Filtering Null Values

1.
```
update orders
set city=null
where order_id in ('CA-2020-161389','US-2021-156909')
```
2. Here first we have updated 2 order id's city to null
3.
```
select * from orders
where order_id in ('CA-2020-161389','US-2021-156909')
```

	e	ship_date	ship_mode	customer_id	customer_name	segment	country	city
1	05 00:00:00.000	2020-12-10 00:00:00.000	Standard Class	IM-15070	Irene Maddox	Consumer	United States	NULL
2	16 00:00:00.000	2021-07-18 00:00:00.000	Second Class	SF-20065	Sandra Flanagan	Consumer	United States	NULL

4. Now if we want to find the null values in the city column… we cant use this

```
select * from orders
where city=null
```

row_id	order_id	order_date	ship_date	ship_mode	customer_id	customer_name	segment	country	city	state

we cant compare city = null

5. Instead of that we have **is** keyword …we can use this to find the null values in the column

```
select * from orders
where city is null
```

	row_id	order_id	order_date	ship_date	ship_mode	customer_id	customer_name	segment	country	city
1	14	CA-2020-161389	2020-12-05 00:00:00.000	2020-12-10 00:00:00.000	Standard Class	IM-15070	Irene Maddox	Consumer	United States	NULL
2	24	US-2021-156909	2021-07-18 00:00:00.000	2021-07-18 00:00:00.000	Second Class	SF-20065	Sandra Flanagan	Consumer	United States	NULL

6. Similarly we have not null as well

```
select * from orders
where city is not null;
```

100 %

Results | Messages

	ship_mode	customer_id	customer_name	segment	country	city
1	00:00.000	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson
2	00:00.000	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson
3	00:00.000	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles
4	00:00.000	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale
5	00:00.000	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale
6	00:00.000	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles
7	00:00.000	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles
8	00:00.000	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles

Aggregation

1.
```
select count(*) as cnt
from orders
```

100 %

Results | Messages

	cnt
1	9994

our first aggregation function is count..it counts and returns number of rows in our table

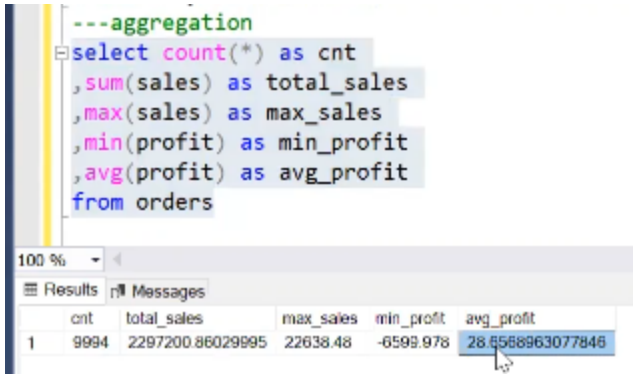
2.
```
select count(*) as cnt
,sum(sales) as total_sales
from orders
```

100 %

Results | Messages

	cnt	total_sales
1	9994	2297200.86029995

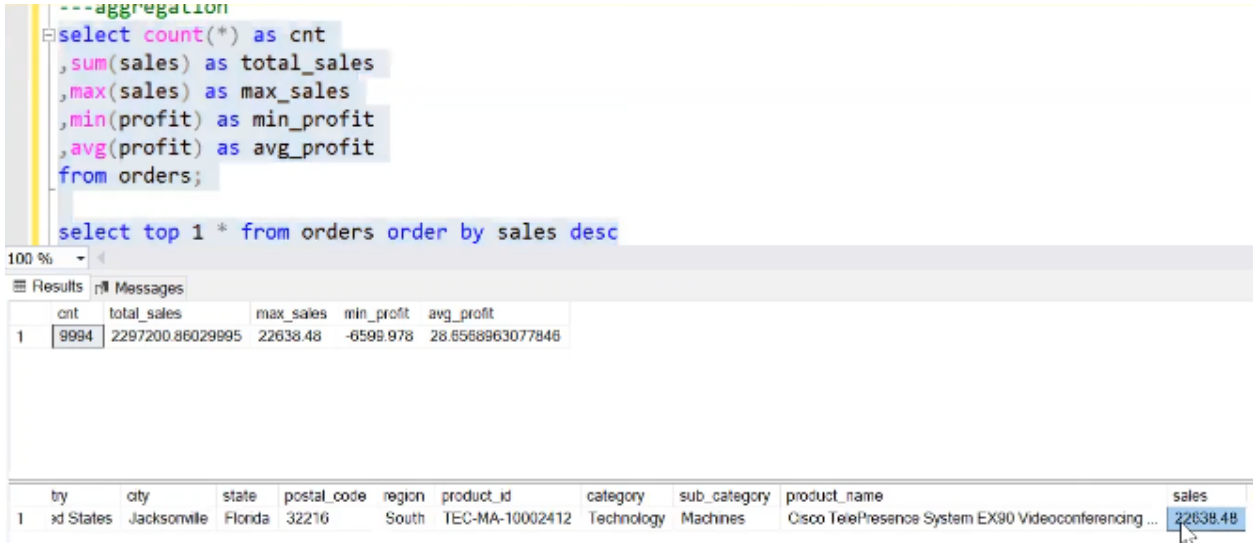
next we have sum..it aggregates the entire sale columns and return its value

3. 
```
---aggregation
select count(*) as cnt
,sum(sales) as total_sales
,max(sales) as max_sales
,min(profit) as min_profit
,avg(profit) as avg_profit
from orders
```
100 %

Results | Messages

	cnt	total_sales	max_sales	min_profit	avg_profit
1	9994	2297200.86029995	22638.48	-6599.978	28.6568963077846

..here we have implemented **max** of sales column and **min** of profit and **avg** of profit..see pic

4. 
```
---aggregation
select count(*) as cnt
,sum(sales) as total_sales
,max(sales) as max_sales
,min(profit) as min_profit
,avg(profit) as avg_profit
from orders;

select top 1 * from orders order by sales desc
```
100 %

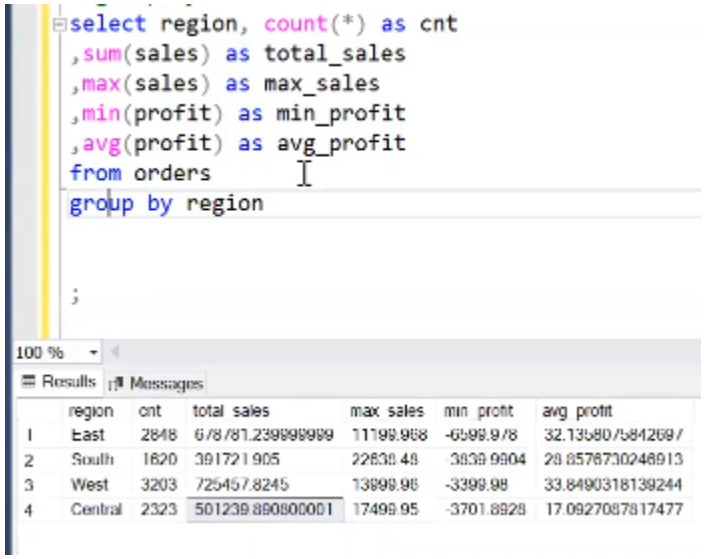
Results | Messages

	cnt	total_sales	max_sales	min_profit	avg_profit
1	9994	2297200.86029995	22638.48	-6599.978	28.6568963077846

	try	city	state	postal_code	region	product_id	category	sub_category	product_name	sales
1	ıd States	Jacksonville	Florida	32216	South	TEC-MA-10002412	Technology	Machines	Cisco TelePresence System EX90 Videoconferencing ...	22638.48

just a proof that max is working…we retrieved max of sale column using order by desc

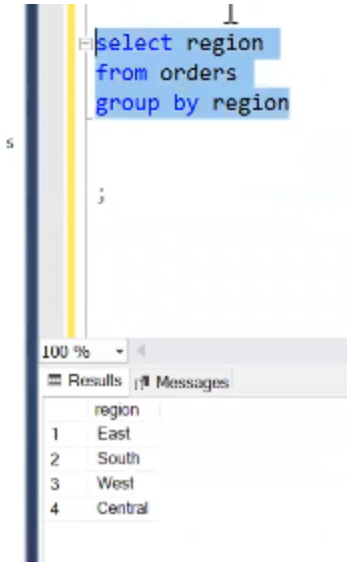
5. Group by

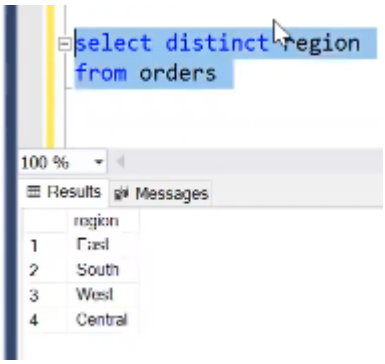
6. 
```
select region, count(*) as cnt
,sum(sales) as total_sales
,max(sales) as max_sales
,min(profit) as min_profit
,avg(profit) as avg_profit
from orders
group by region

;
```
100 %

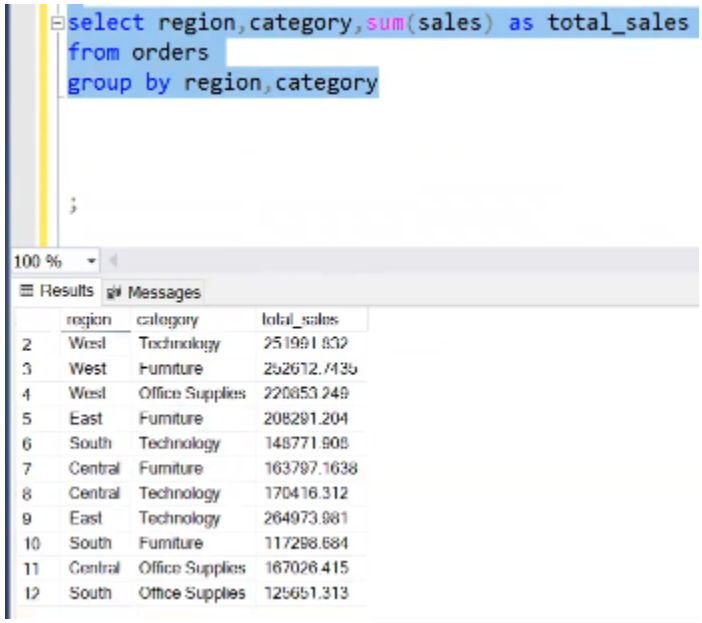
Results | Messages

	region	cnt	total_sales	max_sales	min_profit	avg_profit
1	East	2848	678781.2399999999	11199.968	-6599.978	32.13580756842697
2	South	1620	391721.905	22638.48	-3839.9904	28.857673024691
3	West	3203	725457.8245	13999.96	-3399.98	33.8490318139244
4	Central	2323	501239.8908000001	17499.95	-3701.8928	17.092708781747

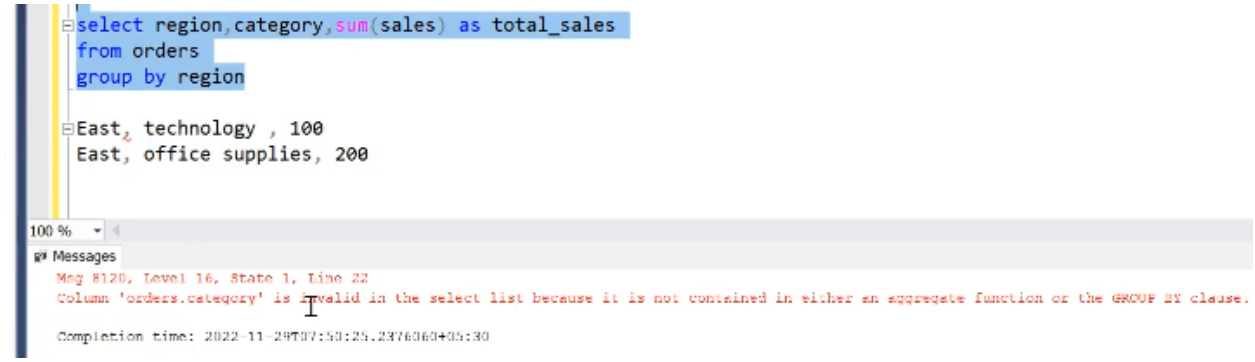
7. Here we have calculated aggregate functions on each region present in our table..by selecting region and grouping it by region..see pic

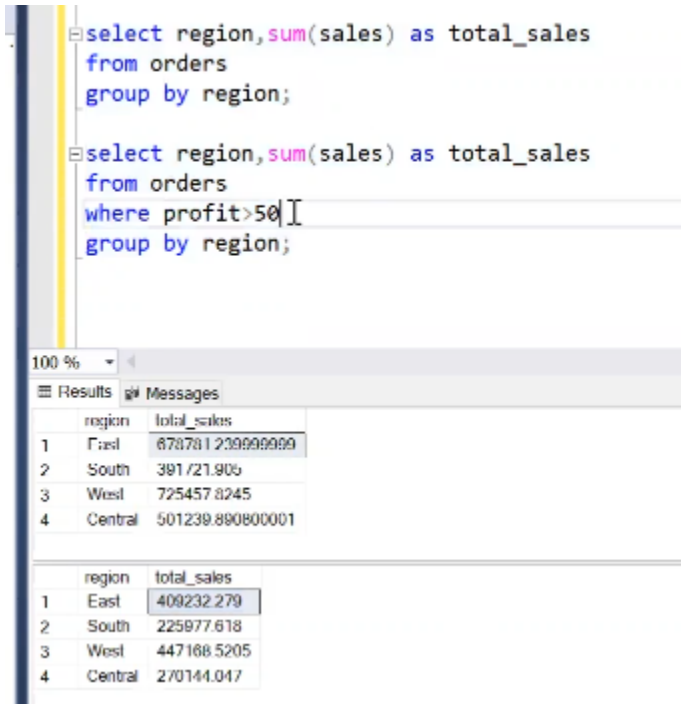
8. 

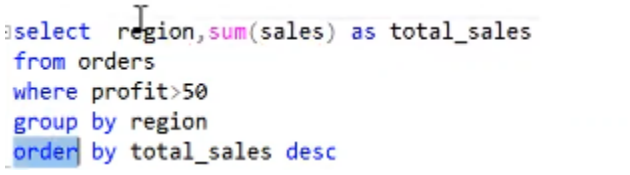
This commands give results same as when we use distinct on the region 

9. 

10. Here we have retrieved the total sales for each category in its region

11. 

this is an interview question
12. We cannot select two columns and group by one column
13. 

14. Here in the first sql statement ..we have sum up all the sales and grouped it by region..and in 2nd statement we have filtered using where command..see oic
15. 

here point of execution..is first from,2nd where, 3rd grp by,4th select and last order by

16.
```
select sub_category,sum(sales) as total_sales
from orders
group by sub_category
```

100 %

Results | Messages

	sub_category	total_sales
1	Supplies	46673.538
2	Storage	223843.608
3	Phones	330007.054
4	Fasteners	3024.28
5	Copiers	149528.03
6	Chairs	328449.103
7	Bookcases	114879.9963
8	Machines	189238.631
9	Art	27118.792
10	Envelopes	16476.402
11	Binders	203412.733

17. Here we have calculated total sum for each category..using aggregate function SUM and group by keyword
18.
```
select sub_category,sum(sales) as total_sales
from orders
group by sub_category
order by total_sales desc
```

00 %

Results | Messages

	sub_category	total_sales
7	Accessories	167380.318
8	Copiers	149528.03
9	Bookcases	114879.9963
10	Appliances	107532.161
11	Furnishings	91705.1639999999
12	Paper	78479.2060000001
13	Supplies	46673.538
14	Art	27118.792
15	Envelopes	16476.402

19. Total sales for each category in desc order
20.
```
select sub_category,sum(sales) as total_sales
from orders
where total_sales > 100000
group by sub_category
order by total_sales desc
```
as we know about the order of execution…here where command doesn't know what total_sales column is and it cannot perform the filtering
21. So the solution to it is we use having after group by

```
select sub_category,sum(sales) as total_sales
from orders
group by sub_category
having sum(sales) > 100000
order by total_sales desc
```

22. Also try with where by just giving sum(sales) --(just for practice)

23.
```
select sub_category,sum(sales) as total_sales
from orders
group by sub_category
having sum(sales) > 100000
order by total_sales desc
```

100 %

Results | Messages

	sub_category	total_sales
1	Phones	330007.054
2	Chairs	328449.103
3	Storage	223843.608
4	Tables	206965.532
5	Binders	203412.733
6	Machines	189238.631
7	Accessories	167380.318
8	Copiers	149528.03
9	Bookcases	114879.9963
10	Appliances	107532.161

here we have retrieved the total sales of each sub category where sum(sales) > 100000

24.
```
select sub_category,sum(sales) as total_sales
from orders
where profit>50
group by sub_category
having sum(sales) > 100000
order by total_sales desc
```

here the order of execution is **from,where,grp by,having,select and order by**

25.
```
select top 5 sub_category,sum(sales) as total_sales
from orders
group by sub_category
having sub_category = 'Phones'
order by total_sales desc;


select top 5 sub_category,sum(sales) as total_sales
from orders
where sub_category = 'Phones'
group by sub_category
order by total_sales desc
```

100 %

Results | Messages

	sub_category	total_sales
1	Phones	330007.054

26. Having : https://www.w3schools.com/sql/sql_having.asp

27. Try this
28.
```
select sub_category,sum(sales) as total_sales--, max(order_date)
from orders
group by sub_category
having max(order_date) > '2020-01-01'
order by total_sales desc;
chairs,300,'2019,10-10'
bookcases,700,'2020,10-10'
```
29.
```
chairs, '2019-01-01', 100
chairs, '2019,10-10', 200
bookcases, '2019-01-01', 300
bookcases, '2020,10-10',400
bookcases,400
bookcases,700
```
30.
```
select sub_category,sum(sales) as total_sales, max(order_date)
from orders
group by sub_category
having max(order_date) > '2020-01-01'
order by total_sales desc;
chairs,300,'2019,10-10'
bookcases,700,'2020,10-10'
```
31. Solve some questions on where and having ..to clear confusion
32. And a rule of thumb ..if we are using group by..we can have only those columns in select statemnt.. Which we using in group by

```
select sub_category,order_date,sum(sales) as total_sales--, max(order_date)
from orders
group by sub_category
having max(order_date) > '2020-01-01'
order by total_sales desc;
```

this will not execute..because select statement having order_date column which is not used in group by

33.
```
select sub_category,min(order_date),sum(sales) as total_sales--, max(order_date)
from orders
group by sub_category
having max(order_date) > '2020-01-01'
order by total_sales desc;
```

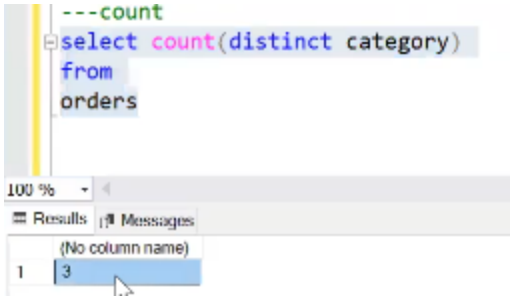
This will execute..because we are using same columns for groupby and select ..
34. In the above statement we have grouped data based on sub category and for each sub category..we got its min date and total sales
35. Count
36.
```
select count(*)
from
orders
```

100 %

Results | Messages

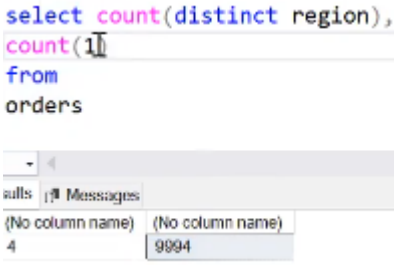
	(No column name)
1	9994

here we got the total number of rows in our table

37. 
```
---count
select count(distinct category)
from
orders
```
100 %
Results | Messages

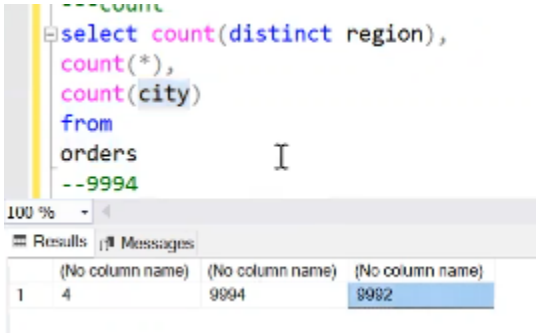
	(No column name)
1	3

we have retrieved the number of disticnt categories in our orders table

38. 
```
select count(distinct region),
count(1)
from
orders
```
Results | Messages

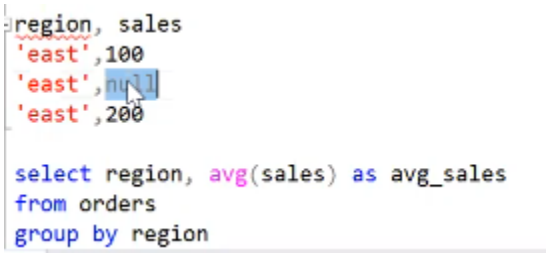
(No column name)	(No column name)
4	9994

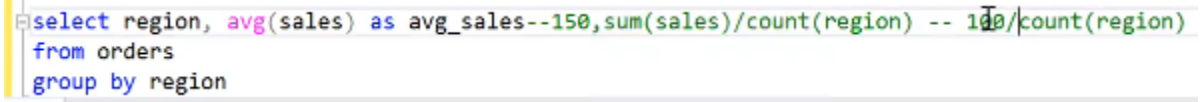
and count(*) is same as count(1)..and we can keep any value

39. 
```
select count(distinct region),
count(*),
count(city)
from
orders
--9994
```
100 %
Results | Messages

	(No column name)	(No column name)	(No column name)
1	4	9994	9992

count does not count null values..as we have 2 null values in the city..we got 9992 as city count

40. 
```
region, sales
'east',100
'east',null
'east',200

select region, avg(sales) as avg_sales
from orders
group by region
```
..if we execute this statement we get 150 as an answer..here our aggregate function(avg) just ignores which has null value..it is an interview ques

41. 
```
select region, avg(sales) as avg_sales--150,sum(sales)/count(region) -- 100/count(region)
from orders
group by region
```

42.