ function is all mathematical and calculates the median of a given slice of unsigned 32-bit integers. The ‘parse_year_from_date’ function takes the year from the data string in the format YYYY-MM-DD. It splits the data string and takes only the year to parse the string into an unsigned 32-bit integer. However, if the parsing fails or string does not have a valid year, it returns a 0.

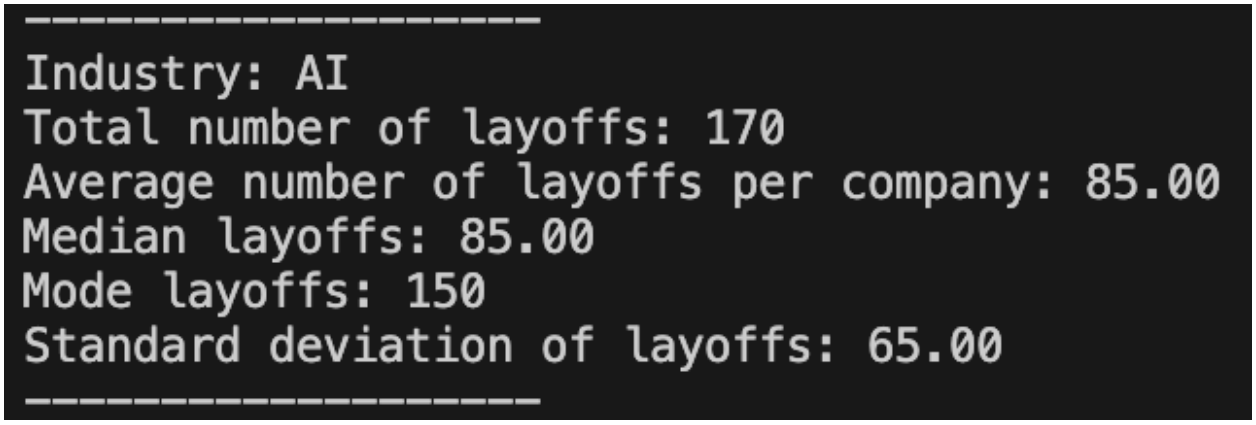
The second and third tests relate to checking the accuracy of parsing dates from the function parse_year_from_date. The second checks how it handles valid ISO 8601 formatted data strings and the third checks how it handles strings that do not fit to the expected date format.

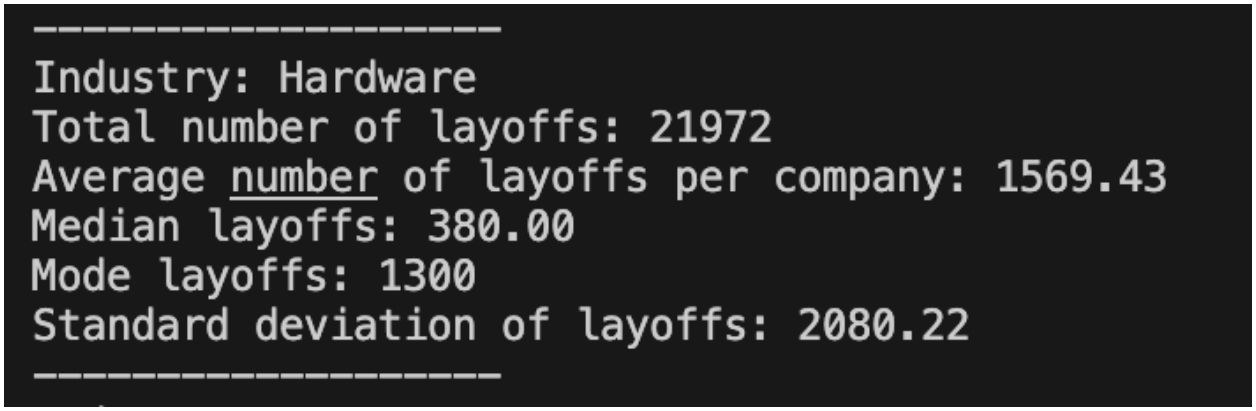
Output

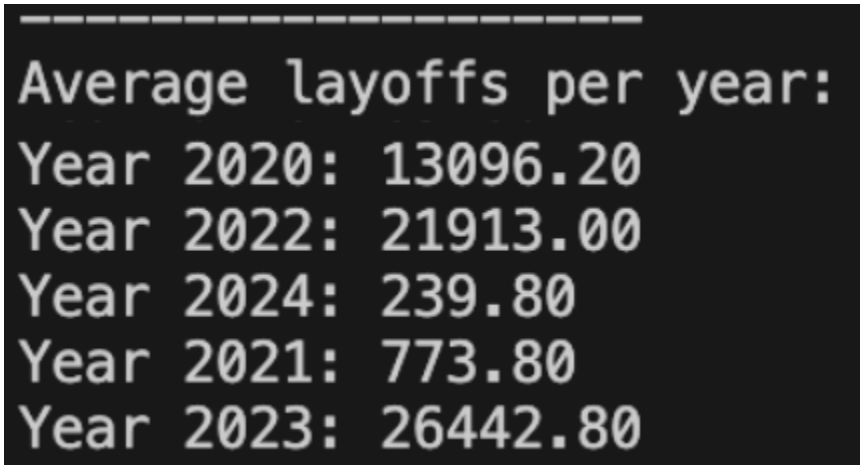
The output consists of 31 clusters, sorted by the different industries in the dataset and produced the statistical calculations for each cluster. The ones in this report are the ones relevant to the conclusion. Additionally, the average number of layoffs per year are produced to provide more information. The degree of centrality or number of companies are spread apart and also consist of 31 outputs for the industries, the ones shown here are relevant to the conclusion as well.

```
--------------------
Industry: Product
Total number of layoffs: 1265
Average number of layoffs per company: 52.71
Median layoffs: 35.50
Mode layoffs: 100
Standard deviation of layoffs: 53.58
--------------------
```

```
--------------------
Industry: Consumer
Total number of layoffs: 40207
Average number of layoffs per company: 670.12
Median layoffs: 100.50
Mode layoffs: 250
Standard deviation of layoffs: 2087.46
--------------------
```







Degree Centrality for each industry:

Industry: Product, Degree (Number of Companies): 24

Industry: Hardware, Degree (Number of Companies): 14

Industry: AI, Degree (Number of Companies): 2

Industry: Consumer, Degree (Number of Companies): 60

Conclusion

Based on the clusters of industries, product and AI had the lowest average number of layoffs per company. The industries, hardware and consumer had the largest average number of layoffs per company. Following with the degree of centrality for each industry, it is shown that the number of companies do play a significant role in the impact of layoffs. For example, AI has 2 companies with a total combination of 170 layoffs. This does not present that much of a significance in terms of layoffs compared to hardware. The hardware industry had only 14 companies with a total number of 21,972 layoffs. This provided greater scope of the number of layoffs within each company and how the ratio of layoffs are larger compared to the AI industry.

Lastly based on the average number of layoffs in each year, 2020 had around 13,096 layoffs. However in 2021, the layoffs drastically decreased to around 773 and increased significantly again to 21,913 layoffs in 2022. This demonstrates inconsistency as the years following do not follow an upward or downward trend. It is uncertain whether the pandemic placed a significant impact on the number of layoffs in the technology field. It would be helpful to analyze data about layoffs before the pandemic to further understand and determine whether the pandemic developed a relationship with the number of technology layoffs.