SQL Final Project Report

Participant Name: Engy Mabrouk

USAID-Carerha Scholarship “Data Analysis Track”

Group 3

Under the Supervision of

Ms. Asmaa El Taher

Firstly: Customer-Oriented Analysis

A. Counting Customers and Countries

```
SELECT
count (DISTINCT customerid) NumCustomers,
count (DISTINCT country) NumCountries
from Customers
```

B. Analyzing Customers:

Customers will be analyzed by categorizing them into Three Groups Based on an RFM Analysis namely (Recency, Frequency, and Monetary Analysis)

The three customer categories are as follows:

1) Champions: Bought most Recently and most Often, and spent most
2) Potential Loyalists: Buy Most frequently or Spent Most
3) At Risk: Anyone else

First: A test will run to identify the Number of days since last order, Revenue and Number of Orders Per customer for the sole purpose of viewing the records to be able to set different intervals for each Customer Group.

Note: when testing the records, it was found out that the number of days since last order showed a minimum of about 400 days as the last order was booked in 2023. Therefore, the Number of days since last order was calculated based on the last day in 2023 to create a meaningful analysis.

```
SELECT
companyname,
round(Julianday('2023-12-31')-julianday(MAX (o.orderdate)),1) as Recency,
COUNT (o.OrderID) as Frequency,
round(SUM (unitprice*quantity*(1-discount)),0) as Revenue
from Customers c
join Orders o
on c.CustomerID = o.CustomerID
join "Order Details" AS od
```

```
on o.OrderID = od."OrderID"
group by c.companyname
```

-After viewing the records, A Minimum and Maximum of every Metric was noted

For instance:

-Recency per customers ranges from 64 days to 233 days

- Number of Orders Per Customer ranges from from 5325 to 13283

-Revenue Per Customer ranges from 3,965,465 to 9,745,371

-Accordingly, the intervals are set as follows;

1. Champions:

Recency will range from 60 days to 120

Frequency will be more than 7000 Orders

Revenue will be more than 6,000,000

2. Potential Loyalists:

Frequency will range from 5000 to 7000 Orders

Revenue ranges from 4,555,555 and 6,000,000

3. At Risk:

Any Customer that doesn't match the above criteria.

To be able to retrieve this data again, a view is created for each group to be used in a CASE WHEN Statement.

The Champions View:	**The Potential Loyalists View:**
create view Champions as SELECT companyname, round(Julianday('2023-12-31')-julianday(MAX (o.orderdate)),1) as Recency, COUNT (o.OrderID) as Frequency, round(SUM (unitprice*quantity*(1-discount)),0) as Revenue from Customers c join Orders o on c.CustomerID = o.CustomerID join "Order Details" AS od on o.OrderID = od."OrderID" group by c.companyname HAVING Recency BETWEEN 60 and 120 and Frequency > 7000 and Revenue > 6000000	CREATE view Potential_Loyalists as SELECT companyname, round(Julianday('2023-12-31')-julianday(MAX (o.orderdate)),1) as Recency, COUNT (o.OrderID) as Frequency, round(SUM (unitprice*quantity*(1-discount)),0) as Revenue from Customers c join Orders o on c.CustomerID = o.CustomerID join "Order Details" AS od on o.OrderID = od."OrderID" group by c.companyname HAVING Frequency BETWEEN 5000 and 7000 and Revenue BETWEEN 4555555 and 6000000

To segment customers into the above three groups, a case when statement was used.

```
SELECT
   c.companyname,
   CASE
      WHEN c.companyname IN
      (SELECT companyname FROM champions) THEN 'champions'
      WHEN c.companyname IN
      (SELECT companyname FROM Potential_Loyalists) THEN 'Potential_Loyalists'
      ELSE 'At_Risk'
   END AS Customer_Gp
FROM
   Customers c;
```

-The above statement categorizes all customers into the three groups; however, it doesn’t show the number of customers in each category to show identify the percentage of each.

-Accordingly, another statement will run using the “Common Table Expression CTE”

```
WITH CustomerGroups AS (
   SELECT
      c.companyname,
  case
      WHEN c.companyname IN
      (SELECT companyname FROM champions) THEN 'champions'
      WHEN c.companyname IN
      (SELECT companyname FROM Potential_Loyalists) THEN 'Potential_Loyalists'
      ELSE 'At_Risk'
   END AS Customer_Gp
FROM
   Customers c
)
SELECT
   Customer_Gp,
   COUNT(*) AS 'Count'
FROM
   CustomerGroups
GROUP BY
   Customer_Gp;
```

C. Analyzing Average Order Value for each customer

```
SELECT
 companyname,
round(AVG (TotalV),0) AvgOrder_Customer
from Customers c
join Orders o on c.CustomerID= o.CustomerID
join (
  SELECT
  orderid,
  sum (quantity*unitprice*(1-discount)) TotalV
  from "Order Details"
  group by 1) od
  on od.orderID= O.OrderID
  group by 1
  order by 2
```

The Findings of the Customer-Oriented Analysis:

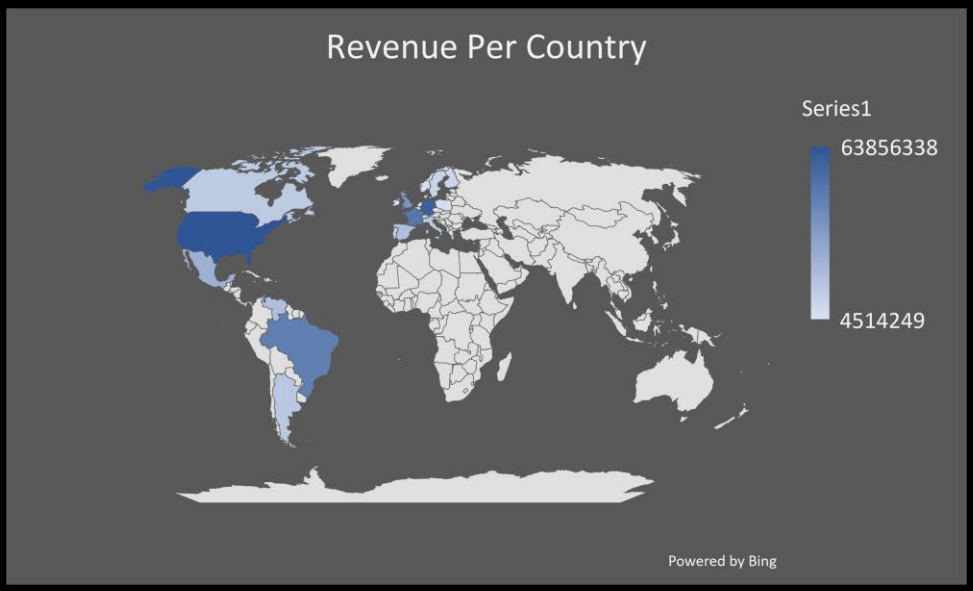
1. The Business has 93 customers in 21 countries
2. There are only three customers who are considered champions; representing 3.3% of the customers
3. There are 52 customers who are considered Potential Loyalists; representing 56.5% of the customers
4. There are 39 customers who are considered At Risk; representing 42.4% of the Customers.
5. Total Revenue Generated is 448,386,633
6. Top 5 Revenue-Generating Countries are
   USA, Germany, France, Brazil, UK
7. Least 5 Revenue-Generating Countries are
   Finland, Belgium, Poland, Norway, Ireland
8. Average Order Value Per Customer ranges from 23790 to 30449 Order

Insight:

More than half of the customers are considered Potential Loyalists, which is great. However, there's a considerable percentage of customers who are at Risk.

Recommendation:

Improving customer retention is a key player here since the "At Risk" percentage is not small. Accordingly, it's either we can win these customers or lose them. Since we have the phone numbers of these Customers, SMS messages could be sent to them offering interesting vouchers to encourage them to re-purchase new items. This could improve Sales and Brand Awareness as well.



Secondly: Product-Oriented Analysis

Identifying the top 10 revenue-generating Products, and the top 10 frequently ordered Products were generated in one query that could run depending on the “Order By” Statement.

In other words, to answer the first question, the query will be ordered by the product revenue as shown below. To answer the second question, the query will be ordered by the Quantity Ordered.

```
SELECT
od.productid, productname,
round(Sum (od.unitprice*od.quantity* (1-discount)),1) ProductRevenue,
SUM(od.quantity) AS QuantityOrdered
from "Order Details" od
JOIN Products as p
ON od.ProductID= p.ProductID
group by 1
order by 4 DESC
LIMIT 10
```

Identifying the least 5 ordered Products was generated in a separate query using a Left Join in order to find out if a product hasn’t generated any sales. To avoid any nulls in that case, The COALESCE Function was also used to return zero values for these products.

```
SELECT
od.ProductID,
p.ProductName,
COALESCE(SUM(od.Quantity), 0) AS TotalQuantityOrdered
FROM
Products p
LEFT JOIN
"Order Details" od
```

```
ON p.ProductID = od.ProductID

GROUP BY

p.ProductID, p.ProductName

ORDER BY

TotalQuantityOrdered ASC

LIMIT 5
```

The Findings of the Product-Oriented Analysis:

1. The business offers 77 products coming from 29 different suppliers
2. Max Number of Products for each supplier reaches 5 and the Least is 2
3. The Top 10 Revenue Generating Products are
   Chartreuse verte
   Original Frankfurter grune Sobe
   Pavlova
   Longlife Tofu
   Sir Rodney's Marmalade
   Uncle Bob's Organic Dried Pears
   Manjimup Dried Apples
   Camembert Pierrot
   Schoggi Schokolade
   Louisiana Fiery Hot Pepper Sauce

4. Top 5 Suppliers are
   Aux Joueaux
   Plutzer
   Pavlova, Ltd.
   Tokyo Traders
   Speciality Biscuits, Ltd.

5. Top 5 Frequently ordered Products
   Pavlova
   Original Frankfurter
   Sir Rodney’s marmalade
   Louisiana Hot Pepper
   Steeleye stout

Insight:

The Top Revenue Generating Products are almost the same as the Most Ordered Products, despite few Products like Charteuese Verte, which is the highest in generating Revenue, Longlife Tofu, and Camembret Pierrot.

Recommendations:

1. It's preferable to lay focus on Products like Charteuese Verte and Tofu that generated the highest revenue despite the fact that they weren't ordered in huge quantities.

2. Since the major goal is to increase income, it's recommended to raise investments in the top revenue-generating products by always ensuring the availability of enough stock.
3. Another suggestion is to expand the Business Collaboration with the suppliers of these products (who are mentioned above) to introduce new products to our customers who will likely show interest

Thirdly: Orders-Oriented Analysis

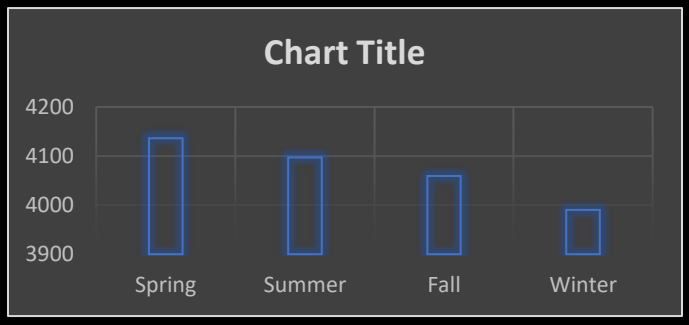
On Analyzing Seasonal Trends, a seasonal Analysis was conducted using a CASE WHEN Statement	Another Periodical analysis was conducted to test Any fluctuations on the days of the week and the distribution of the quantities ordered.	Another Query was introduced to identify the Most Popular days in the Month.
SELECT case when STRFTIME ('%m', orderdate) in ('03','04','05') then 'Spring' when STRFTIME ('%m', orderdate) in ('06','07','08') then 'Summer' when STRFTIME ('%m', orderdate) in ('09','10','11') THEN 'Fall' eLSE 'Winter' END Season, COUNT (orderid) NumOrders from Orders GROUP by Season order by 2 DESC	SELECT case when strftime ('%w', orderdate) = '0' then 'Sunday' when strftime ('%w', orderdate) = '1' then 'Monday' when strftime ('%w', orderdate) = '2' then 'Tuesday' when strftime ('%w', orderdate) = '3' then 'Wednesday' when strftime ('%w', orderdate) = '4' then 'Thursday' when strftime ('%w', orderdate) = '5' then 'Friday' when strftime ('%w', orderdate) = '6' then 'Saturday' END AS Day, COUNT (o.orderid) NumofOrders, SUM (od.quantity) TotalQuantity from Orders o join "Order Details" od on o.OrderID = od.orderID group by 1 order by 3 DESC	SELECT strftime ('%d',orderdate) DayNumber, COUNT (orderid) NumOrders from orders group by 1 order by 2 DESC

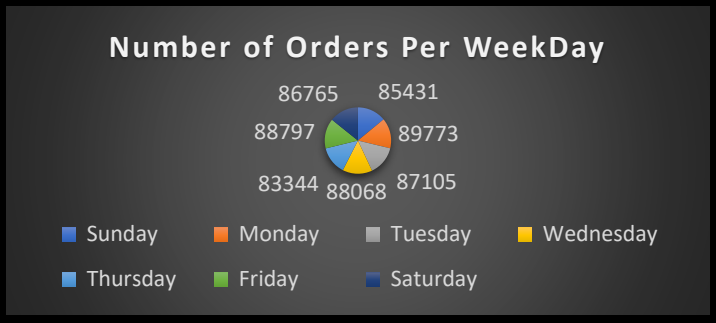
The Order-Oriented Analysis Findings:

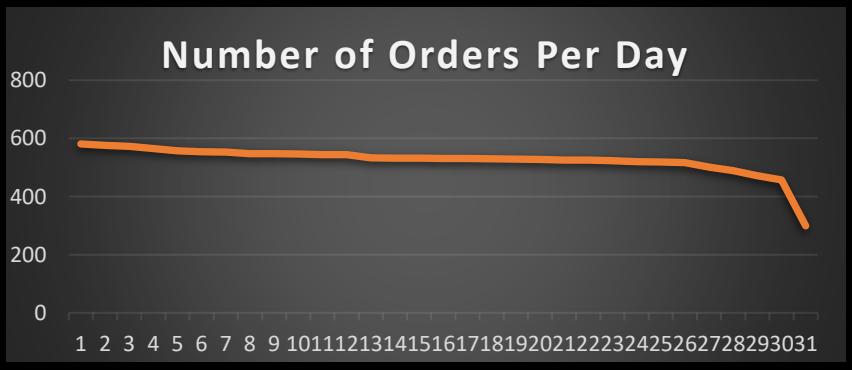
1. The Business Generated the highest Revenue in Spring and the lowest in Winter
2. Monday has shown the highest record when it comes to both the Quantity Ordered and the revenue generated as well
3. In contrast, Thursday has shown the least record when it comes to both the Quantity Ordered and the revenue generated as well.
4. The first day of the month always shows the most popularity in terms of the Number of orders.
5. In contrast the last three days of the month always show a decline in the Number of Orders

Recommendations:

1. To Promote sales on the unpopular days like the end of the month or Thursday, vouchers could run via the website to encourage customers to purchase. Gestures like free shipping or a 10% discount can let customers, especially the ones AT Risk, start their shopping process.
2. Since Winter shows the lowest Season, this in fact may be an opportunity for the business to start offering winter products like Blankets and heaters with reasonable prices and Profit Margin Ratio to test their profitability at first and gain popularity as well.







Fourthly: Employee-Oriented Analysis:

Evaluating the employee performance based on the revenue they generated and the number of orders processed was identified through one query.	Another statement was also used to identify the Average order value per employee	Analyzing Employee Gender
SELECT e.lastname, round(Sum (od.unitprice*od.quantity* (1-od.discount)),0) RevenuePerEmployee, COUNT (o.orderid) NumOrders from Employees e left join Orders o on e.EmployeeID= o.EmployeeID join "Order Details" od on o.OrderID = od.orderID GROUP by 1 order by 2 DESC	SELECT e.LastName, round(AVG(order_total),0) AS AverageOrderValue FROM Employees e JOIN Orders o ON e.EmployeeID = o.EmployeeID JOIN (SELECT OrderID, SUM(UnitPrice * Quantity) AS order_total FROM "Order Details" GROUP BY OrderID) od ON o.OrderID = od.OrderID GROUP BY e.LastName ORDER BY AverageOrderValue DESC;	SELECT Gender, COUNT(Gender) from (SELECT titleofcourtesy, case when titleofcourtesy in ('Mrs.','Ms.') THEN 'Female' when titleofcourtesy = 'Mr.' then 'Male' else 'NA' end 'Gender' from Employees) group by 1

Analyzing Years of Experience	Analyzing Age
SELECT Min(hiredate), max (hiredate) from Employees	Select AVG(Age) from (SELECT date('2023-12-31') - date(birthdate) Age from Employees)

The Findings of the Employee-Oriented Analysis:

1. The Employee Top Performer is Peacock as he generated 51488395
2. The top 10 Employee Performers have each contributed to about 11 % to the total revenue
3. More than 50% of the Employees are Females
4. The Employees years of experience range from 9 to 11 Years
5. The Average Age of Employees reach 47 Years

Recommendations:

1. A ceremony could be held as an event showing appreciation for the Top 10 Performers. This also works as a motivation for the least performers to put more effort and passion into the Business ensuring that it will pay off at the end.

2. Hiring younger Employees may result in an enhanced innovation and productivity for the business even if they require less than 9 years of experience, especially for positions like Marketing that incredibly contribute to the Brand Awareness and subsequently the overall sales

THANK YOU!

E N G Y M A B R O U K