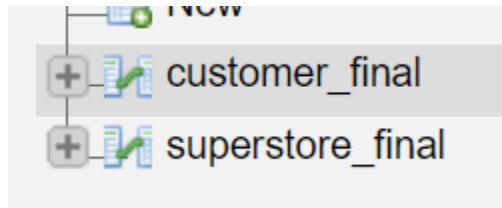


Pada Dataset kali saya menggunakan *dataset* superstore yang saya dapatkan dari Kaggle. Terdapat dua table yang terdiri dari customer_final dan superstore_final yang masing terdiri dari 199 rows.



<code>SELECT Count(*) FROM `customer_final`</code>	Count(*)
	199
<code>SELECT count(*) FROM `superstore_final`</code>	count(*)
	199


A. **customer_final** memiliki 4 kolom yang dapat dilihat pada gambar dibawah ini.

	#	Name	Type
<input type="checkbox"/>	1	Customer ID	varchar(8)
<input type="checkbox"/>	2	Customer Name	varchar(18)
<input type="checkbox"/>	3	Segment	varchar(11)
<input type="checkbox"/>	4	City	varchar(15)
<input type="checkbox"/>	5	State	varchar(14)

B. **superstore_final** memiliki 18 kolom yang dapat dilihat pada gambar dibawah ini.

1	Row ID	int(3)	10	Product ID	varchar(15)
2	Order ID	varchar(14)	11	Category	varchar(15)
3	Order Date	varchar(10)	12	Sub_Category	varchar(11)
4	Ship Date	varchar(10)	13	Product Name	varchar(109)
5	Ship Mode	varchar(14)	14	Sales	decimal(7,4)
6	Customer ID	varchar(8)	15	Quantity	int(2)
7	Country_Region	varchar(13)	16	Discount	decimal(3,2)
8	Postal Code	int(5)	17	Profit	decimal(9,4)
9	Region	varchar(7)	18	Amount	decimal(9,6)

1. customer_final

Customer ID	Customer Name	Segment	City	State
CG-12520	Claire Gute	Consumer	Henderson	Kentucky
CG-12520	Claire Gute	Consumer	Henderson	Kentucky
DV-13045	Darrin Van Huff	Corporate	Los Angeles	California
SO-20335	Sean O'Donnell	Consumer	Fort Lauderdale	Florida
SO-20335	Sean O'Donnell	Consumer	Fort Lauderdale	Florida
BH-11710	Brosina Hoffman	Consumer	Los Angeles	California
BH-11710	Brosina Hoffman	Consumer	Los Angeles	California
BH-11710	Brosina Hoffman	Consumer	Los Angeles	California
BH-11710	Brosina Hoffman	Consumer	Los Angeles	California
BH-11710	Brosina Hoffman	Consumer	Los Angeles	California
BH-11710	Brosina Hoffman	Consumer	Los Angeles	California
BH-11710	Brosina Hoffman	Consumer	Los Angeles	California
 Console	Andrew Allen	Consumer	Concord	North Carolina

2. superstore_final

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Country_Region	Postal Code	Region	Product ID	Category	Sub_Category	Product Name	Sales	Quantity	Discount	Profit	Amount
1	CA-2019-152156	43688	11/11/2019	Second Class	CG-12520	United States	42420	South	FUR-BO-10001798	Furniture	Bookcases	Bush Somerset Collection Bookcase	251.9600	2	0.00	41.9136	0.000000
2	CA-2019-152156	43688	11/11/2019	Second Class	CG-12520	United States	42420	South	FUR-CH-10000454	Furniture	Chairs	Hon Deluxe Fabric Upholstered Stacking Chairs, Rou...	731.9400	3	0.00	219.5820	0.000000
3	CA-2019-138888	43805	6/16/2019	Second Class	DV-13045	United States	90036	West	OFF-LA-10000240	Office Supplies	Labels	Self-Adhesive Address Labels for Typewriters by Un...	14.8200	2	0.00	6.8714	0.000000
4	US-2018-108996	43414	10/18/2018	Standard Class	SO-20335	United States	33311	South	FUR-TA-10000577	Furniture	Tables	Bretford CR4500 Series Slim Rectangular Table	957.5775	5	0.45	-383.0310	430.909875
5	US-2018-108996	43414	10/18/2018	Standard Class	SO-20335	United States	33311	South	OFF-ST-10000760	Office Supplies	Storage	Eldon Fold 'N Roll Cart System	22.3680	2	0.20	2.5164	4.473600
6	CA-2017-115812	42984	6/14/2017	Standard Class	BH-11710	United States	90032	West	FUR-FU-10001487	Furniture	Furnishings	Eldon Expressions Wood and Plastic Desk Accessorie...	48.8600	7	0.00	14.1694	0.000000
7	CA-2017-115812	42984	6/14/2017	Standard Class	BH-11710	United States	90032	West	OFF-AR-10002833	Office Supplies	Art	Newell 322	7.2800	4	0.00	1.9656	0.000000
8	CA-2017-115812	42984	6/14/2017	Standard Class	BH-11710	United States	90032	West	TEC-PH-10002275	Technology	Phones	Mitel 5320 IP Phone VoIP phone	907.1520	6	0.20	90.7152	181.430400
9	CA-2017-115812	42984	6/14/2017	Standard Class	BH-11710	United States	90032	West	OFF-BI-10003910	Office Supplies	Binders	DXL Angle-View Binders with Locking Rings by Sams...	18.5040	3	0.20	5.7825	3.700800
10	CA-2017-115812	42984	6/14/2017	Standard Class	BH-11710	United States	90032	West	OFF-AP-10002892	Office Supplies	Appliances	Belkin F5C208VTEL 8 Outlet Surge	114.9000	5	0.00	34.4700	0.000000
11	CA-2017-115812	42984	6/14/2017	Standard Class	BH-11710	United States	90032	West	FUR-TA-10001539	Furniture	Tables	Chromcraft Rectangular Conference Tables	999.9999	9	0.20	85.3092	341.236800
12	CA-2017-115812	42984	6/14/2017	Standard Class	BH-11710	United States	90032	West	TEC-PH-10002033	Technology	Phones	Konitel 250 Conference?? phone??- Charcoal	911.4240	4	0.20	68.3568	182.284800

I. Hal yang pertama saya lakukan adalah, **berapa jumlah customer yang terdapat pada setiap kota?**

Untuk mendapatkan jumlah customer tertinggi saya menggunakan patokan kolom `Customer ID` sebagai patokan penjumlahan (`Customer ID` adalah kolom yang berisi data ID pelanggan yang terdaftar). Mencegah hasil yang duplicate (mengantisipasi pembelian > 1 dari customer yang sama) dari `Customer ID`, maka saya melakukan fungsi DISTINCT didalam Fungsi COUNT.

```
SELECT `City`, COUNT(DISTINCT `Customer ID`) AS Total FROM `customer_final` GROUP BY City LIMIT 5
```

City	Total
Aurora	1
Bloomington	1
Bristol	1
Carlsbad	1
Charlotte	1

Untuk mendapatkan 5 kota yang memiliki customer tertinggi, maka saya melakukan Descending terhadap kolom total.

```
SELECT `City`, COUNT(DISTINCT `Customer ID`) AS Total FROM `customer_final` GROUP BY City ORDER BY `Total` DESC LIMIT 5
```

City	Total ▾ 1
Los Angeles	10
Houston	7
New York City	6
Philadelphia	5
Seattle	5

Ternyata Los Angeles menepati kota sebagai customer terbanyak diikuti Houston dan New York dengan jumlah customer 10, 7 dan 6.

II. Menghitung jumlah order per kota dengan **jumlah duplikasi customer** dan **tanpa duplikasi customer**

Pada gambar awal, bahwa penghubung antara table 1 dan table 2 adalah `Customer ID`, maka dapat disimpulkan untuk menghubungkan dua table tersebut menggunakan kolom `Customer ID`. Untuk mendapatkan jumlah duplikasi dan tanpa duplikasi order per kota maka saya menggunakan INNER JOIN (INNER JOIN merupakan penghubung yang mendapatkan kedua kolom yang benar-benar sama isinya) dan LEFT JOIN (LEFT JOIN merupakan rows yang hanya ada di table bagian kiri saja tetapi tidak ada di table sebelah kanan) dan untuk tanpa duplikasi, saya menggunakan DISTINCT.

Order perkota dengan duplikasi Customer.

```
SELECT customer_final.City, COUNT(superstore_final.`Order ID`) AS Total FROM customer_final INNER JOIN superstore_final ON customer_final.`Customer ID` = superstore_final.`Customer ID` GROUP BY City ORDER BY Total DESC LIMIT 5
```

City	Total ▾ 1
Los Angeles	95
Philadelphia	60
San Antonio	49
Roseville	49
Troy	49

Order perkota tanpa duplikasi

```
SELECT b.city, COUNT(a.`Order ID`) AS Total FROM superstore_final a LEFT JOIN (SELECT DISTINCT `Customer ID`, `City` FROM customer_final) b ON a.`Customer ID` = b.`Customer ID` GROUP BY City ORDER BY Total DESC LIMIT 5
```

city	Total ▾ 1
Los Angeles	25
Houston	17
Philadelphia	12
New York City	11
San Francisco	9

Alasan mengapa saya menggunakan SUBQUERY untuk **tanpa duplikasi** adalah yang ingin dihitung adalah ORDER dan yang ingin ditampilkan hanya kolom City dan Total dari orderan dan kita harus menghilangkan duplikasi yang ada di table customer_final terlebih dahulu untuk mendapatkan data yang akurat.

III. Menemukan **tanggal pertama order** pada setiap customer

```
SELECT `Customer ID`, MIN(`Order Date`) FROM `superstore_final` GROUP BY `Customer ID`
```

Customer ID	Tanggal
AA-10480	4/15/2020
AD-10180	12/26/2017
AG-10270	5/13/2017
AM-10360	7/16/2019
BH-11710	06/09/2017
BM-11140	09/08/2017
BS-11590	12/05/2017
CB-12025	3/21/2017

Untuk menemukan tanggal pertama bagi si Customer, maka saya menggunakan MIN untuk mendapatkan tanggal pertama order setiap customer yang saya kelompokkan Customer ID.