CMPE-255 Project Proposal

Analyzing Bike Share Data

Group Members:

Anshul Shandilya (016039894)
Girish Bisane (016650348)
Parth Saraiya (016715608)
Stavan Patel (016622398)

Team Name: Loud Logic

Project Title: Analyzing Bike Share Data

Project description:

In this project, we plan to analyze the SF Bay Area Bike Share dataset. The dataset has information on bikes, stations, weather, and bike trips. The goal is to use techniques learnt in the class to better understand the dataset. We plan to use techniques like clustering, regression, visualization, and dimensionality reduction.

Proposed methodology:

- Dataset: The Bay Area Bike Share enables quick, easy, and affordable bike trips around the San Francisco Bay Area. They make regular open data releases (this dataset is a transformed version of the data from this link), plus maintain a real-time API.

  The Dataset contains following files:

  - station.csv - Contains data that represents a station where users can pickup or return bikes.
  - status.csv - data about the number of bikes and docks available for given station and minute.
  - trips.csv - Data about individual bike trips
  - weather.csv - Data about the weather on a specific day for certain zip codes

  Dataset Source: SF Bay Area Bike Share | Kaggle (https://www.kaggle.com/datasets/benhamner/sf-bay-area-bike-share?datasetId=57&sortBy=voteCount)

- Problems: Performing Exploratory Data Analysis (EDA) to find insights about the data like:
  - Longest & Shortest trips
  - Most popular routes (source and destination stations)
  - Clustering different user groups
  - Difference in usage pattern of registered and unregistered users
  - Forecasting demand based on past data
  - Using Dimensionality Reduction to visually analyze the data
  - Analyze how various factors like weather, etc. affect number of trips

- Techniques planned to be used:
  - Regression
  - Clustering
  - Data Visualization
  - Dimensionality Reduction

Resources:

- Scikit-learn: machine learning in Python documentation
- Bike Sharing Demand-Exploratory Data Analysis | by Anugya Shaw | Medium
- An Analysis of Bike Sharing Usage: Explaining Trip Generation 2 and Attraction from Observed Demand (nacto.org)
- Bike-share research (bikeshare-research.org)