Massive Data Analysis Project

**Taxi Data Analysis**

*Project Group Members:-*

**Shivali Sohoni**

**Chirag Chauhan**

**Amruta Barretto**

NYC Taxi Data needs to be analysed. The goal of our project to build a comprehensive report of the “State of the NYC Taxis”.Taxi data can be analysed to infer important statistics about the different aspects of NYC taxis and this can be visualized to provide more comprehensive and helpful results about the NYC taxis.

The area that we are focusing on in our project is to analyse the taxi data to determine the various aspects of the taxi drivers and also to analyse how the weather affects the taxi frequency.

### Understanding driver behavior-

Some of the issues that we are investigating and listed below:-

* How do different drivers work? Do drivers (or group of drivers) have a preferred neighborhood (or set of neighborhoods)? What does the pickup/dropoff distribution looks like? Does this preference change over time?
* Are there patterns shared among different drivers?
* Do some drivers get higher tips on average than others?
* Do some drivers take longer routes than others?

### Understanding taxi usage

* What are the most popular destinations? Do these change over time? For example, summer vs the other seasons?
* What are the most popular trips (source, destination)?
* Is the number of trips and taxis affected by weather? E.g., are there fewer cabs when it is raining/snowing?