Northeastern University
College *of* Engineering

Data Warehousing & Integration
IE 6750
FALL 2024

BikeFlow Analytics

Milestone 1

Group 1

Layashree Adepu

Pavithra Moorthy

Adepu.l@northeastern.edu

Moorthy.p@northeastern.edu

Submission Date: 10/02/24

Problem Statement:

The Bay Wheels bike-sharing service faces challenges in maintaining bike availability across its network of stations. Some stations frequently run out of bikes, while others have too many, leading to user dissatisfaction and inefficiencies in the system. Currently, the service does not have a consolidated way to analyse key factors like ride patterns, peak times, and user behaviour to better manage bike distribution.

To address this, we will create a centralised on-premise database that collects trip data, station usage, and user information. This will help the company analyse patterns, improve bike reallocation, and optimize station management, ensuring bikes are available when and where they are needed most.

Database Overview:

The operational database for the Bay Wheels bike-sharing service is designed to collect and organize key information from every trip, allowing for detailed analysis and decision-making. Each column in the database captures specific aspects of the bike-sharing system, which can be leveraged to optimize operations and improve user experience

Columns:

- **duration_sec**: Total trip duration in seconds, useful for analyzing ride patterns and trip lengths.
- **start_time / end_time**: Start and end timestamps for each trip, essential for identifying peak usage periods.
- **start_station_id / start_station_name**: Unique ID and name of the station where the trip starts, used for tracking station popularity.
- **start_station_latitude / start_station_longitude**: Geographic coordinates of the start station, important for location-based analysis.
- **end_station_id / end_station_name**: Unique ID and name of the station where the trip ends, helping to analyze route patterns and station demand.
- **end_station_latitude / end_station_longitude**: Geographic coordinates of the end station, useful for geospatial studies.
- **bike_id**: Unique identifier for each bike, allowing for maintenance tracking and usage patterns.
- **user_type**: Indicates whether the rider is a Subscriber (member) or Customer (casual user), for behavior analysis.

- **bike_share_for_all_trip**: Identifies trips under the Bike Share for All program, important for evaluating program participation.

Link: Dataset website

We have an extensive dataset that covers key aspects such as trip duration, start and end times, station details, bike IDs, and user types. To enhance the analysis and provide more insights, we plan to add additional columns like **user details** (e.g., age group and subscription type), **distance traveled** (calculated between start and end stations). These new columns will be generated or calculated based on existing data, allowing for deeper analysis of user behaviour, trip patterns, and operational efficiencies within the bike-sharing system.

What We Hope to Analyze:

In the future, we hope to analyze user behaviour patterns to gain insights into rental trends, peak usage times, and demographic preferences for Lyft Bikes. This analysis will help optimize fleet management by assessing bike distribution and maintenance needs, ensuring availability aligns with demand. Additionally, I aim to evaluate the effectiveness of integration with public transportation systems, measuring the impact on rental rates and user satisfaction. Overall, these insights will guide improvements in service delivery and support Lyft's commitment to enhancing urban mobility.

- **Optimize Bike Reallocation**: Predict high-demand stations and recommend rebalancing of bikes across stations.
- **Peak Time Analysis**: Identify the busiest times of day/week to optimize bike and station management.
- **User Behaviour**: Analyse differences between casual users and subscribers to tailor marketing or service strategies.
- **Station Usage Efficiency**: Evaluate which stations are under- or over-utilized and plan for capacity adjustments.