Part_I_exploration_FordGo Bike

October 23, 2022

1 Part I - FORD GO BIKE DATASET EXPLORATION

1.1 by ODINIGWE CHINYERE ROSEMARY

1.2 Introduction

Ford GoBike, like other bike share systems, consists of a fleet of specially designed, sturdy and durable bikes that are locked into a network of docking stations throughout the city. The bikes can be unlocked from one station and returned to any other station in the system, making them ideal for one-way trips. The bikes are available for use 24 hours/day, 7 days/week, 365 days/year and riders have access to all bikes in the network when they become a member or purchase a pass. This dataset includes information about individual rides made in a bike-sharing/rental system covering the greater San Francisco Bay area.This dataset basically contains 183412 bike ride records with the features below;

1.3 Variables

- **duration_sec** : Trip duration
- **start_time** : Trip starting time
- **end_time** : Trip end time
- **start_station_id** : Unique ID of trip start station
- **start_station_name** : name of trip start station
- **start_station_latitude** : Latitude of trip start station
- **start_station_longitude** : Longitude of trip start station
- **end_station_id** : Unique ID of trip end station
- **end_station_name** : name of trip end station
- **end_station_latitude** : Latitude of trip end station
- **end_station_longitude** : Longitude of trip end station
- **bike_id** : Unique ID of rented bikes

- **user_type** : Bike user type
  - Subscriber -member (rents bike regulary based on subscription)
  - Customer - a person that rented a bike but might not come back for more rentals)
- **member_birth_year** : Birth year of bike user/member
- **member_gender** : Sex of bike user
- **bike_share_for_all_trip** :Boolean to track members who are enrolled in the "Bike Share for all" program.

1.4 Preliminary Wrangling

```
In [1]: # import all packages and set plots to be embedded inline
        import numpy as np
        import pandas as pd
        import matplotlib.pyplot as plt
        import seaborn as sb

        %matplotlib inline


        import warnings
        warnings.filterwarnings('ignore')
```

1.5 Gathering Data

```
In [2]: #load datasets for flight records in the 19s
        bike_2019=pd.read_csv('201902-fordgobike-tripdata.csv')
```

1.6 Assessing Data

```
In [3]: #check the first 5 observations of the dataset
        bike_2019.head()
```

```
Out[3]:    duration_sec                start_time                  end_time  \
        0         52185  2019-02-28 17:32:10.1450  2019-03-01 08:01:55.9750
        1         42521  2019-02-28 18:53:21.7890  2019-03-01 06:42:03.0560
        2         61854  2019-02-28 12:13:13.2180  2019-03-01 05:24:08.1460
        3         36490  2019-02-28 17:54:26.0100  2019-03-01 04:02:36.8420
        4          1585  2019-02-28 23:54:18.5490  2019-03-01 00:20:44.0740

           start_station_id                                start_station_name  \
        0              21.0  Montgomery St BART Station (Market St at 2nd St)
        1              23.0                      The Embarcadero at Steuart St
        2              86.0                            Market St at Dolores St
        3             375.0                            Grove St at Masonic Ave
        4               7.0                                Frank H Ogawa Plaza
```

```
   start_station_latitude  start_station_longitude  end_station_id  \
0               37.789625              -122.400811            13.0
1               37.791464              -122.391034            81.0
2               37.769305              -122.426826             3.0
3               37.774836              -122.446546            70.0
4               37.804562              -122.271738           222.0

                             end_station_name  end_station_latitude  \
0               Commercial St at Montgomery St             37.794231
1                         Berry St at 4th St             37.775880
2  Powell St BART Station (Market St at 4th St)             37.786375
3                      Central Ave at Fell St             37.773311
4                       10th Ave at E 15th St             37.792714

   end_station_longitude  bike_id   user_type  member_birth_year  \
0            -122.402923     4902    Customer             1984.0
1            -122.393170     2535    Customer                NaN
2            -122.404904     5905    Customer             1972.0
3            -122.444293     6638  Subscriber             1989.0
4            -122.248780     4898  Subscriber             1974.0

  member_gender bike_share_for_all_trip
0          Male                     No
1           NaN                     No
2          Male                     No
3         Other                     No
4          Male                    Yes
```

In [4]:
```
#check the last 5 observations of the dataset
bike_2019.tail()
```

Out[4]:
```
        duration_sec               start_time                   end_time  \
183407           480  2019-02-01 00:04:49.7240  2019-02-01 00:12:50.0340
183408           313  2019-02-01 00:05:34.7440  2019-02-01 00:10:48.5020
183409           141  2019-02-01 00:06:05.5490  2019-02-01 00:08:27.2200
183410           139  2019-02-01 00:05:34.3600  2019-02-01 00:07:54.2870
183411           271  2019-02-01 00:00:20.6360  2019-02-01 00:04:52.0580

        start_station_id                                start_station_name  \
183407              27.0                           Beale St at Harrison St
183408              21.0  Montgomery St BART Station (Market St at 2nd St)
183409             278.0                             The Alameda at Bush St
183410             220.0                       San Pablo Ave at MLK Jr Way
183411              24.0                             Spear St at Folsom St

        start_station_latitude  start_station_longitude  end_station_id  \
183407               37.788059              -122.391865           324.0
```

```
183408              37.789625              -122.400811            66.0
183409              37.331932              -121.904888           277.0
183410              37.811351              -122.273422           216.0
183411              37.789677              -122.390428            37.0

                          end_station_name  end_station_latitude  \
183407  Union Square (Powell St at Post St)             37.788300
183408               3rd St at Townsend St             37.778742
183409           Morrison Ave at Julian St             37.333658
183410            San Pablo Ave at 27th St             37.817827
183411                 2nd St at Folsom St             37.785000

        end_station_longitude  bike_id   user_type  member_birth_year  \
183407            -122.408531     4832  Subscriber             1996.0
183408            -122.392741     4960  Subscriber             1984.0
183409            -121.908586     3824  Subscriber             1990.0
183410            -122.275698     5095  Subscriber             1988.0
183411            -122.395936     1057  Subscriber             1989.0

       member_gender bike_share_for_all_trip
183407          Male                      No
183408          Male                      No
183409          Male                     Yes
183410          Male                      No
183411          Male                      No
```

In [5]:
```
#check random 5 observations of the dataset
bike_2019.sample()
```

Out[5]:
```
       duration_sec                start_time                  end_time  \
43607           809  2019-02-22 16:15:10.8090  2019-02-22 16:28:40.1590

       start_station_id start_station_name  start_station_latitude  \
43607             104.0  4th St at 16th St               37.767045

       start_station_longitude  end_station_id        end_station_name  \
43607              -122.390833           133.0  Valencia St at 22nd St

       end_station_latitude  end_station_longitude  bike_id   user_type  \
43607             37.755213            -122.420975     4617  Subscriber

       member_birth_year member_gender bike_share_for_all_trip
43607             1980.0         Other                     No
```

In [6]:
```
#get info of the various features(variables)
bike_2019.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 183412 entries, 0 to 183411
```

```
Data columns (total 16 columns):
duration_sec               183412 non-null int64
start_time                 183412 non-null object
end_time                   183412 non-null object
start_station_id           183215 non-null float64
start_station_name         183215 non-null object
start_station_latitude     183412 non-null float64
start_station_longitude    183412 non-null float64
end_station_id             183215 non-null float64
end_station_name           183215 non-null object
end_station_latitude       183412 non-null float64
end_station_longitude      183412 non-null float64
bike_id                    183412 non-null int64
user_type                  183412 non-null object
member_birth_year          175147 non-null float64
member_gender              175147 non-null object
bike_share_for_all_trip    183412 non-null object
dtypes: float64(7), int64(2), object(7)
memory usage: 22.4+ MB
```

- The start and end time datatype needs to be changed to datetime format
- The start,end station id and bike id needs to be changed to strings
- The user type and member gender needs to be changed to categorical datatype
- The member birth year datatype needs to be changed to integers

```
In [7]: #get shape of dataset
        bike_2019.shape

Out[7]: (183412, 16)

In [8]: # Check for missing values
        #get info of the various features(variables)
        bike_2019.isnull().sum()

Out[8]: duration_sec                0
        start_time                 0
        end_time                   0
        start_station_id         197
        start_station_name       197
        start_station_latitude     0
        start_station_longitude    0
        end_station_id           197
        end_station_name         197
        end_station_latitude       0
        end_station_longitude      0
        bike_id                    0
        user_type                  0
        member_birth_year       8265
```

```
member_gender              8265
bike_share_for_all_trip       0
dtype: int64
```

We have null values present in start station id & name, end station id & name, member birth year and gender

```
In [9]: # lets get the descriptive analysis of the numerical data

        bike_2019.describe()
```

```
Out[9]:         duration_sec  start_station_id  start_station_latitude  \
        count  183412.000000     183215.000000           183412.000000
        mean      726.078435        138.590427               37.771223
        std      1794.389780        111.778864                0.099581
        min        61.000000          3.000000               37.317298
        25%       325.000000         47.000000               37.770083
        50%       514.000000        104.000000               37.780760
        75%       796.000000        239.000000               37.797280
        max     85444.000000        398.000000               37.880222

               start_station_longitude  end_station_id  end_station_latitude  \
        count            183412.000000   183215.000000         183412.000000
        mean               -122.352664      136.249123             37.771427
        std                   0.117097      111.515131              0.099490
        min                -122.453704        3.000000             37.317298
        25%                -122.412408       44.000000             37.770407
        50%                -122.398285      100.000000             37.781010
        75%                -122.286533      235.000000             37.797320
        max                -121.874119      398.000000             37.880222

               end_station_longitude        bike_id  member_birth_year
        count          183412.000000  183412.000000      175147.000000
        mean             -122.352250    4472.906375        1984.806437
        std                 0.116673    1664.383394          10.116689
        min              -122.453704      11.000000        1878.000000
        25%              -122.411726    3777.000000        1980.000000
        50%              -122.398279    4958.000000        1987.000000
        75%              -122.288045    5502.000000        1992.000000
        max              -121.874119    6645.000000        2001.000000
```

```
In [10]: # lets check for duplicates

         bike_2019.duplicated().sum()
```

```
Out[10]: 0
```

```
In [11]: #lets get the value counts for the members birth year
         bike_2019['member_birth_year' ].value_counts()
```

```
Out[11]: 1988.0    10236
         1993.0     9325
         1989.0     8972
         1990.0     8658
         1991.0     8498
         1992.0     8250
         1987.0     8018
         1986.0     7973
         1994.0     7660
         1995.0     7423
         1985.0     7028
         1984.0     6562
         1983.0     5954
         1980.0     5024
         1982.0     4990
         1996.0     4640
         1981.0     4345
         1979.0     3756
         1997.0     3481
         1998.0     3208
         1978.0     2830
         1977.0     2725
         1974.0     2633
         1999.0     2528
         1975.0     2503
         1976.0     2442
         1973.0     2080
         1968.0     1928
         1971.0     1924
         1972.0     1909
                    ...
         1954.0      301
         1952.0      189
         1951.0      180
         1950.0      178
         1953.0      158
         1947.0      135
         1955.0      134
         1945.0      105
         1949.0       99
         1931.0       89
         1900.0       53
         1948.0       51
         2001.0       34
         1943.0       30
         1942.0       21
         1933.0       20
         1946.0       19
```

```
1902.0       11
1939.0       11
1941.0        9
1901.0        6
1938.0        3
1920.0        3
1934.0        2
1944.0        2
1930.0        1
1910.0        1
1927.0        1
1928.0        1
1878.0        1
Name: member_birth_year, Length: 75, dtype: int64
```

From the information above,the **1988** has the highest number of members.

```
In [12]: #lets get the value counts for the members gender
         bike_2019['member_gender' ].value_counts()

Out[12]: Male      130651
         Female     40844
         Other       3652
         Name: member_gender, dtype: int64

In [13]: #lets get the value counts for the user type
         bike_2019['user_type' ].value_counts()

Out[13]: Subscriber    163544
         Customer       19868
         Name: user_type, dtype: int64

In [14]: #lets get the unique start station names
         bike_2019['start_station_name' ].unique()

Out[14]: array(['Montgomery St BART Station (Market St at 2nd St)',
                'The Embarcadero at Steuart St', 'Market St at Dolores St',
                'Grove St at Masonic Ave', 'Frank H Ogawa Plaza',
                '4th St at Mission Bay Blvd S', 'Palm St at Willow St',
                'Washington St at Kearny St', 'Post St at Kearny St',
                'Jones St at Post St',
                'Civic Center/UN Plaza BART Station (Market St at McAllister St)',
                'Valencia St at 21st St', 'Channing Way at Shattuck Ave',
                'Bancroft Way at College Ave', 'Howard St at Mary St',
                '22nd St at Dolores St', 'Laguna St at Hayes St',
                '5th St at Folsom', 'Telegraph Ave at 23rd St',
                'Page St at Scott St', 'Lake Merritt BART Station',
                'West St at 40th St', 'The Embarcadero at Sansome St',
                'Folsom St at 9th St', 'University Ave at Oxford St',
```

```
'MLK Jr Way at University Ave', 'The Embarcadero at Bryant St',
'17th St at Valencia St', 'Valencia St at 16th St',
'Valencia St at 22nd St', 'Franklin Square',
'San Pablo Ave at MLK Jr Way', '19th St at Mission St',
'Market St at 10th St', 'Folsom St at 13th St',
'San Francisco Ferry Building (Harry Bridges Plaza)',
'4th St at 16th St', 'Beale St at Harrison St',
'Broadway at Battery St', 'Cesar Chavez St at Dolores St',
'San Fernando St at 4th St', 'Grove St at Divisadero',
'Sanchez St at 17th St', 'Harmon St at Adeline St',
'Mission Playground', 'Davis St at Jackson St',
'Haste St at Telegraph Ave', 'Howard St at 8th St',
'Folsom St at 3rd St', 'Father Alfred E Boeddeker Park',
'Commercial St at Montgomery St', 'Hubbell St at 16th St',
'San Francisco Public Library (Grove St at Hyde St)',
'Bancroft Way at Telegraph Ave', '19th Street BART Station',
'18th St at Noe St', 'Hyde St at Post St', '24th St at Market St',
'Vine St at Shattuck Ave',
'San Francisco Caltrain (Townsend St at 4th St)',
'Valencia St at Clinton Park',
'Union Square (Powell St at Post St)', 'Broderick St at Oak St',
'San Francisco Caltrain Station 2  (Townsend St at 4th St)',
'North Berkeley BART Station', 'Downtown Berkeley BART',
'Fell St at Stanyan St', 'San Salvador St at 9th St',
'Marston Campbell Park', 'Oregon St at Adeline St',
'11th St at Natoma St', 'Harrison St at 20th St',
'Haste St at College Ave', '24th St at Bartlett St',
'Sanchez St at 15th St', 'Telegraph Ave at 19th St',
'Powell St BART Station (Market St at 5th St)',
'Jersey St at Castro St', 'Pierce St at Haight St',
'MacArthur BART Station', 'El Embarcadero at Grand Ave',
'23rd St at San Bruno Ave', 'Golden Gate Ave at Hyde St',
'S Van Ness Ave at Market St', 'Jackson Playground',
'San Fernando St at 7th St', 'West St at University Ave',
'Myrtle St at Polk St', 'Woolsey St at Sacramento St',
'Townsend St at 7th St', 'Harrison St at 17th St',
'West Oakland BART Station', 'Cyril Magnin St at Ellis St',
'Fulton St at Bancroft Way', '14th St at Mission St',
'San Pedro Square', 'Market St at Franklin St',
'Folsom St at 19th St', 'College Ave at Taft Ave',
'Rhode Island St at 17th St', 'Shattuck Ave at Hearst Ave',
'The Embarcadero at Vallejo St', 'Webster St at Grove St',
'Raymond Kimbell Playground', 'Victoria Manalo Draves Park',
'20th St at Bryant St', 'S Park St at 3rd St',
'Lakeshore Ave at Trestle Glen Rd', 'Channing Way at San Pablo Ave',
'Mission Dolores Park', 'Lombard St at Columbus Ave',
'17th St at Dolores St', 'Precita Park', 'Central Ave at Fell St',
'4th St at Harrison St', 'Horton St at 40th St',
```

```
'Golden Gate Ave at Franklin St',
'Embarcadero BART Station (Beale St at Market St)',
'9th St at San Fernando St', '3rd St at Townsend St',
'McCoppin St at Valencia St', '13th St at Franklin St',
'Mission Bay Kids Park', 'Potrero Ave and Mariposa St',
'Emeryville Public Market', 'Union St at 10th St',
'Jackson St at 11th St', 'Broadway at Kearny',
'Paseo De San Antonio at 2nd St', 'Valencia St at Cesar Chavez St',
'Rockridge BART Station', '8th St at Brannan St',
'College Ave at Alcatraz Ave', '16th St Mission BART Station 2',
'San Jose Diridon Station', 'Masonic Ave at Turk St',
'17th & Folsom Street Park (17th St at Folsom St)',
'Grand Ave at Webster St', '7th St at Brannan St',
'Steuart St at Market St', 'Scott St at Golden Gate Ave',
'Parker St at Fulton St', 'Berkeley Civic Center',
'Clay St at Battery St', '11th St at Bryant St',
'Powell St BART Station (Market St at 4th St)',
'Doyle St at 59th St', '34th St at Telegraph Ave', 'Esprit Park',
'Emeryville Town Hall', 'Division St at Potrero Ave',
'Irwin St at 8th St', 'Pierce Ave at Market St',
'Howard St at Beale St', 'Washington St at 8th St', 'Snow Park',
'Dolores St at 15th St', 'Hearst Ave at Euclid Ave',
'Telegraph Ave at Ashby Ave', '8th St at Ringold St',
'14th St at Mandela Pkwy', 'Morrison Ave at Julian St',
'Church St at Duboce Ave', 'Townsend St at 5th St',
'Valencia St at 24th St', '16th St at Prosper St',
'5th St at Virginia St', "Webster St at O'Farrell St",
'Shattuck Ave at Telegraph Ave', 'Jackson St at 5th St',
'Berry St at 4th St', '2nd St at Townsend St',
'Telegraph Ave at Carleton St', 'Ellsworth St at Russell St',
'Adeline St at 40th St', 'Bay Pl at Vernon St',
'Russell St at College Ave', '22nd St Caltrain Station',
'Folsom St at 15th St', nan, 'Ninth St at Heinz Ave',
'15th St at Potrero Ave', '23rd St at Tennessee St',
'McAllister St at Baker St', 'Bryant St at 2nd St',
'Mississippi St at 17th St', 'Ryland Park',
'Fountain Alley at S 2nd St', 'Turk St at Fillmore St',
'Ashby BART Station', 'Shattuck Ave at 51st St',
'Julian St at The Alameda', '20th St at Dolores St',
'Broadway at Coronado Ave', 'Grand Ave at Santa Clara Ave',
'Eureka Valley Recreation Center', 'Parker Ave at McAllister St',
'Berry St at King St',
'Salesforce Transit Center (Natoma St at 2nd St)',
'San Antonio Park', 'Lakeside Dr at 14th St',
'16th St Mission BART', 'Stanford Ave at Hollis St',
'Broadway at 40th St',
'Mechanics Monument Plaza (Market St at Bush St)',
'Madison St at 17th St', 'Grand Ave at Perkins St',
```

```
'Garfield Square (25th St at Harrison St)', '53rd St at Hollis St',
'2nd St at Julian St', 'Telegraph Ave at Alcatraz Ave',
'San Francisco City Hall (Polk St at Grove St)',
'5th St at Brannan St', '10th St at Fallon St',
'Yerba Buena Center for the Arts (Howard St at 3rd St)',
'30th St at San Jose Ave', '29th St at Tiffany Ave',
'Webster St at 2nd St', 'Koshland Park', 'Jersey St at Church St',
'Santa Clara St at 7th St', 'Telegraph Ave at 58th St',
'Fruitvale BART Station', 'Addison St at Fourth St',
'Leavenworth St at Broadway', 'Telegraph Ave at 27th St',
'Potrero del Sol Park (25th St at Utah St)',
'Spear St at Folsom St', 'College Ave at Harwood Ave',
"O'Farrell St at Divisadero St", '1st St at Folsom St',
'Bryant St at 15th St', 'Golden Gate Ave at Polk St',
'5th St at San Salvador St', '29th St at Church St',
'Alamo Square (Steiner St at Fulton St)',
'Autumn Parkway at Coleman Ave', 'Fulton St at Ashby Ave',
'Howard St at 2nd St', '19th St at Florida St',
'Market St at 45th St', 'Derby St at College Ave',
'Market St at Brockhurst St', 'California St at University Ave',
'MLK Jr Way at 14th St', 'Market St at 40th St',
'Julian St at 6th St', 'Cahill Park', 'San Jose City Hall',
'Virginia St at Shattuck Ave', 'Jack London Square',
'Milvia St at Derby St', 'Webster St at 19th St',
'24th St at Chattanooga St', 'The Alameda at Bush St',
'49th St at Telegraph Ave', 'Broadway at 30th St',
'Bryant St at 6th St', 'Empire St at 1st St',
'China Basin St at 3rd St', '47th St at San Pablo Ave',
'San Salvador St at 1st St', '45th St at Manila',
'San Carlos St at Market St', 'San Pablo Ave at 27th St',
'Market St at Park St', 'Franklin St at 9th St',
'Almaden Blvd at San Fernando St', 'Oak St at 1st St',
'William St at 10th St', 'Isabella St at San Pablo Ave',
'Guerrero Park', '10th St at University Ave', 'DeFremery Park',
'Fifth St at Delaware St', 'Williams Ave at 3rd St',
'4th Ave at E 12th St (Temporary Location)',
'Shattuck Ave at 55th St', '59th St at Horton St', 'SAP Center',
'37th St at West St', 'Almaden Blvd at Balbach St',
'65th St at Hollis St', 'Santa Clara St at Almaden Blvd',
'Ninth St at Parker St', 'Bushrod Park', 'Empire St at 7th St',
'Mendell St at Fairfax Ave', '16th St Depot',
'Newhall St at 3rd St', 'George St at 1st St',
'Mission St at 1st St', 'Duboce Park', 'Locust St at Grant St',
'32nd St at Adeline St', 'Mosswood Park',
'Delmas Ave and San Fernando St', 'Lane St at Revere Ave',
'2nd Ave at E 18th St', 'San Carlos St at 11th St',
'Williams Ave at Apollo St', 'MacArthur Blvd at Telegraph Ave',
'Bestor Art Park', 'College Ave at Bryant Ave',
```

```
       'Miles Ave at Cavour St', 'Saint James Park',
       '14th St at Filbert St', 'Foothill Blvd at Fruitvale Ave',
       'Market St at 8th St', 'Backesto Park (Jackson St at 13th St)',
       '10th Ave at E 15th St', 'Alcatraz Ave at Shattuck Ave',
       '55th St at Telegraph Ave', 'Genoa St at 55th St',
       'Dover St at 57th St', 'San Pablo Park',
       '6th Ave at E 12th St (Temporary Location)', 'Taylor St at 9th St',
       '27th St at MLK Jr Way', 'Foothill Blvd at Harrington Ave',
       '23rd Ave at Foothill Blvd', 'San Pedro St at Hedding St',
       '45th St at MLK Jr Way', '5th St at Taylor St',
       'Foothill Blvd at 42nd Ave', 'Willow St at Vine St',
       '26th Ave at International Blvd', 'Farnam St at Fruitvale Ave',
       '21st Ave at International Blvd', '2nd St at Folsom St'], dtype=object)
```

In [15]:
```
#lets get the unique end station names
bike_2019['end_station_name' ].unique()
```

Out[15]:
```
array(['Commercial St at Montgomery St', 'Berry St at 4th St',
       'Powell St BART Station (Market St at 4th St)',
       'Central Ave at Fell St', '10th Ave at E 15th St',
       'Broadway at Kearny', 'San Jose Diridon Station',
       'Valencia St at 21st St', 'Mission Playground',
       'San Francisco Public Library (Grove St at Hyde St)',
       'Bryant St at 2nd St', 'Shattuck Ave at Hearst Ave',
       'Channing Way at Shattuck Ave', '8th St at Ringold St',
       'Broderick St at Oak St', 'Potrero Ave and Mariposa St',
       'Market St at Franklin St', 'Telegraph Ave at 23rd St',
       '17th St at Dolores St',
       '6th Ave at E 12th St (Temporary Location)',
       'McAllister St at Baker St', 'Telegraph Ave at Carleton St',
       'Genoa St at 55th St', 'Grand Ave at Perkins St',
       'San Francisco Ferry Building (Harry Bridges Plaza)',
       'Folsom St at 9th St', 'Channing Way at San Pablo Ave',
       '2nd St at Townsend St', 'Pierce St at Haight St',
       'Potrero del Sol Park (25th St at Utah St)',
       'Valencia St at 22nd St', 'Jackson Playground',
       'Dolores St at 15th St', '29th St at Church St',
       '19th St at Mission St', 'Bay Pl at Vernon St',
       'Post St at Kearny St',
       'Yerba Buena Center for the Arts (Howard St at 3rd St)',
       '4th St at Mission Bay Blvd S', 'Father Alfred E Boeddeker Park',
       'Market St at 10th St', '24th St at Chattanooga St',
       'Pierce Ave at Market St', 'Fell St at Stanyan St',
       '17th St at Valencia St', 'San Pablo Ave at 27th St',
       'Howard St at Mary St', 'Davis St at Jackson St',
       'Victoria Manalo Draves Park', 'Jersey St at Church St',
       'Haste St at Telegraph Ave', 'Eureka Valley Recreation Center',
       'Washington St at Kearny St', 'Grove St at Divisadero',
```

```
'Parker St at Fulton St', 'El Embarcadero at Grand Ave',
'Lake Merritt BART Station', 'Hyde St at Post St',
'24th St at Market St', '5th St at Brannan St',
'24th St at Bartlett St', 'Townsend St at 5th St',
'Addison St at Fourth St', 'Broadway at Battery St',
'Market St at Dolores St', '5th St at Virginia St',
'Marston Campbell Park', 'University Ave at Oxford St',
'Valencia St at 24th St', 'Valencia St at Cesar Chavez St',
'Ryland Park', 'Precita Park', 'Derby St at College Ave',
'Jersey St at Castro St', '11th St at Natoma St',
'45th St at MLK Jr Way', 'Valencia St at 16th St',
'San Francisco Caltrain Station 2  (Townsend St at 4th St)',
'Montgomery St BART Station (Market St at 2nd St)',
'18th St at Noe St', '37th St at West St', 'Newhall St at 3rd St',
'Haste St at College Ave', 'Cyril Magnin St at Ellis St',
'Beale St at Harrison St', 'Fulton St at Bancroft Way',
'San Fernando St at 4th St',
'Garfield Square (25th St at Harrison St)',
'29th St at Tiffany Ave', 'Bancroft Way at College Ave',
'Ashby BART Station', '11th St at Bryant St',
'14th St at Mandela Pkwy', 'Howard St at 8th St',
'Leavenworth St at Broadway', 'Locust St at Grant St',
'Lombard St at Columbus Ave', 'Sanchez St at 17th St',
'45th St at Manila', '23rd St at San Bruno Ave',
'Morrison Ave at Julian St', 'Sanchez St at 15th St',
'Koshland Park', 'Harrison St at 20th St', '2nd Ave at E 18th St',
'Steuart St at Market St', 'Church St at Duboce Ave',
'Page St at Scott St', 'Bancroft Way at Telegraph Ave',
'Mission Bay Kids Park', 'Folsom St at 3rd St',
'Valencia St at Clinton Park', 'Grand Ave at Santa Clara Ave',
'19th Street BART Station', 'Folsom St at 19th St',
'West Oakland BART Station', 'S Park St at 3rd St',
'5th St at Folsom',
'Embarcadero BART Station (Beale St at Market St)',
'Howard St at 2nd St', 'The Embarcadero at Sansome St',
'Backesto Park (Jackson St at 13th St)', 'Esprit Park',
'Myrtle St at Polk St', 'Franklin Square', 'Empire St at 7th St',
'Lakeside Dr at 14th St', 'Laguna St at Hayes St',
'65th St at Hollis St', '4th St at 16th St',
'49th St at Telegraph Ave', '16th St Mission BART Station 2',
'Lane St at Revere Ave', 'MLK Jr Way at University Ave',
'2nd St at Julian St', 'Webster St at Grove St',
'Telegraph Ave at Ashby Ave', 'Bryant St at 6th St',
'20th St at Dolores St',
'Powell St BART Station (Market St at 5th St)',
'San Francisco Caltrain (Townsend St at 4th St)',
'8th St at Brannan St', 'Shattuck Ave at Telegraph Ave',
'Downtown Berkeley BART', 'North Berkeley BART Station',
```

```
'Turk St at Fillmore St', 'Woolsey St at Sacramento St',
'4th St at Harrison St', 'The Embarcadero at Bryant St',
"O'Farrell St at Divisadero St", 'Grove St at Masonic Ave',
'Hubbell St at 16th St',
'Civic Center/UN Plaza BART Station (Market St at McAllister St)',
'3rd St at Townsend St', 'Fountain Alley at S 2nd St',
'China Basin St at 3rd St', '59th St at Horton St',
'Mission Dolores Park', 'San Carlos St at 11th St',
'Jackson St at 11th St', '22nd St Caltrain Station',
'Townsend St at 7th St', '7th St at Brannan St',
'Webster St at 2nd St', 'Ellsworth St at Russell St',
"Webster St at O'Farrell St", 'Harmon St at Adeline St',
'1st St at Folsom St', 'Vine St at Shattuck Ave',
'Stanford Ave at Hollis St', 'Jones St at Post St',
'West St at University Ave', 'Paseo De San Antonio at 2nd St',
'Duboce Park', 'The Embarcadero at Steuart St',
'Russell St at College Ave', 'Golden Gate Ave at Hyde St',
'Berkeley Civic Center', '47th St at San Pablo Ave',
'George St at 1st St', '53rd St at Hollis St', 'West St at 40th St',
'15th St at Potrero Ave', 'Division St at Potrero Ave',
'San Pablo Ave at MLK Jr Way', 'Jackson St at 5th St',
'Union Square (Powell St at Post St)',
'4th Ave at E 12th St (Temporary Location)', 'Bushrod Park',
'Rhode Island St at 17th St', 'Folsom St at 13th St',
'Virginia St at Shattuck Ave', '16th St Mission BART',
'Lakeshore Ave at Trestle Glen Rd', 'Masonic Ave at Turk St', nan,
'Harrison St at 17th St', 'McCoppin St at Valencia St',
'17th & Folsom Street Park (17th St at Folsom St)',
'10th St at Fallon St', '34th St at Telegraph Ave',
'The Alameda at Bush St', '9th St at San Fernando St',
'20th St at Bryant St', 'Howard St at Beale St',
'Cesar Chavez St at Dolores St', '55th St at Telegraph Ave',
'S Van Ness Ave at Market St', 'Scott St at Golden Gate Ave',
'14th St at Mission St', 'Mississippi St at 17th St',
'Alamo Square (Steiner St at Fulton St)', 'Shattuck Ave at 51st St',
'MacArthur BART Station', 'Madison St at 17th St',
'Horton St at 40th St', 'Hearst Ave at Euclid Ave',
'Folsom St at 15th St', '19th St at Florida St',
'Alcatraz Ave at Shattuck Ave', 'San Fernando St at 7th St',
'MLK Jr Way at 14th St', 'Milvia St at Derby St',
'College Ave at Alcatraz Ave', 'Washington St at 8th St',
'Guerrero Park', 'Oregon St at Adeline St',
'Parker Ave at McAllister St', '23rd St at Tennessee St',
'Clay St at Battery St', 'Broadway at 40th St',
'Salesforce Transit Center (Natoma St at 2nd St)',
'Telegraph Ave at 19th St', 'Emeryville Public Market',
'Golden Gate Ave at Polk St', 'Telegraph Ave at 58th St',
'Foothill Blvd at Harrington Ave', 'The Embarcadero at Vallejo St',
```

```
'16th St at Prosper St', 'Berry St at King St',
'Broadway at Coronado Ave', 'Market St at 45th St',
'Mechanics Monument Plaza (Market St at Bush St)',
'Dover St at 57th St', 'Miles Ave at Cavour St',
'Rockridge BART Station', 'Fifth St at Delaware St',
'College Ave at Harwood Ave', 'California St at University Ave',
'Union St at 10th St', '5th St at San Salvador St', 'Mosswood Park',
'William St at 10th St', '5th St at Taylor St',
'Julian St at The Alameda', 'Irwin St at 8th St',
'Market St at Brockhurst St', 'Adeline St at 40th St',
'30th St at San Jose Ave', 'Spear St at Folsom St',
'27th St at MLK Jr Way',
'San Francisco City Hall (Polk St at Grove St)',
'22nd St at Dolores St', 'Frank H Ogawa Plaza',
'Shattuck Ave at 55th St', 'Golden Gate Ave at Franklin St',
'Broadway at 30th St', 'Bryant St at 15th St',
'Grand Ave at Webster St', 'Julian St at 6th St',
'Santa Clara St at 7th St', '14th St at Filbert St',
'Emeryville Town Hall', 'Cahill Park', 'Raymond Kimbell Playground',
'Autumn Parkway at Coleman Ave', 'Isabella St at San Pablo Ave',
'San Salvador St at 9th St', 'Telegraph Ave at 27th St',
'13th St at Franklin St', 'Doyle St at 59th St',
'Jack London Square', 'SAP Center', 'Telegraph Ave at Alcatraz Ave',
'San Carlos St at Market St', '10th St at University Ave',
'Ninth St at Heinz Ave', 'Market St at 40th St',
'23rd Ave at Foothill Blvd', 'Bestor Art Park',
'32nd St at Adeline St', 'DeFremery Park', 'San Pedro Square',
'San Salvador St at 1st St', 'Fulton St at Ashby Ave',
'Ninth St at Parker St', 'Taylor St at 9th St',
'Empire St at 1st St', 'Franklin St at 9th St',
'Webster St at 19th St', 'San Pablo Park',
'College Ave at Taft Ave', 'Market St at 8th St', 'Snow Park',
'San Antonio Park', 'San Jose City Hall',
'Delmas Ave and San Fernando St', 'Mendell St at Fairfax Ave',
'Santa Clara St at Almaden Blvd', 'College Ave at Bryant Ave',
'Foothill Blvd at Fruitvale Ave', 'Palm St at Willow St',
'Saint James Park', 'Williams Ave at 3rd St',
'Market St at Park St', 'Almaden Blvd at Balbach St',
'Almaden Blvd at San Fernando St', 'Foothill Blvd at 42nd Ave',
'Fruitvale BART Station', 'MacArthur Blvd at Telegraph Ave',
'Williams Ave at Apollo St', 'Mission St at 1st St',
'San Pedro St at Hedding St', 'Oak St at 1st St',
'Farnam St at Fruitvale Ave', '26th Ave at International Blvd',
'16th St Depot', '21st Ave at International Blvd',
'Willow St at Vine St', '2nd St at Folsom St'], dtype=object)
```

1.7 Cleaning Data

```
In [16]: #Make a copy of the dataframe
         bike_19=bike_2019.copy()
```

1.7.1 Issue 1- Dropping missing values

```
In [17]: bike_19.dropna(inplace = True)

In [18]: bike_19.isnull().sum()

Out[18]: duration_sec              0
         start_time                0
         end_time                  0
         start_station_id          0
         start_station_name        0
         start_station_latitude    0
         start_station_longitude   0
         end_station_id            0
         end_station_name          0
         end_station_latitude      0
         end_station_longitude     0
         bike_id                   0
         user_type                 0
         member_birth_year         0
         member_gender             0
         bike_share_for_all_trip   0
         dtype: int64
```

1.7.2 Issue 2- Changing all incorrect datatype to appropriate datatypes

```
In [19]: bike_19['start_time'] = pd.to_datetime(bike_19['start_time'])
         bike_19['end_time'] = pd.to_datetime(bike_19['end_time'])

         bike_19[['start_station_id','end_station_id','bike_id']] = bike_19[['start_station_id',


         bike_19[['user_type','member_gender','bike_share_for_all_trip']] = bike_19[['user_type'

         bike_19['member_birth_year'] = bike_19['member_birth_year'].astype('int')

In [20]: bike_19.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 174952 entries, 0 to 183411
Data columns (total 16 columns):
duration_sec              174952 non-null int64
start_time                174952 non-null datetime64[ns]
end_time                  174952 non-null datetime64[ns]
```

```
start_station_id           174952 non-null object
start_station_name         174952 non-null object
start_station_latitude     174952 non-null float64
start_station_longitude    174952 non-null float64
end_station_id             174952 non-null object
end_station_name           174952 non-null object
end_station_latitude       174952 non-null float64
end_station_longitude      174952 non-null float64
bike_id                    174952 non-null object
user_type                  174952 non-null category
member_birth_year          174952 non-null int64
member_gender              174952 non-null category
bike_share_for_all_trip    174952 non-null category
dtypes: category(3), datetime64[ns](2), float64(4), int64(2), object(5)
memory usage: 19.2+ MB
```

1.7.3 Issue 3- Derive new columns for better analysis

- Duration in minute
- start and end days of the week
- start and end hr
- age of riders

```
In [21]: # Add new columns for trip duration in minute

         bike_19['duration_min'] = bike_19['duration_sec']/60

In [22]: # Add new columns for days of the week
         bike_19['dayoftheweek'] = bike_19['start_time'].dt.strftime('%a')


         weekday = ['Mon', 'Tue', 'Wed', 'Thu', 'Fri', 'Sat', 'Sun']
         weekday_cat = pd.api.types.CategoricalDtype(ordered=True, categories=weekday)
         bike_19['dayoftheweek'] = bike_19['dayoftheweek'].astype(weekday_cat)

In [23]: # Add new columns for start and end hr of the day
         bike_19['hr_day'] = bike_19['start_time'].dt.hour

In [24]: # Add new columns for members age from birth year
         bike_19['member_age'] = 2019 - bike_19['member_birth_year']

In [25]: bike_19.head(10)

Out[25]:     duration_sec              start_time                end_time  \
         0          52185 2019-02-28 17:32:10.145 2019-03-01 08:01:55.975
         2          61854 2019-02-28 12:13:13.218 2019-03-01 05:24:08.146
         3          36490 2019-02-28 17:54:26.010 2019-03-01 04:02:36.842
         4           1585 2019-02-28 23:54:18.549 2019-03-01 00:20:44.074
```

```
5           1793 2019-02-28 23:49:58.632 2019-03-01 00:19:51.760
6           1147 2019-02-28 23:55:35.104 2019-03-01 00:14:42.588
7           1615 2019-02-28 23:41:06.766 2019-03-01 00:08:02.756
8           1570 2019-02-28 23:41:48.790 2019-03-01 00:07:59.715
9           1049 2019-02-28 23:49:47.699 2019-03-01 00:07:17.025
10           458 2019-02-28 23:57:57.211 2019-03-01 00:05:35.435

   start_station_id                                start_station_name  \
0              21.0  Montgomery St BART Station (Market St at 2nd St)
2              86.0                            Market St at Dolores St
3             375.0                            Grove St at Masonic Ave
4               7.0                                Frank H Ogawa Plaza
5              93.0                       4th St at Mission Bay Blvd S
6             300.0                               Palm St at Willow St
7              10.0                         Washington St at Kearny St
8              10.0                         Washington St at Kearny St
9              19.0                               Post St at Kearny St
10            370.0                                Jones St at Post St

    start_station_latitude  start_station_longitude end_station_id  \
0                37.789625              -122.400811           13.0
2                37.769305              -122.426826            3.0
3                37.774836              -122.446546           70.0
4                37.804562              -122.271738          222.0
5                37.770407              -122.391198          323.0
6                37.317298              -121.884995          312.0
7                37.795393              -122.404770          127.0
8                37.795393              -122.404770          127.0
9                37.788975              -122.403452          121.0
10               37.787327              -122.413278           43.0

                                  end_station_name  end_station_latitude  \
0                    Commercial St at Montgomery St             37.794231
2       Powell St BART Station (Market St at 4th St)             37.786375
3                           Central Ave at Fell St             37.773311
4                            10th Ave at E 15th St             37.792714
5                               Broadway at Kearny             37.798014
6                          San Jose Diridon Station             37.329732
7                            Valencia St at 21st St             37.756708
8                            Valencia St at 21st St             37.756708
9                               Mission Playground             37.759210
10  San Francisco Public Library (Grove St at Hyde...             37.778768

    end_station_longitude bike_id   user_type  member_birth_year  \
0              -122.402923    4902    Customer               1984
2              -122.404904    5905    Customer               1972
3              -122.444293    6638  Subscriber               1989
4              -122.248780    4898  Subscriber               1974
```

```
5             -122.405950    5200  Subscriber              1959
6             -121.901782    3803  Subscriber              1983
7             -122.421025    6329  Subscriber              1989
8             -122.421025    6548  Subscriber              1988
9             -122.421339    6488  Subscriber              1992
10            -122.415929    5318  Subscriber              1996

   member_gender bike_share_for_all_trip  duration_min dayoftheweek  hr_day  \
0           Male                     No    869.750000          Thu      17
2           Male                     No   1030.900000          Thu      12
3          Other                     No    608.166667          Thu      17
4           Male                    Yes     26.416667          Thu      23
5           Male                     No     29.883333          Thu      23
6         Female                     No     19.116667          Thu      23
7           Male                     No     26.916667          Thu      23
8          Other                     No     26.166667          Thu      23
9           Male                     No     17.483333          Thu      23
10        Female                    Yes      7.633333          Thu      23

    member_age
0           35
2           47
3           30
4           45
5           60
6           36
7           30
8           31
9           27
10          23
```

In [26]:
```
#Drop columns not necessary for analysis
bike_19= bike_19.drop(columns = ['start_time','end_time','start_station_id','end_statio
```

In [27]:
```
bike_19.sample(10)
```

Out[27]:
```
                                      start_station_name  \
18986                          Fulton St at Bancroft Way
102234                             Spear St at Folsom St
82854                                 18th St at Noe St
40234              Union Square (Powell St at Post St)
163699                     North Berkeley BART Station
24408                         Beale St at Harrison St
20055                      Bancroft Way at College Ave
29253                             Folsom St at 19th St
144432  San Francisco Ferry Building (Harry Bridges Pl...
18330                             Post St at Kearny St
```

```
                                          end_station_name   user_type  \
18986                       MLK Jr Way at University Ave  Subscriber
102234                             Post St at Kearny St  Subscriber
82854                         Mississippi St at 17th St  Subscriber
40234                              5th St at Brannan St    Customer
163699                           Fifth St at Delaware St  Subscriber
24408                              3rd St at Townsend St  Subscriber
20055                        Bancroft Way at Telegraph Ave  Subscriber
29253    Montgomery St BART Station (Market St at 2nd St)  Subscriber
144432                             3rd St at Townsend St  Subscriber
18330   San Francisco Caltrain Station 2  (Townsend St...  Subscriber

       member_gender bike_share_for_all_trip  duration_min dayoftheweek  \
18986           Male                      No      3.766667          Tue
102234          Male                      No     10.583333          Thu
82854           Male                      No     12.166667          Mon
40234           Male                      No      5.400000          Fri
163699        Female                      No      6.000000          Tue
24408           Male                      No      7.983333          Mon
20055           Male                      No      1.583333          Tue
29253         Female                      No     14.266667          Mon
144432          Male                      No     10.200000          Thu
18330           Male                      No      9.500000          Tue

        hr_day  member_age
18986       17          19
102234      16          60
82854        8          32
40234       21          22
163699       9          33
24408       18          31
20055       14          24
29253        7          26
144432      12          39
18330       18          31
```

In [28]:
```
bike_19.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 174952 entries, 0 to 183411
Data columns (total 9 columns):
start_station_name         174952 non-null object
end_station_name           174952 non-null object
user_type                  174952 non-null category
member_gender              174952 non-null category
bike_share_for_all_trip    174952 non-null category
duration_min               174952 non-null float64
dayoftheweek               174952 non-null category
```

```
hr_day                      174952 non-null int64
member_age                  174952 non-null int64
dtypes: category(4), float64(1), int64(2), object(2)
memory usage: 8.7+ MB
```

1.7.4 What is the structure of your dataset?

The original dataset is made up of 183412 rows and 16 columns with discription above

After wrangling,the resulting dataset is made up of 174952 rows and 9 columns with the following new columns generated for analysis purpose - duration_minute- extracted from duration_sec - member-age- extracted from the birth year - dayofthe week & hr_day - extracted from start and end time.

1.7.5 What is/are the main feature(s) of interest in your dataset?

The main features of interest are - The users demographics with respect to gender, age and user type - The number of trips and ride duration as it differs with age ,gender ,time and day of the week.

1.7.6 What features in the dataset do you think will help support your investigation into your feature(s) of interest?

The following features will help support my investigation of interest - members age and gender - hr of the day - day of the week - trip duration - user_type

1.8 Univariate Exploration

1.8.1 Question- What is the Bike Rental system usage for Feb 2019

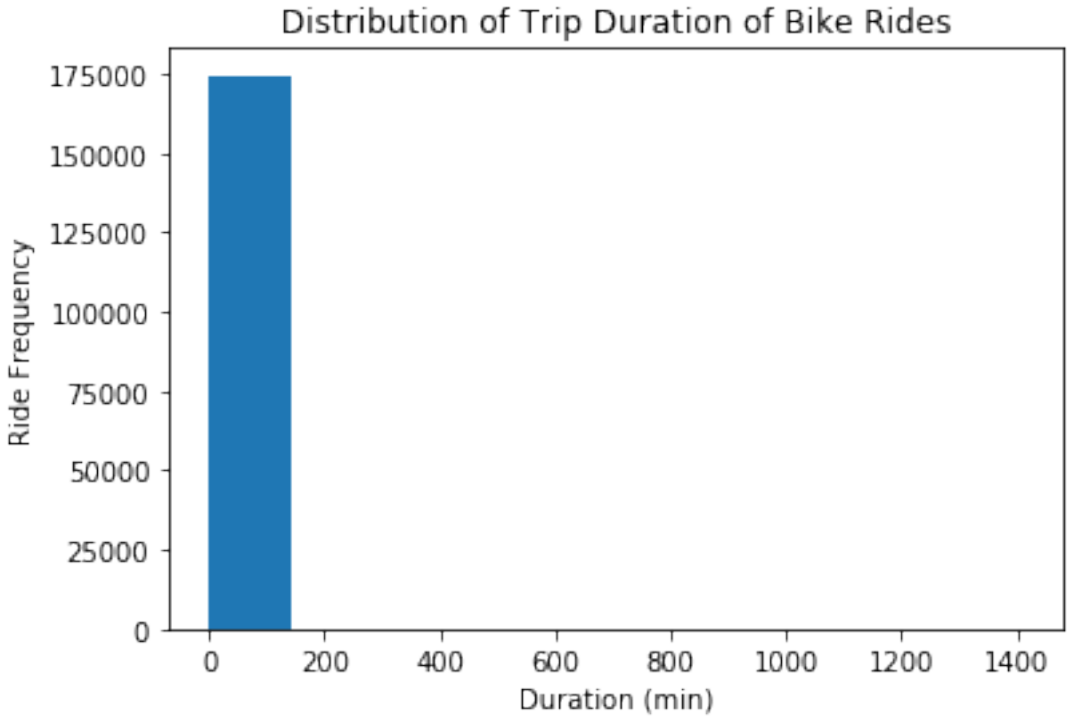
- Based on trip duration
- Based on weekdays
- Based on hours of the day

1.8.2 Visualization-The Distribution of trip duration

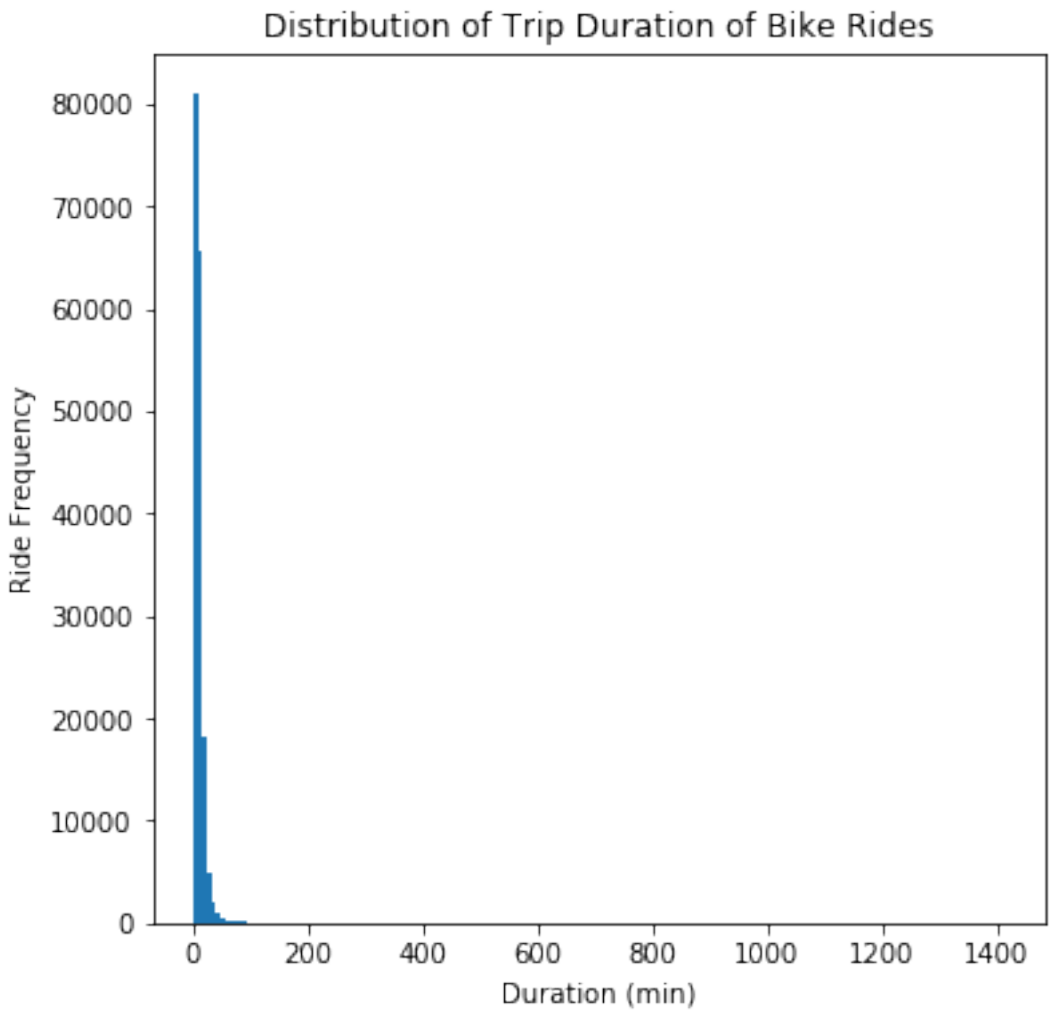
```
In [29]: bike_19['duration_min'].describe()
```

```
Out[29]: count    174952.000000
         mean         11.733379
         std          27.370082
         min           1.016667
         25%           5.383333
         50%           8.500000
         75%          13.150000
         max        1409.133333
         Name: duration_min, dtype: float64
```

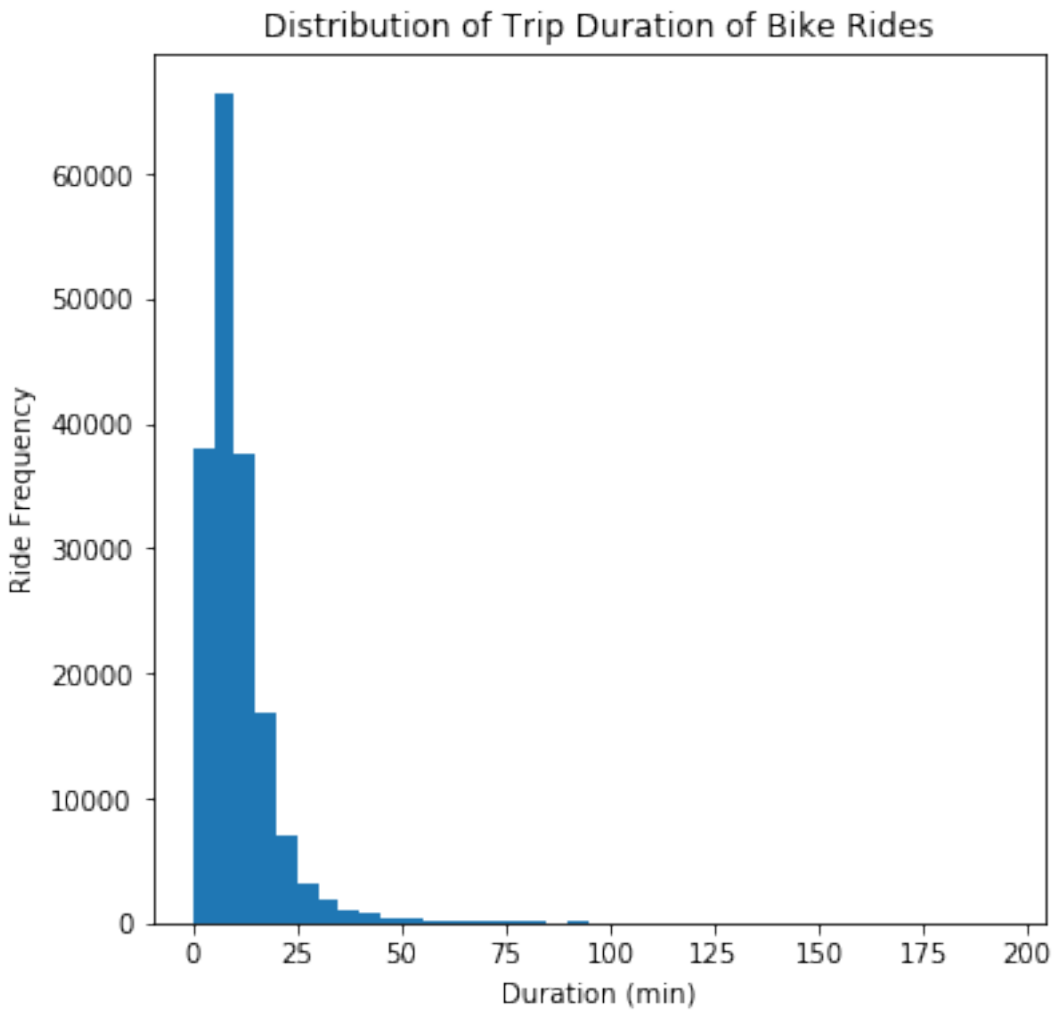
In [30]:
```
# A simple histogram plot of the variable
plt.hist(data=bike_19, x='duration_min');
plt.title('Distribution of Trip Duration of Bike Rides')
plt.xlabel('Duration (min)')
plt.ylabel('Ride Frequency');
```



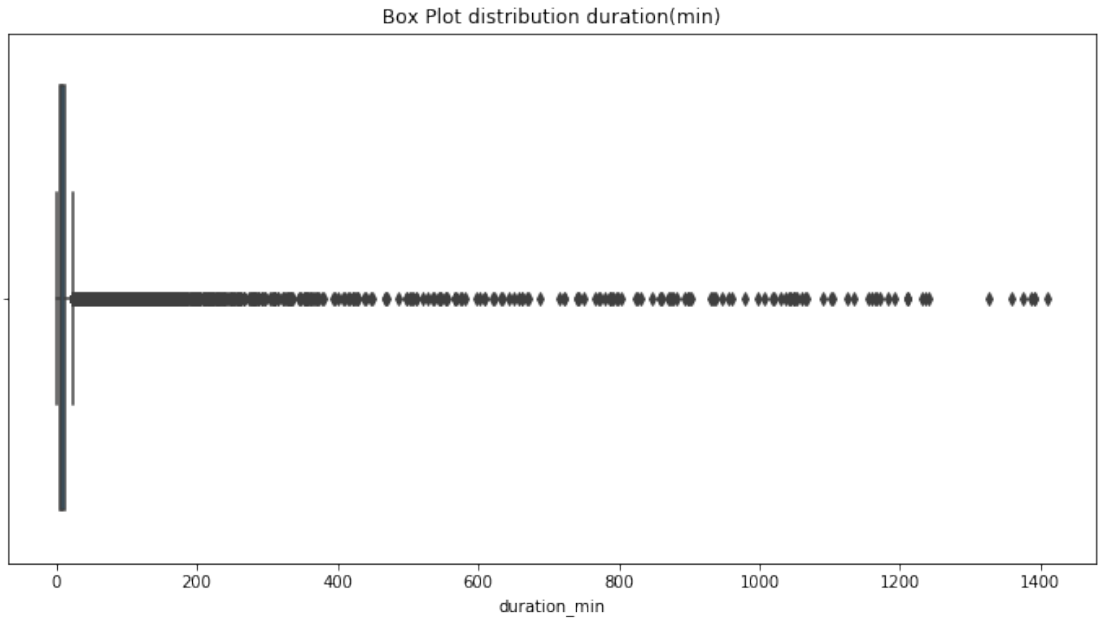
In [31]:
```
# plotting with a standard-scaled plot
binsize = 8
bins = np.arange(0,bike_19['duration_min'].max()+binsize, binsize)

plt.figure(figsize=[6, 6])
plt.hist(data = bike_19, x = 'duration_min', bins=bins)
plt.title('Distribution of Trip Duration of Bike Rides')
plt.xlabel('Duration (min)')
plt.ylabel('Ride Frequency');
```



In [32]:
```
# Investigating further on an even smaller bin size, zooming into the peak region(xlim-
bins = np.arange(0,200, 5)

plt.figure(figsize=[6, 6])

plt.hist(data = bike_19, x = 'duration_min', bins=bins)
plt.title('Distribution of Trip Duration of Bike Rides')
plt.xlabel('Duration (min)')
plt.ylabel('Ride Frequency');
```



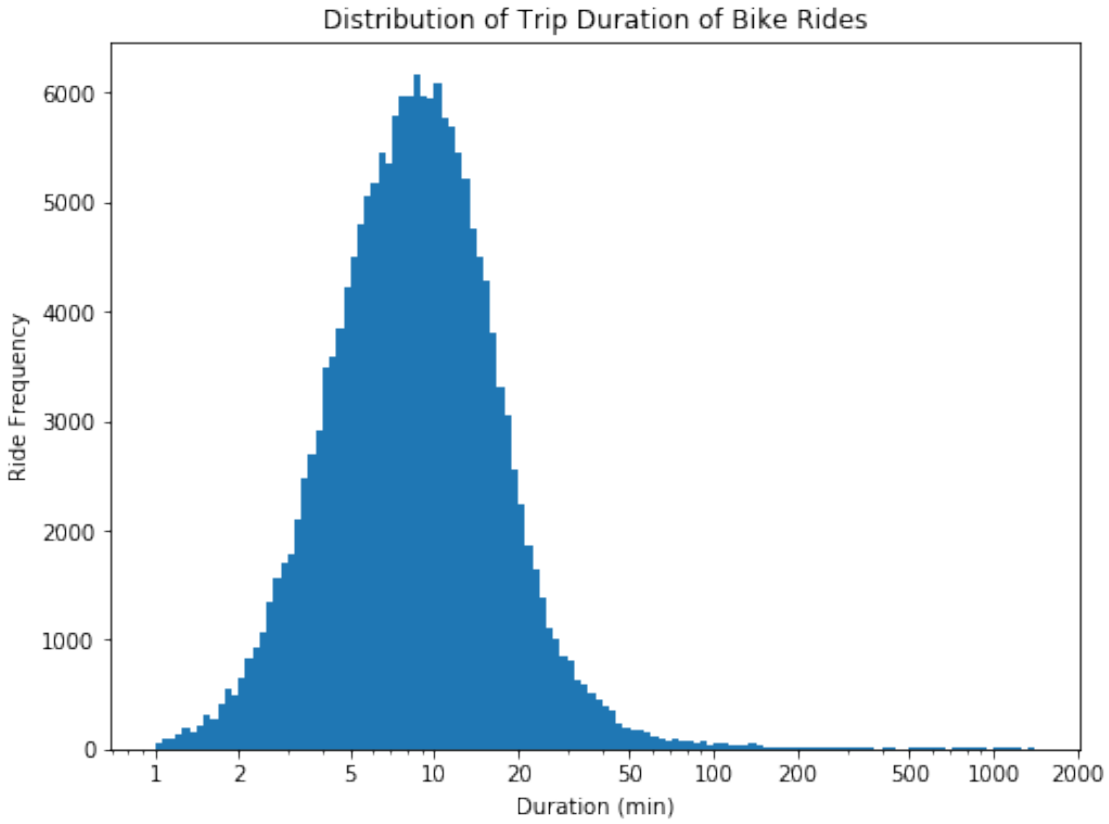
In [33]:
```
#lets check for outliers with a boxplot

plt.figure(figsize = (12,6))

sb.boxplot(data=bike_19,x='duration_min')
plt.title('Box Plot distribution duration(min)');
```



In [34]:
```
# Using IQR method to detect outliers
Q1 = bike_19['duration_min'].quantile(0.25)
Q3 = bike_19['duration_min'].quantile(0.75)
IQR = Q3 - Q1
LowerLimit = Q1 - 1.5*IQR  #Data which falls below this is an outlier
UpperLimit = Q3 + 1.5*IQR  #Data which falls above this is an outlier
print(LowerLimit, UpperLimit)
```

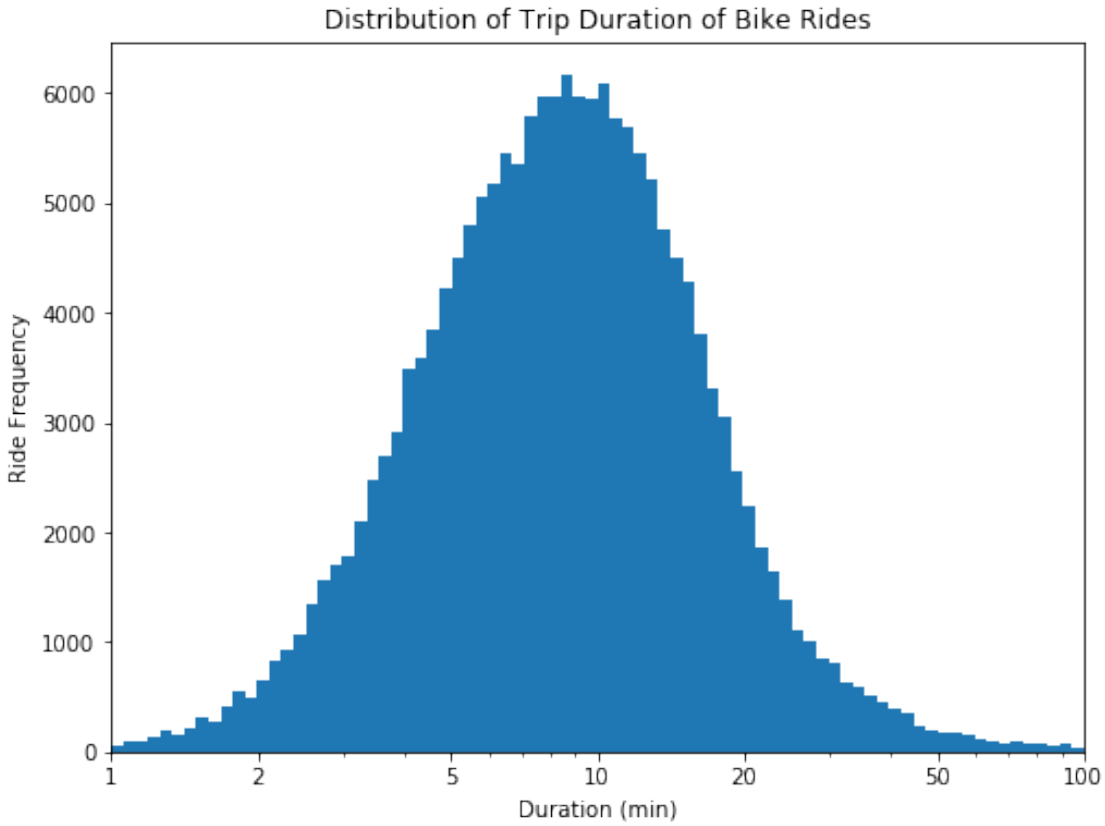
```
-6.266666666666667 24.8
```

In [35]:
```
# from plot with smaller bin size, there's a long tail in the distribution, so let's pu
log_binsize = 0.025
bins = 10 ** np.arange(0, np.log10(bike_19['duration_min'].max())+log_binsize, log_bins

plt.figure(figsize=[8, 6])
plt.hist(data = bike_19, x = 'duration_min', bins = bins)
plt.xscale('log')
plt.xticks([1, 2, 5, 10, 20, 50, 100, 200, 500, 1000, 2000], [1, 2, 5, 10, 20, 50, 100,
plt.title('Distribution of Trip Duration of Bike Rides')
plt.xlabel('Duration (min)')
plt.ylabel('Ride Frequency');
```



The distribution is still positively skewed due to the presence of the outliers as indicated by the IQR. lets plot another histogram using 100minutes as the Xlim to focus on the most common trips

```
In [36]: log_binsize = 0.025
         bins = 10 ** np.arange(0, np.log10(bike_19['duration_min'].max())+log_binsize, log_bins

         plt.figure(figsize=[8, 6])
         plt.hist(data = bike_19, x = 'duration_min', bins = bins)
         plt.xscale('log')
         plt.xticks([1, 2, 5, 10, 20, 50, 100, 200, 500, 1000, 2000], [1, 2, 5, 10, 20, 50, 100,
         plt.xlim([0,100])
         plt.title('Distribution of Trip Duration of Bike Rides')
         plt.xlabel('Duration (min)')
         plt.ylabel('Ride Frequency');
```



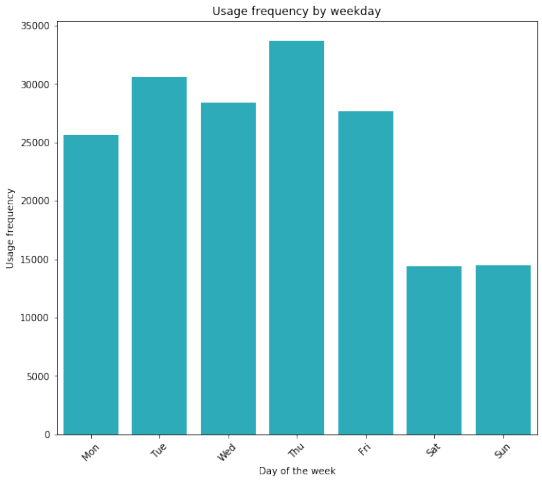
1.8.3 Observation

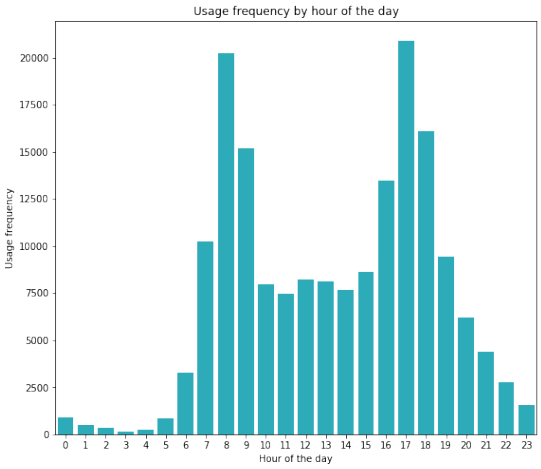
From the plot above, it is observed that - The distribution is more clear and its not skewed to the right rather exhibits approximately a normal distribution. - Most trips duration was between 5-25mins with the average peak duration being approximately 9mins. - There is presence of outliers as shown in IQR calculation

1.8.4 Visualization- The weekday and hr distribution

In [37]:
```
# let's plot the bike usage frequency based on weekday and hours of the day
plt.figure(figsize=[20,8])
plotcolor = sb.color_palette()[9]

plt.subplot(1,2,1)

sb.countplot(data = bike_19, x = 'dayoftheweek', color = plotcolor)
plt.xticks(rotation = 45)
plt.title('Usage frequency by weekday')
plt.xlabel('Day of the week')
plt.ylabel('Usage frequency');
```

```
plt.subplot(1,2,2)

sb.countplot(data = bike_19, x = 'hr_day', color = plotcolor)
plt.title('Usage frequency by hour of the day')
plt.xlabel('Hour of the day')
plt.ylabel('Usage frequency');
```





1.8.5 Observation

From the plot above, it is observed that - The usage frequency peaks during the week is on **Thursday** and **Tuesday**. - The peak hours are between **(8-9)hrs** for the morning hours of the day and **(17-18)hrs** for the evening hrs of the day.This can be attributed to the work timeline in san francisco.

1.8.6 Question- What is the demographics of the FordGo customers for Feb 2019

- The age distribution
- The gender distribution
- The user type distribution
- Are they members of the **Bike share for all Program**

1.8.7 Visualization- The age distribution

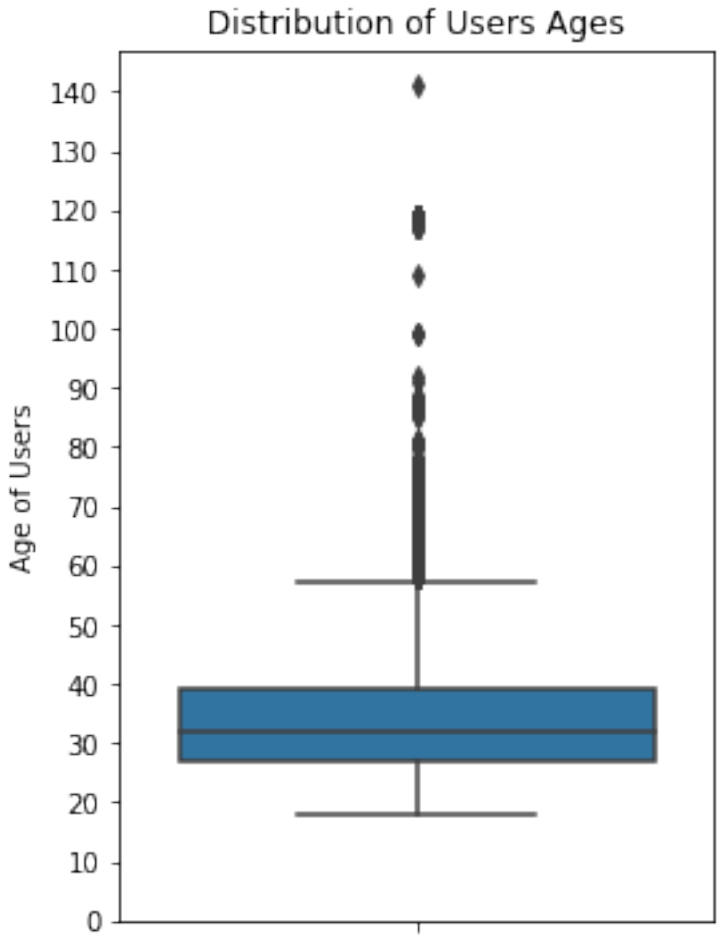
```
In [38]: bike_19['member_age'].describe()
```

```
Out[38]: count    174952.000000
         mean         34.196865
         std          10.118731
         min          18.000000
```

```
25%           27.000000
50%           32.000000
75%           39.000000
max          141.000000
Name: member_age, dtype: float64
```

Max age from above shows **141**, this indicates presence of outliers.

```
In [39]: #lets plot a boxplot to show the distribution of ages
         plt.figure(figsize = (4,6))

         ytick = np.arange(0,bike_19['member_age'].max(),10)
         sb.boxplot(data=bike_19,y='member_age')
         plt.yticks(ytick,ytick)
         plt.title('Distribution of Users Ages')
         plt.ylabel('Age of Users');
```



The data is not distributed evenly and there are outliers. Let's calculate of IQR:

In [40]:
```
# Using IQR method to detect outliers
Q1 = bike_19['member_age'].quantile(0.25)
Q3 = bike_19['member_age'].quantile(0.75)
IQR = Q3 - Q1
LowerLimit = Q1 - 1.5*IQR  #Data which falls below this is an outlier
UpperLimit = Q3 + 1.5*IQR  #Data which falls above this is an outlier
print(LowerLimit, UpperLimit)
```

```
9.0 57.0
```

In [41]:
```
# let's plot a histogram to display the riders age distribution without the outliers i.

plt.figure(figsize=[8,6])

bins = np.arange(10, bike_19['member_age'].max()+5, 5)
plt.hist(data=bike_19, x='member_age', bins=bins)
plt.title('Users age distribution')
plt.xticks(bins, bins)
plt.xlabel('Member Age')
plt.xlim([10,65])
plt.ylabel('Count');
```