**Analytical SQL Case Study Q1**

**All these queries are using table called onlineretail not online\_retail**

-- Ranking the countries based on most sales per years

-- it is noticed that United Kingdom have the most sales

-- and the countries that recently expanded in 2011 are in the late ranking

**SELECT \***

**,RANK() OVER (**

**PARTITION BY YEAR ORDER BY SALES DESC**

**) RNK**

**FROM (**

**SELECT COUNTRY**

**,EXTRACT(YEAR FROM INVOICEDATE) AS YEAR**

**,ROUND(SUM(QUANTITY \* UNITPRICE)) AS SALES**

**FROM ONLINERETAIL**

**GROUP BY 2**

**,COUNTRY**

**) SUB**

**ORDER BY COUNTRY**

**,RNK;**

-- This query shows the best month of sales in each country and different between each top 3 months of sales

-- Country like Greece have a significant difference of the top and second months of sales that's mean a problem

**SELECT \***

**, SALES - LEAD(SALES, 1, SALES) OVER (**

**PARTITION BY COUNTRY ORDER BY RNK**

**) DIFFERNT**

**FROM (**

**SELECT \***

**, RANK() OVER (**

**PARTITION BY COUNTRY ORDER BY SALES DESC**

**) RNK**

**, FIRST\_VALUE(MONTH) OVER (**

**PARTITION BY COUNTRY ORDER BY SALES DESC**

**) BEST\_MONTH**

**FROM (**

**SELECT COUNTRY**

**, EXTRACT(MONTH FROM INVOICEDATE) AS MONTH**

**, ROUND(SUM(QUANTITY \* UNITPRICE)) AS SALES**

**FROM ONLINERETAIL**

**GROUP BY EXTRACT(MONTH FROM INVOICEDATE)**

**, COUNTRY**

**) SUB**

**) SUB2**

**WHERE RNK <= 3**

**ORDER BY COUNTRY;**

-- after known that United Kingdom is the top country of sales

-- I analysis the customers behavior and get the relations between the products by times of buy it together

**WITH TEMP\_ONLINE\_RETAIL**

**AS (**

**SELECT \***

**FROM ONLINERETAIL**

**WHERE COUNTRY = 'United Kingdom'**

**)**

**SELECT \***

**, RANK() OVER (**

**ORDER BY COUNT\_OF\_SALE\_TOGETHER DESC**

**) RNK**

**FROM (**

**SELECT T1.DESCRIPTION**

**, T2.DESCRIPTION**

**, COUNT(\*) COUNT\_OF\_SALE\_TOGETHER**

**FROM TEMP\_ONLINE\_RETAIL T1**

**, TEMP\_ONLINE\_RETAIL T2**

**WHERE T1.INVOICENO = T2.INVOICENO**

**AND T1.DESCRIPTION != T2.DESCRIPTION**

**GROUP BY T1.DESCRIPTION**

**, T2.DESCRIPTION**

**) SUB**

**ORDER BY RNK LIMIT 20;**

-- By this query exploring country sales to know which top and lowest products sales in each country

-- there are some negative sales so that refer to a problem in these products

**SELECT COUNTRY**

**, DESCRIPTION**

**, ( CASE**

**WHEN RNK = 0**

**THEN 'Top sales in ' || COUNTRY**

**WHEN RNK = 1**

**THEN 'Lowest sales in ' || COUNTRY**

**END**

**)**

**, TOTAL\_SALES**

**FROM ( SELECT \***

**, PERCENT\_RANK() OVER (**

**PARTITION BY COUNTRY ORDER BY TOTAL\_SALES DESC**

**) RNK**

**FROM (**

**SELECT COUNTRY**

**, DESCRIPTION**

**, SUM(QUANTITY \* UNITPRICE) AS TOTAL\_SALES**

**FROM ONLINERETAIL**

**GROUP BY COUNTRY**

**, DESCRIPTION**

**) SUB**

**) SUB2**

**WHERE RNK IN (0,1);**

-- By this query exploring country negative sales to know how much negative sales and which one make big lost and which country

**SELECT \***

**, RANK() OVER (**

**ORDER BY TOTAL\_SALES**

**) RNK**

**FROM (**

**SELECT COUNTRY**

**, DESCRIPTION**

**, SUM(QUANTITY \* UNITPRICE) AS TOTAL\_SALES**

**FROM ONLINERETAIL**

**GROUP BY COUNTRY**

**, DESCRIPTION**

**) SUB**

**WHERE TOTAL\_SALES < 0;**

As Data preparation In online retail table I Removed null values in customerid column and change the datatype of INVOICEDATE from text to date

**CREATE TABLE ONLINERETAIL AS**

**SELECT INVOICENO**

**, STOCKCODE**

**, DESCRIPTION**

**, QUANTITY**

**, TO\_TIMESTAMP(INVOICEDATE, 'MM,DD,YYYY HH24:MI') AS INVOICEDATE**

**, UNITPRICE**

**, CUSTOMERID**

**, COUNTRY**

**FROM ONLINE\_RETAIL**

**WHERE CUSTOMERID != ''**

**ORDER BY CUSTOMERID;**