dataset-analysed-using-sql

December 6, 2023

[3]:
```sql
/* Lets import the Sales , Customer, and Product Datasets.*/
SELECT *
FROM SALES
```

[4]:
```sql
SELECT *
FROM Customer

select *
from Product
```

[5]:
```sql
/* lets explore the basic info of these datasets*/

select  count(*) as Total_Count
from Sales
select  count(*) as Total_Count
from Product
select  count(*) as Total_Count
from Customer
```

[6]:
```sql
/* what are the different categories of sales table 's export class*/

select distinct Export_Class
from Sales
```

[1]:
```sql
/* how many product ordered from respective class*/

select  Export_Class,count(distinct Product_ID) as␣
 ↪Total_Unique_Product,count(Product_ID) as Total_Product,␣
 ↪SUM(Ordered_Quantity) as Total_Ordered_Quantity
from Sales
group by Export_Class
order by Total_Product desc
```

It means that the maximum no. of product ordered from Premium Class and Low Class used less among others.

[8]:
```
/* which respective classes offered greater profit*/

select  Export_Class, round(sum(Price-Cost-Tax_15),2) as Profit
from Sales
group by Export_Class
order by Profit desc
```

It is showing that only Premium Class covers more than 3 times in profit as comapare to Middle Class and First Class and 11 times as compare to Low Class.

[9]:
```
/* what is the first and last date of sales transaction*/
select top 1 *
from Sales
order by Order_Date asc

select top 1  x.*
from
(select *, Row_Number() over(order by Order_No) as rank
from Sales) x
order by rank desc
```

[11]:
```
select *, datediff(day,Order_Date,Delivery_Date) as Delivery_days
into Sales4
from Sales
```

[12]:
```
select *
from Sales4
```

[13]:
```
/* which export class took minimum days to deliver.*/

select x.Order_No,x.Export_Class, x.Delivery_days,
case
when     1 <=Delivery_days and  Delivery_days <=5 THEN 'Excellent'
when     6 <=Delivery_days and  Delivery_days <=10 THEN 'Good'
when     11 <=Delivery_days and  Delivery_days <=15 THEN 'Average'
when     16 <=Delivery_days and  Delivery_days <=20 THEN 'Below Average'
end 'Rate'
into Delivery_Rate5
from
(select Order_No,Export_Class,Delivery_days, count(Delivery_days) as total
from Sales4
group by Order_No,Export_Class, Delivery_days
) as x
order by Rate asc
```

[14]:
```
Select Export_Class, Rate,count(Rate) as Frequency
from Delivery_Rate5
```

```
group by Export_Class,Rate
order by Frequency desc
```

So, it concluded that Below Average rating has occurred frequently in Premium Class compared to other Export Classes.

Key Takeaways:

1. Premium Class needs to improve its performance by reducing the occurrence of (Below Average and Average) ratings because it is significantly affecting the overall performance.
2. First class should focus on (Below Average and Average) ratings.

[19]:
```
/* on which date we got the highest sales and vice versa */

select Order_Date, round(sum(Price*Ordered_Quantity),2) as Total_Sale
from Sales
group by Order_Date
order by Total_Sale desc
```

So we have a data of 20 days which started from 13-Nov-2023 and ended on 02-Dec-2023. The highest and lowest Sales occurred on **17-Nov-2023 and 16-Nov-2023** that was $ **5902.7 and $ 4540.7 respectively.**

[3]:
```
/* What are the top 10 products, who contained the highest Benefit cost ratio*/

Select top 10 p.Product_Name,p.Class, round(sum((s.Price-s.Cost-s.Tax_15)/(s.
 ↪Cost)),2) as BCR
from Sales s
left join Product p
on s.Product_ID = p.Product_ID
group by p.Product_Name,p.Class
order by BCR desc
```

So, the top 4 Products belong to Product Class (Office Supplies) which contained the highest BCR (Benefit Cost Ratio) among other Classes.

[4]:
```
select *
from Product
```

[12]:
```
/* WE WANT TO CHECK THE MARKET BEHAVIOR OF EXPIRED PRODUCTS (ZERO PROMOTION)␣
 ↪DURING THE SALE AND
THOSE ONES WHICH HAD BEEN GONE THROUGH ADVERTISING PHASE  (2022 MEANS NO␣
 ↪PROMOTION, 2023 MEANS
DURING SALES IN 2023 THESE PRODUTS ARE PROMOTED AND YEAR AFTER 2023 MEANS THE␣
 ↪PRODUCT WILL BE PROMOTED) */

select p.Expiry_Year, round(sum(s.Price*s.Ordered_Quantity),3) as Total_Sales ,␣
 ↪round(sum((s.Price-s.Cost-s.Tax_15)*(Ordered_Quantity)),3)as  Total_Profit
```

```
from Sales s
left join Product p
on s.Product_ID = p.Product_ID
group by p.Expiry_Year
order by Total_Sales desc
```

The best thing is that the 2022 year's product which were not been promoted during the year of 2023 still found the third highest in Total Sales and Profit. It shows the importance of those products. We should focus on these products to make a better marketing campaign which will improve the performace of ABC company.

[14]:
```
/* LETS DEEP DIVE IN 2022 NOT PROMOTED PRODUCTS.*/
select p.Class, p.Sub_Class ,count(p.Product_ID) AS Total_Product
into salesprod1
from Sales s
left join Product p
on s.Product_ID = p.Product_ID
where Expiry_Year = '2022'
group by Class, Sub_Class


SELECT Class, Sub_Class, Total_Product, (Total_Product*100/(select␣
 ↪sum(Total_Product) from Salesprod )) as Product_Percentage
FROM Salesprod1
order by Product_Percentage desc
```

So, its clear that the products that belong to Class (Office_Supplies and Furniture) with Sub_class(Papers, Furnishing, and Binders) possessed the 40% among other ones, we should focus on these in the upcoming year during the campaign.

[17]:
```
/* WHAT IS THE MAXIMUM AND MINIMUM AGE OF CUSTOMERS*/

select min(Age) as Min_Age, max(Age) as Max_Age
from Customer
```

[32]:
```
/* Which age group has more responsive during the sale.*/

select
case
when c.Age >=18 and c.Age <= 31 then 'Adult'
when c.Age >=32 and c.Age <= 45 then 'Middle Adult'
when c.Age >=46 and c.Age <= 59 then 'Middle Old'
when c.Age >=60  then 'Old'
end Age_Category,s.Rating ,count(s.Rating) as Total_Rating
from Sales s
left join Customer c
on s.Customer_ID = c.Customer_ID
```

```
group by
case
when c.Age >=18 and c.Age <= 31 then 'Adult'
when c.Age >=32 and c.Age <= 45 then 'Middle Adult'
when c.Age >=46 and c.Age <= 59 then 'Middle Old'
when c.Age >=60  then 'Old'
end , s.Rating
order by Total_Rating desc
```

So, it shows that the age group (Middle Adult) has provided the high number of Best responses among the Age Category. The adult group has also an equal occurrence of Best responses in Age Category.

[1]:
```
/* THERE ARE DIIFERENT ALLOTMENT OF BADGES TO PARTICULAR CUSTOMERS, NOW WE HAVE␣
 ↪TO CHECK THE BEGAVIOR OF CUSTOMER ON DIFFERENT BADGES*/

select  c.Category_Badges, round(sum(Price),2) as Price,sum(Ordered_Quantity)␣
 ↪as Quantity,round(sum(s.Price*s.Ordered_Quantity),2) as Total_Quantity
from Sales s
left join Customer c
on s.Customer_ID = c.Customer_ID
group by c.Category_Badges
order by Total_Quantity desc
```

It means that those holds Golden badges have bought in large quantity. But we should focus on Silver badge holder and need to check their rating to improve their experience.

[2]:
```
select  c.Category_Badges,s.Rating, count(Rating) as Customer_Rating
from Sales s
left join Customer c
on s.Customer_ID = c.Customer_ID
group by c.Category_Badges,s.Rating
order by Customer_Rating desc
```

So it concludes that the Silver Badge holder has provided a low number of ratings. The positive thing is that the Bronze badge holders showed a positive response in good number. One question can be raised here, what if the respective no. of Badged customers are not in equal quantity which became a reason for the downfall of Silver badged holders? Let's find out..

[5]:
```
select Category_Badges, count( Category_Badges) as Number_of_Badges
from Customer
GROUP BY Category_Badges
```

Now, it makes sense that the silver badge holders bought less because they were in fewer numbers as compared to other ones but we should also notice that during this campaigning span the Bronze badge holders were in good numbers and we should try

to find the reason of a low number of silver badge holders.

[19]:
```
select  c.Segment, round(sum(Price*Ordered_Quantity),2) as Total_Value
from Sales s
left join Customer c
on s.Customer_ID = c.Customer_ID
group by c.Segment
order by Total_Value desc
```

[15]:
```
/* what are the condition of different customer segment in different countries␣
 ↪in total sales*/

select  c.country,c.Segment, round(sum(Price*Ordered_Quantity),2) as Total_Value
into countryseg1
from Sales s
left join Customer c
on s.Customer_ID = c.Customer_ID
group by c.country,c.Segment
order by Total_Value desc
```

[17]:
```
select *
from countryseg1
order by Total_Value desc
```

[18]:
```
/* what are the condition of different customer segment in different countries␣
 ↪in percentage*/
select *,
case when Segment = 'Corporate' then round((Total_Value*100)/(select␣
 ↪sum(Total_Value) from countryseg1 where Segment ='Corporate'),2)
when Segment = 'Consumer' then round((Total_Value*100)/(select sum(Total_Value)␣
 ↪from countryseg1 where Segment ='Consumer'),2)
when Segment = 'Home Office' then round((Total_Value*100)/(select␣
 ↪sum(Total_Value) from countryseg1 where Segment ='Home Office'),2)
when Segment = 'Government' then round((Total_Value*100)/(select␣
 ↪sum(Total_Value) from countryseg1 where Segment ='Government'),2)
else 'Not Found'
end 'Percentage'
from countryseg1
order by Segment, Percentage desc
```

Now, we can conclude that **the consumer segment** has taken the highest share among different segments. In this segment, **Switzerland** got a significant share among 16 countries which was **8.93%**. **The corporate segment** placed at the second position with the highest share of **9.03%** by **Germany** in this segment. **Canada** got the highest number in **the Government sector** with **9.38%**. Finally, **the UK** found a good market in **the Home Office sector** with **11.98%** in this segment. Dubai is positioned at 16th place in the Consumer and Home office segment and has zero market share in the other two segments.

[52]:
```
/* lets explore the region wise sales in top segment countries*/
Select  c.country,c.Region,c.Segment, round(sum(Price*Ordered_Quantity),2) as␣
 ↪Total_Value
from Sales s
left join Customer c
on s.Customer_ID = c.Customer_ID
group by c.country,c.Region,c.Segment
order by  Segment  asc, Total_Value desc
```

[34]:
```
select  c.country,c.Region,c.Segment, round(sum(Price*Ordered_Quantity),2) as␣
 ↪Total_Value
into countryregseg
from Sales s
left join Customer c
on s.Customer_ID = c.Customer_ID
group by c.country,c.Region,c.Segment
order by Total_Value desc
```

[51]:
```
select *
from
(select x.*, DENSE_RANK() OVER(partition by x.Segment order by x.Percentage␣
 ↪desc) as Rank
from(

select *,
case when Segment = 'Corporate' then round((Total_Value*100)/(select␣
 ↪sum(Total_Value) from countryseg1 where Segment ='Corporate'),2)
when Segment = 'Consumer' then round((Total_Value*100)/(select sum(Total_Value)␣
 ↪from countryseg1 where Segment ='Consumer'),2)
when Segment = 'Home Office' then round((Total_Value*100)/(select␣
 ↪sum(Total_Value) from countryseg1 where Segment ='Home Office'),2)
when Segment = 'Government' then round((Total_Value*100)/(select␣
 ↪sum(Total_Value) from countryseg1 where Segment ='Government'),2)
else 'Not Found'
end 'Percentage'
from countryregseg) as x ) as y
where Rank in (1,2)
```

It means that in Switzerland, the East Region has the highest share among other regions which is 4.21% in the Consumer segment. In Malaysia, we have the highest sales in the Western region as we concluded earlier where Germany had the highest in the Corporate sector. Region-wise Malaysia's western region took over the sales compared to the western region of Germany, but country-wise Germany got the highest sales in which western region achieved more sales among Germany's other regions. On the other hand, in the Government sector, Canada's west region recorded appreciable sales whereas in the home office sector UK's East region carried the upper hand that is 4.63%.