**DASC 5300/CSE 5300 Foundations of Computing**

**Instructor: Sharma Chakravarthy**

**Project II: Python Programming and Data Analysis**

**Group 8 (city of business to be used- ‘Newark’)**

**Team Members**

1. Palavalli, Sai Dinesh Reddy
2. Dmello, Lolita Louis

**Grading Scheme**

|  |  |
| --- | --- |
| Pre-processing a, b, c (including analysis of graph characteristics) | 20 (5 + 5 +10) |
| Analysis 1 Histogram and plotting on the map and analysis | 25 |
| Analysis 2 max pairs and their visualization and analysis | 25 |
| Q/A performance during demo | 25 |
| Challenges encountered and solution | 5 |
| Weight | 15% of total |
| Total Points | 100 |

**Index**

|  |  |
| --- | --- |
| **Topic** | **Page no** |
| Analysis of the achieved graphs |  |
| Analysis 1 - Histogram and plotting on the map and analysis | **3 - 4** |
| Analysis 2- Max pairs and their visualization and analysis | **5** |
| Graph Visualization and properties | **6-7** |
|  |  |
| Overall Status | **8** |
| Division of labour | **9** |
| Problem encountered and solution | **10** |

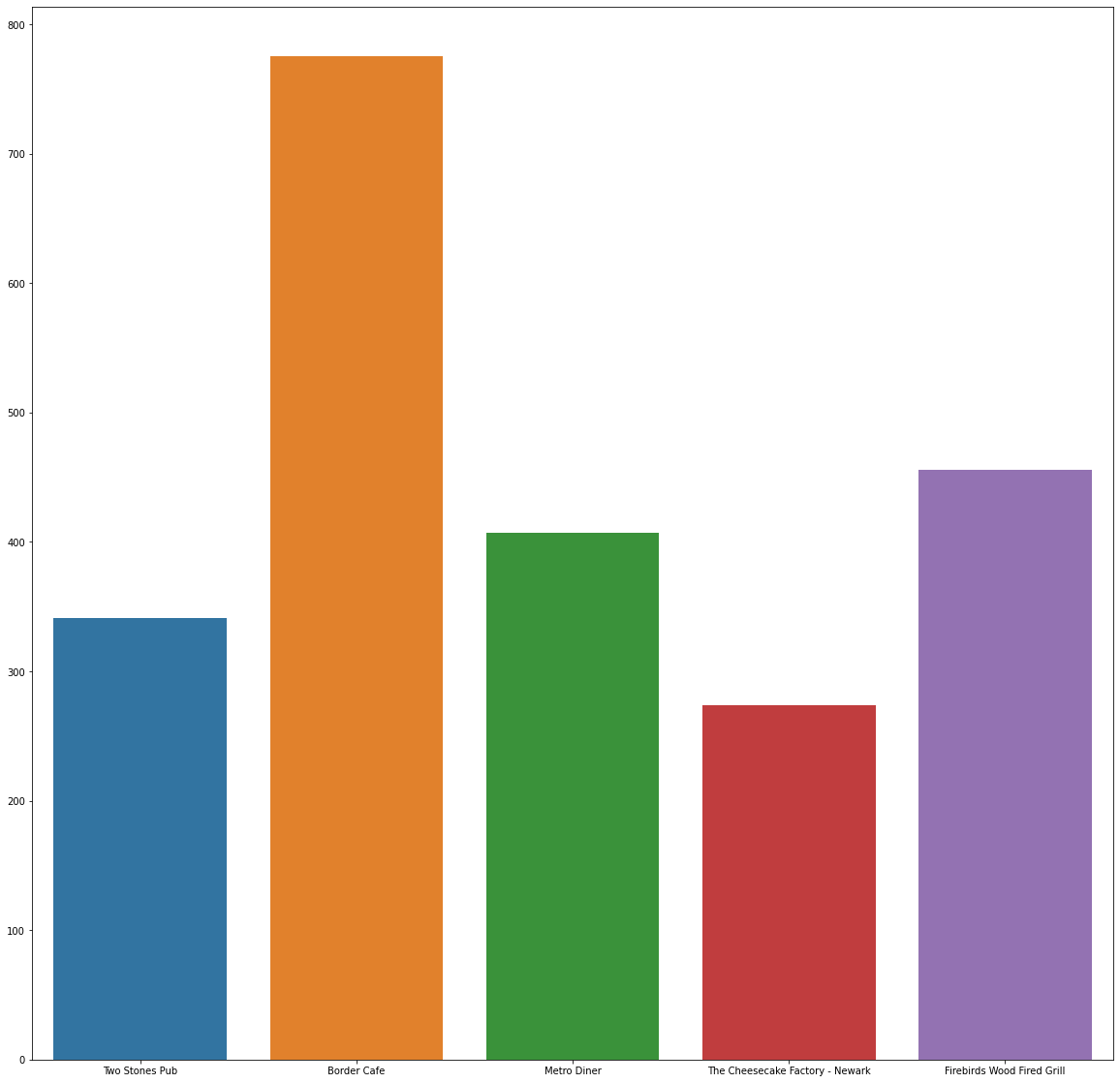
1. **Analysis 1 - Histogram and plotting on the map and analysis**

**Map

Description automatically generated**

Fig 1.1 Google map of top 5 businesses with highest reviews

* + - After plotting the top 5 hub on the map following conclusion can be drawn
    - *The color intensity represents the number of reviews given to business where darker color represents business with a greater number of reviews than lighter color.*
    - The **border café** is in the centre of Christiana Hospital and lake which seems to be a hotspot and can drive more customers from both ends.
    - **‘Firebirds Wood Fired Grill’** and **‘Metro Dinner’** are also nearby Christiana Hospital and are more likely to have a lot of customers and hence are second and third highest businesses to receive a good number of reviews as a greater number of people visit them.

After proper analysis of the graph. Following conclusions may be drawn:

* From the fig 1.2, we can conclude that most numbers of reviews were given to the Border Café as we concluded from the map that it was a hotspot and is more likely to receive a large set of customers which is directly proportional to the number of reviews.
* After reading the reviews of ‘Broder café’ we can conclude that the café is famous for its tasty food, excellent service, and amazing atmosphere.

*Fig 1.2. Bar graph of top 5 business with highest reviews*

* Chart, bar chart

  Description automatically generatedFrom Fig 1.3 it is evident that Sherry has given the highest number of reviews to businesses.
* We can also notice that there is quite a difference in the number of reviews given by sherry and other users.

*Fig 1.3. Bar graph of top 5 users you have given highest number of reviews*

* A possibility could be that Sherry is a paid influencer whose review can draw customers to the business.

1. **Analysis 2- max pairs and their visualization and analysis**

* After proper analysis of the graph using the Edmond carps max-flow algorithm the most important business and user id pair is U180 and B13
* We got the max flow for the pair u180 and b13 with the value 131.0

1. **Graph Visualization and properties**

**Properties of the graph**

Number of nodes present in the graph 9396

Number of edges present in the graph 13455

Density of the graph is 0.0003048413703384249

Number of connected components for the graph 1

Minimum Degree of the graph is 1

Average Degree of the graph is 2.863984674329502

Standard Deviation of the graph is 15.08150896869607

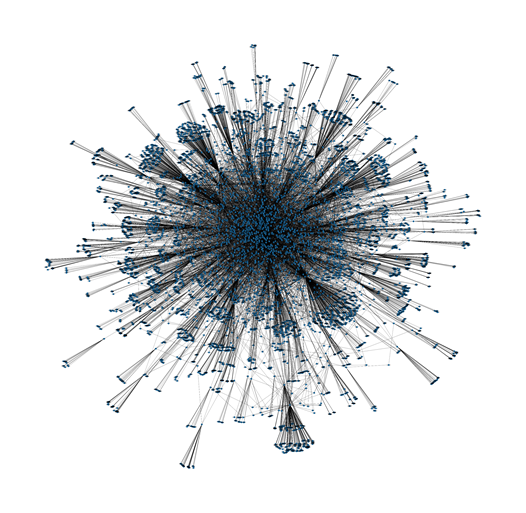
****

Fig 3.1 Graph without weights

Fig 3.2 Graph without weights

* Better image can be found [here](https://drive.google.com/file/d/19wmV0nLylZtMfLjxRX1T9yUCRUSGq7eM/view?usp=sharing) .

**A picture containing echinoderm, tree, plant

Description automatically generated**

Fig 3.2 Graph with weights

* Better image can be found [here](https://drive.google.com/file/d/1auac0JC3jqKBkn2N3j9x7q2mvECfWVv7/view?usp=sharing).

**Overall Status**

* Assigned project was completed on time with proper division of the work between the teammates.
* The dataset consisted of 3 files namely business. json, review. json, and user. json files. As the data size is very large. Newark-city data was filtered out from business.json file for the analysis as per the city of business assigned for (group 8) it had 359 unique businesses and since the graph was disconnected, we Generate a dummy user ‘ux’ for the graph to be a connected graph and with Networkx module we generated a graph and visualized it. Exploratory data analysis was performed on the filtered data set. Histogram and Graphs were plotted for better insights that can be found in this report.

**Division of labour**

To carry out the project and continuously monitor the progress of the work we carried it on google colab [(Link for the code)](https://colab.research.google.com/drive/1AdFuRbM5YJGZjIPLUvevpicIUua3BPG4?usp=sharing)

|  |  |  |
| --- | --- | --- |
| **Time spent** | **Task** | **Member name (Hours dedicated)** |
| 7 hr | Plotting the graphs | Dinesh Reddy(5), Lolita (3) |
| 6 hr | Analysing the graphs | Dinesh Reddy(3), Lolita(3) |
| 10 hr | Report | Dinesh Reddy(5), Lolita (5) |

**Problem Encountered**

1. While analyzing the graph and map (Fig 1.1, 1.2) we noticed that ‘Firebirds Wood Fired Grill’ was closer to Christiana hospital and still was less popular as compared to ‘Border Cafe’ but after reading the reviews of the user we could conclude that ‘Border Cafe’ not only had a better location but for their food taste and service is better too.
2. While analyzing the graph (Fig 1.3) the number of reviews between Sherry and other user had quite a difference we couldn’t understand why but there might be several reasons for it one of which we have mentioned in the report above.
3. By using networkx for doing analysis 1 We were missing out on a repeated node pair. There are a few cases in which one user is writing multiple reviews of a business. In networkx, it is considered as 1 but not as the actual number of reviews.

\*\*\*\*End\*\*\*\*