San Fransisco Bikeshare Study Case

With SQL

Filzahanti Nuha Ramadhani

San Fransisco Ford GoBike Share

- Data ini berasal dari Ford GoBike Share

(https://www.fordgobike.com/system-data)

- Data ini akan dianalisis menggunakan BigQuery

Persiapan Data

- Gabungkan tabel bikeshare_trips, bikeshare_station_info, dan bikeshire_regions.
- Buat tabel baru bernama age_group untuk mengelompokkan umur dari customer ke dalam kelompoknya masing-masing
- Buat tabel Bernama Route yang berisi nama stasiun mulai, ‘ – ‘, dan nama stasiun akhir
- Batasi data dari 2017-04-01 sampai 2018-04-30
- Tidak menggunakan data yang id stasiun mulai, nomor sepeda, tahun lahir member, nama region, dan gender member hilang.

```
SELECT
trip_id,
duration_sec,
start_date AS start_date_time,
DATE(start_date) AS start_date,
start_station_id,
start_station_name,
CONCAT(start_station_name, ' - ', end_station_name) AS route,
bike_number,
start_station_latitude,
start_station_longitude,
subscriber_type,
member_gender,
(2018-member_birth_year) as age,
CASE  WHEN (2018-member_birth_year) BETWEEN 0 AND 2 THEN '0-2'
     WHEN (2018-member_birth_year) BETWEEN 3 AND 16 THEN '3-16'
     WHEN (2018-member_birth_year) BETWEEN 17 AND 30 THEN '17-30'
     WHEN (2018-member_birth_year) BETWEEN 31 AND 45 THEN '31-45'
     ELSE '>45'
     END AS age_group,
c.name AS region_name
FROM bigquery-public-data.san_francisco_bikeshare.bikeshare_trips a
LEFT JOIN bigquery-public-data.san_francisco_bikeshare.bikeshare_station_info b
ON a.start_station_name=b.name
LEFT JOIN bigquery-public-data.san_francisco_bikeshare.bikeshare_regions c
ON b.region_id=c.region_id
WHERE start_date BETWEEN '2017-04-01' AND '2018-04-30'
AND start_station_id iS NOT NULL
AND bike_number IS NOT NULL
AND member_birth_year IS NOT NULL
AND b.name IS NOT NULL
AND member_gender IS NOT NULL
```

Preview Data

Row	trip_id	duration_sec	start_date_time	start_date	start_statio...	start_station_name
1	20171009132723.8390001479	2164	2017-10-09 13:27:23.839000 U...	2017-10-09	198	Snow Park
2	20180225141829.4560002073	935	2018-02-25 14:18:29.456000 U...	2018-02-25	198	Snow Park
3	20171017120627.436000130	1957	2017-10-17 12:06:27.436000 U...	2017-10-17	198	Snow Park
4	20171227114932.69100021	2005	2017-12-27 11:49:32.691000 U...	2017-12-27	198	Snow Park
5	20180117113729.0850003485	1950	2018-01-17 11:37:29.085000 U...	2018-01-17	198	Snow Park
6	20180215143839.8880003701	1988	2018-02-15 14:38:39.888000 U...	2018-02-15	198	Snow Park
7	20180426113259.511000843	2036	2018-04-26 11:32:59.511000 U...	2018-04-26	198	Snow Park
8	20180116113309.2780003637	1991	2018-01-16 11:33:09.278000 U...	2018-01-16	198	Snow Park
9	20180213113517.2930003701	2085	2018-02-13 11:35:17.293000 U...	2018-02-13	198	Snow Park

route	bike_number	start_statio...	start_statio...	subscriber_type	member_gender
Snow Park - Snow Park	1479	37.8078131...	-122.26449...	Subscriber	Male
Snow Park - Snow Park	2073	37.8078131...	-122.26449...	Subscriber	Male
Snow Park - Snow Park	130	37.8078131...	-122.26449...	Subscriber	Male
Snow Park - Snow Park	21	37.8078131...	-122.26449...	Subscriber	Male
Snow Park - Snow Park	3485	37.8078131...	-122.26449...	Subscriber	Male
Snow Park - Snow Park	3701	37.8078131...	-122.26449...	Subscriber	Male
Snow Park - Snow Park	843	37.8078131...	-122.26449...	Subscriber	Male
Snow Park - Snow Park	3637	37.8078131...	-122.26449...	Subscriber	Male
Snow Park - Snow Park	3701	37.8078131...	-122.26449...	Subscriber	Male

age	age_group	region_name
56	>45	Oakland
28	17-30	Oakland
56	>45	Oakland
56	>45	Oakland
56	>45	Oakland
54	>45	Oakland
56	>45	Oakland
56	>45	Oakland
56	>45	Oakland

Eksplorasi Data

Total Baris

```
SELECT COUNT(*) AS total_baris
FROM filzahantisql.bikeshare.trips
```

Row	total_baris
1	749772

Terdapat 749772 baris

Nilai maksimum dan rata-rata

pada kolom age dan duration_sec

```
SELECT
MAX(age) AS max_age, AVG(age) AS avg_age,
MAX(duration_sec) AS max_duration,
AVG(duration_sec) AS avg_duration
FROM filzahantisql.bikeshare.trips
```

Row	max_age	avg_age	max_duration	avg_duration
1	132	36.9638956...	86252	795.893152...

Distribusi Tipe Langganan dan Kelompok Umur

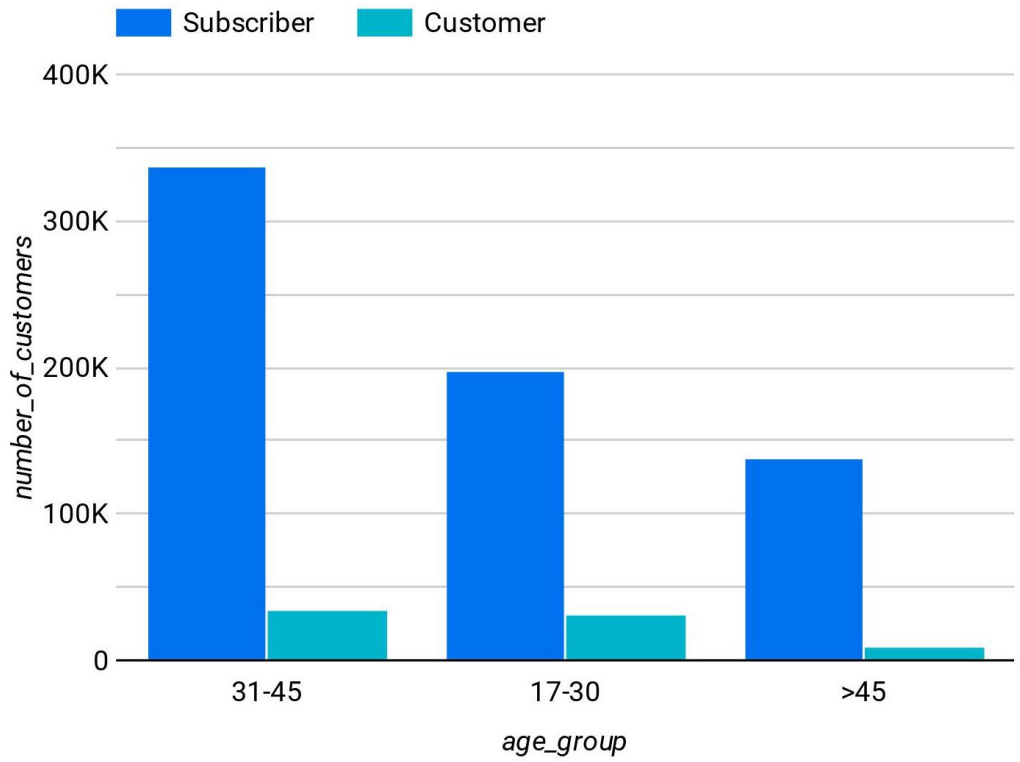
```
SELECT age_group, subscriber_type, COUNT(subscriber_type) AS number_of_customers
FROM filzahantisql.bikeshare.trips
GROUP BY subscriber_type, age_group
ORDER BY number_of_customers
```

Row	age_group	subscriber_type	number_of_...
1	>45	Customer	9831
2	17-30	Customer	32063
3	31-45	Customer	35043
4	>45	Subscriber	137755
5	17-30	Subscriber	197938
6	31-45	Subscriber	337142

Distribusi Tipe Langganan dan Kelompok Umur

Subscriber Type Description:
- Subscriber = annual or 30-day member
- Customer = 24-hour or 3-day member

age_group	subscriber_type	number_of_customers
31-45	Subscriber	337,142
	Customer	35,043
17-30	Subscriber	197,938
	Customer	32,063
>45	Subscriber	137,755
	Customer	9,831

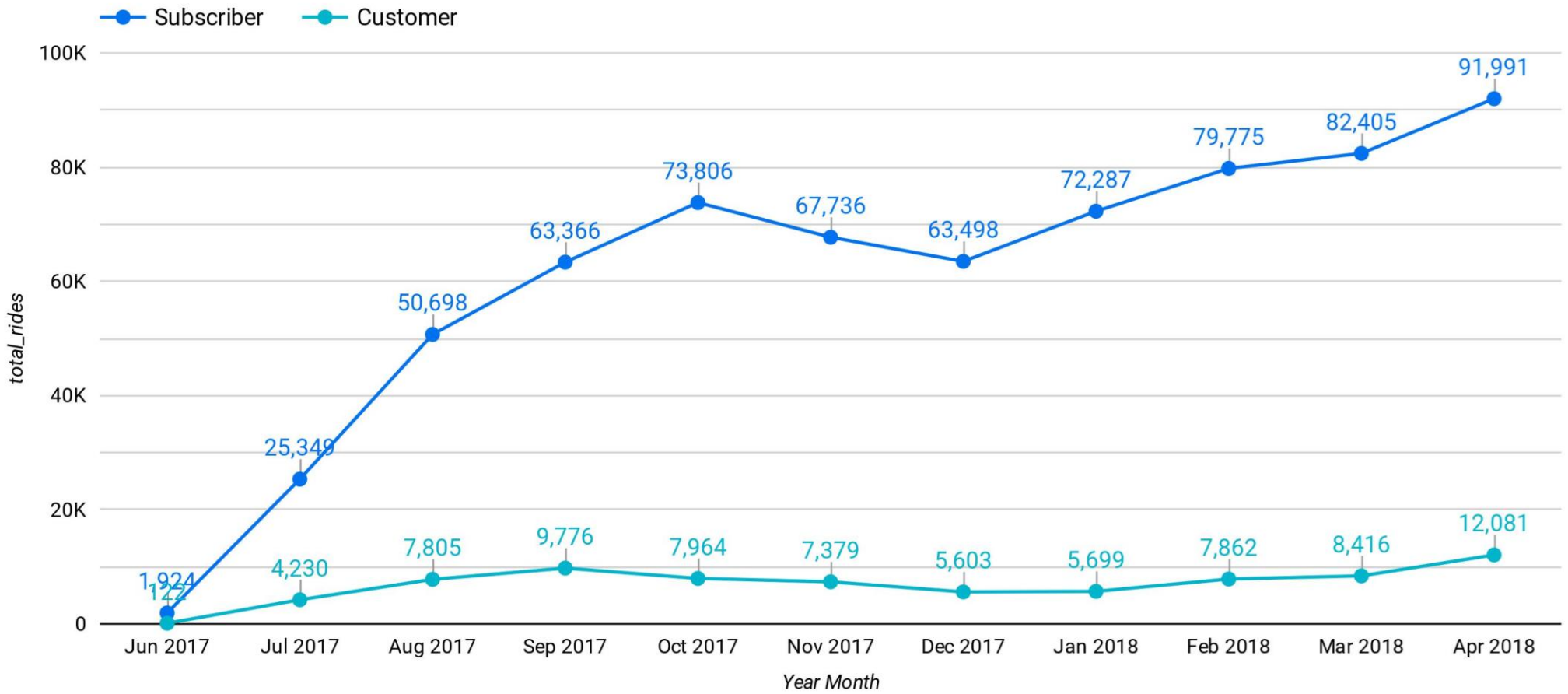


Pengguna bikeshare kebanyakan berusia 31-45 tahun dengan tipe langganan subscriber yaitu tipe yang memiliki member untuk 30 hari atau tahunan.

Jumlah Riders per Bulan

```
SELECT subscriber_type,
DATE(DATE_TRUNC(start_date, MONTH)) AS year_month,
COUNT(DATE(DATE_TRUNC(start_date, MONTH))) AS total_rides
FROM filzahantisql.bikeshare.trips
GROUP BY year_month, subscriber_type
ORDER BY year_month ASC
```

Row	subscriber_type	year_month	total_rides
2	Customer	2017-06-01	122
3	Subscriber	2017-07-01	25349
4	Customer	2017-07-01	4230
5	Subscriber	2017-08-01	50698
6	Customer	2017-08-01	7805
7	Subscriber	2017-09-01	63366



Pengguna bikeshare dengan tipe langganan subscriber mengalami kenaikan yang signifikan pada Juni 2017 sampai Oktober 2017 dan mengalami penurunan pada Oktober 2017 sampai Desember 2017, kemudian mengalami kenaikan sampai April 2018. Sedangkan, pengguna dengan tipe langganan customer memiliki tren yang stabil.

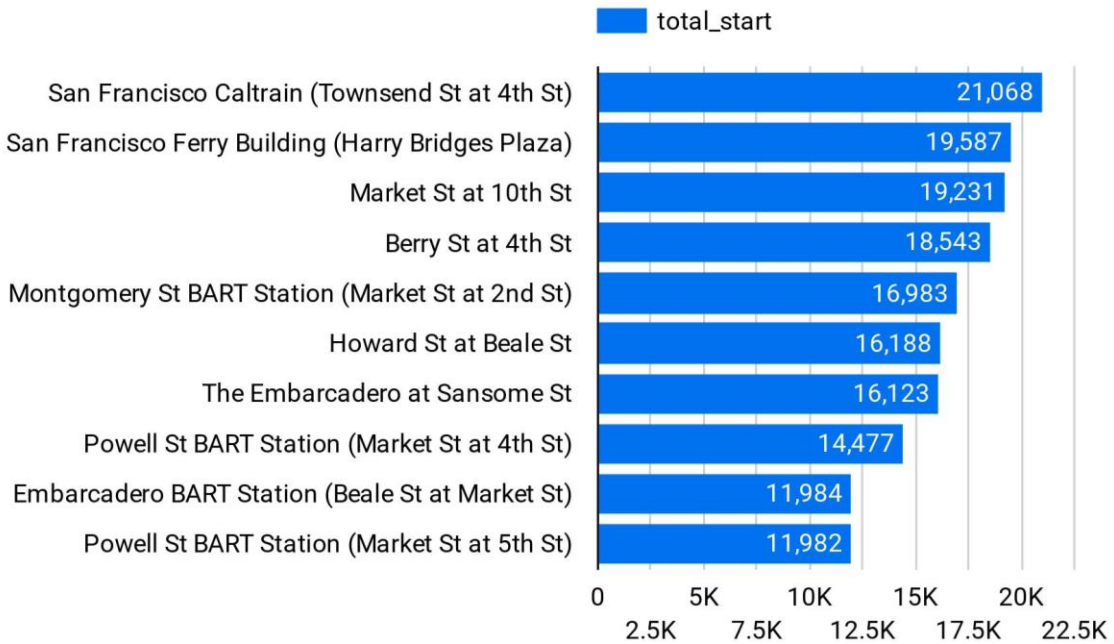
Total Perjalanan dari Stasiun Awal dan Region

```
SELECT start_station_name, region_name, CONCAT(start_station_latitude,',', start_station_longitude)
AS latitude_longitude, COUNT(start_station_name) AS total_start
FROM filzahantisql.bikeshare.trips
GROUP BY 1, 2, 3
```

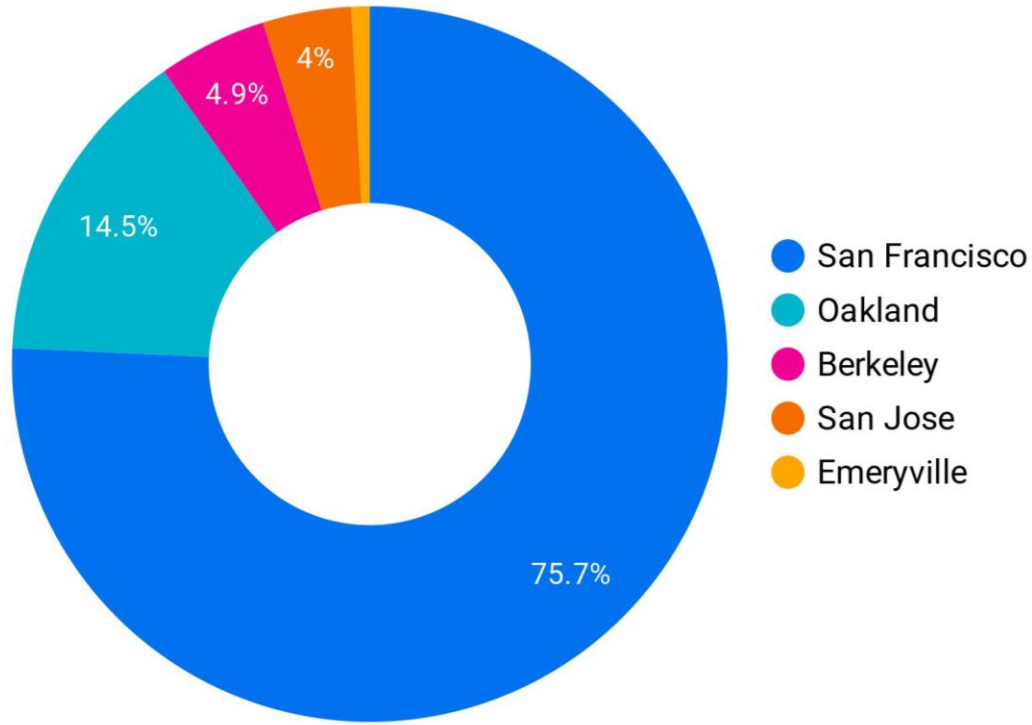
Preview

Row	start_station_name	region_name	latitude_longitude	total_start
1	Snow Park	Oakland	37.807813182179032,-122.264...	1663
2	Mosswood Park	Oakland	37.8249311,-122.2604787	1966
3	San Antonio Park	Oakland	37.790139851853638,-122.242...	455
4	Broadway at 30th St	Oakland	37.8193814,-122.2619284	1994
5	Broadway at 40th St	Oakland	37.8277573,-122.2567156	2713
6	24th St at Market St	Oakland	37.8160598,-122.2782444	941

Top 10 Total Rides by Start Stations



Total Rides Distribution by Start Stations



Perjalanan dimulai paling banyak di San Fransisco. San Fransisco Caltrain merupakan station dengan jumlah perjalanan awal paling banyak.

Kesimpulan

- Pengguna bikeshare kebanyakan berusia 31-45 tahun dan memiliki tipe langganan subscriber (member 30 hari atau tahunan)
- Member dengan tipe langganan subscriber mengalami kenaikan yang signifikan dalam penggunaan bikeshare, sementara member dengan tipe langganan customer memiliki tren yang stabil.
- Perjalanan dimulai kebanyakan dari region San Fransisco, dan stasiun dengan perjalanan awal terbanyak adalah San Fransisco Caltrain.