The University of Chicago
CAPP 30123 Project AHDA Proposal *

An Analysis of the Chicago Taxi Industry: How can Traditional Taxi Industry Survive in the Age of the Sharing Economy?

Chung, Alice `Alicechung` | Liu, Huanye `huanye` | Xu, Ningyin `sixisxu` | Zhang, Dongping `dpzhang`

April 21, 2017

1 Data

The dataset we plan to use for our project is *the Chicago Taxi Trips Dataset*, which could be acquired from *the Chicago Open Data Portal*. The dataset is approximately ***42.76 GBs***, which consists of precisely ***108,521,750*** unique taxi trips from January 1st, 2013 to April 1st, 2017. Each unique taxi trip has following variables:

- Trip ID
- Taxi ID
- Trip Start Timestamp
- Trip End Timestamp
- Trip Seconds
- Trip Miles
- Pickup Census Tract
- Dropoff Census Tract
- Pickup Community Area
- Dropoff Community Area
- Fare
- Tips
- Tolls
- Extra
- Trip Fare Total
- Payment Type
- Company
- Pickup Latitude
- Pickup Longitude
- Pickup Centroid
- Dropoff Latitude
- Dropoff Longitude
- Dropoff Centroid

  - Using "Trip Miles" and "Trip Seconds", we would construct a variable called "Avg Trip Velocity", which could show the traffic condition and route choices of cab drivers
  - We are also able to use "Tips", comparing with "Fare", to get a sense of the quality of service that each unique cab drivers provide to customers

*Project repository is hosted on Github at `https://github.com/dpzhang/Project_AHDA`

2 Hypotheses

- **Questions on Spatial Units:**
  - Where are the top spatial units (census tracts/neighborhoods) that have the highest daily pick-ups and drop-offs in Chicago from the past five years?
  - In certain dates of a year (e.g. U.S. holidays, random workdays, or random weekends), what are the pick-up and drop-off distributions of those identified busy spatial units (busy spatial units: spatial units that have the highest daily pickups or drop-offs)?
  - Would it be possible to simulate taxi pick-ups and drop-offs using statistical distributions in any spatial-temporal unit? (Maybe Poisson?)
- **Questions on Taxis:**
  - Which month/holiday/period of any given year is the busiest for taxis?
  - What is the average duration of one taxi trip?
  - What is the average trip miles of one taxi trip?
  - What is the relationship between average velocity and daily profit / averaged daily profit?
- **Questions on Taxi Profit:**
  - What is the market share/market competition of the Chicago taxi industry by taxi companies?
    - What are the major players in the Chicago taxi industry?
    - What are the daily/monthly/yearly profits of major taxi companies by aggregating revenues of all taxis owned by that company?
  - Identified by taxi ID, which taxis would drive the longest distance daily on average?
  - Identified by taxi ID, which taxis would drive across the most spatial units on average in the city of Chicago?
  - What are the top ten taxis who make the most profit daily/yearly on average in the city of Chicago?
  - When are the most profitable times of the week / of the year?
  - At what time of the day or from which spatial unit of the city would people prefer to give tips?
  - In certain dates of a year (e.g. U.S. holidays, random workdays, and random weekends), which temporal-spatial unit of the city would have the greatest probability to make taxi drivers have a relatively profitable trip?