1) Team Details

a) Pranjali Mehta (801255574),

Harsh Raval (801257980),

Urvashi Murari (801205124)

b) Github Repo: https://github.com/pmehta16/nyc-trip-data-analysis/tree/main

2) Business Problem or Opportunity, Domain Knowledge (link to information on domain relative to data, problem or opportunity)

Business Problem:
It is a priority of the New York City Taxi and Limousine Commission to provide safe, reliable transportation options for all New Yorkers and to recognize and address the needs of our licensees. TLC supports and contributes to city-wide efforts of traffic safety, accessibility and technological improvements. Moreover, their current initiatives include driver fatigue, pay, FHV wheelchair accessibility, language access, etc. Therefore it becomes imperative to analyze these broad aspects of data and draw some meaningful conclusions.

Domain Knowledge:

https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page
https://www1.nyc.gov/site/tlc/about/tlc-initiatives.page
https://www1.nyc.gov/site/tlc/about/data-and-research.page

3) Selection of relevant data :

We have chosen this dataset from Open Data Registry:
https://registry.opendata.aws/nyc-tlc-trip-records-pds/

About Dataset: Yellow and green taxi trip records include fields capturing pick-up and drop-off dates/times, pick-up and drop-off locations, trip distances, itemized fares, rate types, payment types, and driver-reported passenger counts. The data used in the attached datasets were collected and provided to the NYC Taxi and Limousine Commission (TLC) by technology providers authorized under the Taxicab & Livery Passenger Enhancement Programs (TPEP/LPEP).

4) Research Objectives and Question(s) (what you are trying to describe or predict with the data)

Research Objective: We will analyze the geospatial taxi data to predict the features which can help in optimizing the taxi operations currently taking place in NYC. We plan to perform the following analysis on the provided dataset:

1.) Identifying the hot spots - areas that are currently in high demand for taxis.
2.) Approximating trip duration for customers by tracking current traffic conditions.
3.) Visualization of Trips by Industry sectors.
4.) Visualizations of several metrics, including monthly trips, drivers and vehicles, trip locations, and crashes. Additionally we can compare the pickup, drop off and trip data across different locations, industries and periods.
5.) Analyze the changing trends in the industry based on decisions made by the agency and the City.
6.) Finding out the trips that took the maximum amount of time to complete.