How Does Crime Affect Bike-Sharing Trips in San Francisco?

Group 11 CSPB 4502 Project Part 1

Team members:

- Zack Cheng (Tsai-Tzu Cheng)

Description:

- Thanks to the modern software technology, bike-sharing is getting more and more popular in many large metropolitans around the globe, yet one of big challenges to the bike-sharing economy is the notoriously high crime rate in those cities.
- The team is interested in data mining the user trip data of baywheels in the bay area (including San Francisco), with the San Francisco crime report to discover whether the historical / recent criminal incident locations has taken effects to the biker route decisions.

Prior work

- Researches and news about bike sharing and crime:
  - Bicycle Share Systems: A Predictor of Crime?
  - Bike sharing scheme scrapped due to theft

Datasets:

- Multiple trip data csv by months in 2024:
  - Baywheels bike-sharing trip data
- Instant crime statistics report to calculate crime rate, filtered by months and dates:
  - San Francisco Crime Report | Live Map

Proposed work

- Data cleaning:
  - Remove null-valued or missing-value items.
  - Remove unnecessary attributes.
- Data preprocessing:
  - Reorganize / rearrange into chronological order and group by crime types / incident locations.
- Data integration:
  - Combine route data and crime data.
- Data mining:
  - Pattern discovery, clustering, classification.
- Post-processing:
  - Develop predictive models, interpretation, visualization.

List of tool(s):

- Excel
- Weka
- SQL
- Python
- Python libraries:
  - numpy
  - scikit
  - pandas

Evaluation:

- Calculate the percentage of correct predictions by testing with immense amount of crime incidents and trip data in any random given time periods.
- Calculate the classification accuracy on the derived pattern.
- Exploring and applying other professional analysis and metrics.