

## Analyzing Chicago Taxi Trips Behaviour via Data Mining.

### **Who is in your group?**

Team Members: Shaurya Sahai, Abishek Krishnan and Sushmitha Sunkurdi Nataraj

### **What data you plan to use and where you plan to get it from?**

We plan to use the Chicago Taxi Trips data collected from kaggle  
[<https://www.kaggle.com/chicago/chicago-taxi-trips-bq>]

### **What structure you want to mine from the data?**

The dataset contains taxi trips' information from 2013 to 2018 with details like pick-up community, drop-off community, payment methods, tips, extra charges, etc. From this selected data set we aim to mine interesting insights like - which payment method is most used, is there a correlation between drop-off points and tips to the drivers, most popular (private) taxi companies, and more, using various data mining techniques.

### **Why this problem is interesting?**

We find the problem interesting because of various reasons. Firstly, this BigQuery dataset would give us a chance to experiment with the BigQuery API, which we think would be both challenging and informative. Secondly, with such a dataset, we'd be able to analyze the behaviour of taxis and passengers in Chicago. According to kaggle.com "*There are about seven thousand licensed cabs operating within the city limits.*" Hence, we believe substantial insights can be interpreted from mining this dataset.

### **What is new, or what I (the instructor) will learn?**

The results from this project can help Taxi aggregators interpret passenger behavior to reduce cancellations, also improve their service by deploying more taxis in areas of high demand during particular hours to reduce the response time and understand other interesting patterns observed.