



Visualisasi Data pada Dataset City Bike Trips

Takdir Zulhaq Dessiaming

Case Study Instructions

Table of interest :

``bigquery-public-data.new_york_citibike.citibike_stations``
``bigquery-public-data.new_york_citibike.citibike_trips``

Please answer the questions below.

1. Look at this data and start thinking. List down 3 trends/points that you think need to be shown.
2. From here, try to explore the data and make changes, filter, and prepare the data that you need.
3. Create some visualization 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 findings to the stakeholders.

Menentukan Pertanyaan

Menentukan Pertanyaan untuk dijawab

Pertama-tama, berdasarkan dari dataset, terlihat bahwa ada tabel usertype, birth_year, gender data, yang menunjukkan yang mana tipe pelanggan yang umum menjadi target. jika kita dapat menjawab pertanyaan ini, kita dapat membantu tim marketing untuk menjalankan kampanye menuju puncak acara. Kita bisa menuju ke tahun 2018 saja.

Siapa grup yang terbesar dalam pelanggan di tahun 2018?

Kedua, terdapat relasi antar waktu dalam tabel, yang berarti kita dapat melihat tren perjalanan dari waktu ke waktu yang dapat divisualisasikan.

Apa tren jumlah perjalanan dan jumlah pelanggan bulan ke bulan di 2018?

Menentukan Pertanyaan untuk dijawab

Ketiga, terdapat tabel starttime, dan birth_year, dan gender, dimana kita dapat melihat kelahiran mana saja yang banyak melakukan trip, dan di tahun berapa trip terbanyak terjadi dalam rentang waktu 2013-2018, dan kelahiran anak muda dan jenis kelaminnya (dari tahun 1990-2002).


Tahun berapa yang menjadi puncak trip terbanyak dan jenis kelamin apa saja yang sering melakukan trip di rentang tahun 2013-2018?

Eksplorasi Data

Eksplorasi Data

Terdapat 2 pendekatan yang dapat dilakukan untuk memecahkan masalah. Dalam kasus ini, kita mencari masalah ayng dapat kita jawab dengan dataset2 berikut. Pastikan bahwa pertama-tama melihat pada tabel skema dan lihat apa info yang didapatkan. kemudian, tentukan mana yang menjadi koneksi antara kedua tabel dan mencari Primary Key.


Tabel Citibikes_Stations

Field name	Type	Mode	Policy Tags 	Description
station_id	INTEGER	REQUIRED		Unique identifier of a station.
name	STRING	NULLABLE		Public name of the station.
short_name	STRING	NULLABLE		Short name or other type of identifier, as used by the data publisher.
latitude	FLOAT	NULLABLE		The lattitude of station. The field value must be a valid WGS 84 latitude in decimal degrees format.
longitude	FLOAT	NULLABLE		The longitude of station. The field value must be a valid WGS 84 longitude in decimal degrees format.
region_id	INTEGER	NULLABLE		ID of the region where station is located.
rental_methods	STRING	NULLABLE		Array of enumerables containing the payment methods accepted at this station.
capacity	INTEGER	NULLABLE		ANumber of total docking points installed at this station, both available and unavailable.
eightd_has_key_dispenser	BOOLEAN	NULLABLE		
num_bikes_available	INTEGER	NULLABLE		Number of bikes available for rental.
num_bikes_disabled	INTEGER	NULLABLE		Number of disabled bikes at the station.
num_docks_available	INTEGER	NULLABLE		Number of docks accepting bike returns.
num_docks_disabled	INTEGER	NULLABLE		Number of empty but disabled dock points at the station.
is_installed	BOOLEAN	NULLABLE		Is the station currently on the street?
is_renting	BOOLEAN	NULLABLE		Is the station currently renting bikes?
is_returning	BOOLEAN	NULLABLE		Is the station accepting bike returns?
eightd_has_available_keys	BOOLEAN	NULLABLE		
last_reported	DATETIME	NULLABLE		Timestamp indicating the last time this station reported its status to the backend, in NYC local time.

EDIT SCHEMA

VIEW ROW ACCESS POLICIES

Tabel Citibikes_Trips

Field name	Type	Mode	Policy Tags 	Description
tripduration	INTEGER	NULLABLE		Trip Duration (in seconds)
starttime	DATETIME	NULLABLE		Start Time, in NYC local time.
stoptime	DATETIME	NULLABLE		Stop Time, in NYC local time.
start_station_id	INTEGER	NULLABLE		Start Station ID
start_station_name	STRING	NULLABLE		Start Station Name
start_station_latitude	FLOAT	NULLABLE		Start Station Latitude
start_station_longitude	FLOAT	NULLABLE		Start Station Longitude
end_station_id	INTEGER	NULLABLE		End Station ID
end_station_name	STRING	NULLABLE		End Station Name
end_station_latitude	FLOAT	NULLABLE		End Station Latitude
end_station_longitude	FLOAT	NULLABLE		End Station Longitude
bikeid	INTEGER	NULLABLE		Bike ID
usertype	STRING	NULLABLE		User Type (Customer = 24-hour pass or 7-day pass user, Subscriber = Annual Member)
birth_year	INTEGER	NULLABLE		Year of Birth
gender	STRING	NULLABLE		Gender (unknown, male, female)
customer_plan	STRING	NULLABLE		The name of the plan that determines the rate charged for the trip

EDIT SCHEMA

VIEW ROW ACCESS POLICIES

Siapa grup yang terbesar dalam pelanggan di tahun 2018?

Untuk menjawab ini, kita dapat menyiapkan query sederhana. Dapat dilihat bahwa 3 penghubung untuk User Type : usertype, birth_year , gender data dan kita hanya ingin grup pelanggan di 2018. Untuk mengetahui grup terbesar, kita butuh jumlah dari total customer dan total dari trips untuk setiap unique group of customer.

```
1  SELECT
2  distinct (usertype),
3      birth_year,
4      gender,
5  count (distinct(bikeid)) as unique_users,
6  count (bikeid) as trips
7  FROM `bigquery-public-data.new_york_citibike.citibike_trips`
8  where
9      Date(starttime) > '2018-01-01'
10     and date(stoptime) < '2018-12-31'
11  group by
12      1,
13      2,
14      3
```

Query complete (4.3 sec elapsed, 2.5 GB processed)

Job information [Results](#) JSON Execution details

Row	usertype	birth_year	gender	unique_users	trips
1	Subscriber	1993	female	11673	42550
2	Subscriber	1993	male	12576	111160
3	Customer	1993	male	3260	4358
4	Customer	1993	female	2560	3324
5	Subscriber	1993	unknown	379	402
6	Subscriber	1994	male	12363	76583
7	Customer	1994	female	2423	3059
8	Subscriber	1994	female	10765	29692
9	Customer	1994	male	3084	3934
10	Subscriber	1995	male	11846	51480
11	Customer	1995	male	2813	3638
12	Subscriber	1995	unknown	455	489

Siapa grup yang terbesar dalam pelanggan di tahun 2018?

Untuk menjawab pertanyaan pertama dan kedua, kita dapat menyiapkan query sederhana. Dapat dilihat bahwa 3 penghubung untuk User Type : usertype, birth_year , gender data dan kita hanya ingin grup pelanggan di 2018. Untuk mengetahui grup terbesar, kita butuh jumlah dari total customer dan total dari trips untuk setiap unique group c

```
1 SELECT
2   distinct (format_datetime("%Y%m",
3     starttime)) as month_year,
4   count (starttime) as count_orders,
5   count (distinct(bikeid)) as count_customers,
6   avg(tripduration) as avg_duration
7 FROM `bigquery-public-data.new_york_citibike.citibike_trips`
8 where
9   starttime IS NOT NULL and starttime > '2017-01-01'
10 group by
11   month_year
12 order by
13   month_year
```

Query complete (5.7 sec elapsed, 1.2 GB processed)

Job information Results JSON Execution details

Row	month_year	count_orders	count_customers	avg_duration
1	201704	1315404	9962	1085.8632237700358
2	201705	1523268	10060	983.2848441639983
3	201706	1731594	9968	1072.8265482555403
4	201707	1735599	10423	1060.1433597276746
5	201708	1816498	10813	1156.7878247044557
6	201709	1878098	11615	997.0888920599434
7	201710	1897592	11528	1067.8202068727157
8	201711	1330649	12776	892.7143356362212
9	201712	889967	11119	764.997161692512
10	201801	718994	10449	826.810654887246
11	201802	843114	10534	877.3977908088353
12	201803	976672	10935	881.9937205121054
13	201804	1307543	11913	970.9073223595703
14	201805	1824710	12335	1121.914173759115

Tahun berapa yang menjadi puncak trip terbanyak dan jenis kelamin apa saja yang sering melakukan trip di rentang tahun 2013-2018?

Untuk menjawab ini, pertanyaan ketiga, kita dapat menyiapkan query sederhana. Dapat dilihat bahwa 3 penghubung untuk Birth Year : birth_year , gender, starttime , dan kita ingin melihat tahun berapa yang dilakukan banyak trip oleh anak muda.

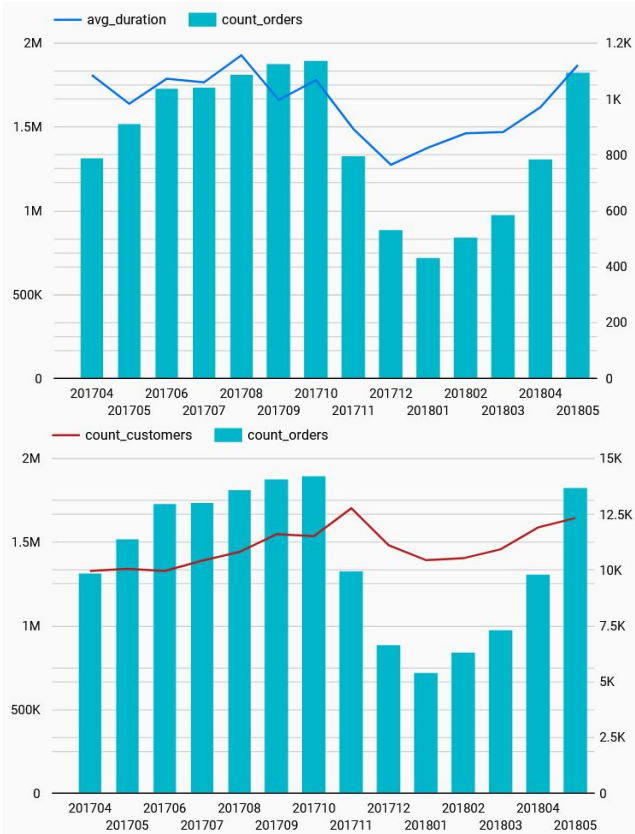
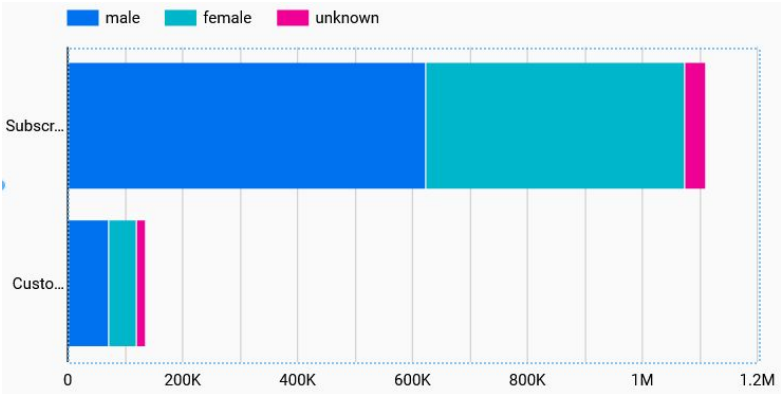
```
SELECT
birth_year, gender,
start_station_name,
starttime, tripduration
FROM `bigquery-public-data.new_york_citibike.citibike_trips`
where tripduration > 3600
```

Row	birth_year	gender	start_station_name	starttime	tripduration
1	1952	female	W 106 St & Central Park West	2018-03-11T11:44:06.044000	4042
2	1999	male	Pioneer St & Richards St	2017-09-23T20:08:18	9847
3	1951	male	W 22 St & 8 Ave	2018-04-24T13:00:51.623000	3851
4	1998	unknown	Throop Ave & Myrtle Ave	2017-08-17T10:31:46	167228
5	1946	male	E 51 St & 1 Ave	2016-05-29T10:26:51	6255
6	1998	unknown	Tompkins Ave & Hopkins St	2017-06-16T14:29:25	10794
7	1998	male	Humboldt St & Varet St	2017-10-06T16:59:50	5301
8	1997	female	South End Ave & Liberty St	2017-07-28T11:00:59	5444
9	1949	female	Clark St & Henry St	2013-10-12T09:20:39	12640

Visualisasi Data

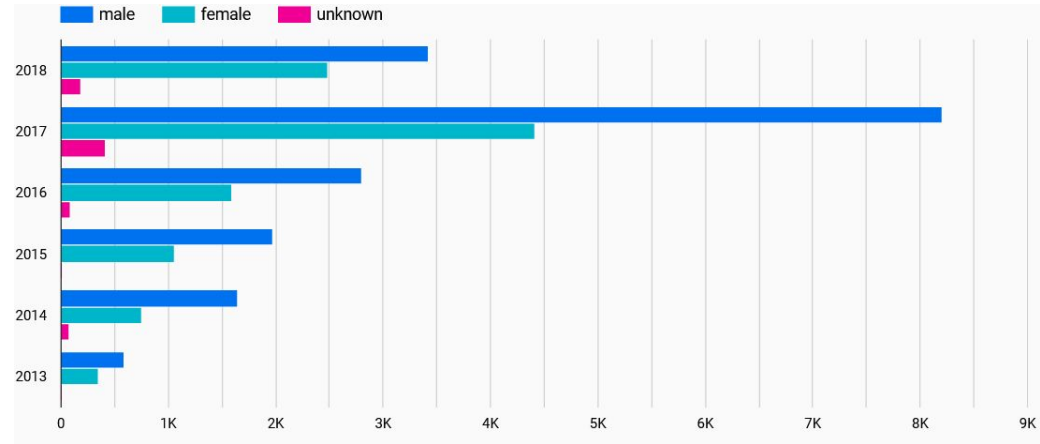
Dapat dilakukan dengan visualisasi apapun, tapi untuk kasus ini, idealnya menggunakan tabel with bar, bar chart dan line chart, hingga insightnya akan diletakkan bersamaan dibagian akhir (menjawab pertanyaan sebelumnya). Ini menggunakan Google Data Studio

	usertype	gender	unique_users ▾	trips
1	Subscriber	male	<div></div>	<div></div>
2	Subscriber	female	<div></div>	<div></div>
3	Customer	male	<div></div>	<div></div>
4	Customer	female	<div></div>	<div></div>
5	Subscriber	unknown	<div></div>	<div></div>
6	Customer	unknown	<div></div>	<div></div>



Dapat dilakukan dengan visualisasi apapun, tapi untuk kasus ini, idealnya menggunakan tabel with heatmap, bar chart dan line chart, hingga insightnya akan diletakkan bersamaan dibagian akhir (menjawab pertanyaan sebelumnya). Ini menggunakan Google Data Studio

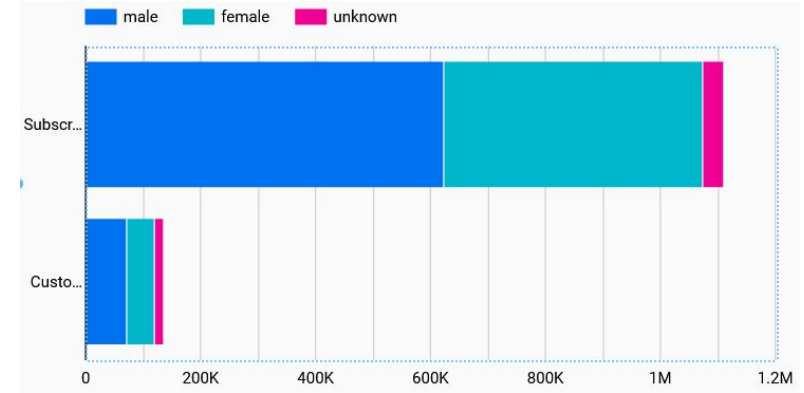
starttime (Year) ▾		Record Count
1.	2018	6,088
2.	2017	13,022
3.	2016	4,474
4.	2015	3,037
5.	2014	2,466
6.	2013	931



Penjelasan dan Insight

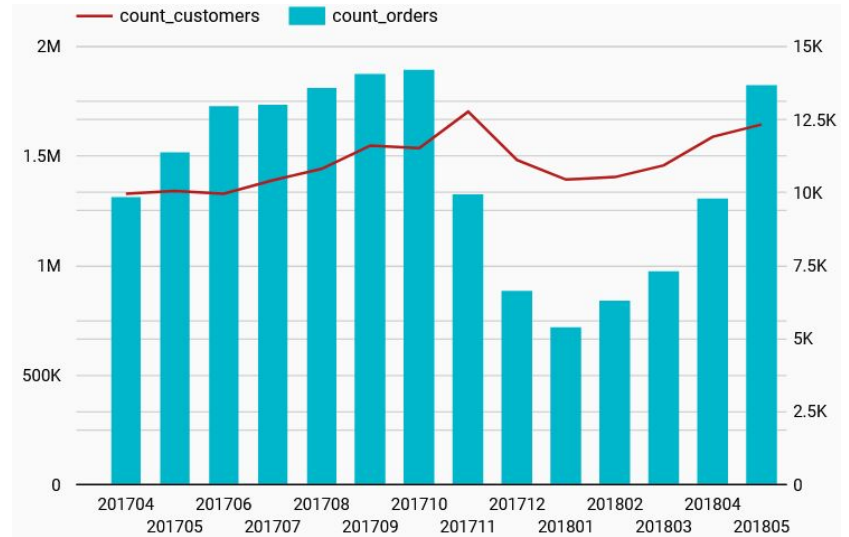
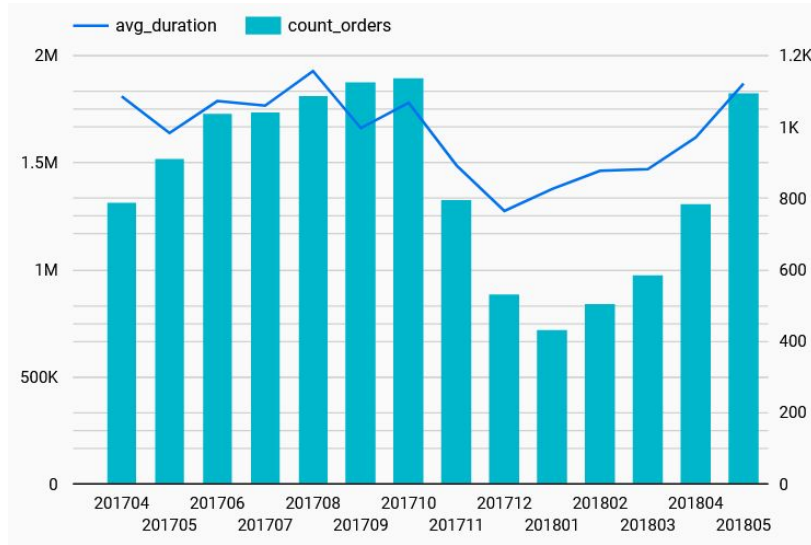
Tipe pengguna tertinggi menurut data pada tahun 2018 adalah pelanggan **PRIA**, yang mewakili 68% dari semua perjalanan yang dilakukan

	usertype	gender	unique_users ▾	trips
1	Subscriber	male	<div></div>	<div></div>
2	Subscriber	female	<div></div>	<div></div>
3	Customer	male	<div></div>	<div></div>
4	Customer	female	<div></div>	<div></div>
5	Subscriber	unknown	<div></div>	<div></div>
6	Customer	unknown	<div></div>	<div></div>



- Tertinggi kedua yaitu pelanggan **PEREMPUAN** (21%), dan pelanggan (gender tidak diketahui)
- Oleh karena itu, tim pemasaran harus terus fokus untuk menargetkan pelanggan pria, sambil memikirkan cara untuk lebih meningkatkan pelanggan wanita dengan strategi baru.

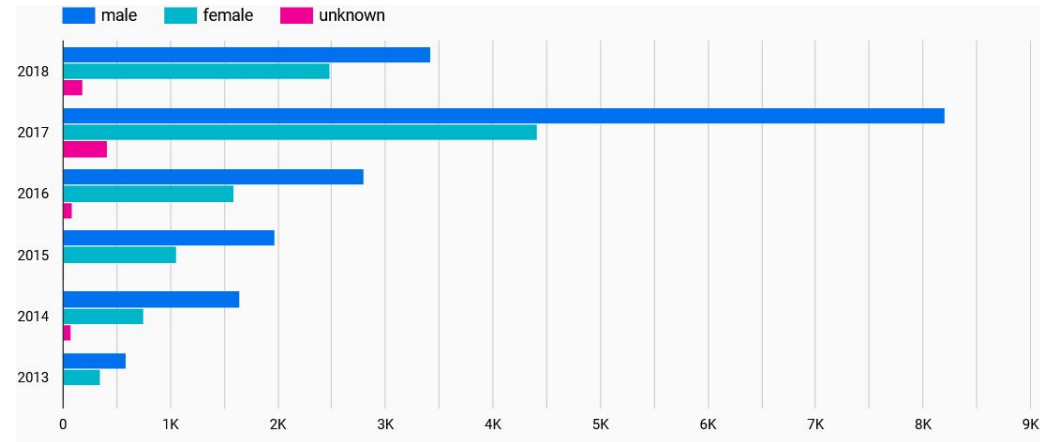
Bisnis telah menunjukkan pertumbuhan inkremental pada tahun 2017, tetapi menurun pada kuartal 4 - 2017, hingga kuartal 1 - 2018



- Terjadi penurunan rata-rata durasi per trip yang juga berkorelasi dengan rendahnya jumlah pesanan pada 17 Desember hingga 18 Maret, namun jumlahnya meningkat kembali ke sebelumnya.
- Ada sedikit penurunan jumlah pelanggan unik di Kuartal 1 - 2018, tetapi penjualan tampaknya banyak menurun. Dengan demikian, ada faktor eksternal yang perlu dianalisis untuk mendiagnosis penurunan lebih lanjut

Tahun 2017 menjadi puncak orang-orang melakukan trip. Jumlah pria dan wanita bertambah secara konsisten, namun mulai turun di tahun 2018

	starttime (Year) ▾	Record Count
1.	2018	6,088
2.	2017	13,022
3.	2016	4,474
4.	2015	3,037
5.	2014	2,466
6.	2013	931



- Terlihat bahwa dari tahun 2013 orang melakukan trip sampai 2018, orang-orang secara konsisten bertambah hingga naik secara signifikan dan menjadi puncak banyaknya trip dilakukan di tahun 2017 sebanyak **13022 trip (pria 8201 orang, wanita 4410 orang, dan unknown 411 orang)**. Namun **terjadi penurunan 50%** lebih trip yang dilakukan orang-orang di tahun 2018.
- Terlihat juga bahwa **Pria** lebih sering melakukan trip dibandingkan Wanita, dan Unknown dari tahun 2013-2018.
- Jenis kelamin bersamaan meningkat dengan total jumlah orang yang melakukan trip selama 2013 - 2017, dan bersamaan turun di tahun 2018.



Thank You!