Data Cleansing Steps

In a single query, perform the following operations and generate a new table in the data_mart schema named clean_weekly_sales:

Add a week_number as the second column for each week_date value, for example any value from the 1st of January to 7th of January will be 1, 8th to 14th will be 2, etc.

Add a month_number with the calendar month for each week_date value as the 3rd column

Add a calendar_year column as the 4th column containing either 2018, 2019 or 2020 values

Add a new column called age_band after the original segment column using the following mapping on the number inside the segment value

Add a new demographic column using the following mapping for the first letter in the segment values:

segment | demographic |

C | Couples |

F | Families |

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

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

---

ANS----------

```
use case1;

create  table clean_weekly_sales as select week_date,week(week_date) as
week_number,month(week_date) as month_number,year(week_date) as calendar_year,

region,platform ,case when segment='null' then 'unknown' else segment end as
segment,case when right(segment,1)='1' then 'Young Adults'
```

```
when right(segment,1)='2' then 'Middle Aged' when right(segment,1)='3' or '4'then 'Retires'
else 'unknown' end as age_band ,case when left(segment,1)='C'

then 'Couples' when left(segment,1)='F' then 'Families' else 'unknown' end as demographic
,customer_type,transactions,sales,round(sales / transactions,2)

as avg_transaction from  weekly_sales;
```

Data Exploration

Which week numbers are missing from the dataset?

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

What are the total sales for each region for each month?

What is the total count of transactions for each platform

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

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

Which age_band and demographic values contribute the most to Retail sales?

```
Q1-------------------------------------------------

create table seq100(x int auto_increment primary key);

insert into seq100 values (),(),(),(),(),(),(),(),(),();

insert into seq100 values (),(),(),(),(),(),(),(),(),();

insert into seq100 values (),(),(),(),(),(),(),(),(),();

insert into seq100 values (),(),(),(),(),(),(),(),(),();

insert into seq100 values (),(),(),(),(),(),(),(),(),();

insert into seq100  select x+50 from seq100;

select * from seq100;

drop table seq100;

create table seq52 as select x from seq100 limit 52;
```

```
select x as week_number from seq52 where x not in (select  distinct week_number as
dis_wk from clean_weekly_sales);
```

Q2----------

```
select  calendar_year,sum(transactions)as total_transactions from clean_weekly_sales group
by calendar_year order by calendar_year;
```

Q3.----------

```
select region,month_number,sum(sales) as total_sales   from clean_weekly_sales group by
region,month_number order by month_number,region;
```

Q4.----------

```
select platform,count(transactions) as tot_cnt from clean_weekly_sales group by platform;
```

Q5.----------

```
create view v as (select calendar_year,month_number ,sum(sales) as total from
clean_weekly_sales where platform='retail' group by month_number,calendar_year order by
month_number);

create view v1 as (select calendar_year, month_number ,sum(sales) as total1 from
clean_weekly_sales where platform='shopify' group by month_number,calendar_year order
by month_number);

select month_number,calendar_year,round((select total from v where
month_number=c.month_number and calendar_year=c.calendar_year)/sum(sales)*100,2)
as retail_per,

round((select total1 from v1 where month_number=c.month_number and
calendar_year=c.calendar_year)/sum(sales)*100,2) as shopify_per

from clean_weekly_sales as c group by month_number,calendar_year order by
month_number;
```

Q6.----------

```
select calendar_year,demographic ,sum(sales) as yearly_sales,round((sum(sales)/(select t
from vw where calendar_year=cws.calendar_year))*100,2)  as percentage_in_sales from
clean_weekly_sales as cws group by calendar_year,demographic order by calendar_year;

create view vw as (select sum(sales) as t ,calendar_year   from clean_weekly_sales group by
calendar_year);
```

Q.7----------

```
 with cte as (select sum(sales) as tot,age_band,demographic from clean_weekly_sales where
platform='retail' group by age_band,demographic)

 select age_band,demographic from cte where tot=(select max(tot) from cte);
```