

# **ANALYSIS AND VISUALIZE: SAN FRANCISCO BIKE SHARE**

**Jihan Rana Ayunda Dewana**  
**RevoU Mini Course - Data Analyst Case Study**

# **LIST OF CONTENT PORTFOLIO**

## **01. Main Instruction**

About the instruction or a guide from RevoU about doing analyze and visualize for the data

## **03. Analyze Point**

Main Analyst / Business Question for the basis of data processing.

## **02. Table of Interest**

List of data (table) such as table name and source for analyze and visualize

## **04. Data Process and Visualization**

Data Process step by step until becomes visualize and easy to read for getting the point of view (insight)

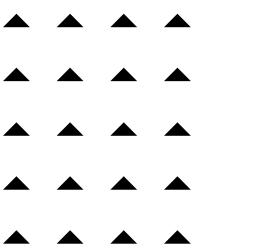
## **Tools**



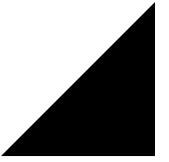
Google  
BigQuery



# 01. MAIN INSTRUCTION



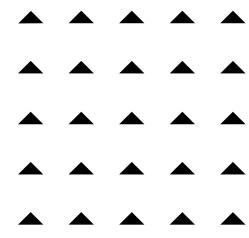
RevoU gives the instruction to explore the data and make changes, filter and prepare the data needs from "San Francisco Bikeshare - Bigquery Public Data". Then, make 1-2 slides from the graphs with the insights you got to present the findings to the stakeholders.



**Lets Get Started**



# 02. TABLE OF INTEREST

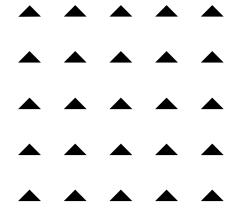


## How We Work ?

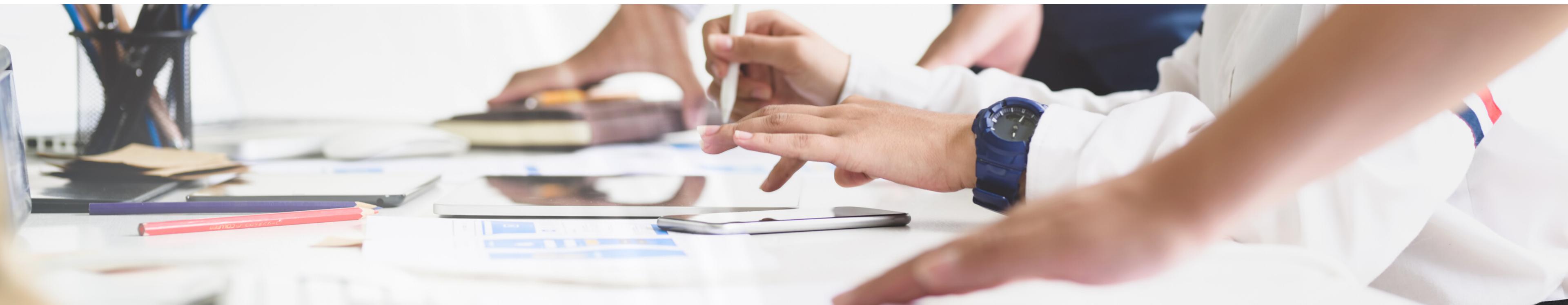
Using SQL Querying to explore four (4) data set of  
"San Francisco Bike Share"

```
`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`
```

# 03. ANALYZE POINT



1. What is the distribution or characteristic of Bike users?
2. What time of the day do the bikers typically have the trips?
3. What was the monthly trend of the trips in H2 2017?
4. How about intensity levels of health cycling of subscribers in 2017? What level are the most (and least)?
5. How is the distribution trip by region? What are the most (and least) frequently visited bike stations?



# 04. DATA PROCESS AND VISUALIZATION

## How does this Work?

1. Using SQL Querying to explore four (4) data sets of "San Francisco Bike Share" based on Analyze Point
2. Visualize the analysis
3. More information needs like:
  - Age Classification/ Age Group  
[https://www.indexmundi.com/united\\_states/age\\_structure.html](https://www.indexmundi.com/united_states/age_structure.html)
  - Intensity Levels of Health Cycling  
<https://www.welovecycling.com/wide/2021/11/09/how-many-hours-per-week-should-you-ride-for-health/>
  - Parts of the Day/Time  
<https://www.britannica.com/dictionary/eb/qa/parts-of-the-day-early-morning-late-morning-etc>

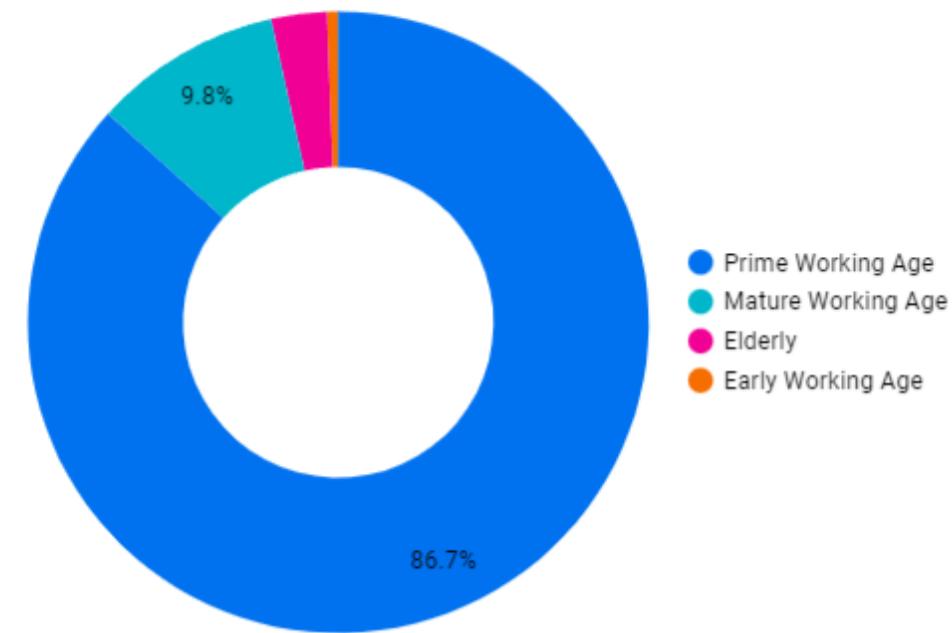
# SQL QUERYING

```
1 select
2   trip_id,
3   duration_sec,
4   case when duration_sec = 1800 then 'Shorter Ride with Sprints'
5   |||   when duration_sec > 1800 and duration_sec <= 3600 then 'Hilly Ride'
6   |||   when duration_sec > 3600 then 'Longer Ride'
7   else 'Other'
8 end as intensity_level,
9 date(start_date) start_date,
10 start_station_name,
11 concat(start_station_name, "-", end_station_name) route,
12 time (start_date) time,
13 case when time (start_date) between '5:00:00' and '11:59:00' then 'Morning (5 AM to 11:59 AM)'
14 |||   when time (start_date) between '12:00:00' and '16:59:00' then 'Afternoon (12 PM to 04:59 PM)'
15 |||   when time (start_date) between '17:00:00' and '20:59:00' then 'Evening (5 PM to 08:59 PM)'
16 else 'Night (9 PM to 04:59 AM)'
17 end as day_part,
18 bike_number,
19 subscriber_type,
20 member_birth_year,
21 (2022-member_birth_year) age,
22 case when (2022-member_birth_year) between 15 and 24 then 'Early Working Age'
23 |||   when (2022-member_birth_year) between 25 and 54 then 'Prime Working Age'
24 |||   when (2022-member_birth_year) between 55 and 64 then 'Mature Working Age'
25 else 'Elderly'
26 end as age_class,
27 member_gender,
28 c.name region_name
29 from `bigquery-public-data.san_francisco_bikeshare.bikeshare_trips` a
30 left join `bigquery-public-data.san_francisco_bikeshare.bikeshare_station_info` b on a.start_station_name = b.name
31 left join `bigquery-public-data.san_francisco_bikeshare.bikeshare_regions` c on b.region_id = c.region_id
32 where start_date between '2017-07-01' and '2017-12-31'
33 and b.name is not null
34 and member_birth_year is not null
35 and member_gender is not null
36 and member_gender != 'Other'
```

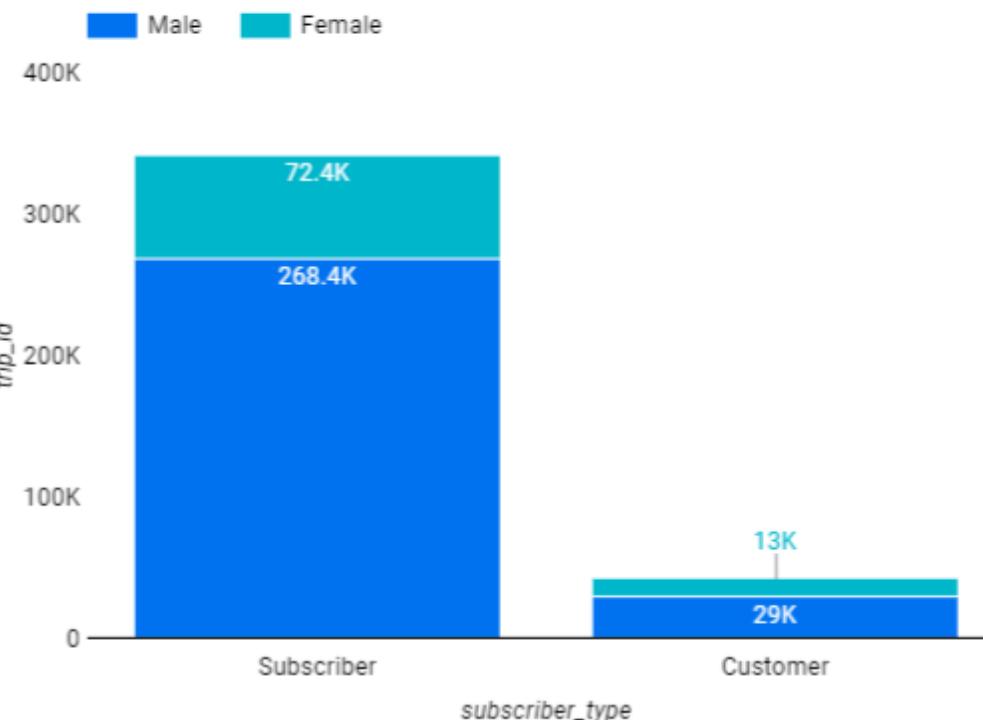


# #1 DEMOGRAPHY OF CYCLIST

Distribution Trip by Age Classification



Distribution Trip by Subscription Type and Gender



Our users are supremacy by prime working age (25-54 years old) by the subscriber type as a subscriber, and the gender is male. It contributed 268.4 of 340.8 (in thousands or 78.8%)

However, when viewed from the user's duration, the longer is dominated by female subscribers (subscriber type: customer) with an average time of 36.4 minutes with a median age of 38.

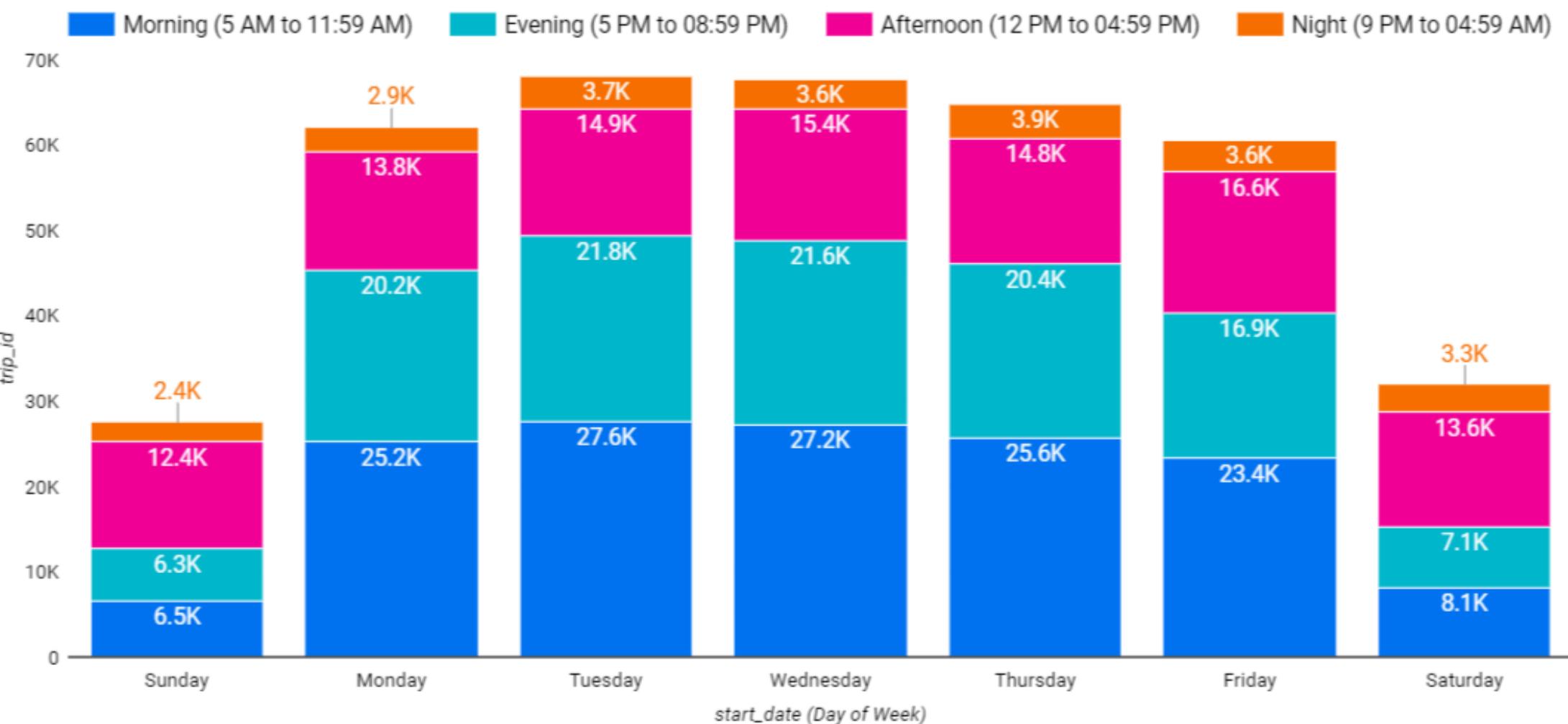
Distribution Trip by Subscription Type and Gender

	Subscriber Type	Gender	Trip_id ▾	Age	Average Duration in Minutes
1.	Subscriber	Male	268.4K	40	11.4
2.	Subscriber	Female	72.4K	38	13.6
3.	Customer	Male	29K	37	28.7
4.	Customer	Female	13K	35	36.4



## #2 DAILY USAGE IN 2017 BASED ON DAY AND TIME

Daily Usage in 2017 based on Day and Time



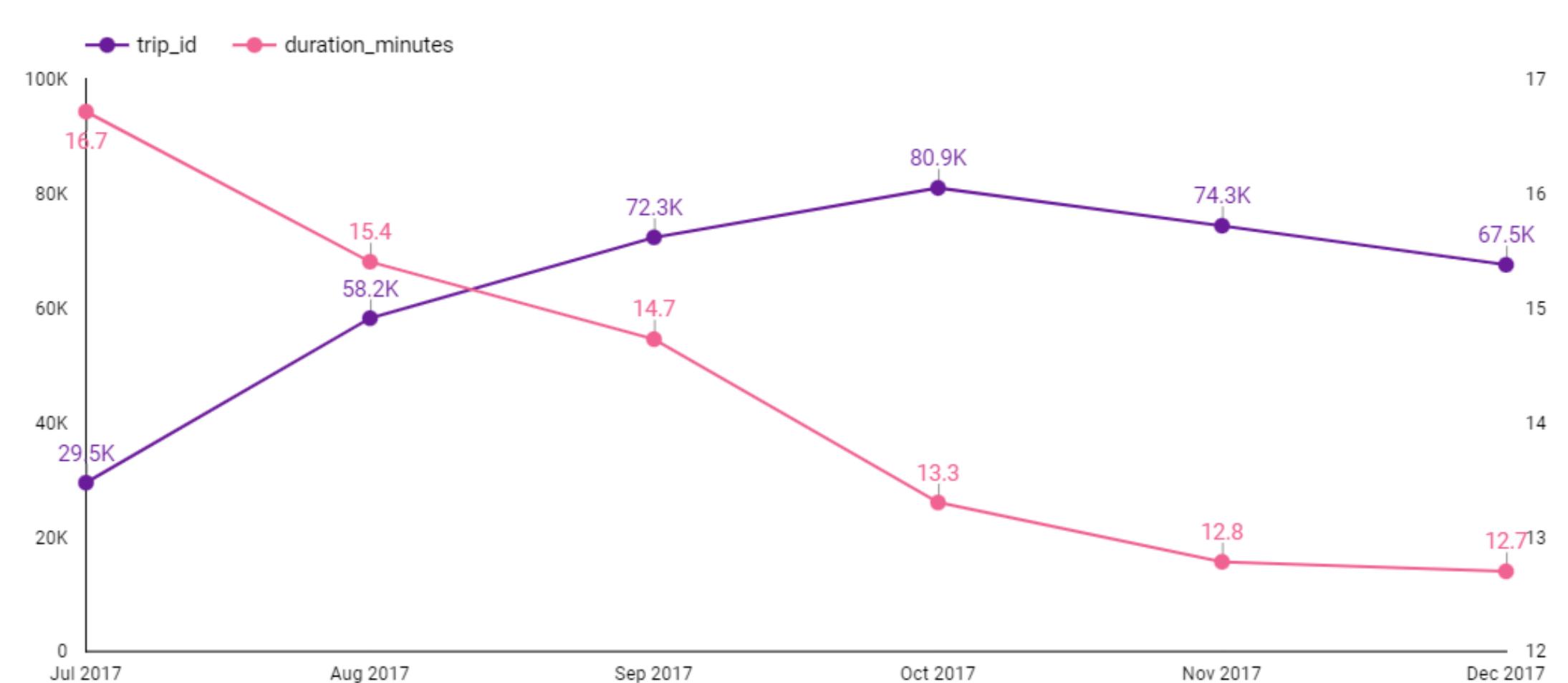
The graph indicates that bicycle users are generally more in the morning (5 AM to 11:59 AM) and evening (5 PM to 08:59 PM).

However, if we look in more detail per day, the highest number of users occurs on weekdays, especially Tuesdays (68 thousand trips or 17.7% of a week), and the lowest on weekends, especially on Sundays (27.6 thousand trips or 7.2% of a week).



## #3 MONTHLY TREND TRIP BY H2 IN 2017

Monthly Trend by Trip and Average Duration in Minutes



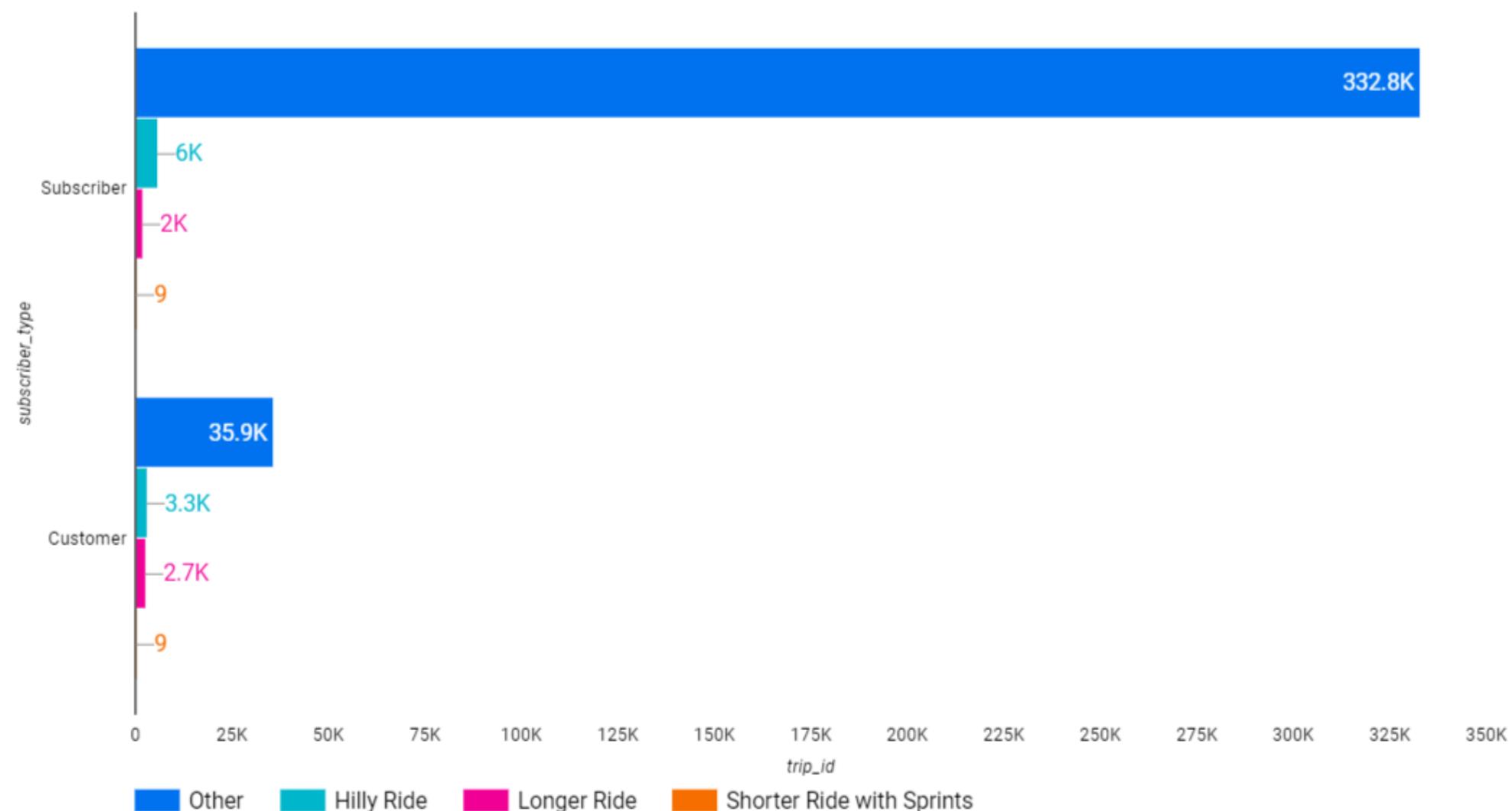
The number of trips from July to December 2017 experienced a sufficiently fluctuating change, where the highest peak of trips was in October (80.9 thousand trips).

However, compared to the average trip in units of time (minutes), it decreased every month. Where between the number of trips by the travel time can be said as an inverse comparison. Or modestly, the route is more of a short-distance trip than a long-distance one.



## #4 INTENSITY LEVEL OF HEALTH CYCLING IN 2017

Intensity Levels for Health Cycling

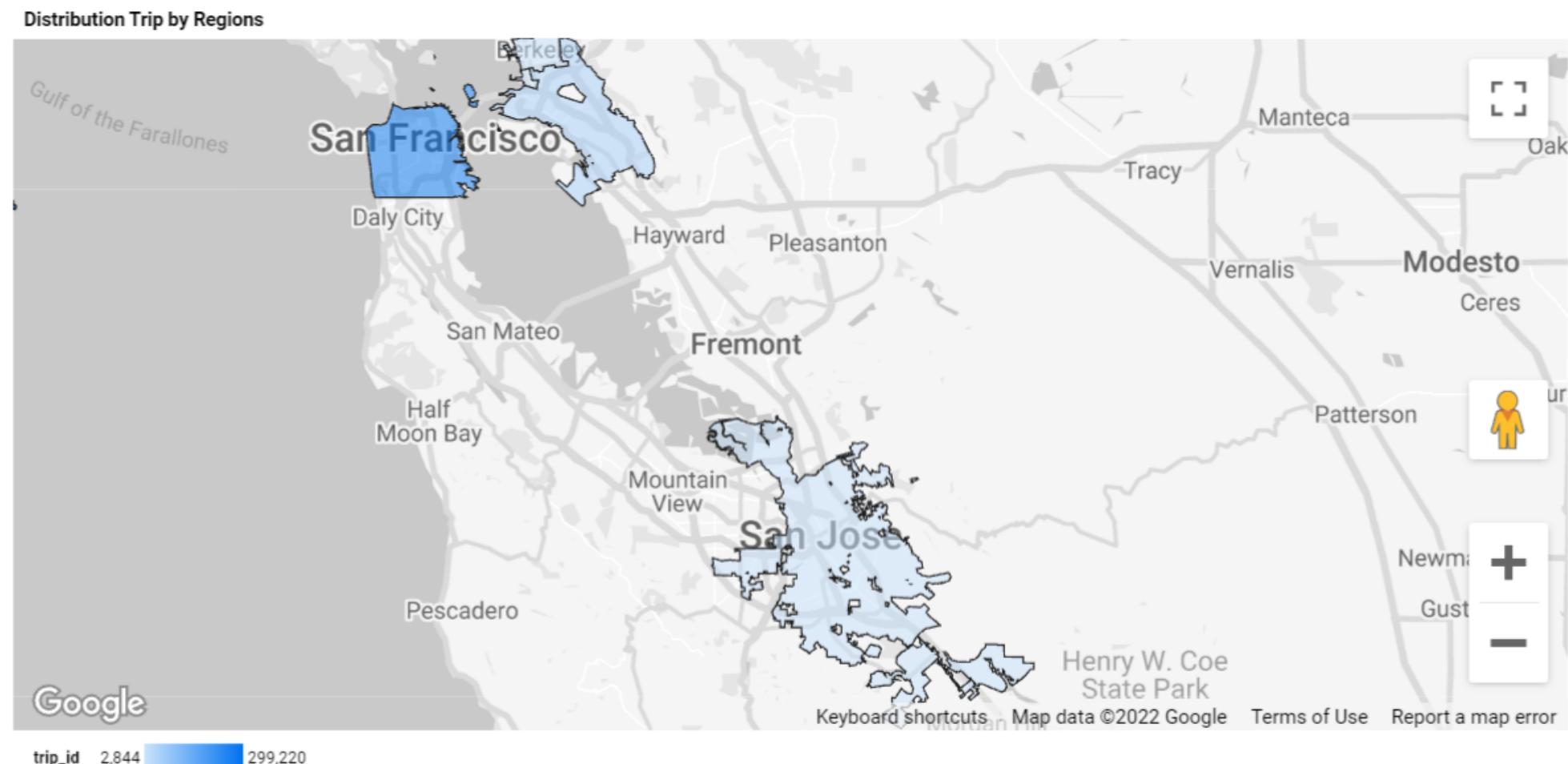


Found that the intensity of cycling that is good for health is divisible into three based on the duration of pedaling, namely Shorter Ride with Sprints (30 minutes), Hilly Ride (30-60 minutes), then Longer Ride (more than 60 minutes).

However, the results show that most of our users are outside this duration, both subscriber (332.8 thousand people) and customer (35.9 thousand people) types.

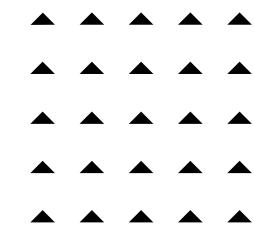


# #5 DISTRIBUTION BY REGION



region_name	trip_id	start_station_name	bike_number	Frequency Bike	Frequency Station	Bike Availability
1. San Francisco	299.2K	107	2.6K	115.8	2.8K	24.1
2. Oakland	55.6K	61	1.6K	33.9	911.3	26.9
3. San Jose	12.7K	34	541	23.5	374	15.9
4. Berkeley	12.4K	25	1.4K	9	496.3	54.9
5. Emeryville	2.8K	9	1K	2.8	316	113.2

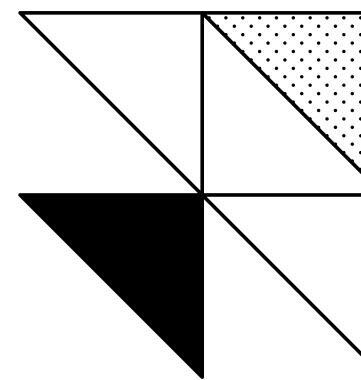
San Francisco is a region with the most number of trips (299.2 thousand), stations (107), and bikes (2.6 thousand). But bike availability in Emeryville is the highest among the others, which means the station in Emeryville has more bikes than any other region (113 or 48.7% of total bike availability).



# SUGGESTION



- The government can hold regular cycling events to increase interest in using bicycles. Even though the break in the latest article, 2021 will still be scheduled. Might after COVID-19 has stopped, this event can be rescheduled.
- The distribution of bicycle rental quotas is evenly distributed in all public areas in San Francisco so that people are not trapped in the minimum allocation wherever they want to rent.
- The government or the bicycle community can increase cycling campaigns related to health issues or energy savings to global warming.



# THANK YOU

**Open Discussion and Feedback,  
We can discuss further through:**



<https://www.linkedin.com/in/jihan-dewana-9b94b5185/>



jihan.dewana99@gmail.com