

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: families

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?

```

17213 • select dayname(week_date) as day_of_week
17214 from weekly_sales
17215 group by day_of_week;
17216

```

Result Grid | Filter Rows: | Export: | Wrap Cell

day_of_week
Monday

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

```

17219 • select week_number from weekly_sales
17220 where week_number not in (select distinct week_number from weekly_sales );
17221



```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: I A

week_number



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

```
l7223 • select calender_year, count(transactions) as total_transactions
l7224       from weekly_sales
l7225       group by calender_year ;
```

<		
Result Grid		
Filter Rows: <input type="text"/>		
Export:  Wrap Cell Content: 		
	calender_year	total_transactions
▶	2020	5711
	2019	5708
	2018	5698

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

```
l7228 • select region, monthname(week_date) as month, sum(sales) total_sales
l7229       from weekly_sales
l7230       group by 1, 2
l7231       order by 1,2 ;
l7232
```

<			
Result Grid			
Filter Rows: <input type="text"/>			
Export:  Wrap Cell Content: 			
	region	month	total_sales
▶	AFRICA	April	1911783504
	AFRICA	August	1809596890
	AFRICA	July	1960219710
	AFRICA	June	1767559760
	AFRICA	March	567767480
	AFRICA	May	1647244738
	AFRICA	September	276320987
	ASIA	April	1804628707
	ASIA	August	1663320609
	ASIA	July	1768844756
	ASIA	June	1619482889
	ASIA	March	529770793
	ASIA	May	1526285399
	ASIA	September	252836807
	CANADA	April	484552594

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

```
select platform, count(transactions) count_of_transactions
from weekly_sales
group by platform;
```

Grid

Filter Rows:

Export:

Wrap Cell Content:

platform	count_of_transactions
Retail	8568
Shopify	8549

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

```
with cte as (
select platform, monthname(week_date) as month, sum(sales) as total_sales
from weekly_sales
group by platform,month),
cte2 as (
select month,sum(case when platform ='Retail' then total_sales end) as retail_sales,
sum(case when platform ='Shopify' then total_sales end) as shopify_sales
from cte
group by month)
select month, (retail_sales/(retail_sales+shopify_sales))*100 as month_retail_sales,
(shopify_sales/(retail_sales+shopify_sales))*100 as month_shopify_sales
from cte2
group by month;
```

Grid

Filter Rows:

Export:

Wrap Cell Content:

month	month_retail_sales	month_shopify_sales
August	97.0824	2.9176
July	97.2889	2.7111
June	97.2713	2.7287
May	97.3047	2.6953
April	97.5939	2.4061
March	97.5403	2.4597
September	97.3754	2.6246

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;

```

Filter Rows: | Export:  | Wrap Cell Content: 

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?

```

select age_band, demographic, sum(sales) as total_sales
  from weekly_sales
  group by age_band, demographic
  order by total_sales desc;

```

  Filter Rows: | Export:  | Wrap Cell Content: 

age_band	demographic	total_sales
unknown	unknown	16338612234
Retriees	Families	6750457132
Retriees	Couples	6531115070
Midle Aged	Families	4556141618
Young Adults	Couples	2679593130
Midle Aged	Couples	1990499351
Young Adults	Families	1897215692

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?

```
select calender_year, platform ,
       round(sum(sales)/sum(transactions),2) as avg_transcation_size
from weekly_sales
group by calender_year, platform;
```

calender_year	platform	avg_transcation_size
2020	Retail	36.56
2020	Shopify	179.03
2019	Retail	36.83
2019	Shopify	183.36
2018	Retail	36.56
2018	Shopify	192.48

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),
cte2 as (
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;
```

before_change	after_change	difference	percentage
2330895615	2334905223	4009608	0.17

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

```
select 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 week_date, week_number),
cte2 as (
select 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
)
select before_change, after_change,
after_change - before_change as difference,
round(100*(after_change - before_change)/before_change,2) as precentage
from cte2;
```

id	before_change	after_change	difference	precentage
7058100989	6403922405	-654178584	-9.27	

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 21 and 28
group by calender_year, week_number),
cte2 as (
select calender_year, 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
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 percentage
from cte2;
```

calender_year	before_change	after_change	difference	percentage
2020	2330895615	2334905223	4009608	0.17
2019	2247980434	2264499542	16519108	0.73
2018	2125140809	2129242914	4102105	0.19

```

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;

```

calender_year	before_change	after_change	difference	precentage
2020	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

```

with cte as (
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 ;

```


Result Grid											
Filter Rows:		Export:		Wrap Cell Content:		Fetch rows:					
region	platform	age_band	demographic	customer_type	week_date	week_number	before_change	after_change	difference	percentage	
ASIA	Shopify	Midle 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	Midle 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	Existing	2020-04-06	14	608634	NULL	NULL	NULL	
OCEANIA	Retail	Midle 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	Midle Aged	Couples	New	2020-03-30	13	32471	NULL	NULL	NULL	
SOUTH ...	Shopify	Midle Aged	Couples	New	2020-03-30	13	5460	NULL	NULL	NULL	