Portfolio Data Analytics
1st

San Francisco Bikeshare Trip Insight

TAKEN FROM CASE STUDY
RevoU Mini Course - Data Analytics

DHEA AMALIA LUTFIANI
(7 FEB – 18 FEB)

Case Study Instructions

QUESTION

Table of interest :

‘bigquery-public-data.san_francisco_bikeshare.bikeshare_trips’

1. Look at this data and start thinking. List down 3 trends/points that you want to show.
2. From here, try to explore the data and make changes, filter, and prepare the data that you need.
3. Create some visualizations or dashboard with the best type of chart you have learned.

   The easiest is with Google Data Studio or Google Sheets.
4. Then, make 1-2 slides from the Graphs with the insights you got to present your findingsto the stakeholders

   (read this article from HBR)

Preview Data

Schema from the Dataset :

Field name	Type	Mode	Collation	Policy Tags	Description
trip_id	STRING	REQUIRED			Numeric ID of bike trip
duration_sec	INTEGER	NULLABLE			Time of trip in seconds
start_date	TIMESTAMP	NULLABLE			Start date of trip with date and time, in PST
start_station_name	STRING	NULLABLE			Station name of start station
start_station_id	INTEGER	NULLABLE			Numeric reference for start station
end_date	TIMESTAMP	NULLABLE			End date of trip with date and time, in PST
end_station_name	STRING	NULLABLE			Station name for end station
end_station_id	INTEGER	NULLABLE			Numeric reference for end station
bike_number	INTEGER	NULLABLE			ID of bike used
zip_code	STRING	NULLABLE			Home zip code of subscriber (customers can choose to manually enter zip at kiosk however data is unreliable)
subscriber_type	STRING	NULLABLE			Subscriber = annual or 30-day member; Customer = 24-hour or 3-day member
c_subscription_type	STRING	NULLABLE			
start_station_latitude	FLOAT	NULLABLE			
start_station_longitude	FLOAT	NULLABLE			
end_station_latitude	FLOAT	NULLABLE			
end_station_longitude	FLOAT	NULLABLE			
member_birth_year	INTEGER	NULLABLE			
member_gender	STRING	NULLABLE			

Preview Data

Schema from the Dataset :

bike_share_for_all_trip	STRING	NULLABLE
start_station_geom	GEOGRAPHY	NULLABLE
end_station_geom	GEOGRAPHY	NULLABLE

From this data preview, we can find out the description of each column and know which columns can be used to answer problems or which columns can be useful for finding new insights.

Preview Data

Preview the Data :

Row	trip_id	duration_sec	start_date	start_station_name	start_station_id	end_date	end_station_name	end_station_id	bike_number	zip_code	subscriber_type
1	4420151206105200	5780	2015-12-06 10:52:00 UTC	Mezes	83	2015-12-06 12:28:00 UTC	Mezes	83	44	94064	Customer
2	65020151013190200	255	2015-10-13 19:02:00 UTC	Mezes	83	2015-10-13 19:07:00 UTC	Mezes	83	650	94063	Subscriber
3	5220160303220000	1377	2016-03-03 22:00:00 UTC	Redwood City Public Library	24	2016-03-03 22:23:00 UTC	Mezes	83	52	94063	Customer
4	12120151024142000	422	2015-10-24 14:20:00 UTC	Redwood City Public Library	24	2015-10-24 14:27:00 UTC	Mezes	83	121	94063	Subscriber
5	12720160218173400	341	2016-02-18 17:34:00 UTC	Redwood City Caltrain Station	22	2016-02-18 17:40:00 UTC	Mezes	83	127	94063	Subscriber
6	8420151214180800	256	2015-12-14 18:08:00 UTC	Redwood City Caltrain Station	22	2015-12-14 18:12:00 UTC	Mezes	83	84	94063	Subscriber
7	57720160201180900	266	2016-02-01 18:09:00 UTC	Redwood City Caltrain Station	22	2016-02-01 18:13:00 UTC	Mezes	83	577	94063	Subscriber
8	57720160326111200	3951	2016-03-26 11:12:00 UTC	Redwood City Caltrain Station	22	2016-03-26 12:18:00 UTC	Mezes	83	577	94117	Customer
9	64220160414170700	256	2016-04-14 17:07:00 UTC	Redwood City Caltrain Station	22	2016-04-14 17:11:00 UTC	Mezes	83	642	94063	Subscriber
10	6420160509173100	326	2016-05-09 17:31:00 UTC	Redwood City Caltrain Station	22	2016-05-09 17:37:00 UTC	Mezes	83	64	94063	Subscriber

Row	c_subscription_type	start_station_latitude	start_station_longitude	end_station_latitude	end_station_longitude	member_birth_year	member_gender	bike_share_for_all_trip	start_station_geom	end_station_geom
1	*null*	*null*	*null*	*null*	*null*	*null*	*null*	*null*	*null*	*null*
2	*null*	*null*	*null*	*null*	*null*	*null*	*null*	*null*	*null*	*null*
3	*null*	*null*	*null*	*null*	*null*	*null*	*null*	*null*	*null*	*null*
4	*null*	*null*	*null*	*null*	*null*	*null*	*null*	*null*	*null*	*null*
5	*null*	*null*	*null*	*null*	*null*	*null*	*null*	*null*	*null*	*null*
6	*null*	*null*	*null*	*null*	*null*	*null*	*null*	*null*	*null*	*null*
7	*null*	*null*	*null*	*null*	*null*	*null*	*null*	*null*	*null*	*null*
8	*null*	*null*	*null*	*null*	*null*	*null*	*null*	*null*	*null*	*null*
9	*null*	*null*	*null*	*null*	*null*	*null*	*null*	*null*	*null*	*null*
10	*null*	*null*	*null*	*null*	*null*	*null*	*null*	*null*	*null*	*null*

Defining Question

List down trends/points that you want to show :

1. Number of trips per month
2. Average duration of the trips per month
3. The monthly trend of the total trips and the average of duration time
4. Who are the most enthusiasts or renters of this San Francisco bikeshare? (Male or Female)
5. From which customers do we get more revenue? (from subscriber or customer)

Exploring Data

Include SQL with BigQuery

- There are 2 queries used in this project.
  The first query is to answer problem questions for numbers 1 and 2.
  And the second query is to answer other problems, namely problems number 3, 4, and 5.
- Why are these two queries separated?
  Because I personally want to answer the first and second problems in the entire span of time. This is done in order to see what happened and how the picture or outline of the company's situation.
  And to answer another problem, I use a more specific time, starting in 2017.
  Why did I choose this range?
  Will be answered in the second probem later.

Exploring Data

Include SQL with BigQuery

1. First Query

```sql
SELECT trip_id,
  duration_sec,
  start_date,
  start_station_id,
  start_station_name,
  subscriber_type,
  member_gender
FROM `bigquery-public-data.san_francisco_bikeshare.bikeshare_trips`
```

And the result :

Query results

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS

Row	trip_id	duration_sec	start_date	start_station_id	start_station_name	subscriber_type	member_gender
1	15520160318103300	6032	2016-03-18 10:33:00 UTC	83	Mezes	Customer	*null*
2	17420160223122500	801	2016-02-23 12:25:00 UTC	83	Mezes	Customer	*null*
3	20820160318103300	6079	2016-03-18 10:33:00 UTC	83	Mezes	Customer	*null*
4	65320150924172200	2618	2015-09-24 17:22:00 UTC	83	Mezes	Customer	*null*
5	15720150926125700	1008	2015-09-26 12:57:00 UTC	83	Mezes	Subscriber	*null*
6	4420151206105200	5780	2015-12-06 10:52:00 UTC	83	Mezes	Customer	*null*
7	24620160214052500	1021	2016-02-14 05:25:00 UTC	26	Kaiser Hospital	Subscriber	*null*
8	65020151005164100	587	2015-10-05 16:41:00 UTC	21	Sequoia Hospital	Subscriber	*null*
9	17420160223120000	1334	2016-02-23 12:00:00 UTC	23	San Mateo County Center	Customer	*null*
10	14920151225131700	1113	2015-12-25 13:17:00 UTC	23	San Mateo County Center	Subscriber	*null*

Exploring Data

Include SQL with BigQuery

2. Second Query

```sql
SELECT trip_id,
  duration_sec,
  start_date,
  start_station_id,
  start_station_name,
  subscriber_type,
  member_gender
FROM `bigquery-public-data.san_francisco_bikeshare.bikeshare_trips`
WHERE start_date > '2017-01-01'
```

And the result :

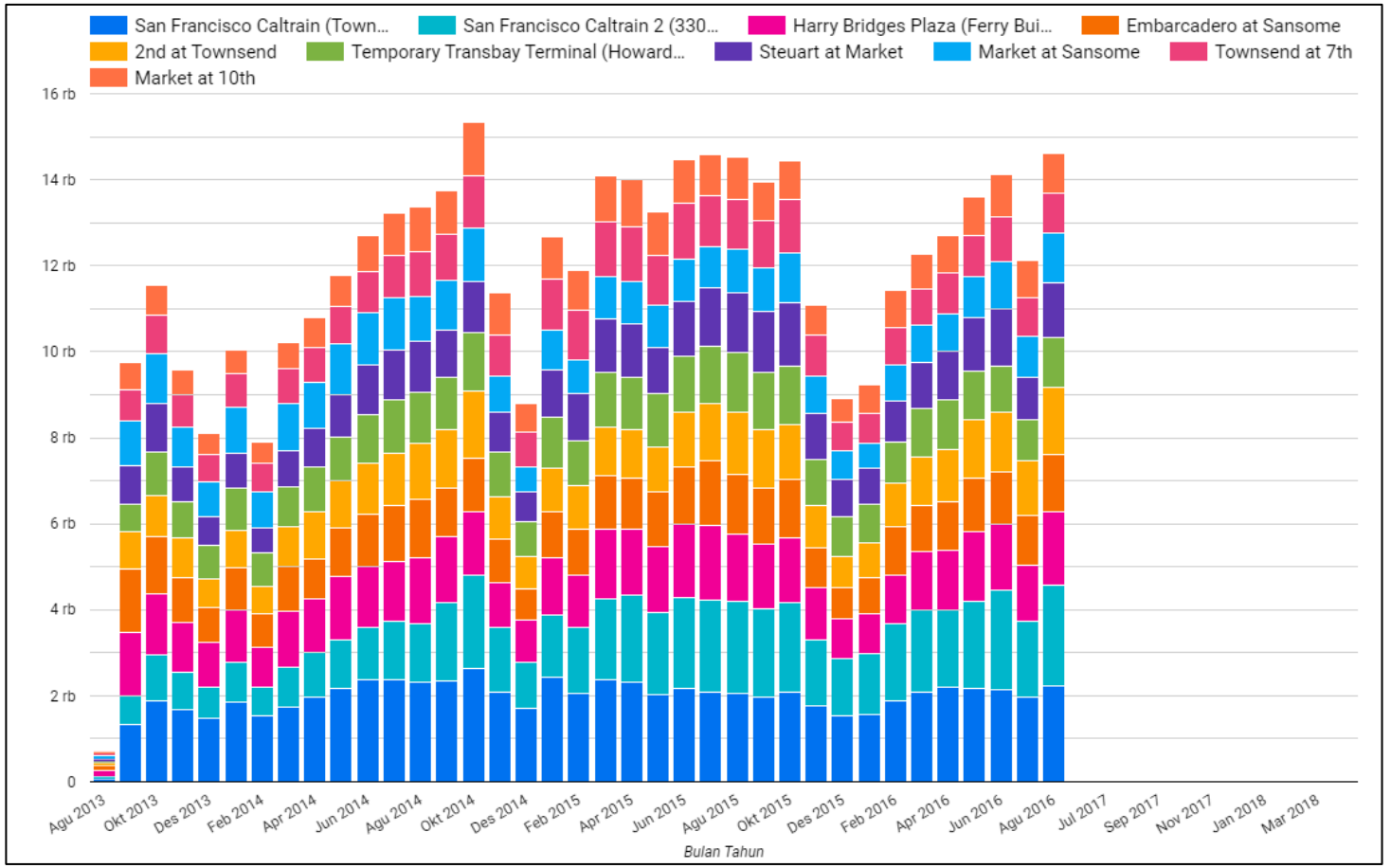
Query results

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS

Row	trip_id	duration_sec	start_date	start_station_id	start_station_name	subscriber_type	member_gender
1	11772018012512354700	1289	2018-01-25 12:35:47 UTC	198	Snow Park	Customer	Male
2	35082017122611364600	2014	2017-12-26 11:36:46 UTC	198	Snow Park	Subscriber	Male
3	18072017112013471000	2032	2017-11-20 13:47:10 UTC	198	Snow Park	Subscriber	Male
4	4962017072412014500	1401	2017-07-24 12:01:45 UTC	198	Snow Park	Subscriber	Male
5	1102017112113071300	2147	2017-11-21 13:07:13 UTC	198	Snow Park	Subscriber	Male
6	11022018012513063900	1537	2018-01-25 13:06:39 UTC	198	Snow Park	Subscriber	Female
7	13892017102016295800	3744	2017-10-20 16:29:58 UTC	198	Snow Park	Customer	*null*
8	9632017102215414300	6242	2017-10-22 15:41:43 UTC	198	Snow Park	Customer	Female

Visualization with Insight

1. Number of trips per month



This diagram shows the number of trips per month at each station. We can say that :

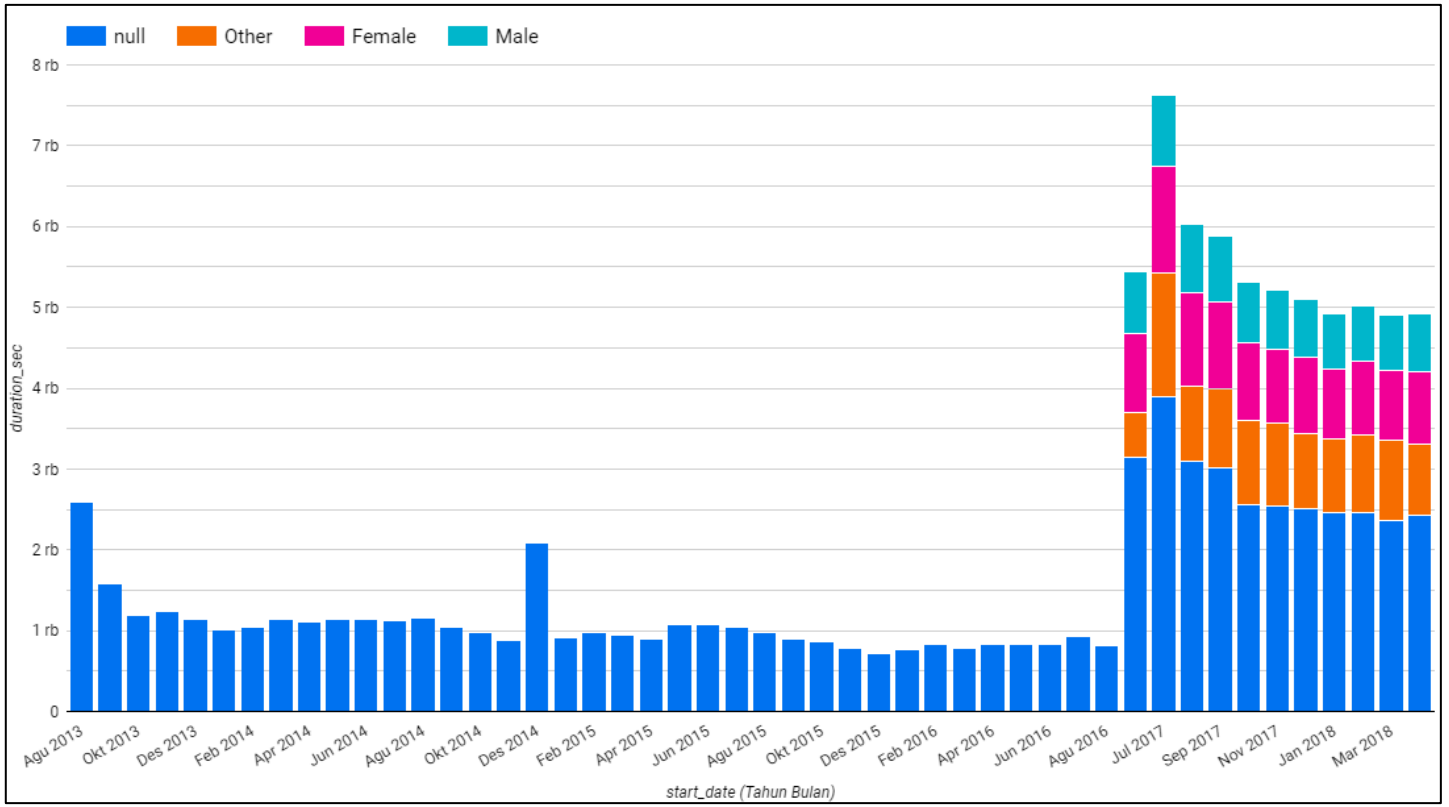
- Where Market at Sansome station has the highest number of trips per month than any other station.
- Every year in the end of the year, the users of this San Francisco bikeshare service are decreasing.
- And in October 2014, had the most number of trips.

Visualization with Insight

2. Average duration of the trips per month

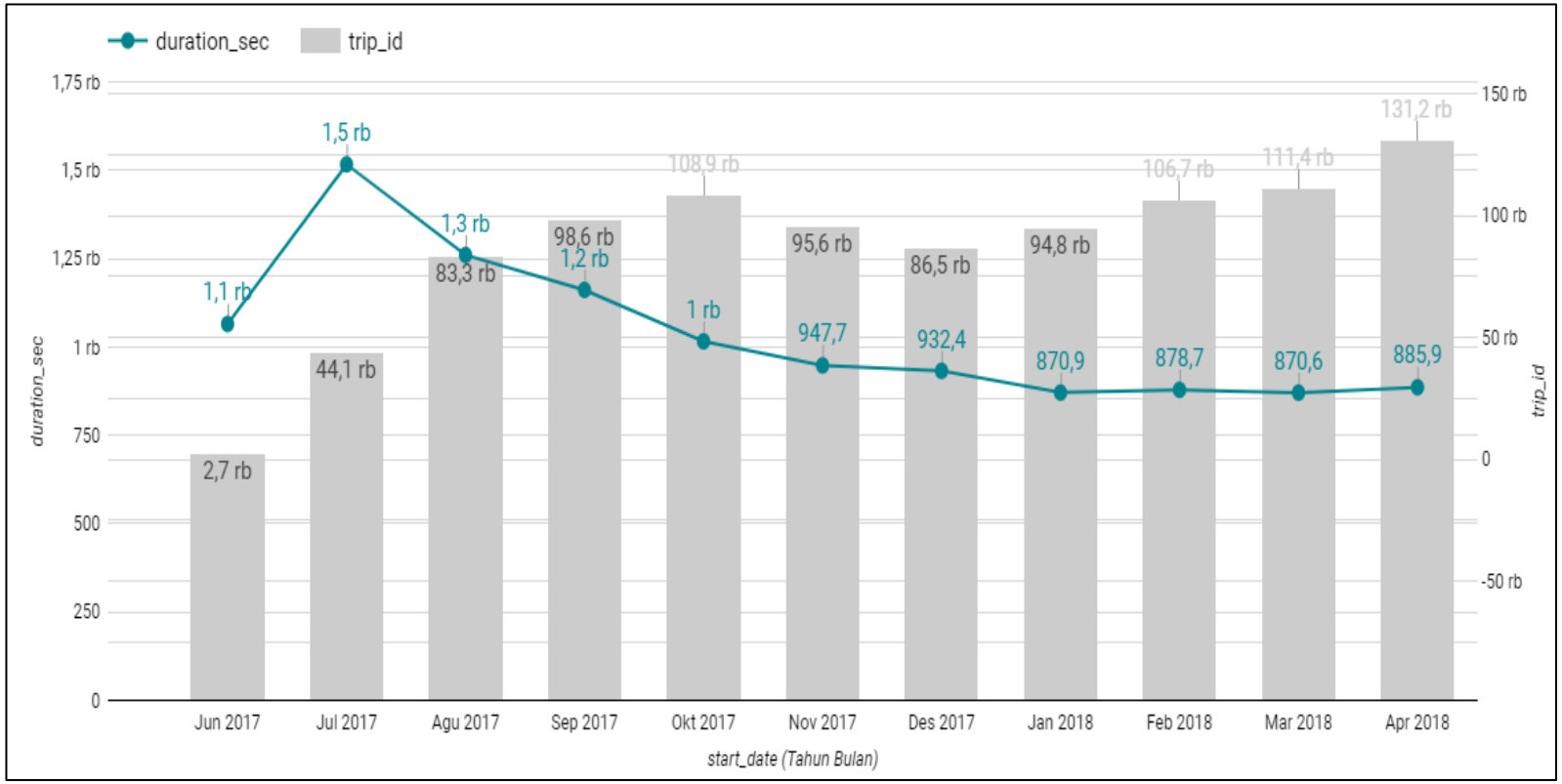
This chart shows the average duration of bikeshare usage/ borrowing (grouped by gender). However, it can be seen that:

- The implementation of this gender grouping started in June 2017.
- From Sept 2016 to May 2017 no data was recorded.
- And the highest average duration of the trips per month occurred in July 2017



Visualization with Insight

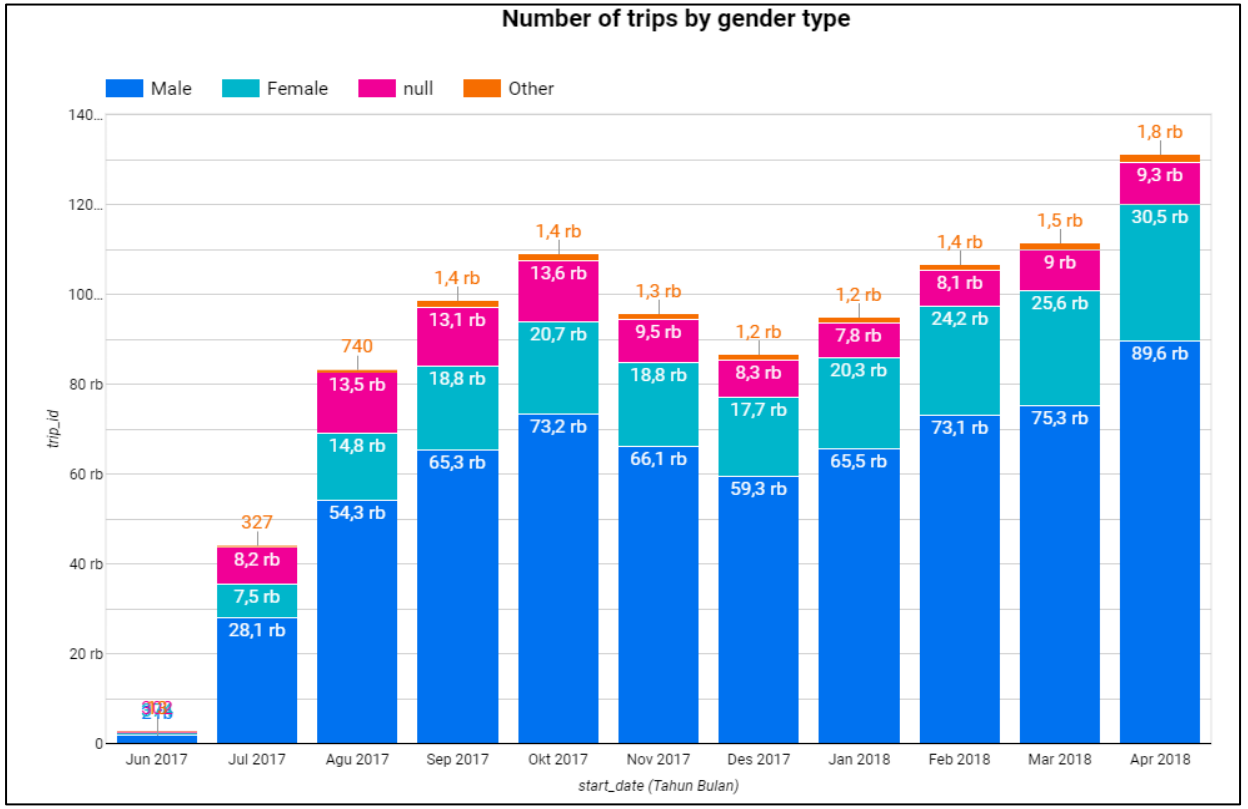
3. The monthly trend of the total trips and the average of duration time (after re-open the bikeshare)



- Means that for a large number of trips, it is not necessarily a lot of time spent on bikeshare for each trip.
- Where in the diagram it can be seen that the trend in the number of trips from the beginning of the bikeshare reopening until April 2018 tends to increase but for the average duration of use of this bikeshare, it tends to decrease.
- Which means the number of trips is a lot but the use is only for a short time.

Visualization with Insight

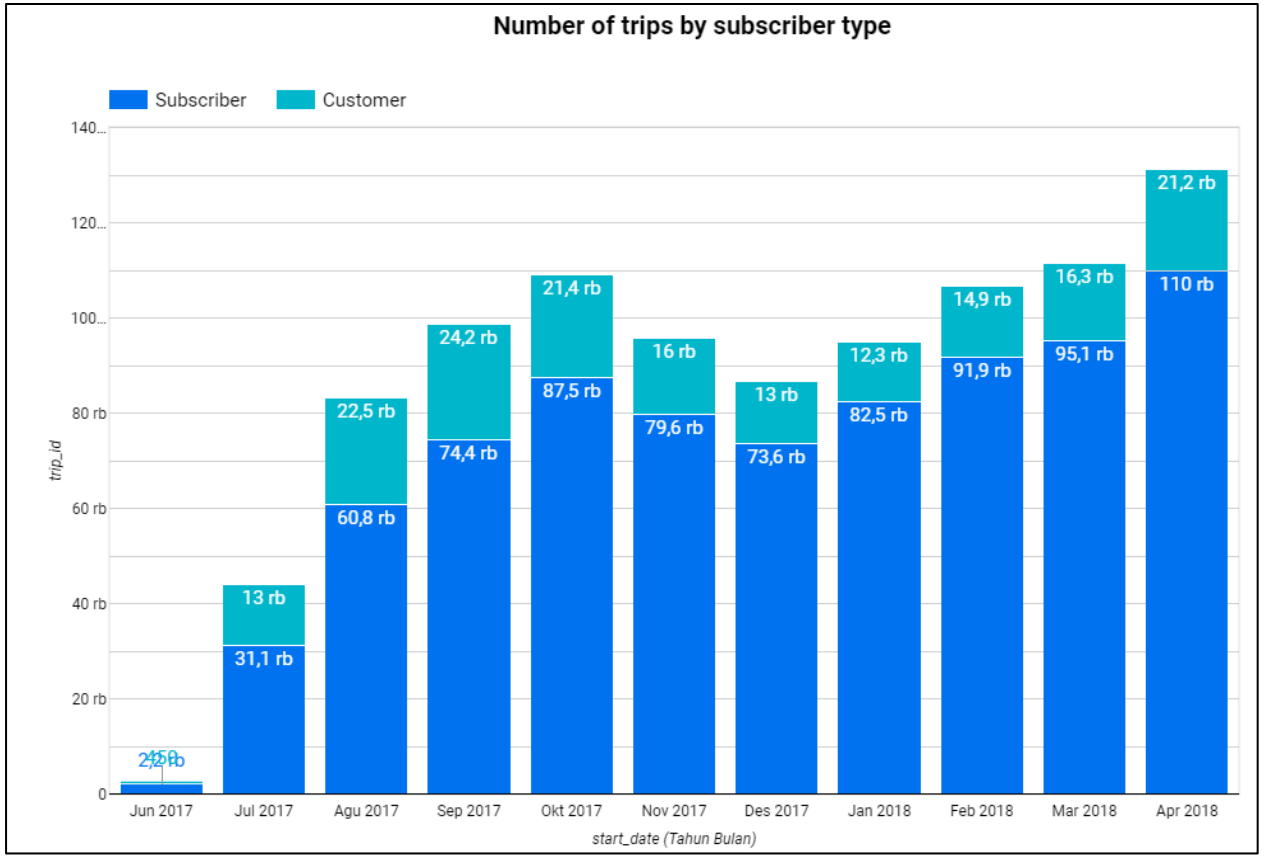
4. Who are the most enthusiasts or renters of this San Francisco bikeshare? (Male or Female)



- In this section we can focus on the data starting in 2017, because in this data there is already a gender distinction between male and female and we can find more detailed insight.
- From the diagram below, it can be seen that the most users of bikeshare are male.
- Which reaches more than half the number of service users each month.
- And it is undeniable that there are still service users who do not want to provide information about their gender.

Visualization with Insight

5. From which customers do we get more revenue? (from subscriber or customer)



From the diagram below, it can be seen that :

- We got more revenue from subscriber.
- Which is a customer segment that becomes a monthly or annual member who has a longer period.

THANK YOU

DHEA AMALIA LUTFIANI
(7 FEB – 18 FEB)