ITCS-6100 Big Data for Computational Advantage

Group 13 Project Deliverable 1

1.A. Members

- Ronik Bodduna
- Mahender Reddy Gaddam
- Pranay Dayal Guddati
- Harish Puranam
- Reetesh Zope

1. B. Communication Plan

- Members shall share insights and discuss through discord whenever needed.
- Meetings will be held on Zoom weekly or frequently as required based on the need to track progress
- Repository created for the project is found on GitHub at URL: https://github.com/Mahi1550/BigData_Group_13

2. Selection of Data

The dataset (New York City Taxi and Limousine Commission (TLC) Trip Record Data) is sourced from the NYC government's website. It is published by TLC. The New York City Taxi and Limousine Commission (TLC), created in 1971, is the agency responsible for licensing and regulating New York City's Medallion (Yellow) taxi cabs, for-hire vehicles (community-based liveries, black cars, and luxury limousines), commuter vans, and paratransit vehicles.

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

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

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

3. Business Problem or Opportunity

An American taxi mobility company based out of New York has a large volume of subscribers using it daily to commute. They provide their services through a mobile app to connect passengers with local drivers. This mobile application is robust and can handle huge volumes of load effortlessly at any point in time. Though the company is performing well, it is falling short in capitalizing on the sudden surge in demand at times, despite the application being capable enough to handle the spurt. The company intends to bank on the demand spikes by increasing

its prices during the surge. The lack of demand forecast is a major roadblock in their plans to execute this.

4. Research Objectives and Question(s)

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. To implement the solution, we plan to employ AWS technologies as part of our research. The key here is to implement appropriate models and algorithms to predict a pattern in the surge in demand. Our question now is to figure out the appropriate models and algorithms that would best fit in analyzing the dataset.