**PROJECT SUBMISSION**

ISTM – 622 605

ADVANCED DATA MANAGEMENT

ASSIGNMENT – 1

**CODE FOR PART 1**

**TEAM MEMBERS:**

Prakriti Sinha

Sidharth Dhanesh Kumar Shah

Gavin Wong

Shubham Mahendra Mundada

Vanshika Varshney

Aravindan Dakshinamoorthy

**Objective**

The objective of the system is to analyze the New York City bus data, and derive insights on the delays in the public transportation during various times of the day. With these insights, we aim to help local cab operators to target areas with frequent delays in buses. This will enable them to identify probable high demand neighborhoods and facilitate their business. The system will be built using New York City Bus Data, compiled by Michael Stone, obtained from Kaggle.

**Impact**

Aims to increase the revenue generation of the local cab owners by providing them critical insights regarding potential target areas. The results from our analysis would facilitate the cab drivers to get more trips as they would be able to estimate the demand for cabs in the neighborhood in an orderly fashion. Overall, our system would help the cab operators take data driven decisions for their business.

**Key features**

The primary feature of this project is to provide the cab agency with a list of potential high demand areas for a given time of the day. The areas would be listed in a prioritized manner on the basis of length and frequency of delays to give a holistic view of target areas for the agency.

**Files and their Purposes:**

1. **Data\_Cleaning.py (Data cleaning script):**

Code is in python for converting the file into a format that can be loaded into the databases in the cluster.

1. **Data\_Cleaning.html:**

The same file as above but allows you to view the code in Jupyter notebook format.

1. **Master\_Script.sql:**

Contains SQL command for creating tables and loading the data.

1. **nyc-busdata-analysis-master (Folder)**:

Contains the API files required for connecting with the database and UI.

(Note: This code is available in github

URL: <https://github.com/smmundada95/nyc-busdata-analysis>)

1. **UI\_Final.ipynb:**

Contains the scripts in python for displaying the contents on the UI along with the business logic.

1. **TransportERD.pdf**

ERD diagram for the database setup that is available.