**Question 1**

**Using OnlineRetail dataset • write at least 5 analytical SQL queries that tells a story about the data • write small description about the business meaning behind each query**

Customers has purchasing transaction that we shall be monitoring to get intuition behind each customer behavior to target the customers in the most efficient and proactive to improve customer retention and decrease churn. way as well as total sales in all countries will be monitoring to increase sales/revenue .

**1-we can observe here that the top country in total sales is "United Kingdom" and with a large difference from the nearest country .**

select country , sum (quantity \* unitprice) as total\_price

from online\_retail

group by country

order by total\_price desc;

the top 5 countries in total sales is :

1. "United Kingdom" 8,004,787 $

2. "Netherlands" 278,740$

3. "EIRE" 258,732 $

4. "Germany" 219,851 $

5. "France" 200,013 $

Also we can see the last 5 countries in total Sales:

1. "Saudi Arabia" 133 $
2. "Bahrain" 502 $
3. "Czech Republic" 744 $
4. "RSA" 991 $
5. "Brazil" 1,164 $

So this countries need new strategy to increase the sales .

And this query retrieve the top country have sales

select \* from (

select country , sum (quantity \* unitprice) as total\_price ,

dense\_rank()over ( order by sum (quantity \* unitprice) desc ) as rnk

from online\_retail

group by country

order by total\_price desc) new\_table

where rnk = 1;

**2-we can observe here the top five commodities in total sales :**

select distinct (stockcode) , description ,

sum(quantity) over (partition by country , stockcode order by country ) as quantities

from online\_retail

order by quantities desc;

1. "POPCORN HOLDER"
2. "WORLD WAR 2 GLIDERS ASSTD DESIGNS"
3. "ASSORTED COLOUR BIRD ORNAMENT"
4. "damaged"
5. "JUMBO BAG RED RETROSPOT"

**3-we can retrieve the best sale commodity and the lowest sale commodity in every country :**

select distinct Country,

first\_value(Description) over (partition by Country order by total\_sales

rows between unbounded preceding and unbounded following) as best\_sale,

last\_value(Description) over (partition by Country order by total\_sales

rows between unbounded preceding and unbounded following) as lowes\_sale

from (select Country, Description,

sum(Quantity\*UnitPrice) as total\_sales

from Online\_Retail

group by Country, Description ) as sub\_query

order by Country

**4-we can observe here the top five return commodities is :**

select distinct (stockcode) , description ,

sum(quantity) over (partition by country , stockcode order by country ) as quantities

from online\_retail

order by quantities ;

1. "printing smudges/thrown away"
2. "TRAVEL CARD WALLET I LOVE LONDON"
3. "Printing smudges/thrown away"
4. "TRAVEL CARD WALLET VINTAGE ROSE "
5. "throw away"

**5-We can retrieve the total count of transactions per month and total sales per month also:**

select Month\_Name,count(distinct(invoiceno)) as transactions\_month , sum(quantity\*unitprice) as totalprice\_month

from (

select \* , TO\_CHAR( to\_timestamp(invoicedate , 'MM/DD/YYYY HH24:MI') , 'Month' ) as Month\_Name

from online\_retail ) nn

group by Month\_Name

order by transactions\_month desc

The observation here that the top three months in sales is

1. "November "
2. "December "
3. "October "

And the last three months in sales is

1. "February "
2. "January "
3. "August "

So we can see some observation about sales in relation with countries , sales in relation with months and sales in relation with the commodity itself but also we need to focus in customers behavior

**6- this Query can retrieve the top 20 percent of the employees who pay the most :**

select \* from (

select customerid,sum (quantity \* unitprice) as Monetary ,

percent\_rank() over (order by sum (quantity \* unitprice) desc)\*100 as rnk

from online\_retail

group by customerid ) asca

where rnk <20

**7- we will see in the next Question ten categories of customers depending on three main variables**

* Recency : how recent the last transaction is
* Frequency : how many times the customer has bought from our store
* Monetary : how much each customer has paid for our products

So there is table created has all customers with categories called **customer\_type** but to observe the count of customers in every category we need to do **Pivoting** (note : here I use **ANSI-CODE** to do Pivoting to run in any DBMS)

select country ,

count (case when group\_name = 'Champions' then c.customerid end ) as Champions ,

count (case when group\_name = 'Potential Loyalists' then c.customerid end ) as Potential\_Loyalists ,

count (case when group\_name = 'Loyal Customers' then c.customerid end ) as Loyal\_Customers ,

count (case when group\_name = 'Recent Customers' then c.customerid end ) as Recent\_Customers ,

count (case when group\_name = 'Promising' then c.customerid end ) as Promising ,

count (case when group\_name = 'Customers Needing Attention' then c.customerid end ) as Customers\_Needing\_Attention ,

count (case when group\_name = 'At Risk' then c.customerid end ) as At\_Risk ,

count (case when group\_name = 'Cant Lose Them' then c.customerid end ) as Cant\_Lose\_Them ,

count (case when group\_name = 'Hibernating' then c.customerid end ) as Hibernating ,

count (case when group\_name = 'Lost' then c.customerid end ) as Lost

from online\_retail r , customer\_type c

where r.customerid = c.customerid

group by country

order by Champions desc;

After Running this code we can see issue that in almost all countries this issue is

* Few numbers of customers is in safe category this category contain this categories
* Champions
* Potential Loyalists
* Loyal Customers
* Huge numbers of customers in promising category that contain
* Recent Customers
* Promising
* Huge numbers of customers in At Risk category that contaion
* Customers Needing Attention
* At Risk
* Can’t Lose Them
* Hibernating
* Lost