ANALYSIS ON

FORD GOBIKE TRIP DATA

WHAT LIBRARIES DID WE USE

- PANDAS
- NUMPY
- PYSPARK
- MATPLOTLIB
- HAVERSINE

Columns in Dateframe

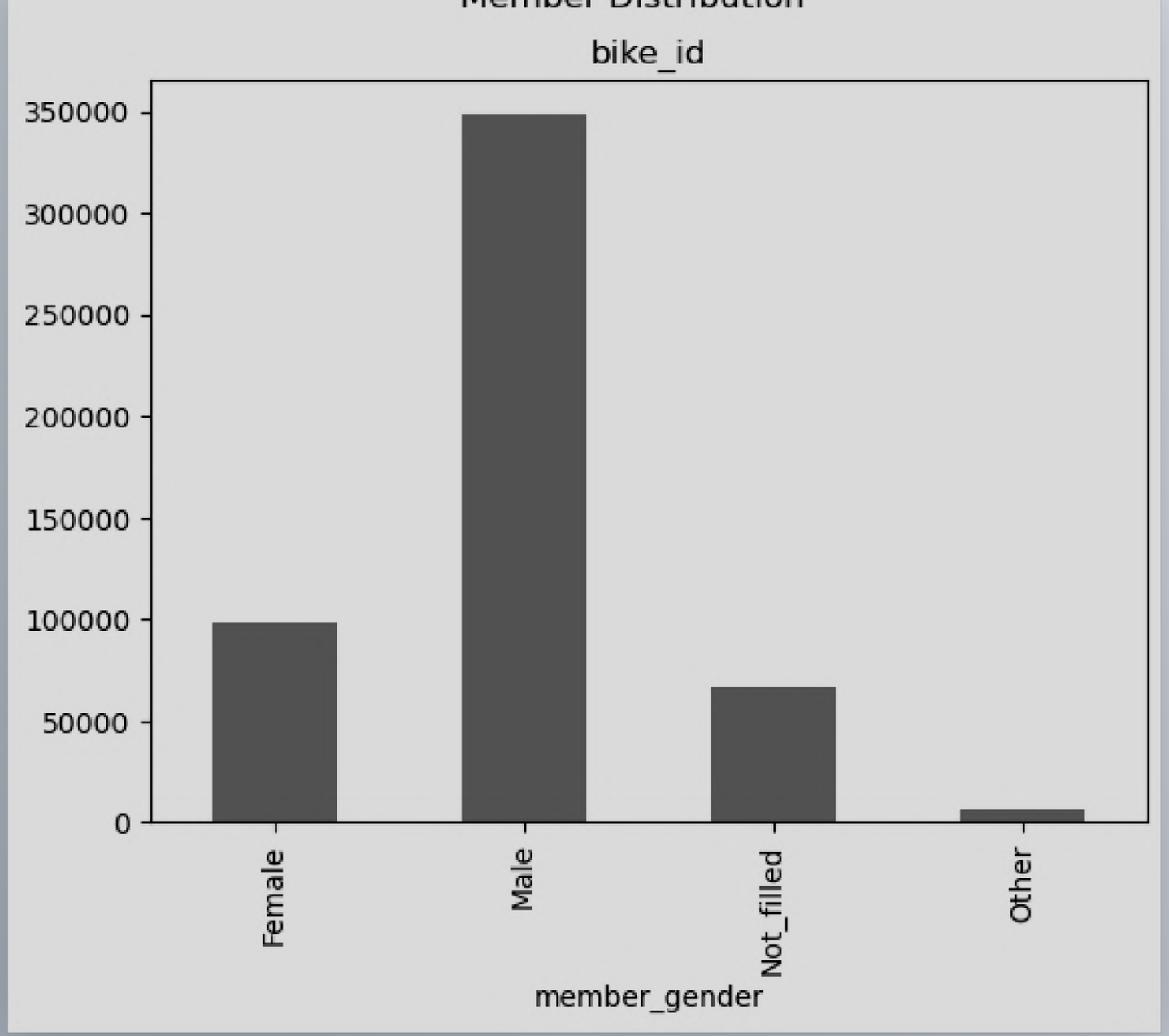
1. start time
2. end time
3. start station id
4. start station name
5. start station latitude
6. start station longitude
7. end station id
8. end station name
9. end station latitude
10. end station longitude
11. bike id
12. user type
13. member birth year
14. member gender
15. payment

Work Done on the Ford gobike Data

- Check the Data description and unique values in each column
- Calculate the distance for each trip
- Calculate the cost of trip using 0.35$ per minute
- Replace the 66,462 null value in the member_gender column with not_filed
- Exploratory Data Analysis (EDA) on the data

DATA DESCRIPTION

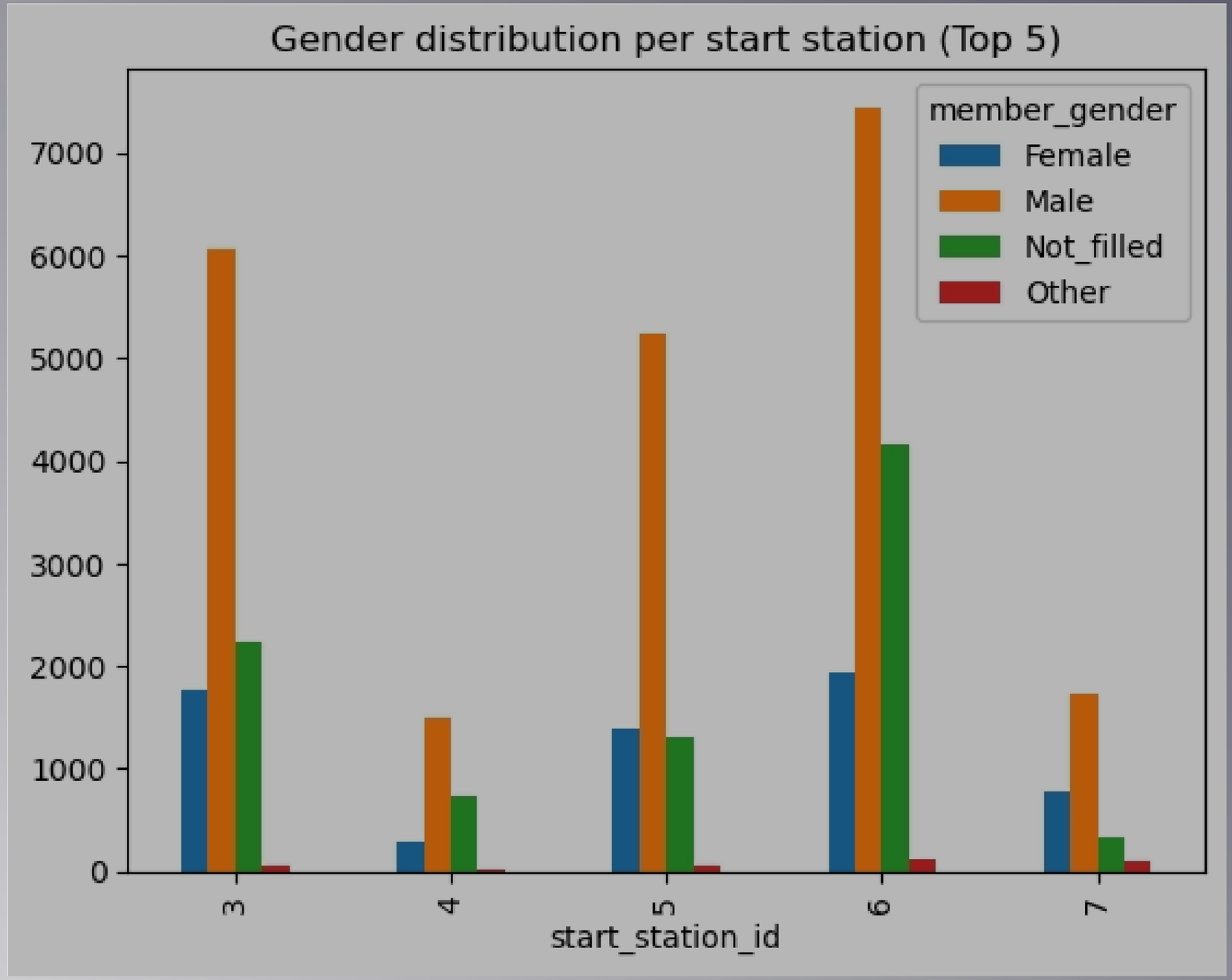
- The dataframe has 12 Columns and 519700 rows.
- It has no duplication data.
- The was no null value in other columns except for the member gender and member birth year.



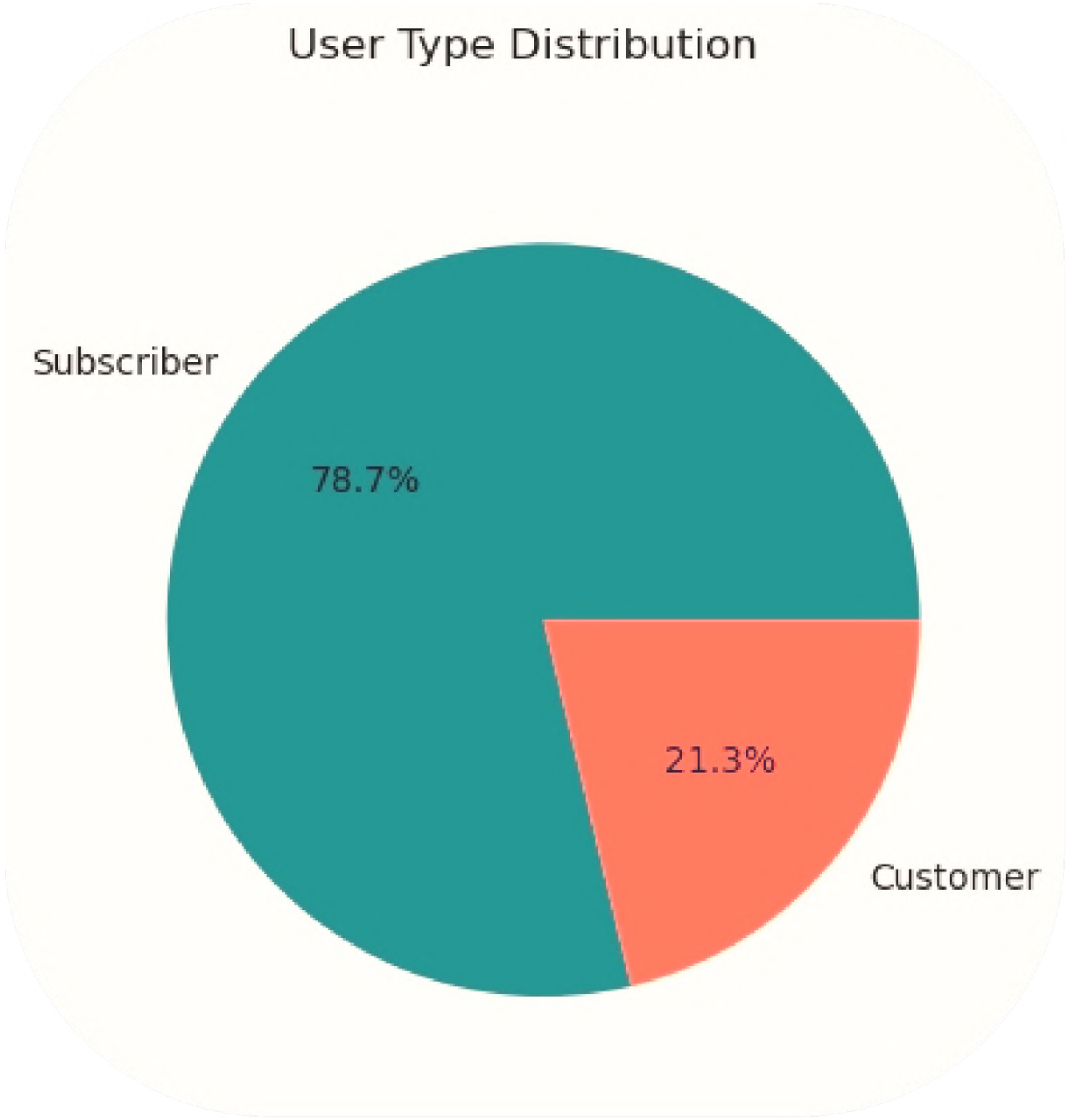
This shows the total gender distribution of ford gobike trip data

A table showing the top 5 gender distribution per station

start_station_id	start_station__name	Female	Male	Other	Not_filled
2	Powell St BART Station (Market St at 4th St)	1768	6070.0	64	2240.0
3	Cyril Magnin St at Ellis St	295	1491	7	735
4	Powell St BART Station (Market St at 5th St)	1393	5240	58	1296
5	The Embarcadero at Sansome St	1938	7444	119	4163
6	Frank H Ogawa Plaza	769	1727		330

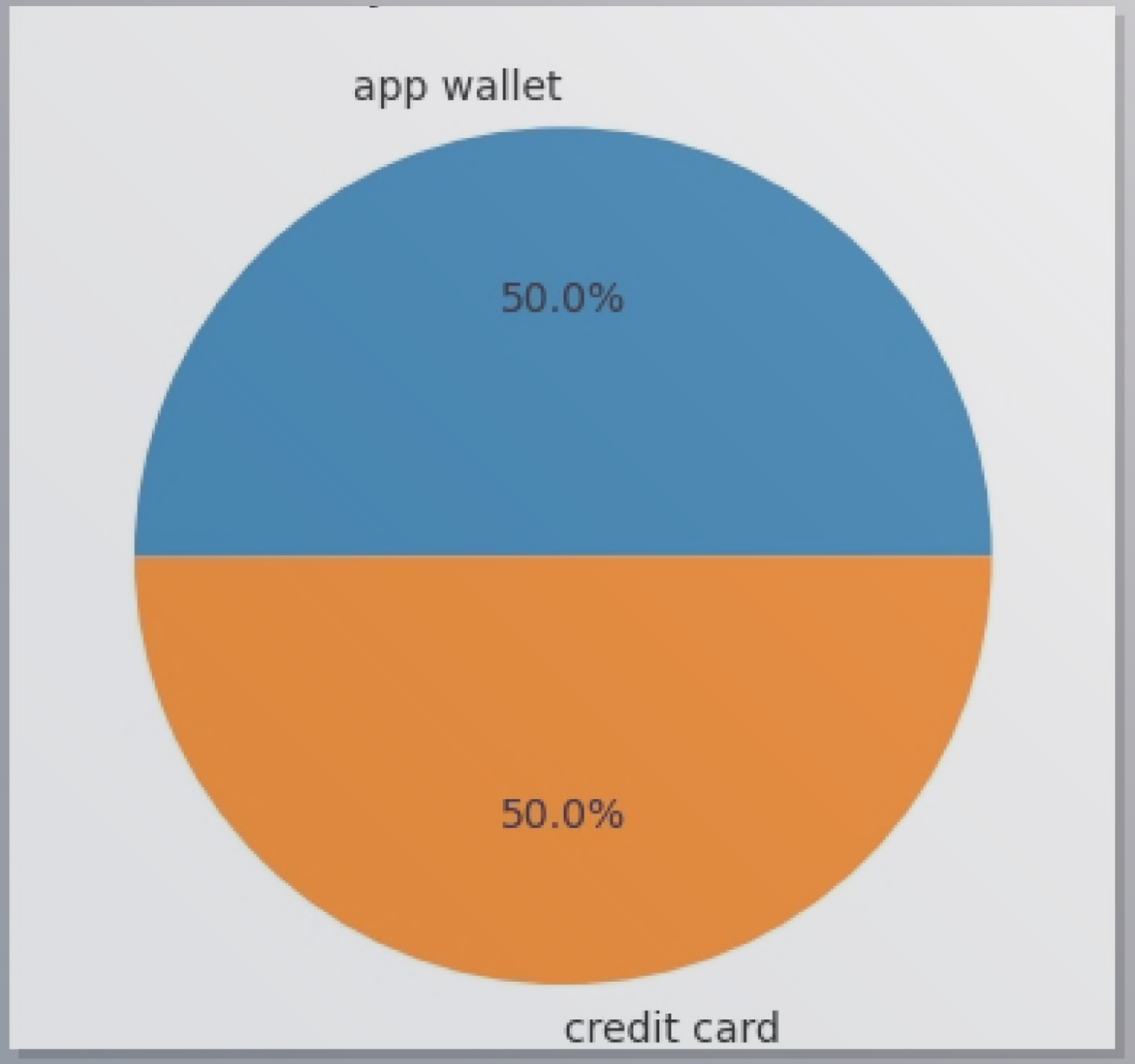


This graph shows the top 5 Gender distribution per station



user type	total cost of trip
Subscriber	1684845.4
Customer	1648191.06

Payment Distribution between App Wallet and Credit card



Count Of Total Trips By Different Time Interval Of The Day