QUAD SQUAD - PROJECT PROPOSAL

Inderjeet Singh(isb5), Naina Thapar(nta65), Nidhi Kantekar(nkk9), Rituraj Ojha(roa11)

Overview

Our project revolves around a bike-sharing firm, and as analysts for that business, we will be using several big data tools to offer insights into various use cases, predicting their future profits and assisting them in expanding their business.

Dataset(s)

We plan to use two datasets -

1. Mobi Shaw Bike dataset, which provides us information about the start time, end time, duration of the ride, pick up station, return station, type of membership, and more from January 2017 to September 2022.
   Link- https://drive.google.com/file/d/18nSmd8szs1abJ-_emITXkjPMhQ0sAA-k/view
2. Vancouver Weather dataset, which provides us with the daily and hourly weather recorded in vancouver.
   Link- https://vancouver.weatherstats.ca/download.html

Proposed Goals

1. To understand how casual riders and annual members use Mobi Shaw bikes differently.
2. To maximize the number of annual members.
3. To understand and plan the repositioning of bikes according to the demand and supply in order to increase the average utilization of bikes.
4. To analyze how the hour-of-day, temperature, wind speed, and seasons affects the bike usage.
5. To create a report containing frontend visualizations of our analysis that can be used by business stakeholders.

Stretch Goals

1. To create a forecast model using historical patterns in order to maximize the earnings.
2. To offer suggestions on potential locations for new bike share stops.

Technologies

1. Languages: Python, SQL, HTML/ CSS
2. Frameworks and Libraries: PySpark, Apache Spark, Hadoop, NumPy, Pandas, Seaborn
3. Software Tools: Amazon S3, Amazon EMR, Git, Tableau
   Note: We shall add more technologies as per our requirements.