# Introduction

There are 183412 rows in the dataset with 16 features. I select few features and want to explore data basic on them. Here are few questions that I will answer.

# Data wrangling:

1. Delete useless columns('start\_time', 'end\_time', 'start\_station\_id','start\_station\_name','start\_station\_latitude', 'start\_station\_longitude','end\_station\_id','end\_station\_name','end\_station\_latitude','end\_station\_longitude', 'bike\_share\_for\_all\_trip'])
2. Drop rows with null values
3. Convert datatype into int
4. Add duration\_mins & age & age group columns
5. Delete duration\_mins longer than 88 mins in duration\_mins column
6. Delete age group older than 80 years old in age\_group column

# Questions:

1.What is the average trip length?

* The average trip length is 10.0 minutes.

2.What proportion of rides are longer than 10 minutes?

* The proportion of rides that longer than 10 minutes is 34%.

3.Which type of user takes longer rides on average?

* Customers take longer rides on average than Subscribers.

4.What is the average trip length for Subscribers or Customers?

* The average trip length of Customers is 15.0 minutes and the average trip length of Subscribers 9.0 minutes.

5.Which age of people use sharing bikes most?

* People at 31 years old are the most age group use sharing bike.

# Visualization

Univariate Exploration

1. Average trip length of each user type

2. Average trip length of each age group

3. Number of users in each age group

Bivariate Exploration

1. How does the trip duration distribution vary between customers and subscribers?

2. How does the trip duration distribution vary between each age group?

3. How does the trip duration distribution vary between male and female?

Multivariate Exploration

1. Compare the correlation between "duration\_mins"， "age\_group" and the other features

2. Average trip Duration of user type in different age group