|  |  |
| --- | --- |
| **Project Title:** | **OptimizeU** |
| **Lab Section Number:** | **L03** |
| **Student Name:** | **Michael Le** |
| **Student Number:** | **400070369** |
| **Student Email:** | **lem12@mcmaster.ca** |

By virtue of submitting this document I electronically sign and date that the work being submitted is my own individual work.

**Abstract**

OptimizeU is designed to tackle to the problem finding an Uber on busy nights, while also trying to optimize the routes of uber drivers. The project will use a dataset of twenty million uber pick ups in New York City to the help drivers find the busiest destinations, routing more cars to more dense locations allowing more users to quickly have access to Uber vehicles. Using k-mediods to create population clusters and minimum spanning trees to minimize the travel distances of drivers, the aim of OptimizeU is to benefit both the consumer and the driver.

**1. Objective**

OptimizeU will help for-hire vehicle drivers find the most popular spots in new york city, maximizing their daily profits, while simultaneously decreasing the consumers’ wait time.

**2. Motivation**

My motivation for creating this project is to help consumers such as myself quickly find a safe ride home on busy nights. By decreasing the density of pedestrian at these locations, it increases the safety of the area. This product is designed to be used by local drivers, directing them to whichever locations are the most popular at the current time. As a result, the increase in drivers at these location will result in a shorter wait time for consumers while also increasing the profits of for-hire drivers.

**3. Prior Work**

Currently Uber is providing their driver with the names of a few of the most popular location in a certain area and the best time to drive in the week [1]. Although useful for their drivers, the locations and time span provide by Uber aren’t very precise as they are just general location names and time spans. The benefit of using OptimizeU is that the program uses a dataset that stores the longitude and latitude of past for-hire vehicle pickups and their corresponding time, this way OptimizeU is able to accurately display the most popular location at the current time. As a result, drivers can plan out their schedules more efficiently by reacting dynamically to the trends in consumer demands.

**4. Input/output and proposed solutions**

The dataset being used for OptimizeU is provided by FiveThiryEight and it contains twenty million entries of Uber pickups in 2014 [2]. Each entry in the dataset contains four elements, the date/time, the longitude, the latitude and the TLC code connected to the pickup. The longitude and latitude in this dataset will be used calculate the most popular locations while the time will be used to keep track of which coordinates are being used to calculate the output.

With the data provided, OptimizeU will use the coordinate points to output hotspot clusters, these clusters will help drivers see which spots are the most popular compared to other clusters based on size and colour intensity. In addition, these clusters will be treated as nodes and connected to each other by an undirected weighted edge. Using these edges OptimizeU will calculate the nearest clusters helping drivers find the most efficient cluster combinations.

With a large set of individual coordinate points, OptimizeU will attempt to combine these points into groups of clusters. Once these clusters are generated they will be represented as a singular object, with individual attributes such as a midpoint, size and density.

After all the clusters have been grouped, OptimizeU will draw edges between each cluster and their neighbours.Each edge will be given a weighted value depending on the absolute distance between the two nodes being connected and their midpoints. Depending on the whichever cluster the driver is in, OptimizeU will point the driver to the nearest cluster and display the next groups of cluster the driver should visit based on distance.

Another important feature of OptimizeU is for the program to dynamically update the map according to the current time. Using the time/date input of the dataset, the program will organize the coordinate points into groups. Depending whatever time interval the user selects, the program will return the corresponding set of coordinate points. By providing the user with a time interval rather than the instantaneous time more data can be displayed at once, allowing for a more visually coherent representation of the data.

**5. Algorithmic challenges:**

The first challenge OptmizeU has to tackle is generating the cluster objects given a set of coordinate points. This will be done by implementing the k-mediods algorithm [3], which given a set of points will group them into clusters and return the center midpoint of each clusters. After that, the program must be able to show the minimized distance between clusters in a group. Since the program is only trying to return the recommended order of cluster visit, we will use a minimum spanning tree [4]. By using an MST we can ensure that the net distance between all the clusters is minimized, optimizing for the worst case if the driver were to travel through all the nodes. An implementation of Kruskal’s Algorithm will be used to generate the tree [5].

**6. Project plan**

|  |  |  |
| --- | --- | --- |
| **Milestone** | **Deliverable** | **Date** |
| Requirement Specification | Professional Requirements Documents produced and submitted. | Feb 17th - Feb 18th |
| First Prototype | Produce a prototype that can map coordinate points, update base of time frame and form cluster objects. | Mar 8th - Mar 9th |
| Second Prototype | Project will have all algorithms functioning properly. | Mar 20th - Mar 21th |
| Final Deployment | Project will have a user interface, documentations, and a finish code base. | Apr 7th- Apr 8th |

**References**

[1]"Where to Drive in Toronto", *Uber.com*, 2018. [Online]. Available: https://www.uber.com/en-CA/drive/toronto/where-to-drive/. [Accessed: 27- Jan- 2018].

[2]"Uber Pickups in New York City | Kaggle", *Kaggle.com*, 2018. [Online]. Available: https://www.kaggle.com/fivethirtyeight/uber-pickups-in-new-york-city. [Accessed: 27- Jan- 2018].

[3]"K-means and K-medoids", *Math.le.ac.uk*, 2018. [Online]. Available: http://www.math.le.ac.uk/people/ag153/homepage/KmeansKmedoids/Kmeans\_Kmedoids.html. [Accessed: 27- Jan- 2018].

[4]"Minimum spanning trees", *Ics.uci.edu*, 2018. [Online]. Available: https://www.ics.uci.edu/~eppstein/161/960206.html. [Accessed: 27- Jan- 2018].

[5]"8.3.3 Kruskal's Algorithm", *Lcm.csa.iisc.ernet.in*, 2018. [Online]. Available: http://lcm.csa.iisc.ernet.in/dsa/node184.html. [Accessed: 27- Jan- 2018].