ITCS-6100 BIG DATA FOR COMPUTATIONAL ADVANTAGE

GROUP-13 PROJECT DELIVERABLE-1

1. a.) TEAM MEMBERS:

- Vineeth Avula
- Srikar Gaddipati
- Likhitha Alla
- Nikhita Somanchi
- Kamala Kumari Karuturi

1.b.) COMMUNICATION PLAN:

- Team members will discuss perspectives through slack and exchange their ideas apparently whenever it's required.
- To monitor and get the required results, all the team members will gather via Zoom or Google meet and will finish the tasks accordingly.
- The project's repository can be accessed on GitHub using URL that's given below: https://github.com/AvulaVineeth/Group13

2. BUSINESS PROBLEM OR OPPORTUNITY:

There are a lot of people who use the services provided by an American cab mobility company on a regular basis whose headquarters is in New York. Services were offered through a mobile application where it can associate its customers with nearby drivers whoever is available. The sturdy nature of our mobile application will allow us to cope with substantial loads whenever we want. Even if the organization is doing well, it is not to meet the sudden raise in the demand at some time even though it has the full capacity to function. The organization decides to make some money by raising the charge based on the ongoing demand. In case there is no demand in the future it would be very difficult to implement.

3. SELECTION OF DATA:

The dataset that's chosen (New York City Taxi and Limousine Commission (TLC) Trip Record Data) is taken from a URL that's related to the NYC government. TLC is the one who published it. The one that is

in charge of giving permits and enforcing regulations over Medallion (Yellow) taxi cabs, in case of for-hire vehicles (including commuter vans, black cars, and opulent limousines), community-based liveries, and paratransit vehicles is the New York City Taxi and Limousine Commission (TLC), which was founded in 1971.

License: : https://www1.nyc.gov/home/terms-of-use.page

Documentation: http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml

Dataset: https://registry.opendata.aws/nyc-tlc-trip-records-pds/

4. PREDICTION ANALYSIS:

Our objective is to delve into the dataset and work on the findings to figure out the optimal solution for the company to rely on, in reaching their demand forecast goals. We believe the research outcomes will provide insights in greater depth into their requirement and enable them to achieve their business objective.

Pick-up hourly distribution:

On examining the given dataset, we thoroughly analyze and predict the number of trips on an hourly basis.

Pick-up weekday distribution:

In addition to this, we also perform analysis on how well the cabs and taxis are performing on a day-to-day basis. It generates a report of the rides booked on each day, through this we can assess which days of the week are the most and least busy.

Vendor pick-up hour density, by weekday:

In addition to this, we also determine the vendor pickup density every single day with a detailed hourly description as well. Through this, a demonstration of busy hours in a day for vendor pick-ups can be examined.

5.RESEARCH OBJECTIVES AND QUESTIONS:

Our plan is to search in the dataset and find a solution to the existing problem that is being faced in the organization in such a way that there will not be any problems in the future. The main aim is to provide a follow-up based on the demand and assist in reaching the business goals. To carry out the solution we decided to use AWS technologies. We are planning to execute relevant designs and algorithms to find out the flow in demand. The main task now is to find out the relevant design and algorithms that are most appropriate in examining the dataset.