🎬 **Communicate Data Findings**

Ford bike trip data

Data description :-

This data set includes information about individual rides made in a bike-sharing system covering the greater San Francisco Bay area

Cleaned data columns :-

start_time	• The detailed date of the trip
start_station_name	• Address of trip start location
end_station_name	• Address of trip end location
bike_id	• An unique ID for every bike
user_type	• (Subscriber or Customer – "Subscriber" = Member or "Customer" = Casual)
member_gender	• Male , female or other
age	• The member age
distance	• The trip distance in KM
duration_min	• Trip duration in minutes
speed(km/hr)	• The member speed in the trip

Insights :-

I LOOK at age , Speed , duration and distance distribution and get some results

Study number of trips in different times as Different hour of the day…etc

User classification counting "gender and type"

Most common start and end station names

Relation between customer type and trip duration and his speed

Relation between customer gender and trip duration and his speed

find the relation between speed , duration and age

relation between gender ad user type

Relation between day and hour

Common start to end combination with more than 200 record

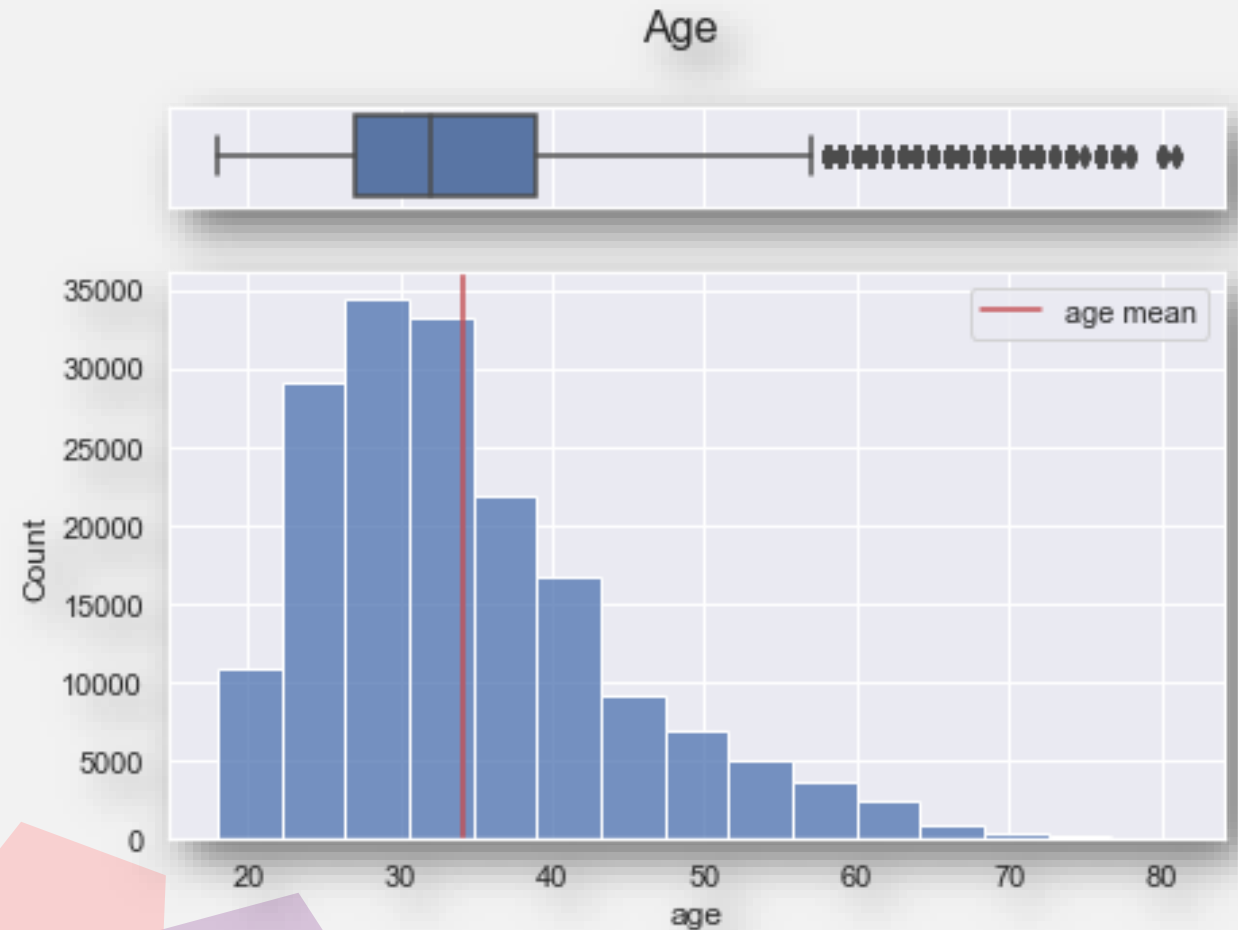relation between distance and other parameter

LOOK at age distribution

Age

Comment :-

age has right skewed distribution with mean of 34 and minimum age of 18 and maximum age of 81
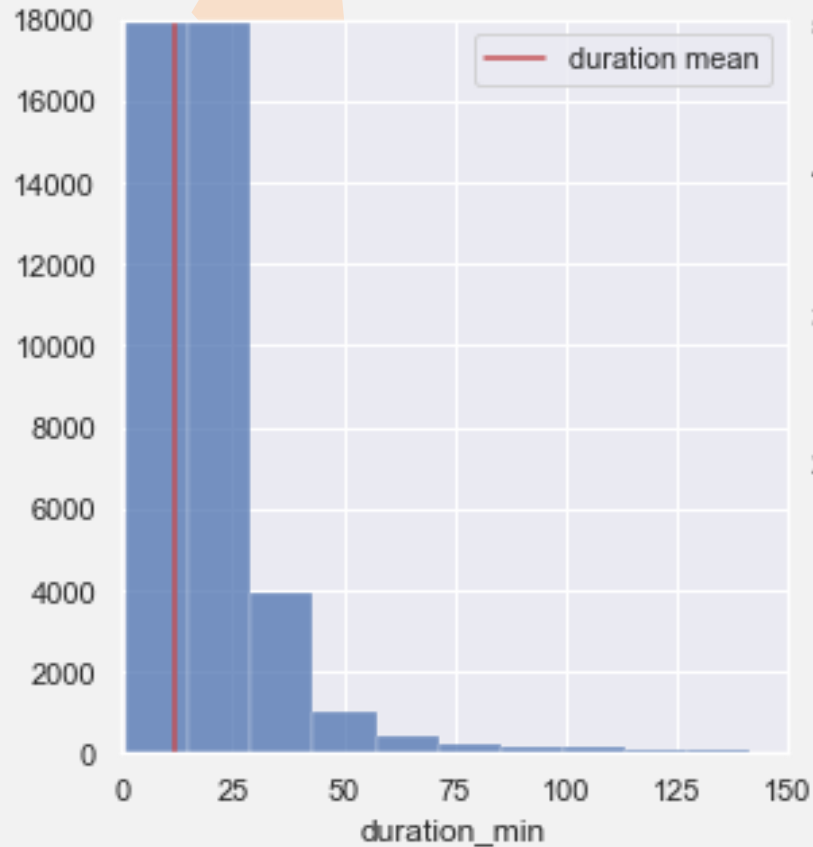
the most common customer age is 31

with the increase of age the number of customer decrease and that makes sense because the health
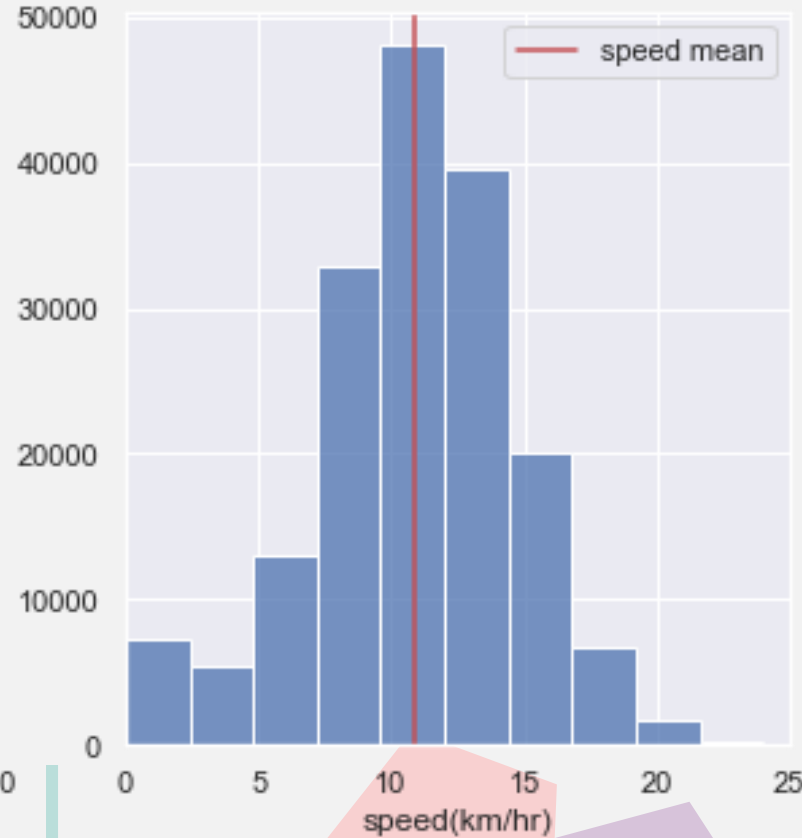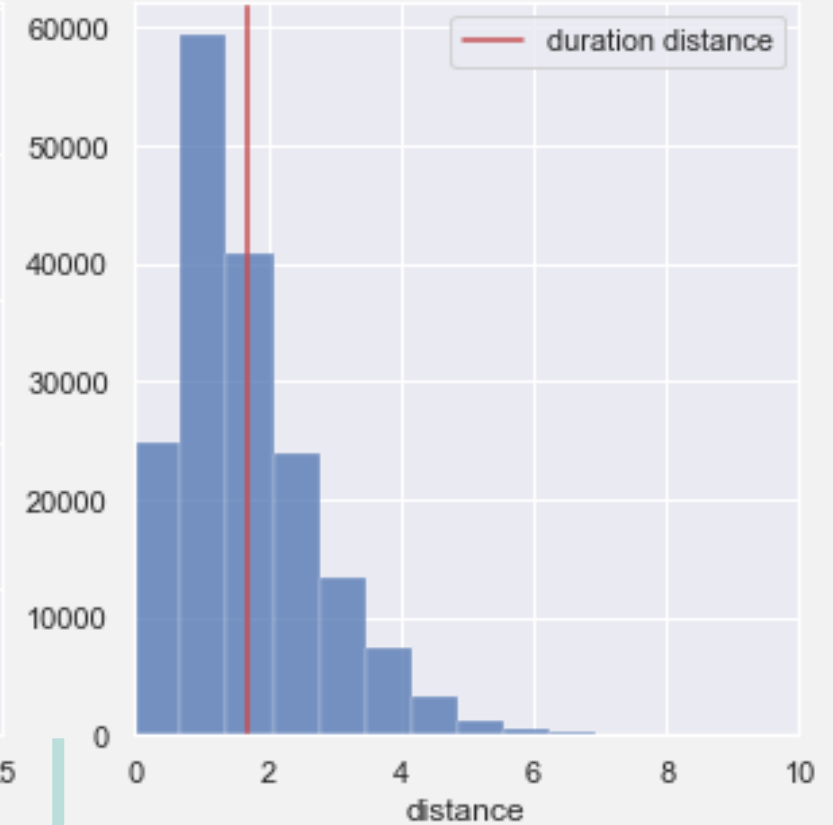**Also the small age has a little number of trips**

Speed , duration distance distributions

duration has a common values from 0 to 25 min after that it has a right skewed distribution , and it has some outliers
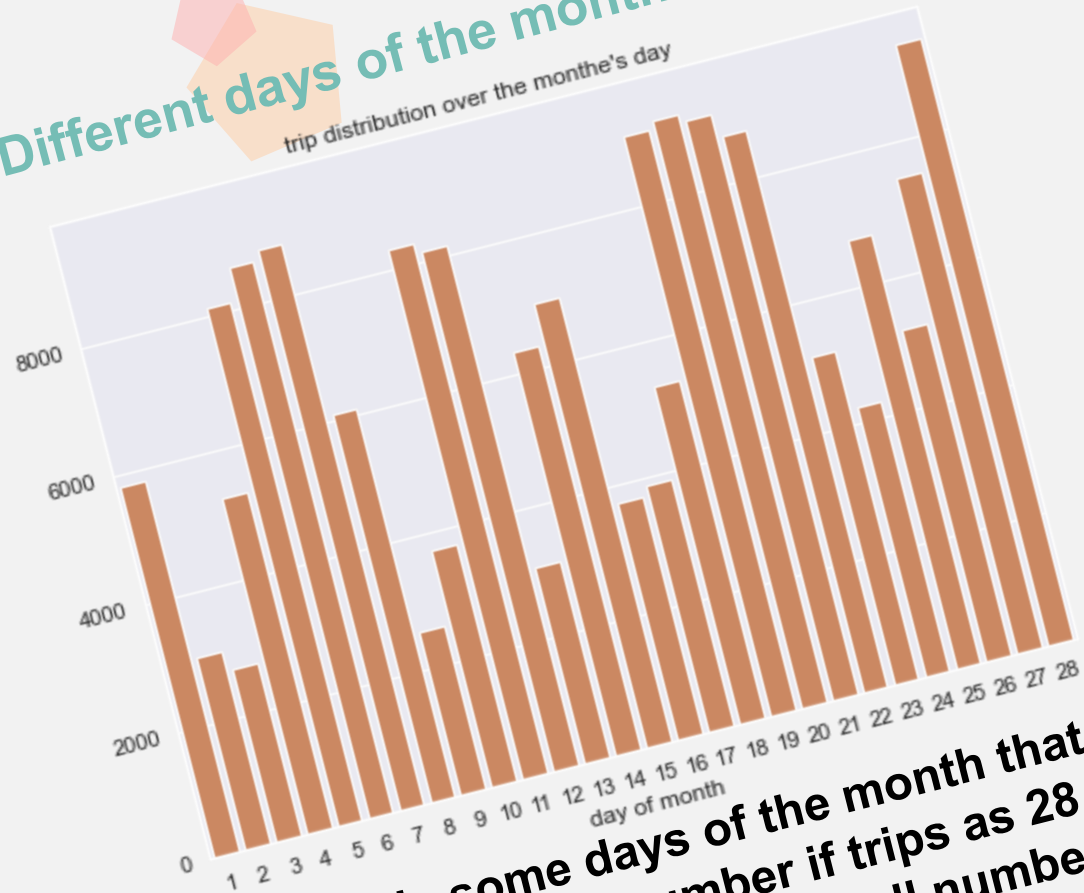
Speed has a normal distribution with a mean 10.8 km/hr and a max speed 36 km/hrs

distance has a mean of `1.7 km` ,the maximum is `69 km` , and it has some outliers
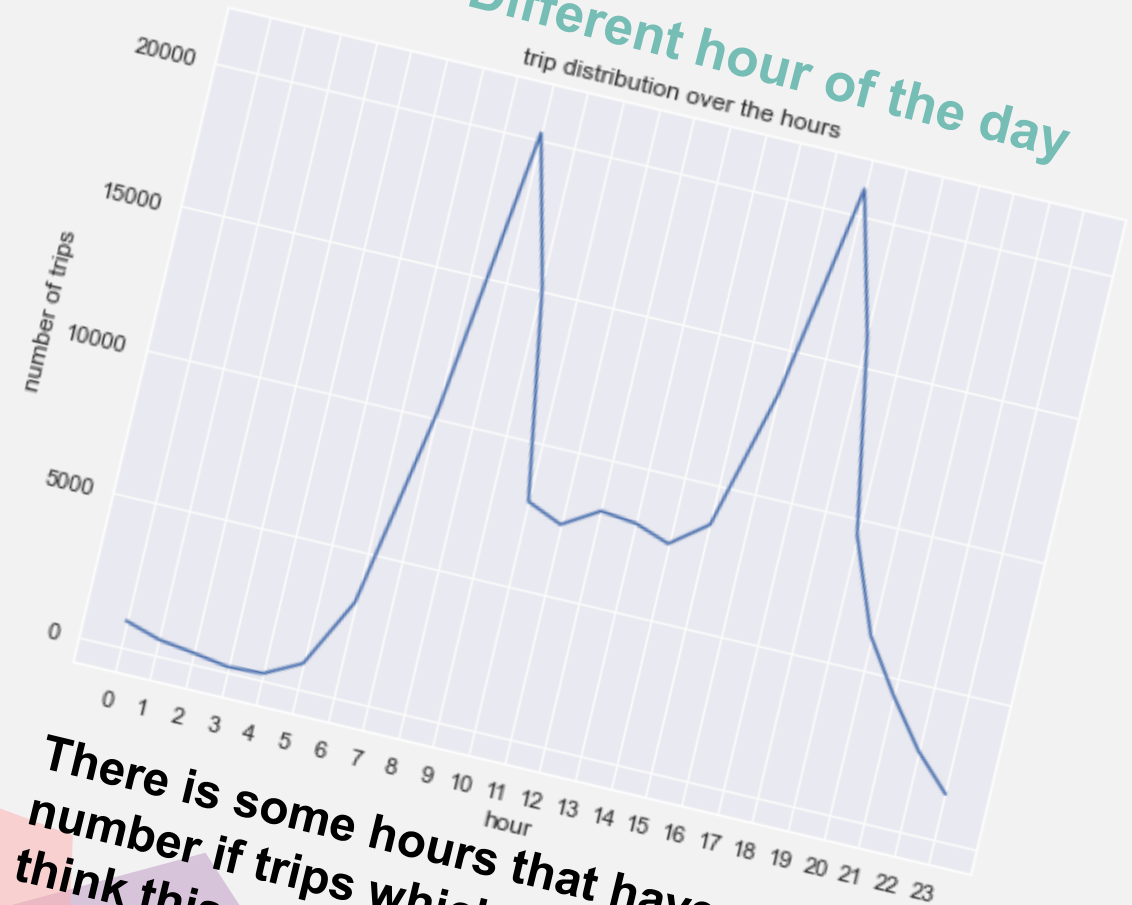
Study number of trips in different times

Different days of the month

trip distribution over the monthe's day

There is some days of the month that have many number if trips as 28 , 30 and other days with small number of trips as 9
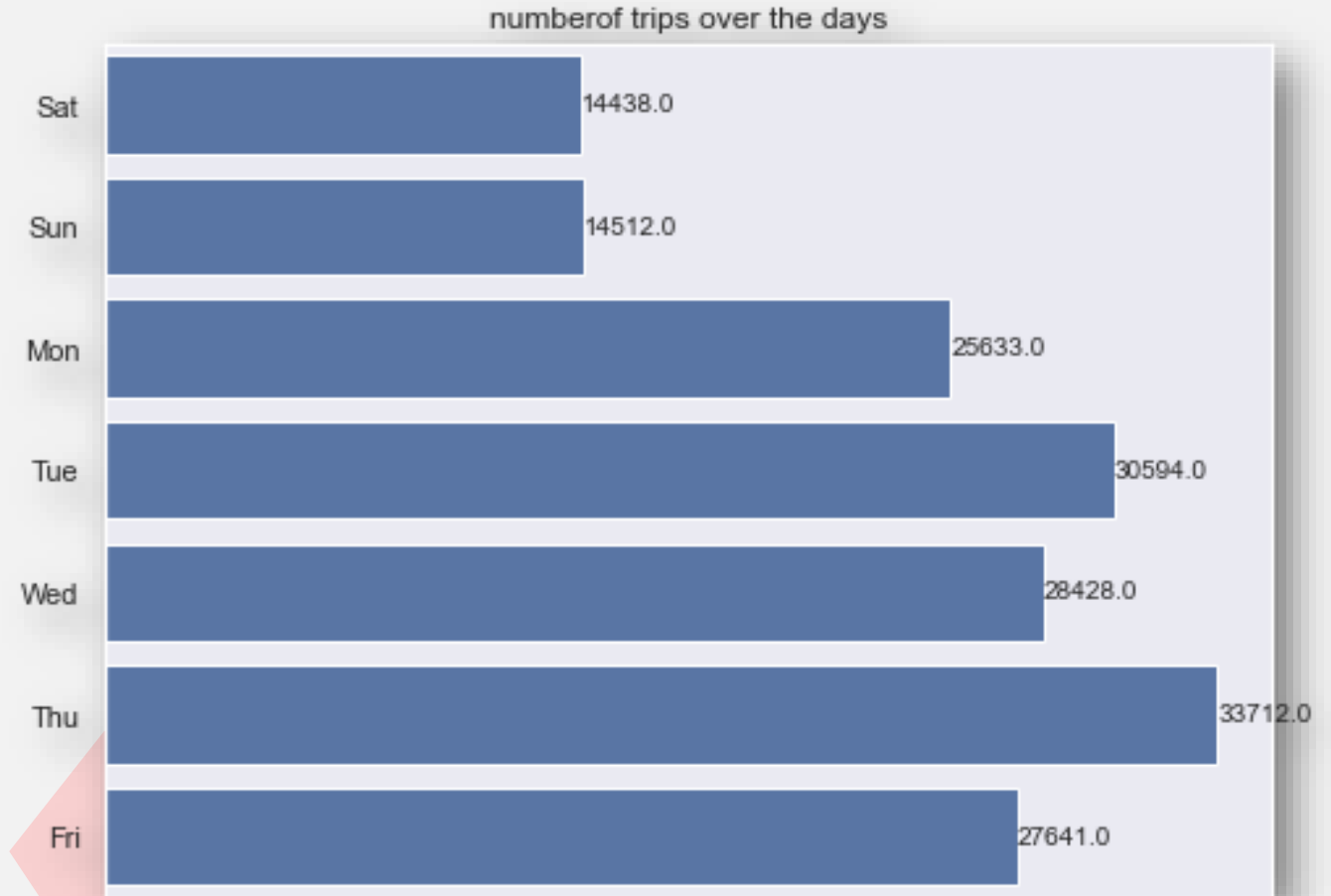
Different hour of the day

trip distribution over the hours

There is some hours that have many number if trips which are 8 ,17 and I think this make sense because it's time for work
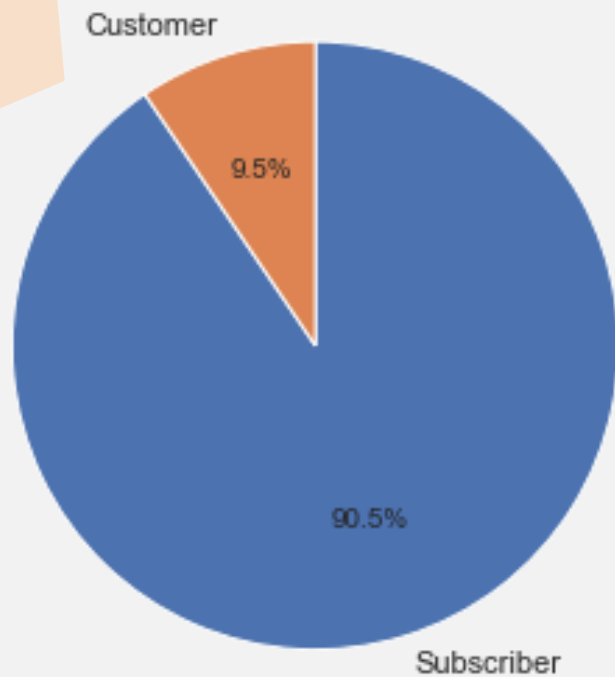
Number of trips per week day

Comment :-

Thursday has the high number of trip in the week and Saturday has the least number of trips .

In my opinion this related to the weekend

numberof trips over the days

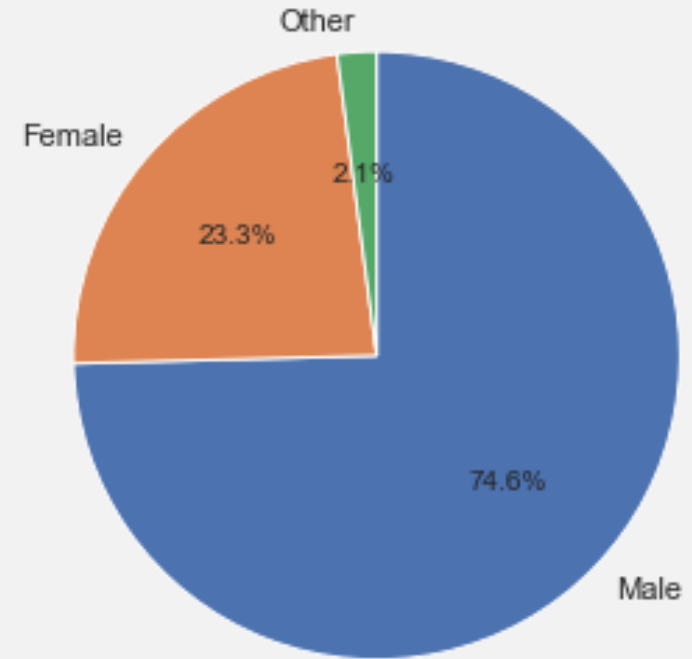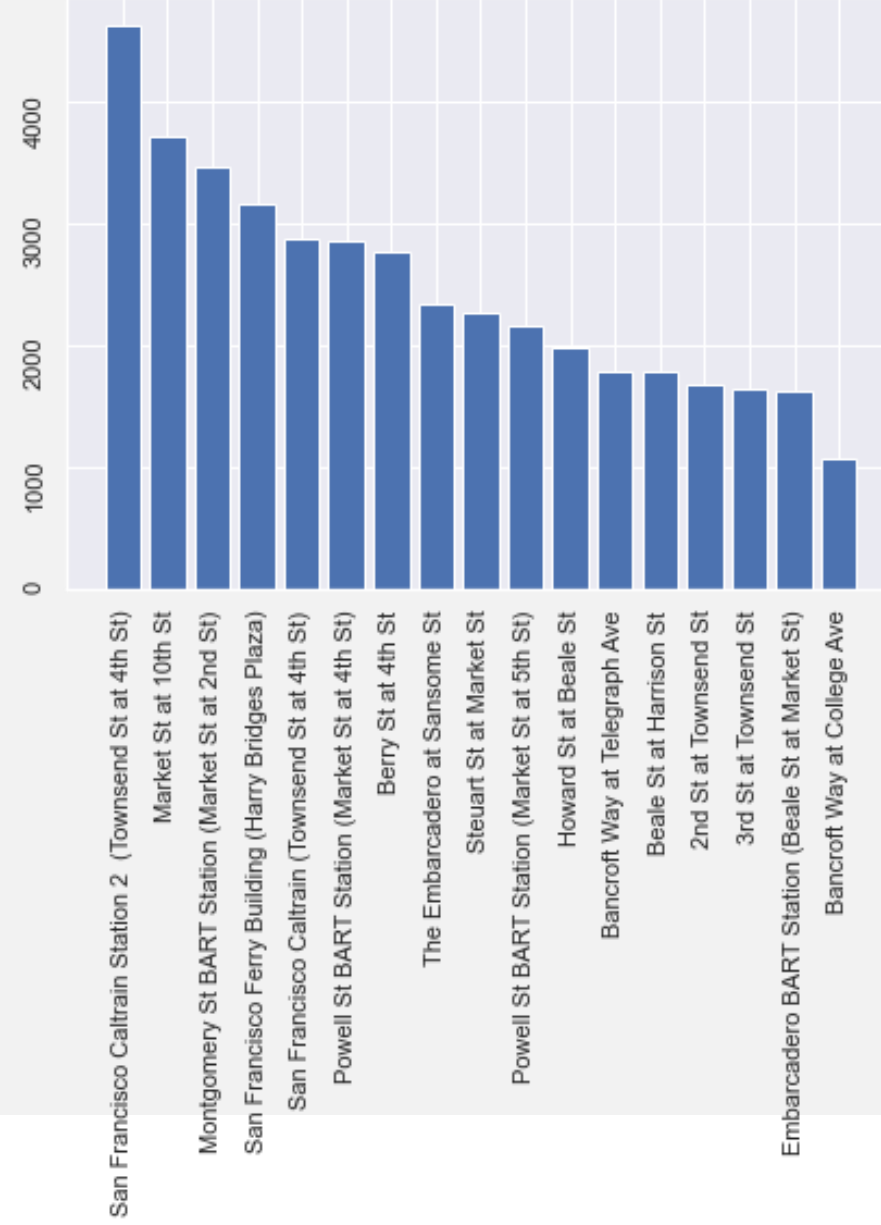| Day	Number of trips
Sat	14438.0
Sun	14512.0
Mon	25633.0
Tue	30594.0
Wed	28428.0
Thu	33712.0
Fri	27641.0

User classification

type
Subscriber type is the most common type more than the customer type

Gender
Men love riding bike more than women so that's normal that the number of men's trips mor than number of women

Most common start and end station names "more than 1500 user "
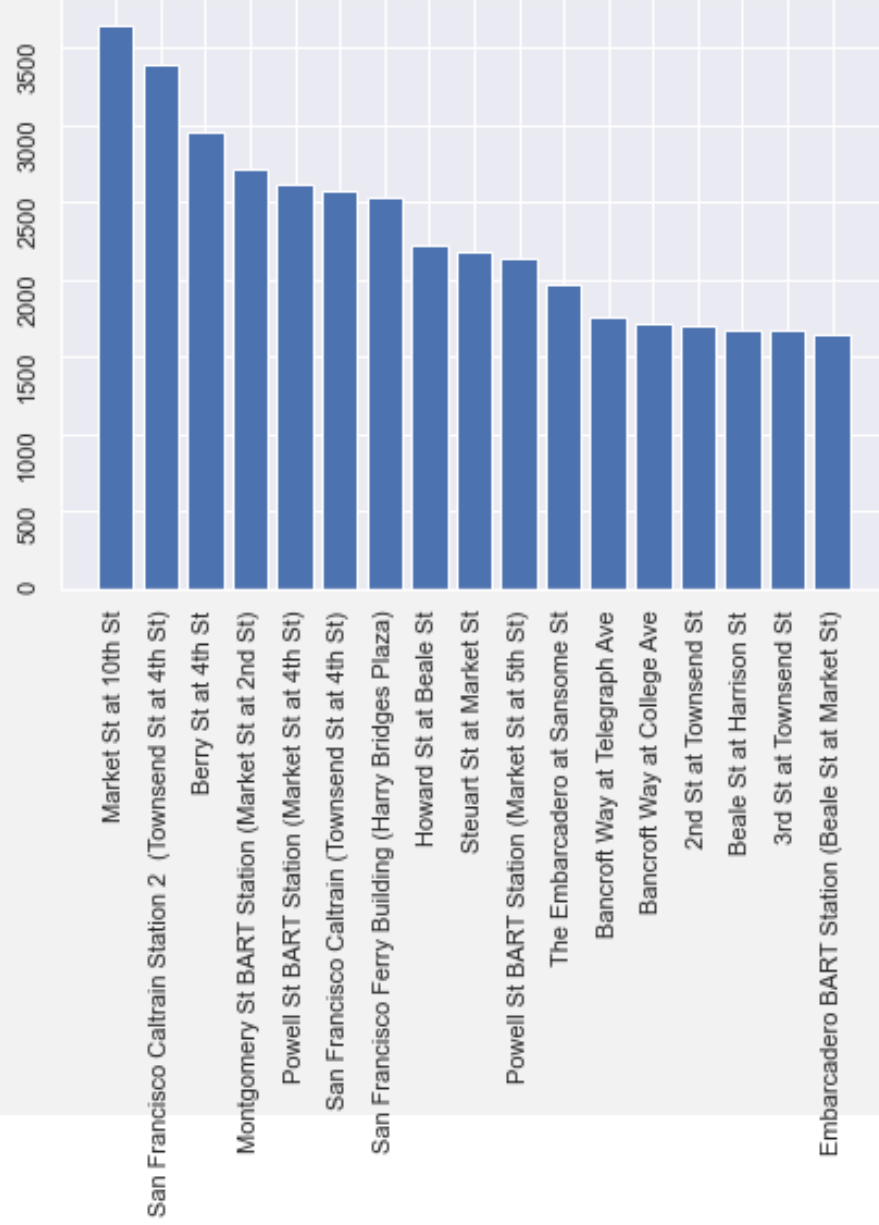
Relation between customer type and trip duration and his speed
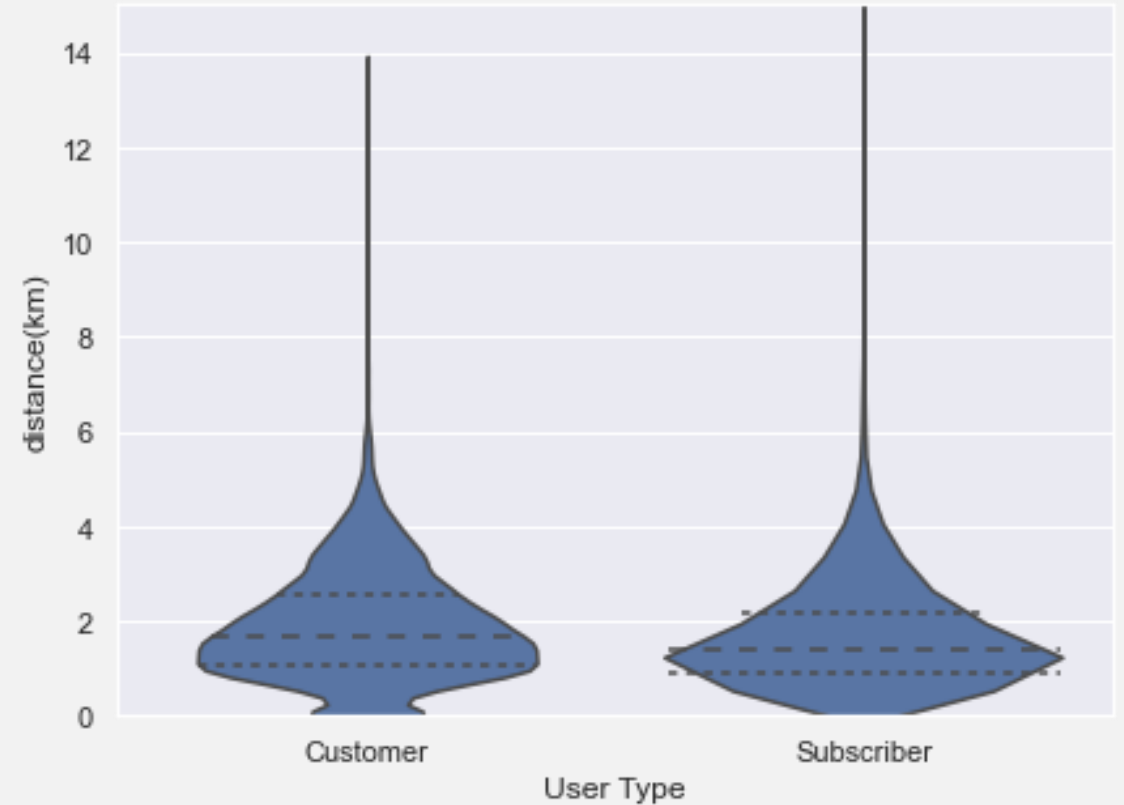
Customer travel in a small average compare to the subscriber and the normal also it has a wide range of speed rather than the customer

Relation between customer gender and trip duration and his speed
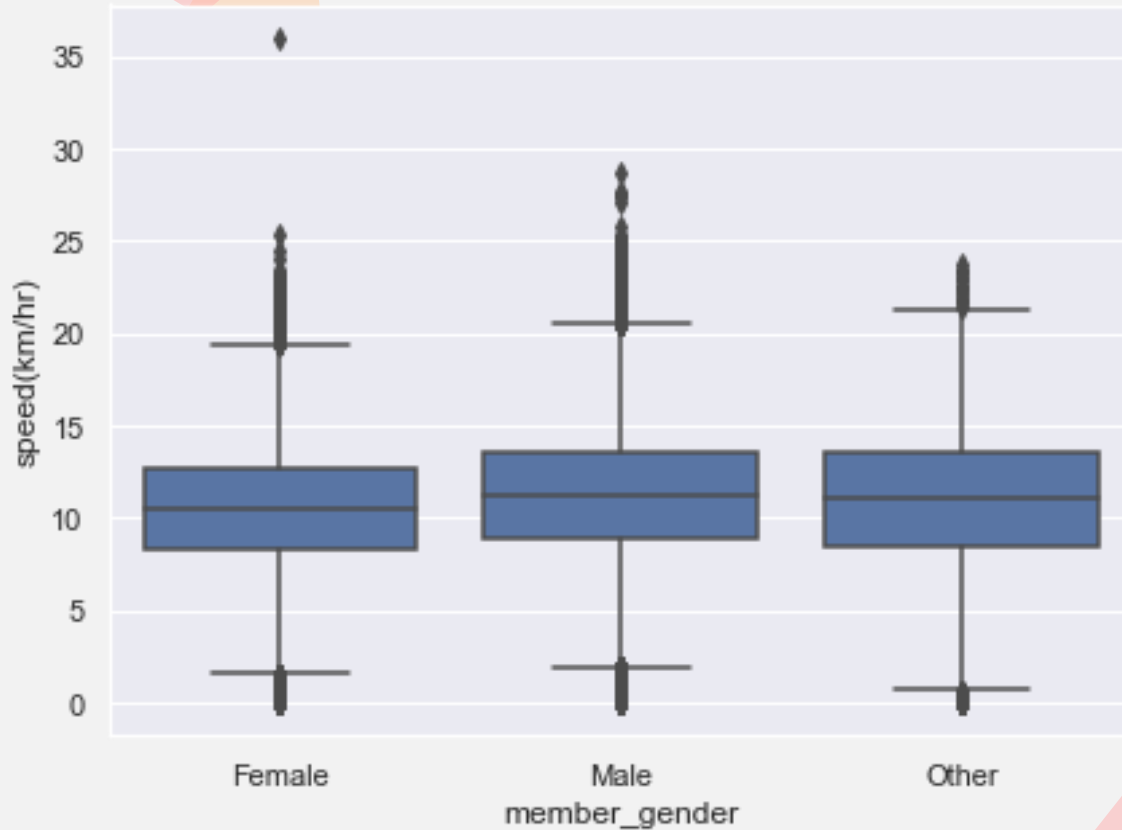
Its clear that male has a bigger speed and longer trips than the female

find the relation between speed , duration and age

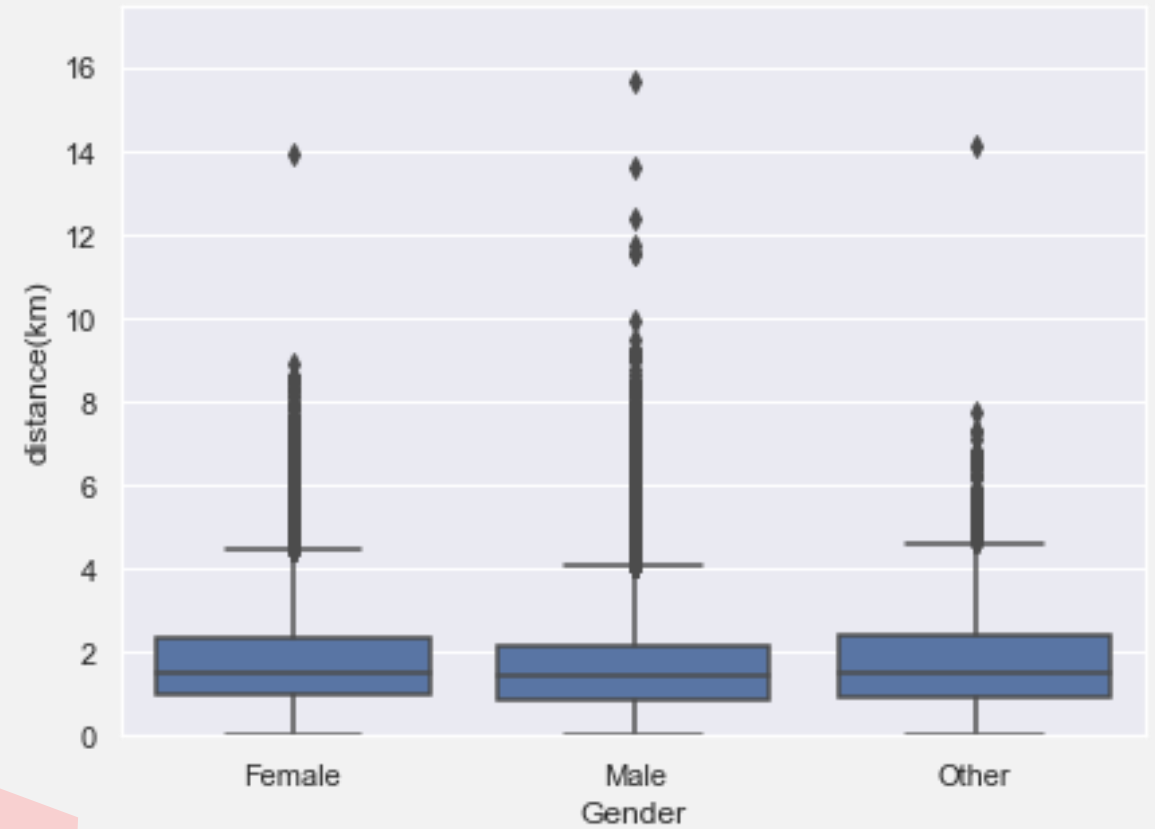relation between duration and age

relation between speed and age
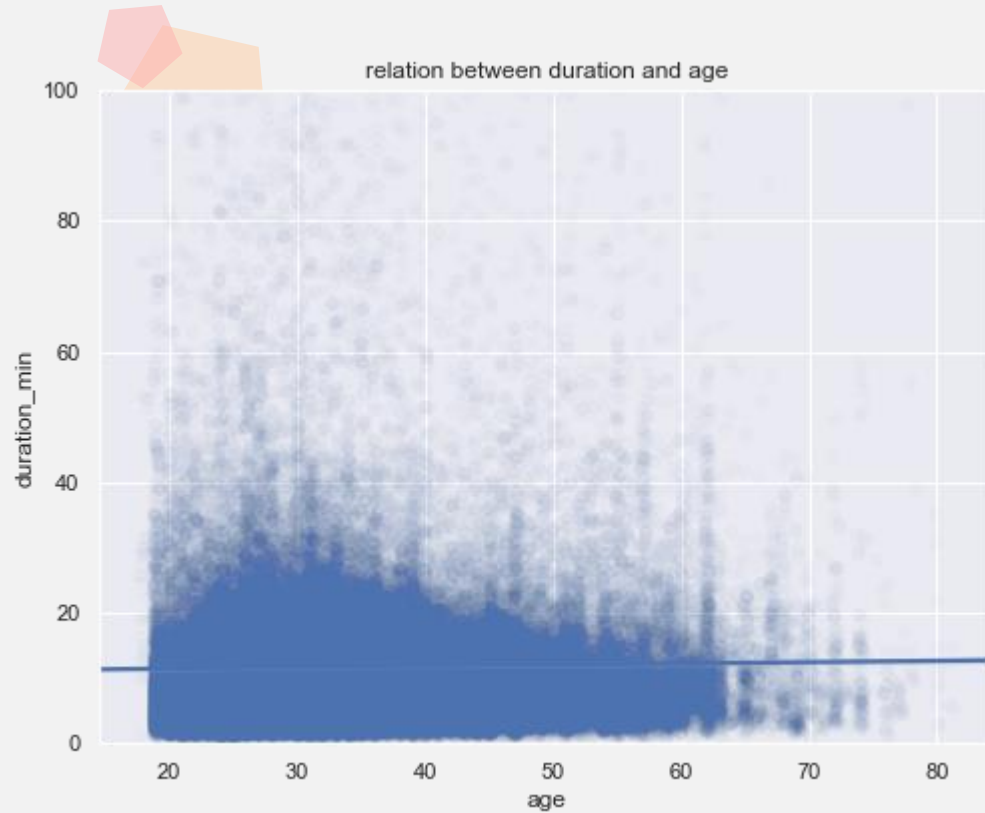
The **duration decrease also** with the increase of age as older people can't ride for long times
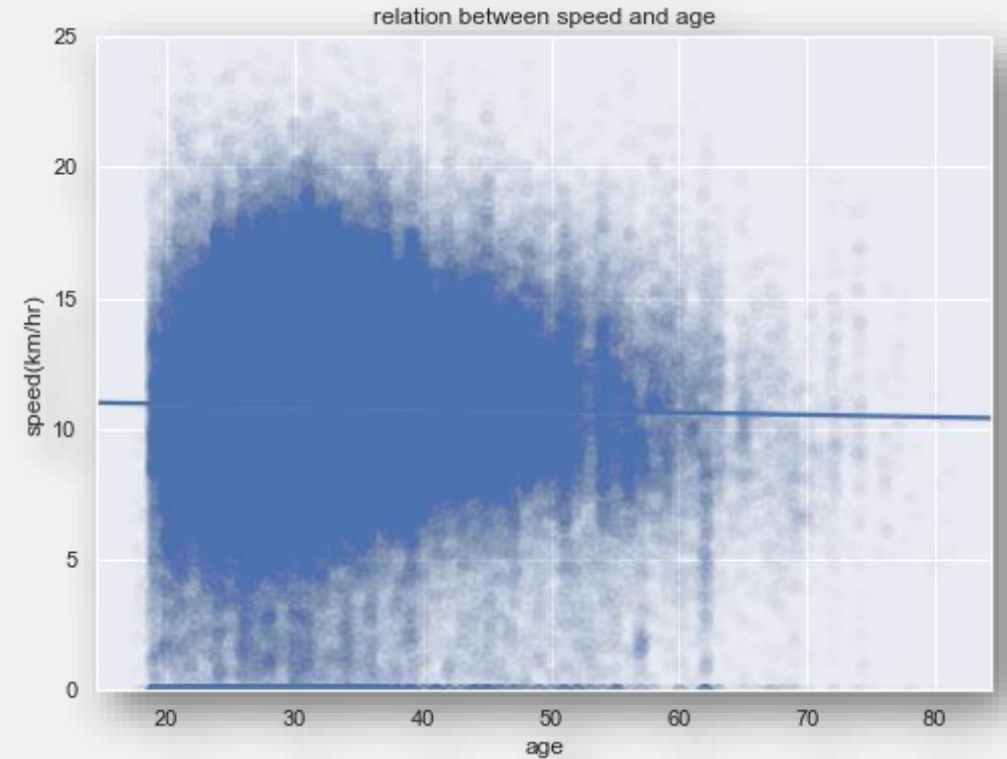
The **speed decrease** with the increase of age and that a normal observation because it is a negative relation between the age and riding bike speed

Relation between day and hour

relation between gender ad user type
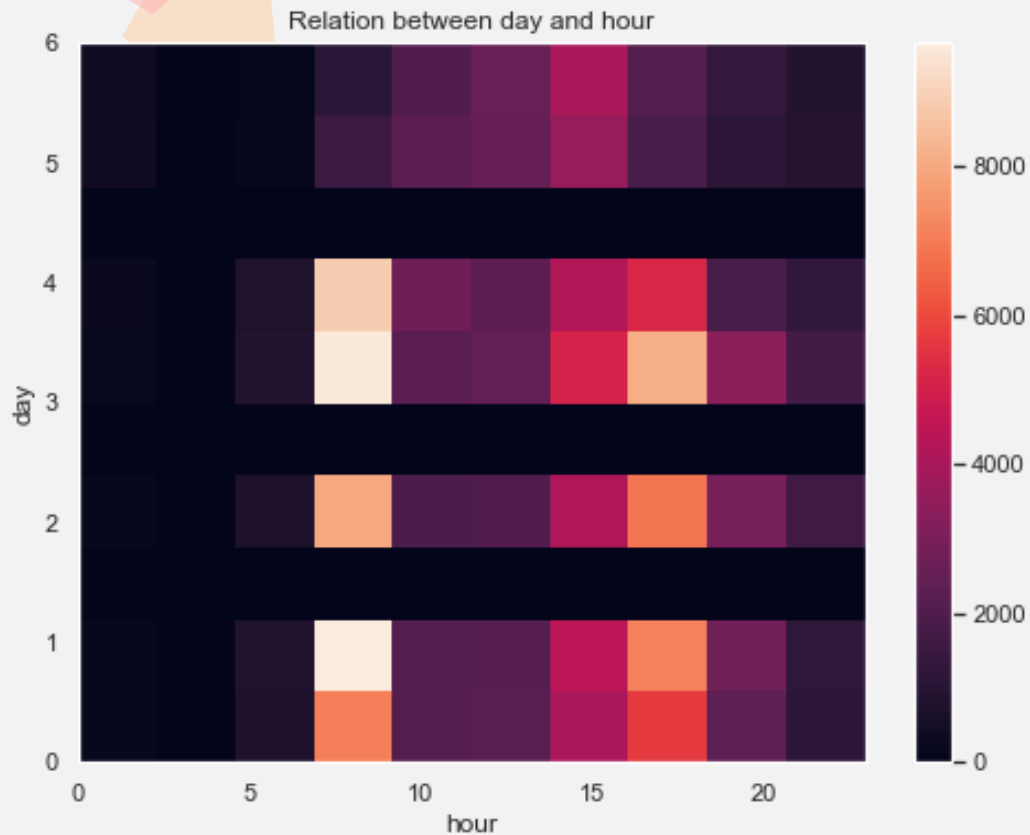
there is some active hour for every day

- there is a relation between gender and type as all gender has many subscriber more than customer

Common start to end combination with more than 200 record

common combination of satrt end station with trip counts

Legend:
- Grand Ave at Perkins St
- Montgomery St BART Station (Market St at 2nd St)
- San Francisco Ferry Building (Harry Bridges Plaza)
- San Fernando St at 4th St
- San Fernando St at 7th St
- Bancroft Way at Telegraph Ave
- San Francisco Caltrain Station 2 (Townsend St at 4th St)
- 19th Street BART Station
- Powell St BART Station (Market St at 4th St)
- 5th St at Virginia St
- The Embarcadero at Sansome St
- Steuart St at Market St

The most common combination is from 19th Street BART Station to Grand Ave at Perkins St
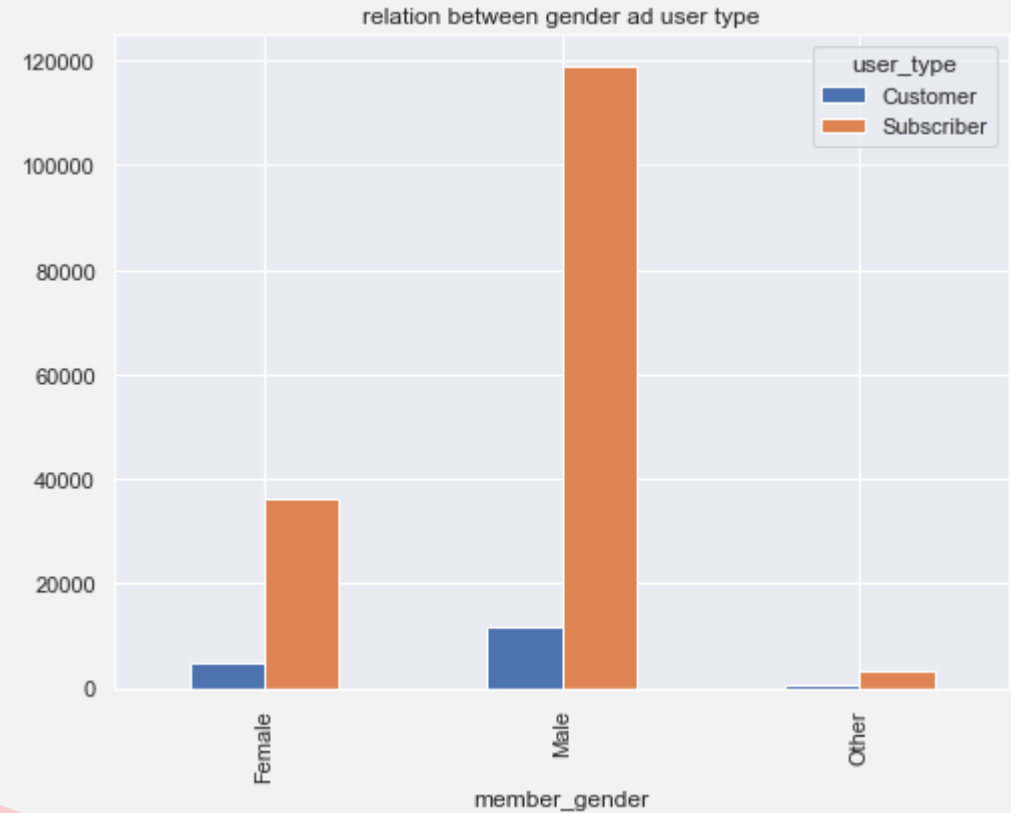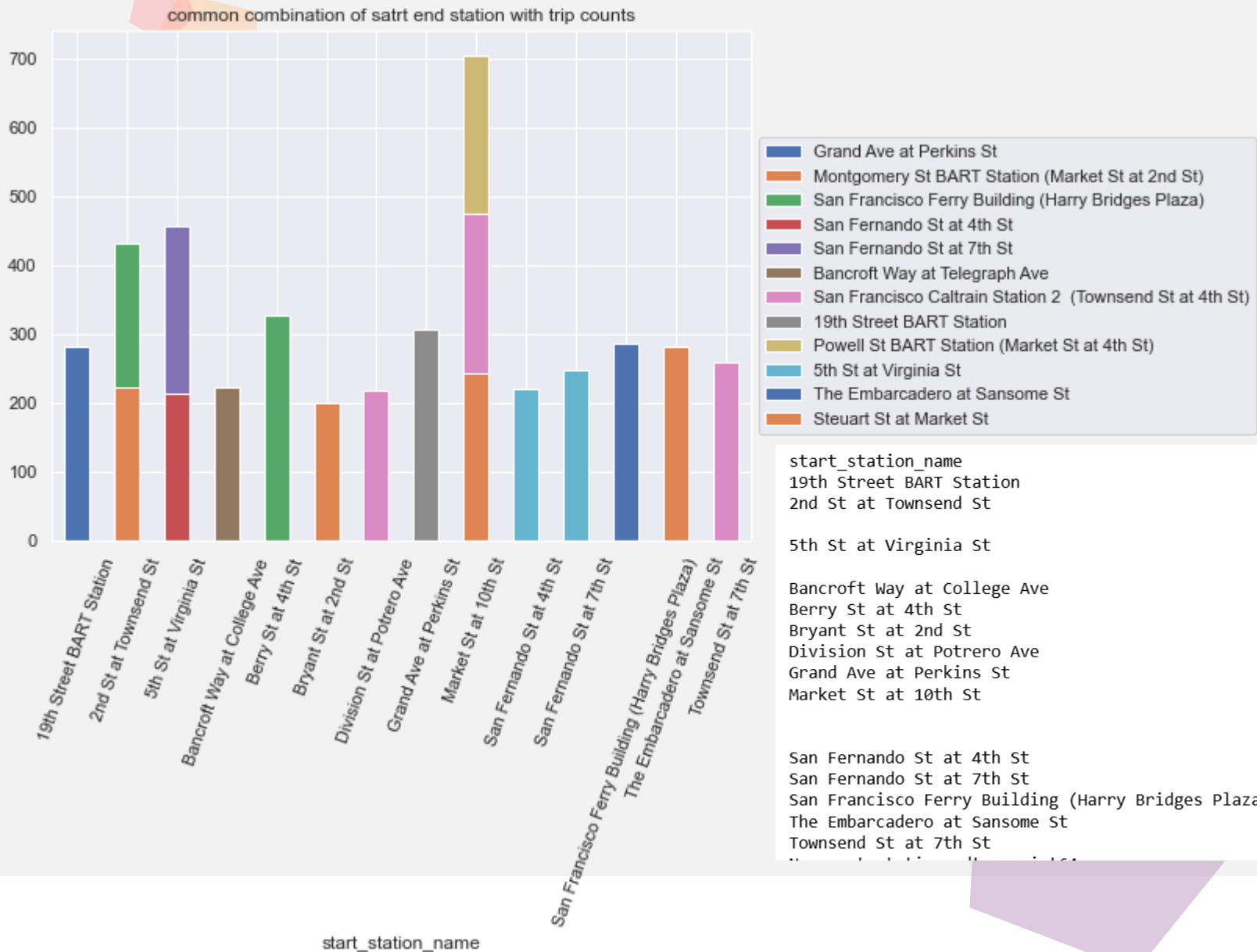
start_station_name	end_station_name	
19th Street BART Station	Grand Ave at Perkins St	283
2nd St at Townsend St	Montgomery St BART Station (Market St at 2nd St)	223
	San Francisco Ferry Building (Harry Bridges Plaza)	210
5th St at Virginia St	San Fernando St at 4th St	215
	San Fernando St at 7th St	242
Bancroft Way at College Ave	Bancroft Way at Telegraph Ave	223
Berry St at 4th St	San Francisco Ferry Building (Harry Bridges Plaza)	327
Bryant St at 2nd St	Montgomery St BART Station (Market St at 2nd St)	201
Division St at Potrero Ave	San Francisco Caltrain Station 2 (Townsend St at 4th St)	218
Grand Ave at Perkins St	19th Street BART Station	308
Market St at 10th St	Montgomery St BART Station (Market St at 2nd St)	243
	Powell St BART Station (Market St at 4th St)	230
	San Francisco Caltrain Station 2 (Townsend St at 4th St)	232
San Fernando St at 4th St	5th St at Virginia St	220
San Fernando St at 7th St	5th St at Virginia St	249
San Francisco Ferry Building (Harry Bridges Plaza)	The Embarcadero at Sansome St	286
The Embarcadero at Sansome St	Steuart St at Market St	282
Townsend St at 7th St	San Francisco Caltrain Station 2 (Townsend St at 4th St)	260

relation between distance and other parameter

distance vs age

In general the trip distance decrease with the increase of age but it's clear that subscribers has big distance than the customer and also male has a bigger distance than females

Data cleaning process

② Convert data type

```
start_time        object
bike_id           int64
user_type         object
member_gender     object
```

```
start_time        datetime64[ns]
bike_id                   object
user_type               category
member_gender           category
```

① Delete unneeded column

'end_time' , 'start_station_id'
,'end_station_id'
,'bike_share_for_all_trip'
I will not use them

④ make new columns

③ invalid birth of year

There is some birth of year smaller than 1934 which mean the customer is older than 85

speed(km/hr)
•Distance / duration

distance
•Calculate by the langitude and longitude

age
•Trip year – birth year

$$d = 2r \arcsin\left(\sqrt{\sin^2\left(\frac{\phi_2 - \phi_1}{2}\right) + \cos(\phi_1)\cos(\phi_2)\sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)}\right)$$