



Northeastern University
College of Engineering

Data Warehousing & Integration
IE 6750
FALL 2024

BikeFlow Analytics

Milestone 2

Group 1

Layashree Adepu

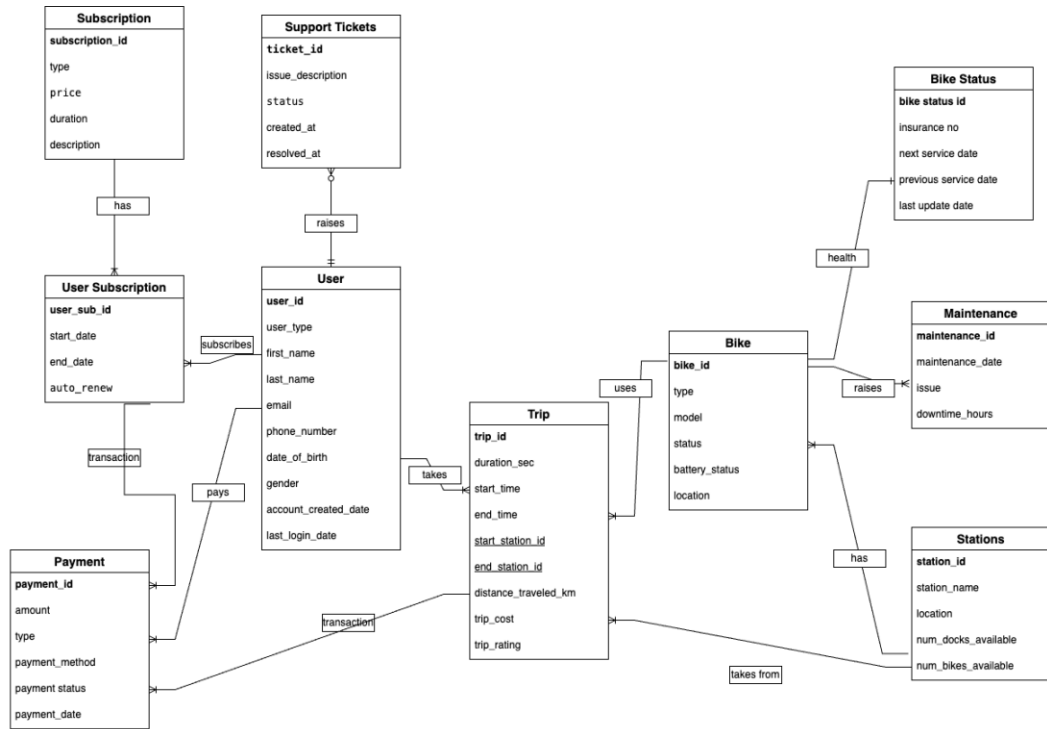
Pavithra Moorthy

Adepu.l@northeastern.edu

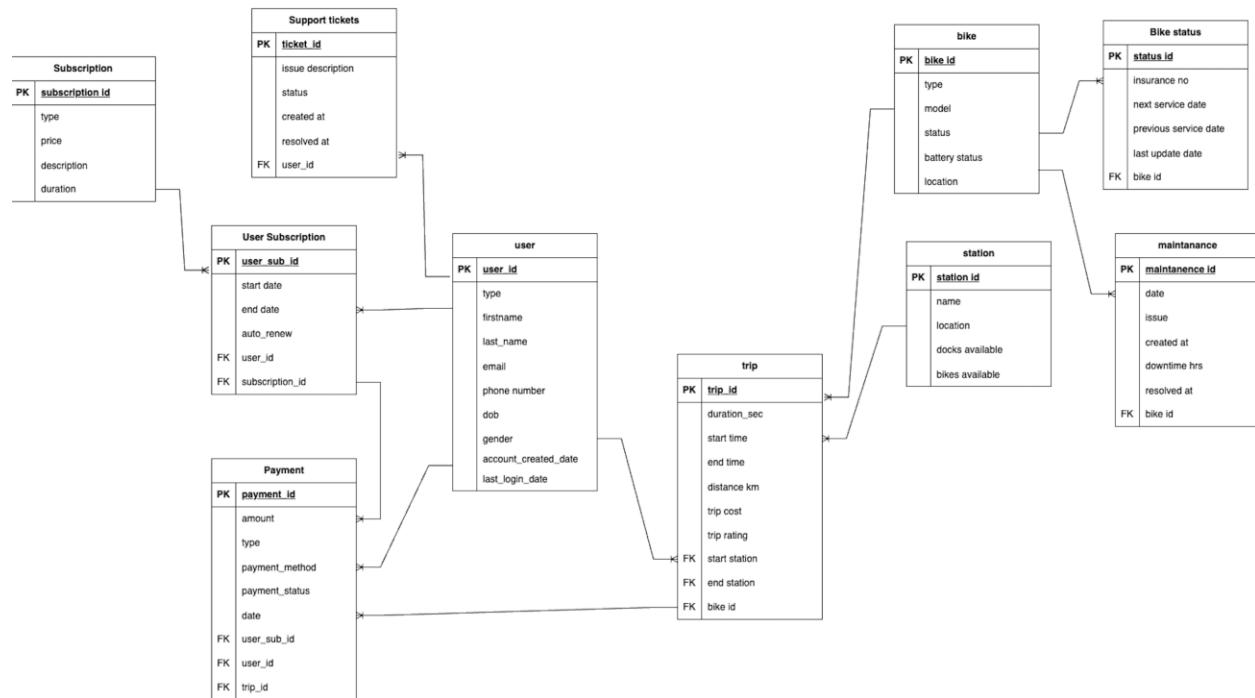
Moorthy.p@northeastern.edu

Submission Date: 10/09/24

ERD



Relational Model



Data Population Methodology

1. Data set upload as a csv file for trip table from Kaggle (90000 rows)
2. For other tables - <https://www.mockaroo.com>

1. Dimensions

- a. Bike
- b. Station
- c. User
- d. Subscription

2. Hierarchies

a. Date/Time Dimension

Hierarchy:

- Year → Quarter → Month → Day
- Year → Month → Week → Day → Hour

b. Location Dimension

Hierarchy:

- Country → State → City → Zipcode
- Country → Region → City → Zipcode

3. Measures

a. Trip Measures

- **Trip Count:** Total number of trips
- **Average Trip Duration:** Sum of all trip durations / Trip Count
- **Total Distance Traveled:** Sum of all distances for each trip
- **Trip Cost:** Sum of costs associated with all trips
- **Average Trip Rating:** Average of all ratings given for trips
- **Trips by Time of Day:** Number of trips during different parts of the day (morning, afternoon, evening, night)

b. User Measures

- **User Count:** Total number of unique users
- **New Users:** Number of users who signed up in a specific time period
- **Retention Rate:** Percentage of users who continue to use the service over a period of time
- **Average Subscription Length:** Average duration of active subscriptions per user
- **User Activity:** Count of trips per user over time

c. Bike Measures

- **Bike Utilization Rate:** Percentage of time a bike is in use vs. available
- **Maintenance Frequency:** Average number of maintenance events per bike
- **Mileage per Bike:** Total distance traveled by each bike
- **Downtime:** Total time each bike is out of service

d. Station Measures

- **Station Usage:** Number of trips starting or ending at each station
- **Capacity Utilization:** Average percentage of docks occupied at a station
- **Peak Hours:** Most frequently used hours at a station
- **Top Stations:** Stations with the highest number of rentals or returns

e. Financial Measures (Payments)

- **Total Revenue:** Sum of all payments
- **Average Payment per Trip:** Total Revenue / Trip Count
- **Subscription Revenue:** Total revenue from subscriptions