Danny Ma – DATA MART 8 WEEK CHALLENGE

1. DATA CLEANING

```
-- Convert the week date to a DATE format
ALTER TABLE weekly_sales
MODIFY COLUMN week_date VARCHAR(10);
UPDATE weekly_sales
SET week_date = DATE_FORMAT(STR_TO_DATE(week_date, '%d/%m/%y'), '%Y-%m-%d');
alter table weekly_sales modify column week_date date;
-- Add a week_number as the second column for each week_date
alter table weekly_sales
add week number int after week date;
update weekly_sales
set week_number = floor((dayofyear(week_date)-1)/7)+1;
-- Add a month_number with the calendar month for each week_date value as the 3rd column
alter table weekly_sales
add column month_number int after week_number;
update weekly_sales
set month_number = month(week_date) ;
-- Add a calendar_year column as the 4th column containing either 2018, 2019 or 2020 values
alter table weekly_sales
add column calender_year int after month_number;
update weekly_sales
set calender_year = year(week_date);
-- Add a new column called age_band after the original segment column
alter table weekly_sales
add column age_band varchar(20) after segment;
update weekly_sales
set age_band =case when right(segment, 1) = '1' then 'Young Adults'
when right(segment, 1) ='2' then 'Midlle Aged'
when right(segment, 1) = '3' then 'Retriees'
when right(segment, 1) = '4' then 'Retriees' else 'unknown' end ;
alter table weekly_sales drop column age_band;
-- Add a new demographic column using the following mapping for the first letter in the segment values
-- c: couples and F: familes
alter table weekly_sales
add column demographic varchar(20) after segment;
update weekly_sales
set demographic = case when left(segment, 1) = 'C' then 'Couples'
when left(segment, 1) ='F' then 'Families' else 'unknown' end;
```

```
/* Ensure all null string values with an "unknown" string
value in the original segment column as well as the new age_band and demographic columns*/
-- increatse the varchar(4) to varchar(10)
ALTER TABLE weekly_sales
MODIFY COLUMN segment VARCHAR(10);

update weekly_sales
set segment = 'unknown'
where segment = 'null';

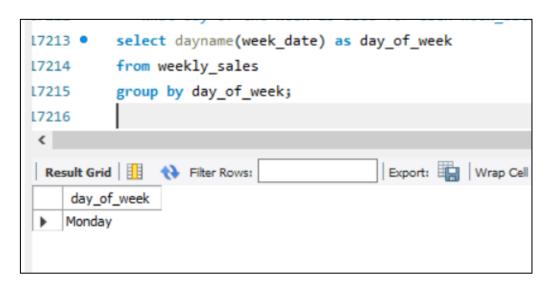
/* Generate a new avg_transaction column as the sales
value divided by transactions rounded to 2 decimal places for each record*/

alter table weekly_sales
add column avg_transaction float;

update weekly_sales
set avg_transaction = round(sales/transactions, 2);
alter table weekly_sales drop column avg_transaction;
```

2. DATA EXPLORATION

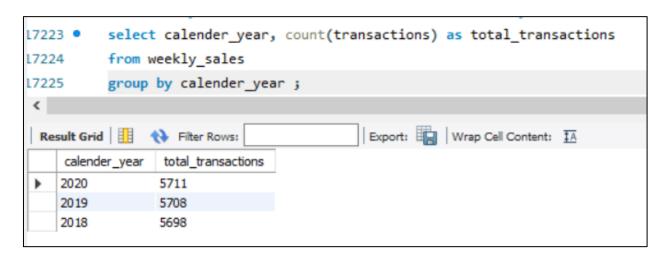
1. What day of the week is used for each week_date value?



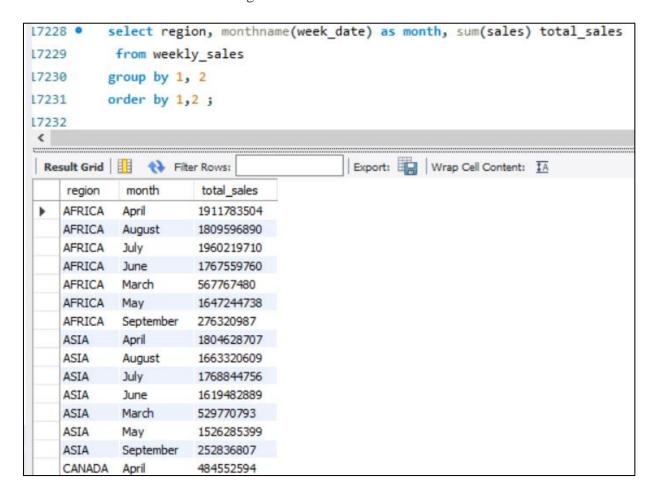
2. What range of week numbers are missing from the dataset?

L7219 •	select week_number from weekly_sales
L7220	where week_number not in (select distinct week_number from weekly_sales);
17221	
<	
Result Grid	Filter Rows: Export: Wrap Cell Content: IA
week_r	number

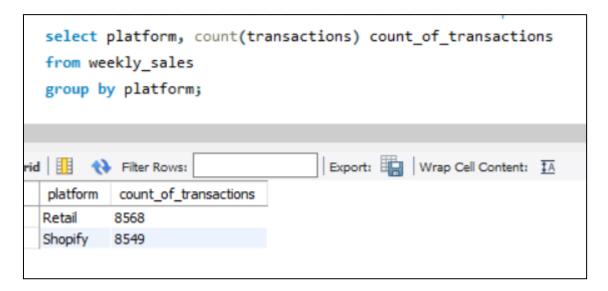
3. How many total transactions were there for each year in the dataset?



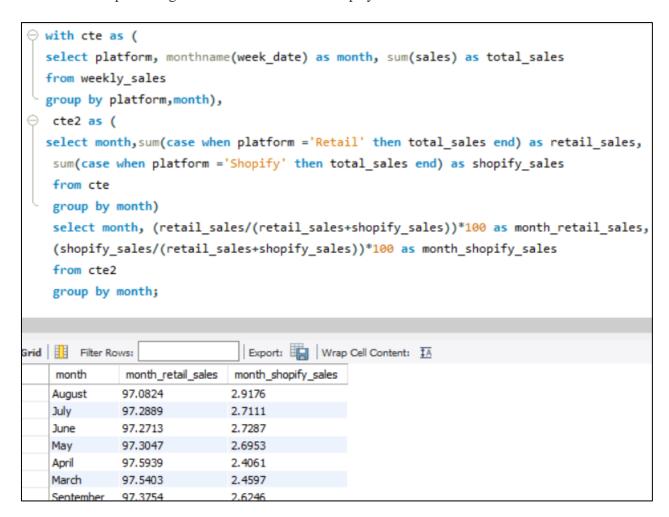
4. What is the total sales for each region for each month?



5. What is the total count of transactions for each platform?



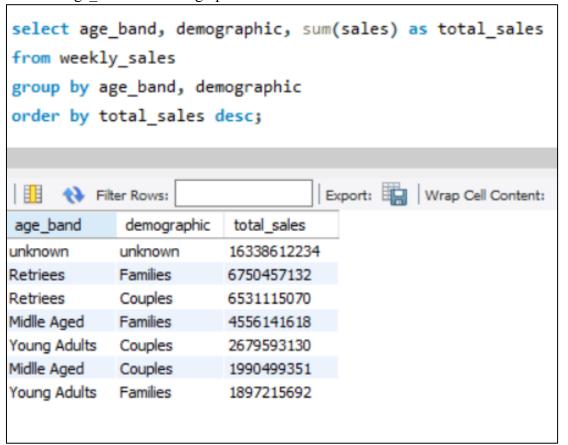
6. What is the percentage of sales for Retail vs Shopify for each month?



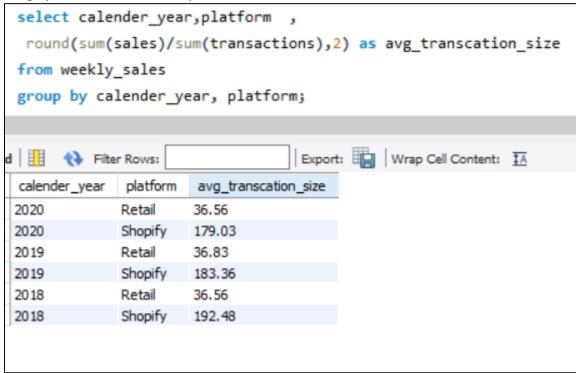
7. What is the percentage of sales by demographic for each year in the dataset?

```
with cte as (
 select calender year, demographic, sum(sales) as total sales
 from weekly_sales
 group by calender_year , demographic),
 cte2 as (select calender_year ,
 sum(case when demographic='couples' then total_sales end) as c_sales,
 sum(case when demographic='families' then total_sales end) as f_sales,
 sum(case when demographic='Unknown' then total_sales end) as un_sales
 from cte
 group by calender year)
 select calender_year , (c_sales /(c_sales+f_sales+un_sales))*100 as couple_sales,
 (f_sales /(c_sales+f_sales+un_sales))*100 as family sales,
 (un_sales /(c_sales+f_sales+un_sales))*100 as unknown_sales
 from cte2;
d Filter Rows:
                             Export: Wrap Cell Content: TA
 calender_year couple_sales family_sales unknown_sales
2020
             28.7199
                        32.7253
                                   38,5548
2019
             27.2752
                        32.4742
                                   40.2506
2018
             26.3805
                        31.9876
                                   41.6320
```

8. Which age band and demographic values contribute the most to Retail sales?



9.Can we use the avg_transaction column to find the average transaction size for each year for Retail vs Shopify? If not - how would you calculate it instead?



This technique is usually used when we inspect an important event and want to inspect the impact before and after a certain point in time.

Taking the week_date value of 2020-06-15 as the baseline week where the Data Mart sustainable packaging changes came into effect.

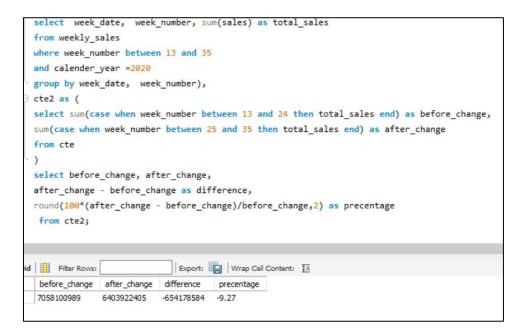
We would include all week_date values for 2020-06-15 as the start of the period after the change and the previous week_date values would be **before**

Using this analysis approach - answer the following questions:

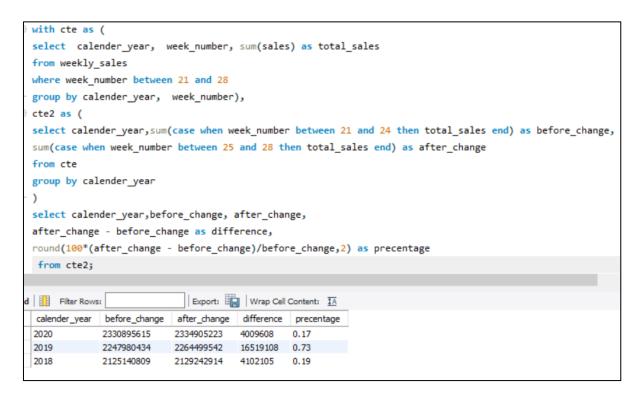
1. What is the total sales for the 4 weeks before and after 2020-06-15? What is the growth or reduction rate in actual values and percentage of sales?

```
with cte as (
select week_date, week_number, sum(sales) as total_sales
from weekly_sales
where week number between 21 and 28
and calender_year =2020
group by week_date, week_number),
select sum(case when week_number between 21 and 24 then total_sales end) as before_change,
sum(case when week_number between 25 and 28 then total_sales end) as after_change
from cte
select before change, after change,
after_change - before_change as difference,
round(100*(after_change - before_change)/before_change,2) as precentage
from cte2;
Filter Rows:
                           Export: Wrap Cell Content: IA
before_change after_change difference precentage
2330895615
             2334905223
                        4009608
```

2. What about the entire 12 weeks before and after?



3. How do the sale metrics for these 2 periods before and after compare with the previous years in 2018 and 2019?



```
with cte as (
  select calender_year, week_number, sum(sales) as total_sales
  from weekly_sales
 where week_number between 13 and 35
group by calender year, week number),
cte2 as (
 select calender_year, sum(case when week_number between 13 and 24 then total_sales end) as before_change,
 sum(case when week_number between 25 and 35 then total_sales end) as after_change
 from cte
 group by calender_year
 )
 select calender year, before change, after change,
 after_change - before_change as difference,
 round(100*(after_change - before_change)/before_change,2) as precentage
  from cte2;
rid Filter Rows:
                             Export: Wrap Cell Content: IA
 calender_year before_change after_change difference precentage
            7058100989
                          6403922405 -654178584 -9.27
 2019
         6861158161 6303557285 -557600876 -8.13
 2018
       6396562317 5947847148 -448715169 -7.01
```

BONUS QUESTION

Which areas of the business have the highest negative impact in sales metrics performance in 2020 for the 12 week before and after period?

- region
- platform
- age_band
- demographic
- customer_type

```
select region, platform, age_band, demographic, customer_type, week_date, week_number,
sum(sales) as total sales
from weekly sales
where week_number between 13 and 35
and calender_year =2020
group by region, platform, age_band, demographic, customer_type, week_date, week_number),
cte2 as (
select region, platform, age band, demographic, customer type, week date, week number,
sum(case when week number between 13 and 24 then total sales end) as before change,
sum(case when week_number between 25 and 35 then total_sales end) as after_change
from cte
group by region, platform, age_band, demographic, customer_type, week_date, week_number
)
select region, platform, age_band, demographic, customer_type, week_date, week_number,
before_change, after_change,
after change - before change as difference,
round(100*(after_change - before_change)/before_change,2) as precentage
from cte2
order by precentage ;
```

region	platform	age_band	demographic	customer type	week_date	week number	before_change	after change	difference	precentage
ASIA	Shopify	Midlle Aged	Families	New	2020-04-06	14	32801	NULL	NULL	NULL
EUROPE	Retail	Young Adults	Families	New	2020-04-06	14	32911	NULL	NULL	NULL
ASIA	Retail	Midlle Aged	Families	Existing	2020-04-06	14	11115259	NULL	NULL	NULL
EUROPE	Shopify	Young Adults	Couples	New	2020-04-06	14	1427	NULL	NULL	NULL
AFRICA	Shopify	unknown	unknown	Guest	2020-04-06	14	847013	NULL	NULL	NULL
EUROPE	Retail	Young Adults	Families	Existing	2020-04-06	14	390357	NULL	NULL	NULL
EUROPE	Retail	Young Adults	Couples	Exis Existing	2020-04-06	14	608634	NULL	NULL	NULL
OCEANIA	Retail	Midlle Aged	Families	Existing	2020-04-06	14	17218066	NULL	NULL	NULL
AFRICA	Shopify	Young Adults	Couples	Existing	2020-03-30	13	236170	NULL	NULL	NULL
OCEANIA	Shopify	Retriees	Couples	Existing	2020-03-30	13	1011520	NULL	NULL	NULL
OCEANIA	Retail	unknown	unknown	New	2020-03-30	13	3249566	NULL	NULL	NULL
OCEANIA	Shopify	Young Adults	Families	New	2020-03-30	13	45230	NULL	NULL	NULL
EUROPE	Retail	Retriees	Couples	New	2020-03-30	13	172033	NULL	NULL	NULL
SOUTH	Shopify	Young Adults	Couples	Existing	2020-03-30	13	12556	NULL	NULL	NULL
USA	Shopify	Retriees	Couples	New	2020-03-30	13	34312	NULL	NULL	NULL
USA	Shopify	Midlle Aged	Couples	New	2020-03-30	13	32471	NULL	NULL	NULL
SOUTH	Shopify	Midlle Aged	Couples	New	2020-03-30	13	5460	NULL	NULL	NULL
sult 78 ×	B 1 4		- +		2022 22 22		45004	NUIU	RUUU	NULL