Project: Exploratory Data Analysis (EDA)

DATS 6101: Intro to Data Science

Date - 10/05/2022

Team 8: Shikha Sharma, Abhishek Pradhan, Chirag Lakhanpal

Demystifying the Components of Riding Hailing: A Data-Driven Approach

### Synopsis:

As the most populous metropolitan in the United States, New York City boasts a vast and complex transportation infrastructure, including one of the world's largest public transport systems and a large fleet of more than 13,000 cabs, which have become favorite subjects in photography, films and unique part of New York Experience.

### Research Topic:

This project's primary goal is to use tools from the Exploratory Data Analysis (EDA) family to analyze datasets by visualizing graphical representations of data, identifying hidden patterns and outliers, testing hypotheses about subgroups of the population within the dataset, and segmenting datasets into meaningful categories for further analysis.

### SMART Question:

1. How and Why do some city areas affect ride-hailing more than others?
2. What statistical properties of New York City Taxis make them so effective in terms of passenger distribution, location, and trip duration?
3. What factors affect the duration of a trip?

### Data Source: [Link](https://www.kaggle.com/competitions/nyc-taxi-trip-duration/data?select=train.zip)

**Number of Observations of Data:** 1458644

**GitHub:** [Link](https://github.com/ChiragLakhanpal/Intro-to-Data-Science-Project-1.git)