RevoU Mini Course

Case Study: San Francisco Bikeshare Insight

Muhammad Rizky Anugraha

in /muhammd-rizky-anugraha

Project Background

- Bikeshare in San Francisco Bay Area, including San Francisco, San Jose, Oakland, Berkeley and Emeryville
- During H2 of 2017 (July 2017 - December 2017), the dataset shows growth in bikeshare users in this city
- This project is made to gain insight related to customer demography and behavior to improve Bikeshare service.

Data Gathering

Dataset

In this case study, i am using dataset from Google BiqQuery Public. The list of dataset i used inculding:

```
`bigquery-public-data.san_francisco_bikeshare.bikeshare_trips`
`bigquery-public-data.san_francisco_bikeshare.bikeshare_station_info`
`bigquery-public-data.san_francisco_bikeshare.bikeshare_regions`
`bigquery-public-data.san_francisco_bikeshare.bikeshare_station_status`
```

Tool Used

Defining the Problem

Defining the Problem

1. How was the monthly trend of the trips in H2 2017?
2. How is characteristic of Bike Users in H2 2017?
3. How is distribution trip by region in H2 2017?
4. At what time of the day do the users tpically have the trips in H2 2017?
5. What are the most (and least) freaquently visited bike stations in H2 2017?

Data Preparation

```sql
1   select
2   trip_id,
3   date(start_date) start_date,
4   FORMAT_TIME("%R",extract(time from start_date)) start_hour,
5   case when EXTRACT(hour FROM start_date) between 4 and 11 then 'Morning'
6        when EXTRACT(hour FROM start_date) between 12 and 18 then 'Afternoon'
7        when EXTRACT(hour FROM start_date) between 19 and 23 then 'Evening'
8        else 'Night'
9   end as Time_Category,
10  duration_sec,
11  start_station_name,
12  concat(start_station_name," - ",end_station_name) route,
13  bike_number,
14  num_bikes_available,
15  subscriber_type,
16  member_birth_year,
17  (2022-member_birth_year) age,
18  case when (2022-member_birth_year) < 40 then 'Young (<40 Y.O)'
19       when (2022-member_birth_year) between 40 and 60 then 'Adult (40-60 Y.O)'
20       else 'Senior (>60 Y.O)'
21  end as age_group,
22  member_gender,
23  reg.name region_name
24  from
25  bigquery-public-data.san_francisco_bikeshare.bikeshare_trips as trip
26  left join bigquery-public-data.san_francisco_bikeshare.bikeshare_station_info as info
27  on trip.start_station_name = info.name
28  left join bigquery-public-data.san_francisco_bikeshare.bikeshare_regions as reg
29  on info.region_id = reg.region_id,
30  bigquery-public-data.san_francisco_bikeshare.bikeshare_station_status as status
31  where start_date between '2017-06-01' and '2017-12-31'
32  and info.name is not null
33  and member_gender is not null
34  and member_gender != 'Other'
35  order by trip.start_date asc
```
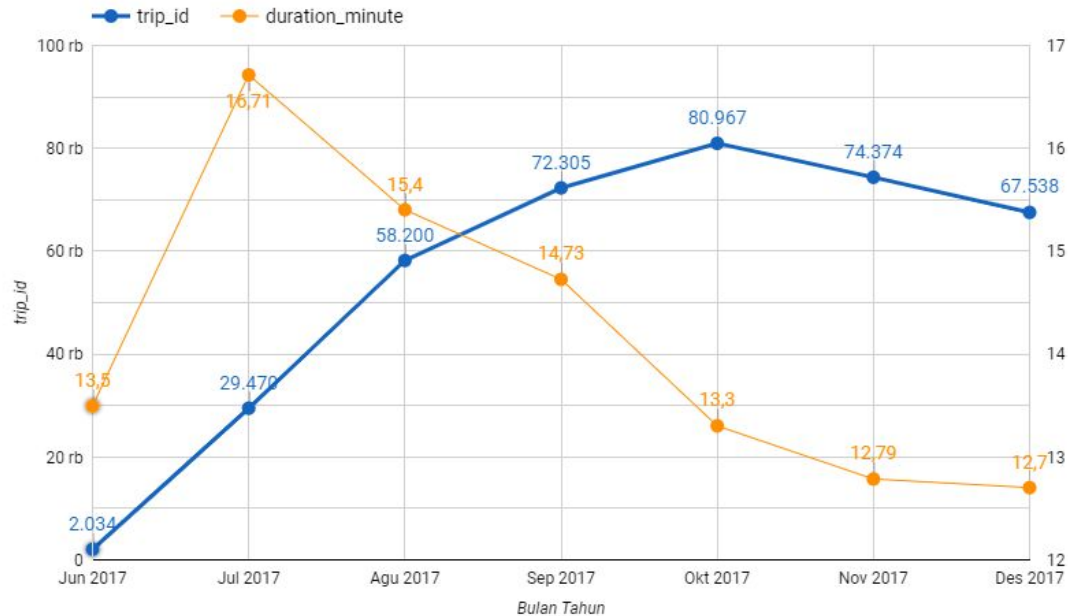
Data Query using Google Big Query
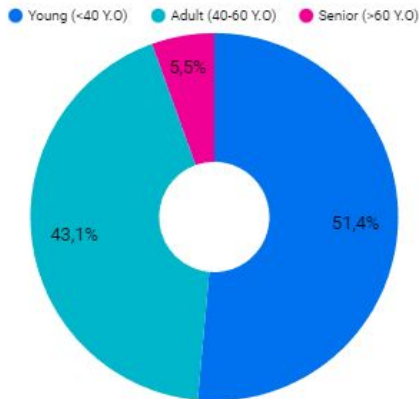
Data Visualization

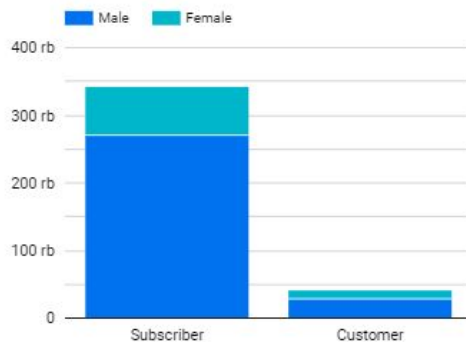# Number of Trips during H2 of 2017 (Jul 2017 - Dec 2017)

There is an increase in the number of trips which peaked in October. Bike usage duration peaked in July. However, there is a decline in the average bike usage duration.

Demography of Bike Users during H2 of 2017 (Jul 2017 - Dec 2017)

Duration Trip by Age Group

- Young (<40 Y.O)
- Adult (40-60 Y.O)
- Senior (>60 Y.O)

5,5%
43,1%
51,4%

Distribution Trip by Subscriber Type and Gender

- Male
- Female

400 rb
300 rb
200 rb
100 rb
0

Subscriber | Customer

Distribution Trip by SubscriberType and Gender

	subscriber_type	member_gender	trip_id ▾	age	duration_minute
1.	Subscriber	Male	270.053	40	11,35
2.	Subscriber	Female	72.780	38	13,66
3.	Customer	Male	29.044	37	28,75
4.	Customer	Female	13.011	35	36,46

1 - 4 / 4   <   >

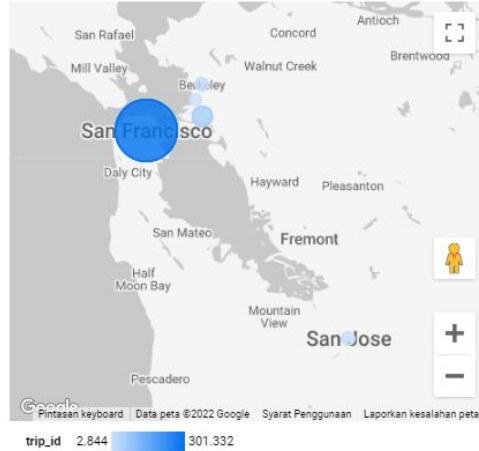Bike users are dominated by young (<40 y.o) and adult (40-60 y.o) with young ages at most.

Most of the users were male compared to female, There are more subscribers* than customers* among them.

*(Subscriber = annual or 30-day member; Customer = 24-hour or 3-day member).

Distribution Trips by Region during H2 of 2017 (Jul 2017 – Dec 2017)


Most of the trips occur in San Francisco compared to other regions. San Francisco also has the most station, bike usage and most bike available among the other regions.

	region_name	trip_id ▾	start_station_name	bike_number	num_bikes_ava...
1.	San Francisco				
2.	Oakland				
3.	San Jose				
4.	Berkeley				
5.	Emeryville				

1 - 5 / 5 ‹ ›

trip_id 2.844 ▬▬ 301.332

Distribution Trip by Time Category during H2 of 2017 (Jul 2017 - Dec 2017)

Duration Trip by Time Category

● Afternoon ● Morning ● Evening ● Night

48%
37,7%
13,4%
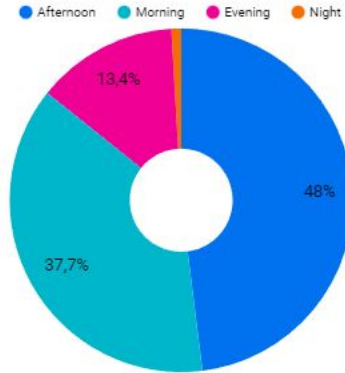
Most trips occur in the morning and afternoon with afternoon trips having the most trips. Although there is a different frequency of trips in all time categories, the average duration of all time categories is relatively close ranging from 13 - 16 minutes.

Distribution Trip by Time Category

	Time_Category	trip_id ▾	duration_minute
1.	Afternoon	184.848	14,71
2.	Morning	145.114	13
3.	Evening	51.397	13,7
4.	Night	3.529	16,54

1 - 4 / 4 ＜ ＞

Bike Station Visit during H2 of 2017
(Jul 2017 - Dec 2017)

Top 5 Most Visited Station

|	start_station_name	region_name	trip_id ▾	duration_minute
1.	San Francisco Caltrain (Townsend St ...	San Francisco	11.843	73.626.264,8
2.	San Francisco Ferry Building (Harry B...	San Francisco	11.393	83.516.769,33
3.	Market St at 10th St	San Francisco	10.853	63.102.946,93
4.	Berry St at 4th St	San Francisco	10.056	66.867.016,53
5.	Montgomery St BART Station (Marke...	San Francisco	9.995	58.449.086,4

1 - 5 / 236 ‹ ›

Top 5 Least Visited Station

	start_station_name	region_name	trip_id ▴	duration_minute
1.	Jackson St at 11th St	Oakland	1	5.381,6
2.	Ellsworth St at Russell St	Berkeley	1	3.992,8
3.	Harmon St at Adeline St	Berkeley	2	23.667,47
4.	Empire St at 1st St	San Jose	3	12.482,67
5.	Pierce Ave at Market St	San Jose	7	34.546,4

1 - 5 / 236 ‹ ›

Correlation between Number of Trip and Trip Duration
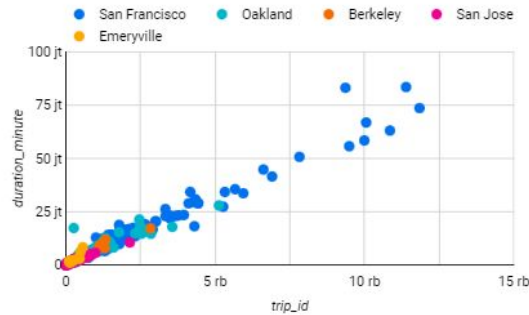
The Most visited station is San Francisco Caltrain from San Francisco and the least visited station is Jackson St at 11th St from Oakland.

The amount of the trip is correlated with an average duration of the trip. San Francisco is the region with the most visited station compared to other regions.

Conclusion and Recomendation

Conclusion

- The users tend to have a trip with a low duration.
- Most users are young and adult age within the range of 0 - 60 years old.
- There are more male users than female users.
- There are more subscribers than customers.
- Male users have more bike usage but less duration. Female users have less bike usage but higher duration.
- San Francisco has the most trips, available bike and bike usage.
- Most of the trips occurred in the morning and afternoon.
- The amount of the trip is correlated with the average duration of the trip.

Recomendation

- Integrating bike stations with the workplace, university and public places.
- Maintain subscribers and attract new users to become new customers.
- Making bicycles with male and female options, to attract more users from both genders.
- Increase available bikes in the other region.
- Charging higher prices during morning and afternoon.

Thank you
Feel free to contact me for feedback

Muhammad Rizky Anugraha

anugrahamrizky@gmail.com

/muhammad-rizky-angraha

0896-4442-7730