

Project Development Phase

Sprint 1

Date	29 October 2022
Team ID	PNT2022TMID18963
Project Name	Project – A New Hint to Transportation-Analysis of the NYC Bike Share System
Sprint 1	Data Pre-processing

Understanding the Dataset :

Let's understand the data we're working with and give a brief overview of what each feature represents or should represent.

1. Trip Duration (seconds) — How long a trip lasted
2. Start Time and Date - Self-explanatory
3. Stop Time and Date - Self-explanatory
4. Start Station Name - Self-explanatory
5. End Station Name - Self-explanatory
6. Station ID - Unique identifier for each station
7. Station Lat/Long - Coordinates
8. Bike ID - unique identifier for each bike
9. User Type (Customer = 24-hour pass or 3-day pass user; Subscriber = Annual Member)
- Customers are usually tourists, subscribers are usually NYC residents
10. Gender (Zero=unknown; 1=male; 2=female) - Usually unknown for customers since they often sign up at a kiosk
11. Year of Birth - Self-entered, not validated by an ID.