CASE STUDY: How does a bike-share company navigate speedy success?

Author: Jai Mirchandani
Date: 15th June, 2025

Summary

This case study explores Cyclistic bike-sharing data over a 12-month period to uncover trends in rider behavior. The objective is to compare casual vs member riders and provide actionable marketing recommendations to convert casual riders into annual members.

Company

In 2016, Cyclistic (fictional company) launched a successful bike-share offering. Since then, the program has grown to a fleet of 5,824 bicycles that are geotracked and locked into a network of 692 stations across Chicago. The bikes can be unlocked from one station and returned to any other station in the system anytime. Until now, Cyclistic's marketing strategy relied on building general awareness and appealing to broad consumer segments. One approach that helped make these things possible was the exibility of its pricing plans: single-ride passes, full-day passes, and annual memberships. Customers who purchase single-ride or full-day passes are referred to as casual riders. Customers who purchase annual memberships are Cyclistic members.

SCENARIO

Marketing director of cyclist believes the company's future success depends on maximizing the number of annual memberships. Hence, a new data-driven marketing strategy is required to focus on increasing annual memberships. The marketing director further believes there is a solid opportunity to convert casual riders into members. Any marketing action must be approved by the executives and primary stakeholder and hence, a data-driven proof of marketing director's hypothesis is essential.

Business Task

As a data analyst at this company my tasks are as follows:

- To compare the usage pattern of annual subscribers vs casual riders
- Prove the success rate of annual membership to the stakeholders of the company
- To reccomend the target audience within the casual riders for conversion into annual membership

Datasets

1. Divy bike-sharing company trip data: https://divvy-tripdata.s3.amazonaws.com/index.html

- Data is stored in monthly tables From 06-2024 to 05-2025
- The data source is the official divvy database making it reliable and credible

2. New York Citibike public dataset: https://console.cloud.google.com/bigquery(cameo:product/city-of-new-york/nyc-citi-bike)

- Data is sourced, cleaned and analysed through big-query and later on imported here to visualize and interpret the insights

DIVY Data Analysis

The data is stored in seperate tables for each month in the directory "divy_cycle_data"

In [1]:
```
#IMPORTING LIBRARIES
%matplotlib inline
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import os
from collections import Counter
```

In [2]:
```
#Reading the data
Directory= 'divy_cycle_data/'
data_list=os.listdir(Directory)
for g in range(len(data_list)):
   globals()["divvy_"+str(g)] =pd.read_csv(Directory+data_list[g])
```

In [3]:
```
#Merging the data into 1 single table
merged_divvy_data=pd.concat(globals()["divvy_"+str(g)] for g in range(len(data_l
#changing the data types
merged_divvy_data.dtypes
merged_divvy_data.started_at=pd.to_datetime(merged_divvy_data.started_at,format=
merged_divvy_data.ended_at=pd.to_datetime(merged_divvy_data.ended_at,format="%Y%
#Calculating the trip duration
merged_divvy_data['tripduration'] = (merged_divvy_data.ended_at-merged_divvy_dat
```

In [4]:
```
#CHECK FOR ERRORS
print('Before removing duplicates=',len(merged_divvy_data))
merged_divvy_data.drop_duplicates
print('After removing duplicates=',len(merged_divvy_data))

print('tripduration_null=',merged_divvy_data.tripduration.isnull().sum())
print('date_null=',merged_divvy_data.started_at.isnull().sum())
print('membership_null=',merged_divvy_data.member_casual.isnull().sum())
```

```
Before removing duplicates= 5628847
After removing duplicates= 5628847
tripduration_null= 0
date_null= 0
membership_null= 0
```

Above Codes shows that no null or duplicates are present in the dataset. Hence, it is safe to assume to assume that the dataset is clean

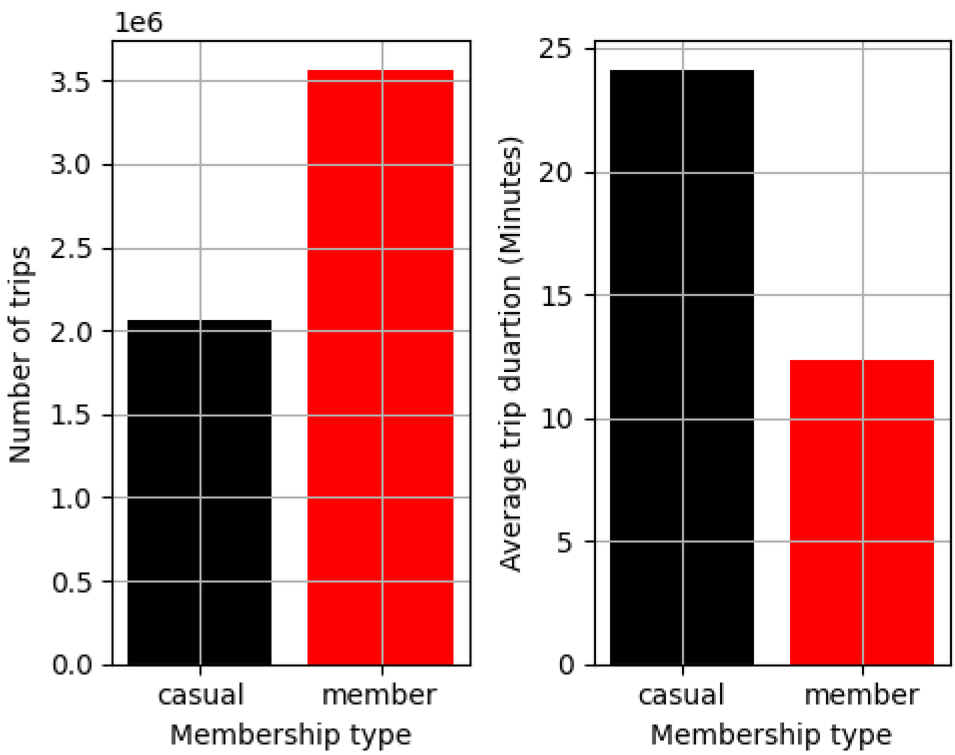
In [5]:

```
#Total Number of trips for each rider type
Count= Counter(merged_divvy_data.member_casual)
print(Count)
count_of_category=list(Count.values())
type_membership= list(Count.keys())

#Average trip duration for each rider type
number_of_trips_per_category = merged_divvy_data.groupby('member_casual')['tripd
number_of_trips_per_category.rename(columns={'tripduration':'average_trip_durati
print(number_of_trips_per_category)

#Visualization
fig,axes=plt.subplots(1,2,figsize=(5,4))
axes[0].bar(type_membership,height=count_of_category,color=['black','red'])
axes[1].bar(number_of_trips_per_category.member_casual,number_of_trips_per_categ
axes[0].grid(),axes[1].grid()
axes[0].set_xlabel('Membership type'), axes[1].set_xlabel('Membership type')
axes[0].set_ylabel('Number of trips'), axes[1].set_ylabel('Average trip duartion
plt.tight_layout()
```

```
Counter({'member': 3564561, 'casual': 2064286})
  member_casual  average_trip_duration
0        casual              24.105187
1        member              12.297787
```



Insights from the above visuals:

- Membership riders took approximately 1.8 time more trips compared to casual riders
- The average trip duration of casual riders is approximately twice as that of membership riders
- Inverse correlation seen between the average duration of trips and frequency of trips for both membership types
- This indicates that:
    - members tend to use bikes for shorter durations because of reduced cost from membership and consequently, leading to higher trip frequency
    - casual riders prefer walking for shorter distances and use bikes only when the distances are reasonably long or take a day pass leading to increase in the average trip duration

In [6]:

```
#Average trip duration grouping by membership type and type of bike
avg_duration_per_ride_type = merged_divvy_data.groupby(['rideable_type','member_
avg_duration_per_ride_type.rename(columns={'tripduration':'Average_trip_duration
avg_duration_per_ride_type
```

Out[6]:

	rideable_type	member_casual	Average_trip_duration
0	classic_bike	casual	37.837699
1	classic_bike	member	14.106986
2	electric_bike	casual	14.158659
3	electric_bike	member	11.005392
4	electric_scooter	casual	11.937651
5	electric_scooter	member	8.239676

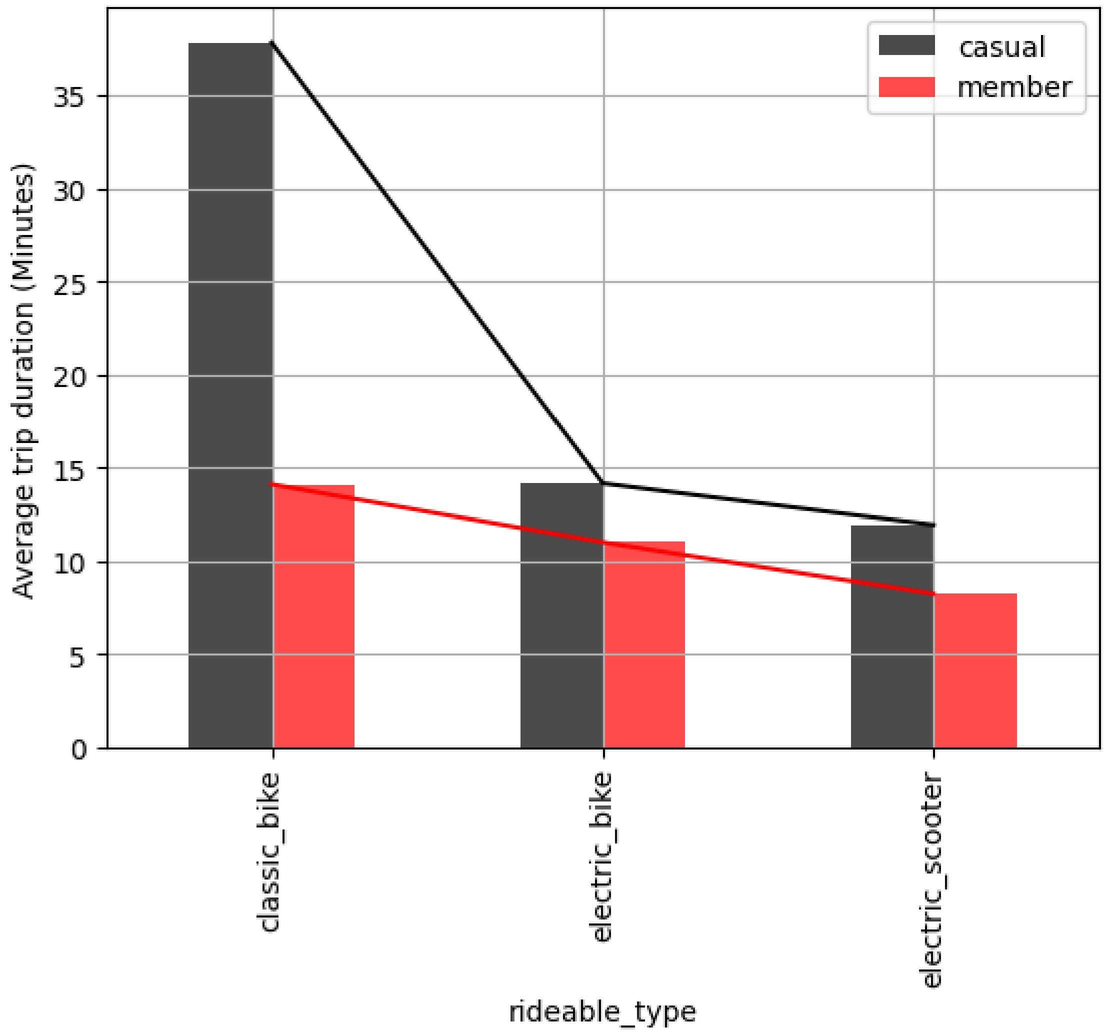
In [7]:

```
##The Total number of trips Grouping by membership type and type of bike
pivot= avg_duration_per_ride_type.pivot(columns='member_casual',index='rideable_
print(pivot)
#Visualization
pivot.plot(kind='bar',color=['black','red'],alpha=0.7)
plt.grid()
plt.plot(pivot.index, pivot.member,color='red')
plt.plot(pivot.index, pivot.casual,color='black')
plt.legend()
plt.ylabel('Average trip duration (Minutes)')
```

```
member_casual         casual     member
rideable_type
classic_bike       37.837699  14.106986
electric_bike      14.158659  11.005392
electric_scooter   11.937651   8.239676
```

Out[7]:

```
Text(0, 0.5, 'Average trip duration (Minutes)')
```

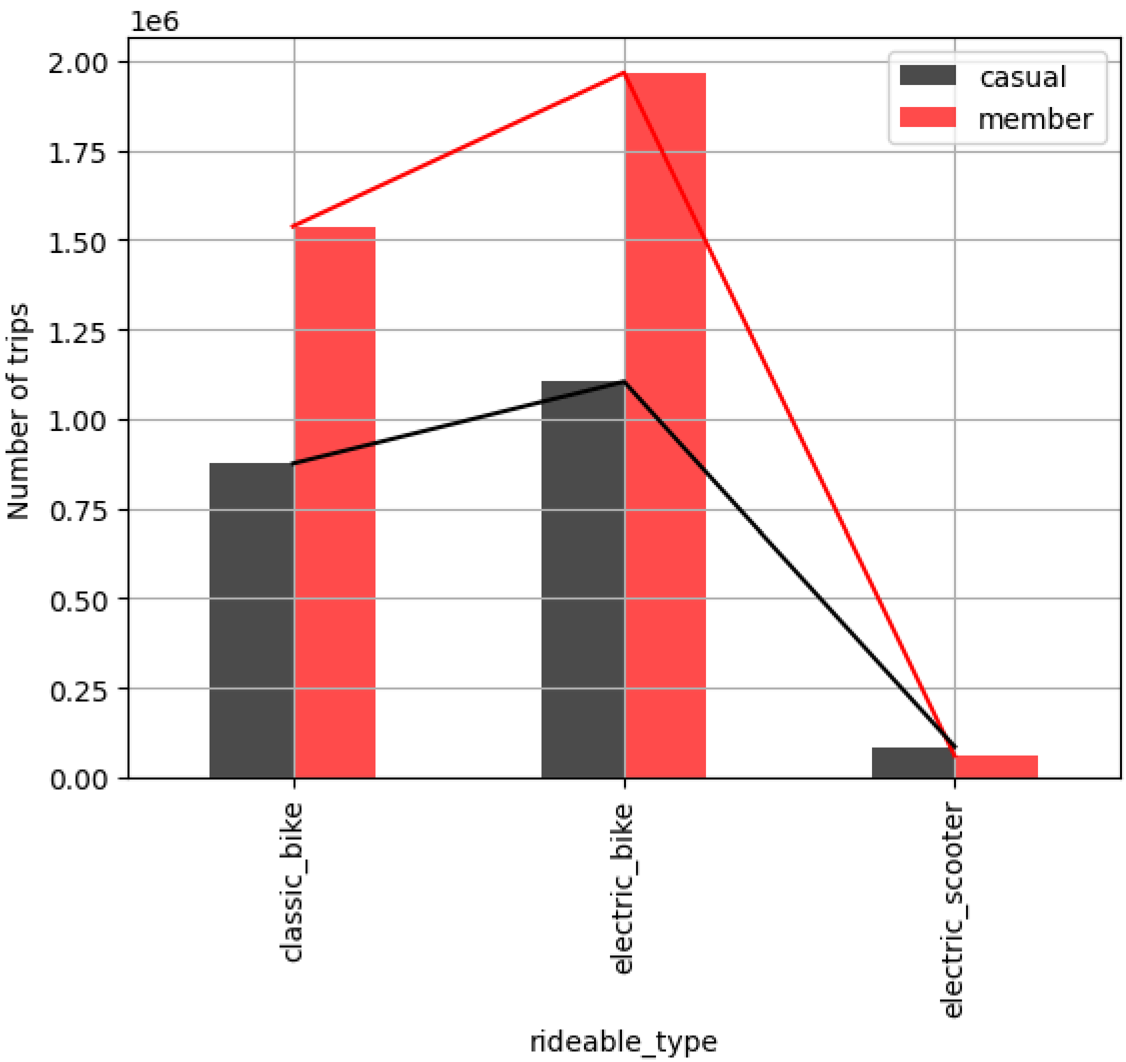


In [8]:
```
#The Total number of trips Grouping by membership type and type of bike
trip_count_per_ride_type = merged_divvy_data.groupby(['rideable_type','member_ca
trip_count_per_ride_type.rename(columns={'tripduration':'trip_count'},inplace=Tr
trip_count_pivot= trip_count_per_ride_type.pivot(columns='member_casual',index='
print(pivot)
#Visualization
trip_count_pivot.plot(kind='bar',color=['black','red'],alpha=0.7)
plt.grid()

plt.plot(trip_count_pivot.index, trip_count_pivot.member,color='red')
plt.plot(trip_count_pivot.index, trip_count_pivot.casual,color='black')
plt.legend()
plt.ylabel('Number of trips')
```

```
member_casual        casual     member
rideable_type
classic_bike      37.837699  14.106986
electric_bike     14.158659  11.005392
electric_scooter  11.937651   8.239676
```

Out[8]: `Text(0, 0.5, 'Number of trips')`



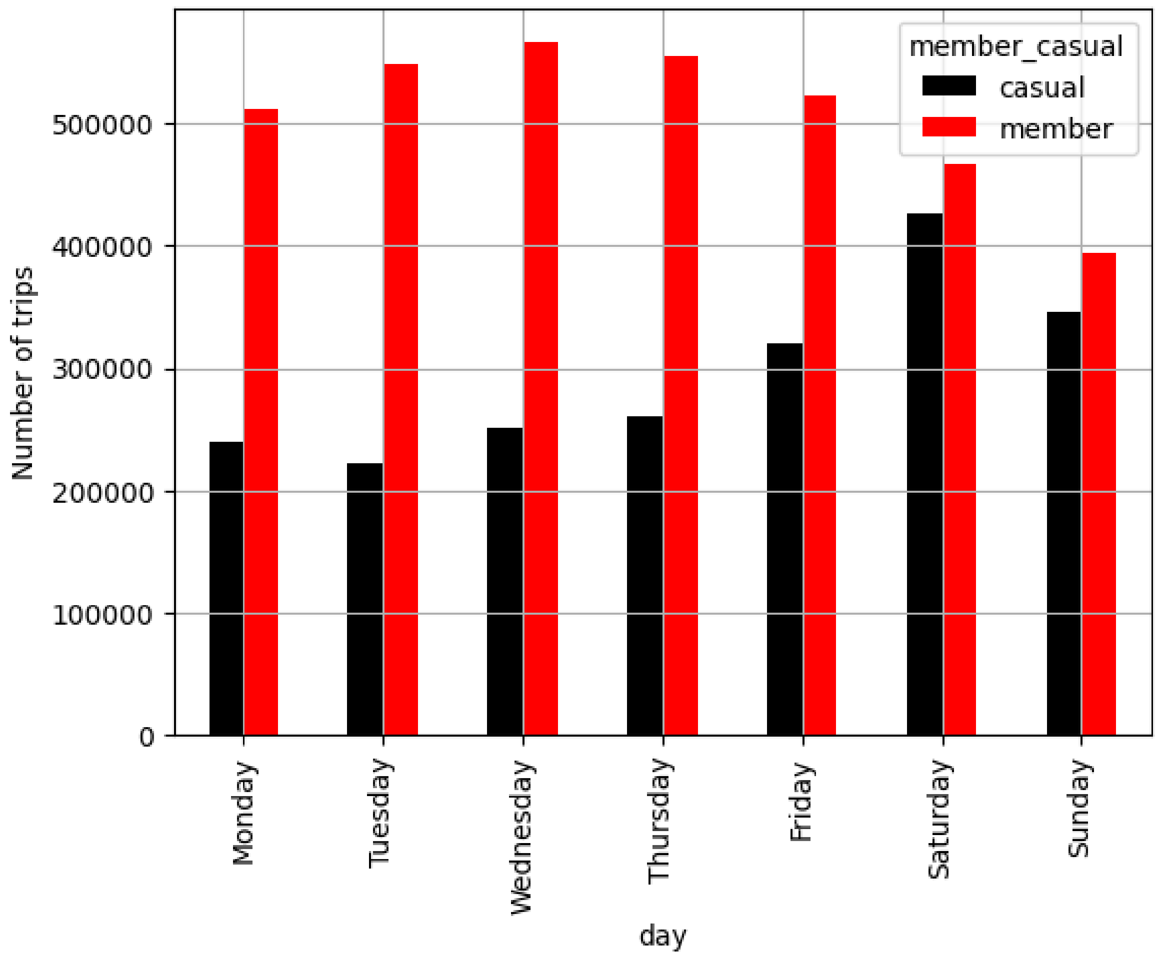
In [9]:

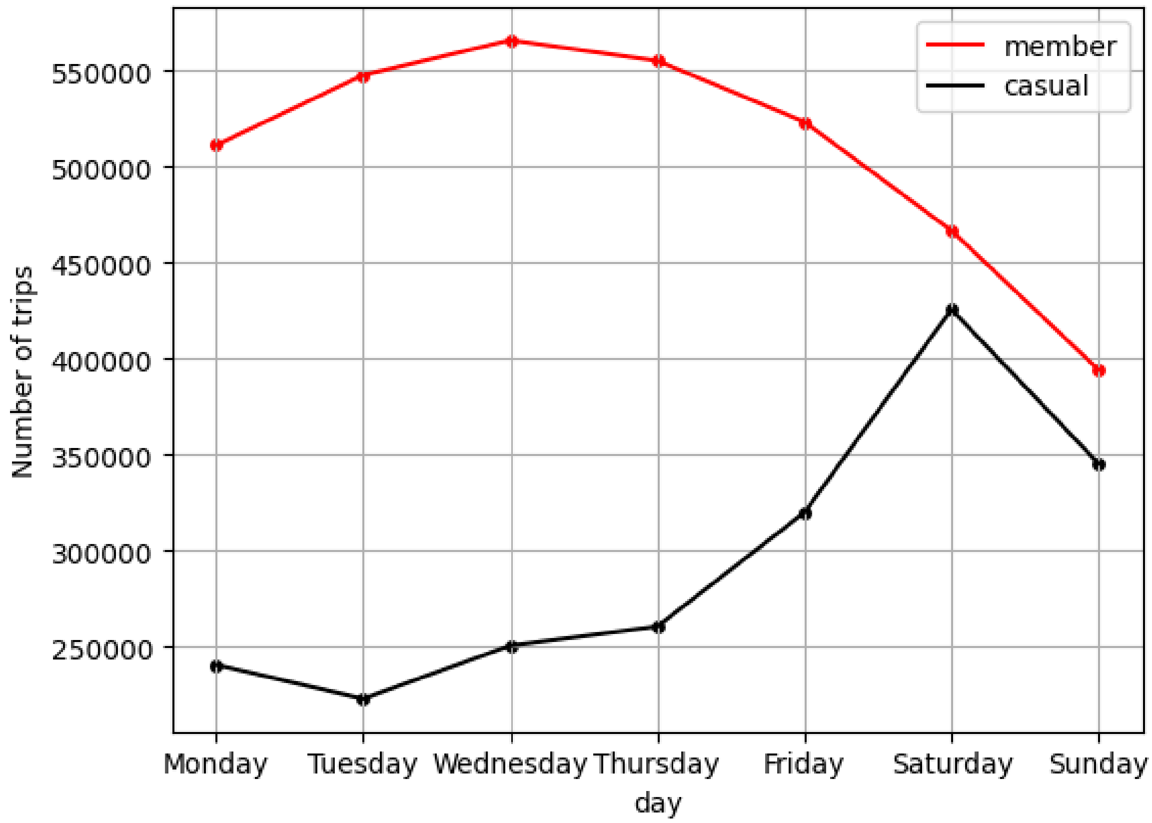
```
#Weekly analysis of the usage patterns
merged_divvy_data['day']= merged_divvy_data['started_at'].dt.day_name()
trips_per_day = merged_divvy_data.groupby(['day','member_casual'])['tripduration
display(trips_per_day)
day_order = ['Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday', 'Saturday',
trips_per_day.day =pd.Categorical(trips_per_day.day,categories=day_order,ordered
trips_per_day.sort_values('day')
pivot_trips_per_day= trips_per_day.pivot(columns='member_casual', values='tripdu
print(pivot_trips_per_day)
#Visualizations
pivot_trips_per_day.plot(kind='bar',color=['black','red'])
plt.ylabel('Number of trips')
plt.grid()
plt.figure()
plt.scatter(pivot_trips_per_day.index,pivot_trips_per_day.member,color='red',s=1
plt.plot(pivot_trips_per_day.index,pivot_trips_per_day.member,label='member',col
plt.scatter(pivot_trips_per_day.index,pivot_trips_per_day.casual,color='black',s
plt.plot(pivot_trips_per_day.index,pivot_trips_per_day.casual,color='black',labe
plt.legend()
plt.grid()
plt.ylabel('Number of trips')
plt.xlabel('day')
```

	day	member_casual	tripduration
0	Friday	casual	319711
1	Friday	member	523417
2	Monday	casual	240310
3	Monday	member	511250
4	Saturday	casual	425623
5	Saturday	member	466721
6	Sunday	casual	345288
7	Sunday	member	394196
8	Thursday	casual	260314
9	Thursday	member	555267
10	Tuesday	casual	222675
11	Tuesday	member	547804
12	Wednesday	casual	250365
13	Wednesday	member	565906

```
member_casual  casual  member
day
Monday         240310  511250
Tuesday        222675  547804
Wednesday      250365  565906
Thursday       260314  555267
Friday         319711  523417
Saturday       425623  466721
Sunday         345288  394196
```

Out[9]:
```
Text(0.5, 0, 'day')
```





Insights from the above visuals:

- Classic bikes are used longest by both rider types whereas electric bikes are used for more number of (shorter) trips by both rider types

- The possible reason for very low use of e-scooters needs to be investigated. However my personal opinions are as follows:
    - Less quantity of e-bikes
    - Less battery power of e-bikes
- An intersting observation from the above visuals is that the drop in average trip duration between classic bike and electric bike is consistent for members whereas it is sharp for the casual riders
- This indicates that casual riders majorly use classic bikes for long distances and with day-pass, probably due to battery limitations of e-bikes.
- This doesn't hold for the member as they have the flexibility to switch bikes during a trip without any extra cost
- Additionally, the earlier established inverse correlation of avergae trip duration and number of trips continues to hold in all three cases for member riders.
- Casual riders have the highest footfall on weekends as rides taken on saturday is 75% higher compared to rides taken on Monday
- Trips by members peak during the weekdays with a reasoanble drop of around 10 to 40 % on the weekend
- An intepretation of this result is:
    - Member riders use the bike-share for work related purposes on weekdays
    - Casual riders use the bike-share services for leisure on weekends

New York Citibike dataset

The Divvy dataset was able to unravel various insights on the usage patterns of members and casual riders. But, the dataset fails to provide any information on the age and gender of thr users. Such inofrmation is required to understand the target audience. Hence, another bike-sharing company:citi's dataset from New York is analysed in the following section. The data was sourced,cleaned and analysed through big query and then imported here. However, a limitation of this dataset is that it ranges from 2013 to 2018 and may not be very accurately represntatitive of 2025 scenario. The SQL QUERIES ARE AS FOLLOWS

```
#QUERY 1
    SELECT
        birth_year,
        count(usertype)as number_of_subscribers,
        count(usertype)/(select count(*) from bigquery-public-
data.new_york_citibike.citibike_trips where is_nan
(tripduration)=FALSE AND usertype="Subscriber") as
subscriber_trip_fraction ,
        (select count(*) from bigquery-public-
data.new_york_citibike.citibike_trips where
```

```
is_nan(tripduration)=FALSE AND usertype="Subscriber") AS
total_number_of_trips_by_subs

FROM
bigquery-public-data.new_york_citibike.citibike_trips
WHERE
 usertype = "Subscriber" AND IS_NAN(tripduration)=FALSE
GROUP BY
 birth_year
ORDER BY
 subscriber_trip_fraction DESC ;


#QUERY 2
SELECT
birth_year, COUNT(birth_year) AS trips_by_birth_year,
COUNT(birth_year)*100/(SELECT COUNT(*) FROM  bigquery-public-
data.new_york_citibike.citibike_trips WHERE
IS_NAN(tripduration)=FALSE) AS percent_of_total_trips
FROM
bigquery-public-data.new_york_citibike.citibike_trips
WHERE
IS_NAN(tripduration)=FALSE
GROUP BY
birth_year
ORDER BY
birth_year;

#QUERY 3
SELECT
usertype,SUM(tripduration)/60 AS
total_trip_duration,AVG(tripduration)/60 AS
Average_trip_duration, COUNT(usertype) AS Number_of_trips
FROM
bigquery-public-data.new_york_citibike.citibike_trips
WHERE
IS_NAN(tripduration)= FALSE
GROUP BY
 usertype;

 #QUERY 4
 SELECT EXTRACT(YEAR FROM starttime)AS year,COUNT(*) AS num_subs
FROM bigquery-public-data.new_york_citibike.citibike_trips
WHERE IS_NAN(tripduration)=FALSE AND usertype= 'Subscriber'
GROUP BY year
order by year;

#QUERY 5
SELECT
        gender,
        count(usertype) as number_of_subscribers,
        (select count(*) from bigquery-public-
data.new_york_citibike.citibike_trips where
is_nan(tripduration)=FALSE AND gender="female") as
total_male_users, (select count(*) from bigquery-public-
data.new_york_citibike.citibike_trips where
```

```
is_nan(tripduration)=FALSE AND gender="male") as
total_female_users, count(usertype)*100/(select count(*) from
bigquery-public-data.new_york_citibike.citibike_trips where
is_nan(tripduration)=FALSE AND usertype="Subscriber") as
subscriber_trip_percent, (select count(*) from bigquery-public-
data.new_york_citibike.citibike_trips where
is_nan(tripduration)=FALSE AND gender="unknown") as
total_uknown_users

FROM
bigquery-public-data.new_york_citibike.citibike_trips
WHERE
 usertype = "Subscriber" AND IS_NAN(tripduration)=FALSE
GROUP BY
gender
```

Note: There are 5 tables corresponding to 5 queries:

In [10]:
```python
#Read and Display the data
citi_list= os.listdir('citi_cycle data/')
for g in range(len(citi_list)):
    globals()['citi'+str(g)]= pd.read_csv('citi_cycle data/'+citi_list[g])
    display(globals()['citi'+str(g)])
```

	gender	number_of_subscribers	total_male_users	total_female_users	subscriber_trip_p
0	male	35308523	35611787	11376412	75.2
1	female	11188711	35611787	11376412	23.8
2	unknown	420338	35611787	11376412	0.8

	birth_year	number_of_subscribers	subscriber_trip_fraction	total_number_of_trips_by_s
0	1985.0	1935170	4.124617e-02	46917
1	1986.0	1866590	3.978445e-02	46917
2	1984.0	1846663	3.935973e-02	46917
3	1988.0	1827755	3.895673e-02	46917
4	1987.0	1814123	3.866617e-02	46917
...	...	...	...	
103	1915.0	24	5.115354e-07	46917
104	1890.0	18	3.836516e-07	46917
105	1909.0	15	3.197096e-07	46917
106	1889.0	13	2.770817e-07	46917
107	1903.0	3	6.394193e-08	46917

108 rows × 4 columns

	usertype	total_trip_duration	Average_trip_duration	Number_of_trips
0	Customer	2.213864e+08	35.758535	6191149
1	Subscriber	6.305580e+08	13.439698	46917572

	birth_year	trips_by_birth_year	percent_of_total_trips
0	NaN	0	0.0000
1	1874.0	6	0.0000
2	1884.0	2	0.0000
3	1885.0	2202	0.0041
4	1886.0	157	0.0003
...	...	...	...
107	1998.0	131231	0.2471
108	1999.0	74922	0.1411
109	2000.0	35246	0.0664
110	2001.0	11150	0.0210
111	2002.0	723	0.0014

112 rows × 3 columns

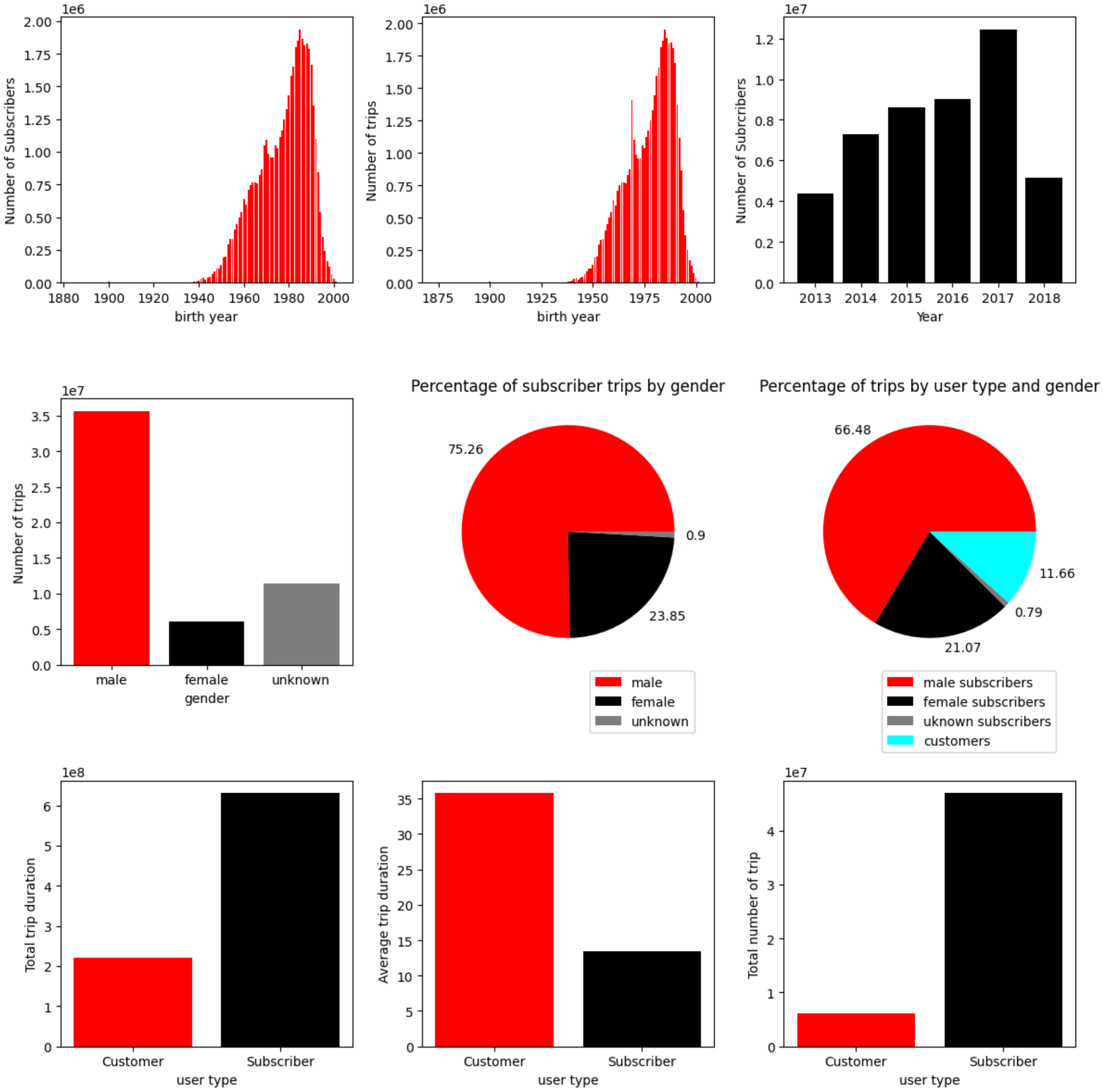
	year	num_of_subs
0	2013	4370245
1	2014	7287721
2	2015	8626638
3	2016	9026384
4	2017	12441957
5	2018	5164627

In [11]:

```
#some calculations
subs_ratio= list(citi0.number_of_subscribers*100/(citi0.total_female_users+citi0
subs_ratio.append(100-sum(subs_ratio))

print(subs_ratio)
subs_ratio= np.array(subs_ratio)
gender_user_type = ['male subscribers','female subscribers','uknown subscribers'
#Visualization
fig,axes=plt.subplots(3,3,figsize=(12,12))
axes[0,0].bar(citi1.birth_year,citi1.number_of_subscribers,color='red')
axes[0,2].bar(citi4.year,citi4.num_of_subs,color='black')
axes[0,1].bar(citi3.birth_year,citi3.trips_by_birth_year,color='red')
axes[1,1].pie(citi0.subscriber_trip_percent,labels=round(citi0.subscriber_trip_p
axes[1,1].legend(citi0.gender,bbox_to_anchor=(1,0))
axes[1,0].bar(citi0.gender, height=[citi0.total_male_users[0],citi0.total_uknown
axes[1,2].pie(subs_ratio,labels=np.round(subs_ratio,2),colors=['red','black','gr
axes[1,2].legend(gender_user_type,bbox_to_anchor=(1,0))
axes[2,0].bar(citi2.usertype,citi2.total_trip_duration, color=['red','black'])
axes[2,1].bar(citi2.usertype,citi2.Average_trip_duration, color=['red','black'])
axes[2,2].bar(citi2.usertype,citi2.Number_of_trips, color=['red','black'])
axes[0,0].set_xlabel('birth year')
axes[0,0].set_ylabel('Number of Subscribers')
axes[0,1].set_xlabel('birth year')
axes[0,1].set_ylabel('Number of trips')
axes[0,2].set_xlabel('Year')
axes[0,2].set_ylabel('Number of Subrcribers')
axes[1,0].set_xlabel('gender')
axes[1,0].set_ylabel('Number of trips')
axes[1,1].set_title('Percentage of subscriber trips by gender')
axes[1,2].set_title('Percentage of trips by user type and gender')
axes[2,0].set_xlabel('user type')
axes[2,1].set_xlabel('user type')
axes[2,2].set_xlabel('user type')
axes[2,0].set_ylabel('Total trip duration')
axes[2,1].set_ylabel('Average trip duration')
axes[2,2].set_ylabel('Total number of trip')
plt.tight_layout()
```

```
[66.48347453142395, 21.067558753674373, 0.7914669984238558, 11.657499716477815]
```



Insights from the Visuals

Row 1 of Visuals:

- Maximum number of people who use bike-sharing are born between 1980 to 1990 i.e age 23 to 33 years (data is from 2013 to 2018)
- Maximum number of subscribers belong to the same age group
- There is an increase in number of trips by subscribers every year indicating that more people are switching towards the membership option

Row 2 of Visuals:

- Male users dominate in terms of number of trips
- Men account for more than 3/4 th of total trips by subscribers are
- 87.4 percent trips are taken by the subscribers out of which 66 % of them are men

Row 3 of Visuals

- The row 3 stats show a good allignment with the Divvy's data strengthening the credibility of insights.
- The inverse correlation observed between average duration and number of trips for both user types also matches with the previous analysis of Divvys data.

Limitations

- The Divvys dataset doesn't include information about the amount spent on each trip making the analysis on cost related benefits of membership out of reach.
- Information on battery duration and battery percentage at the end and start of ride may help in understanding the reason for usage patterns electric bike and electric scooters by both the user types
- Divvys dataset doesn't have any personal information on the age and gender of the customers resulting in usage of other proxy datasets to gain insights on target audience.
- The proxy dataset of NY citi bikes is not representative of the current time period.

Recommendations

- Men in the age between 23 to 33 years are most likely to buy an annual membership amongst the casual riders
- Reason for low bike-sharing trips and subscriptions by Women must be investigated through surveys
- Marketing must focus on the advantages of shorter bike-sharing trips and use it to market annual memberships emphasizing on the low cost.
- A discounted youth/student pass can encourage new target audience. However, more data-driven insights are required to support the hypothesis
- Battery life of electric bikes and scooters must be inspected and enhanced to balance the usage between all the three types of rides
- TCasual riders show high usage on weekends — Cyclistic could promote weekend membership discounts or partner with weekend events.
- Increasing in prices of single ride and full day passes on weekends can increase the advantages of choosing a membership option.

In [ ]:

```

```