Correlations in Cost of Chicago Taxi Rides 2016

Team 10:

Noah Leuthaeuser | Joseph Alsko | Heesuk Jang | Rei Isobe

Description

Analysis of the 2016 Taxi Trips dataset provided by the City of Chicago. The dataset includes information about pickup and dropoff location, fare, trip distance and time, tips, and taxi company. We hope to find answers to questions similar to the following:

- How does fare and ride distance vary over the year?
- How does pickup and dropoff location relate to fare and tip?
- What are frequently traveled taxi trips in or out of low income areas?
- Is there a correlation between certain areas and payment type or tips?

Prior Work

- City of Chicago Business Affairs and Consumer protection study (2014)
  - Focused on fairness of rates versus value of taxi service (fuel, training, etc.), no specific districts
- NYC Data Science Chicago Cabs Analysis (2016)
  - Focused on locations of rides
- Todd W. Schneider (2013)
  - Focused on success of taxi business, best spots for frequent pickups

NYC DATA SCIENCE ACADEMY

BACP BUSINESS AFFAIRS CONSUMER PROTECTION

How Our Study Differs

- Focus on actual price numbers, *as well as* pickup frequency
  - Pair these trends with pickup locations
- Focus less on Chicago taxi business as a whole
  - Look at 2016 alone, no comparison to previous years
- Focus on prices relative to average rates
  - Not looking at cost vs. value of taxi rides

Datasets

- Chicago Taxi Rides in 2016
- https://www.kaggle.com/chicago/chicago-taxi-rides-2016
- Downloaded on every members' computers

kaggle

Proposed Work

- Data Cleaning:
  - Remove entries with missing attributes
- Data Integration:
  - Only use data with certain attributes that we are interested in
- Data Reduction:
  - Reduce to only use what is necessary
    - Pickup and dropoff location, fare, trip distance and time, tips, and taxi company
    - Possibly take out tolls

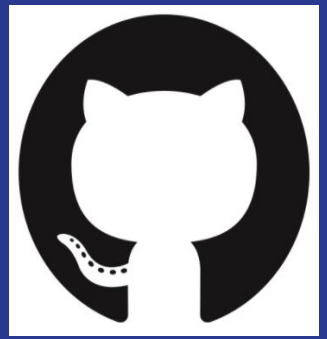
Tools

Python

iPython Notebook

GitHub

Slack



Evaluation

- Use all possible pattern discovery methods along with statistical tests to validate or invalidate our hypothesis of dataset

- Could use our results to provide taxi companies cohesive statistics of the dataset to increase the likelihood of taxi business’ net profit