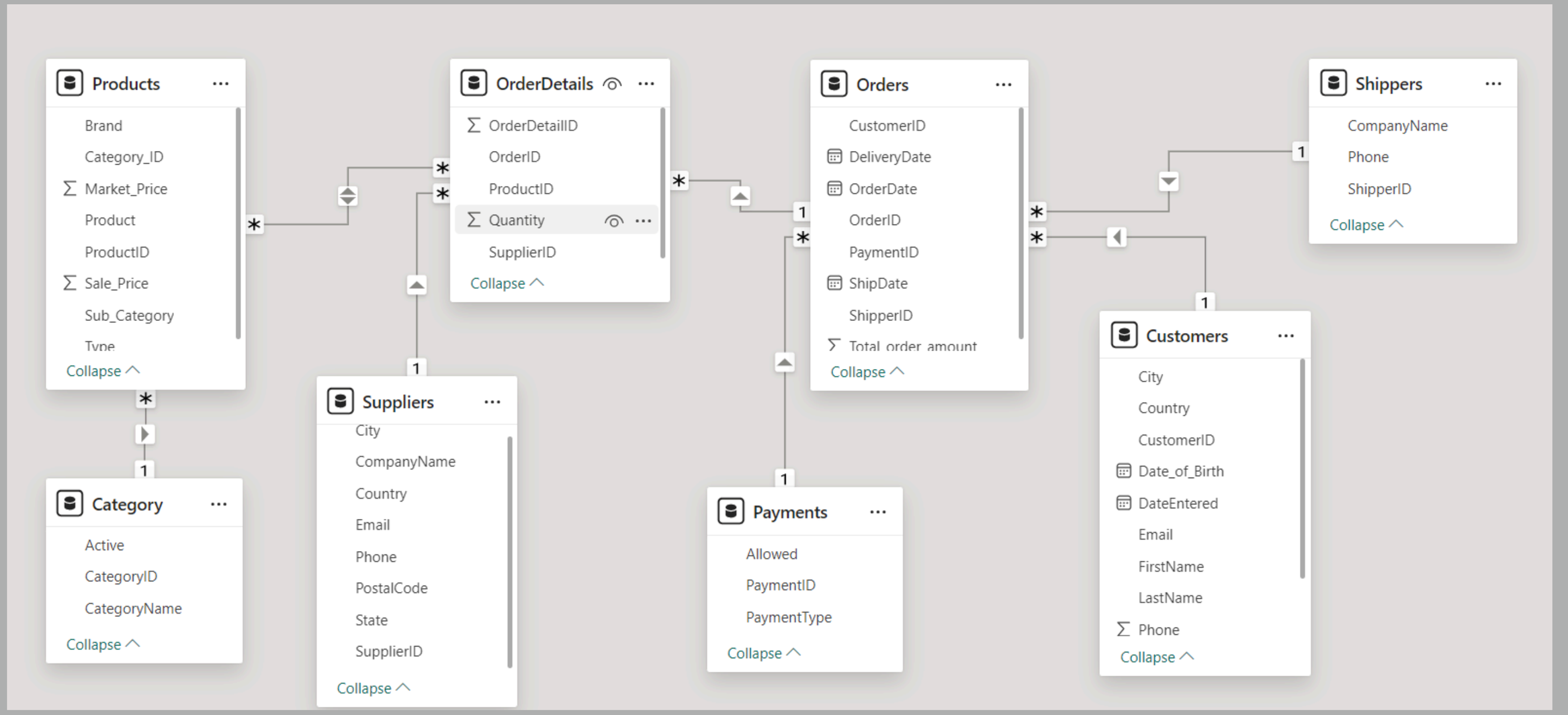
SQL Project
on
E-Commerce Sales

ABOUT THE PROJECT

In this project, extensive analysis of an e-commerce company's sales data was conducted using SQL queries. The data was summarized to reveal overall revenue, order details, customer demographics, and payment method preferences. Key insights were derived by identifying top-spending customers, most popular product categories, and daily top-selling products. The analysis also pinpointed underperforming regions and customer retention issues. Various SQL functions, including aggregate, string, date, join, and conditional functions, were employed to extract and interpret the data, providing valuable information for optimizing marketing strategies, improving customer retention, and enhancing operational efficiency.

DATABASE SCHEMA



OVERALL REVENUE MADE BY THE E-COMMERCE COMPANY

```sql
SELECT
    ROUND(SUM(TOTAL_ORDER_AMOUNT) / 1000000 * 2, 2) AS REVENUE_IN_MILLION
FROM
    ORDERS;
```

	REVENUE_IN_MILLION
▶	183.84

This query calculates the total revenue of an e-commerce company by summing the TOTAL_ORDER_AMOUNT from the ORDERS table and converting the amount into millions. The rounded result is displayed as REVENUE_IN_MILLION, which amounts to 183.84 million.

ORDER DETAILS BETWEEN 10000 AND 20000

```sql
SELECT
    *
FROM
    ORDERS
WHERE
    TOTAL_ORDER_AMOUNT BETWEEN 10000 AND 20000
ORDER BY ORDERID ASC;
```

OrderID	CustomerID	PaymentID	ORDERDATE	ShipperID	ShipDate	DeliveryDate	Total_order_amount
7655502	57086	4	2020-02-06	7	2020-02-11	2020-02-21	13293
7655503	57088	4	2020-02-09	1	2020-02-13	2020-02-26	16063
7655504	57090	4	2020-02-11	3	2020-02-15	2020-02-20	15193
7655505	57094	4	2020-02-15	1	2020-02-24	2020-02-27	13581
7655507	57095	6	2020-02-23	8	2020-03-04	2020-03-11	12092
7655510	57103	3	2020-03-02	7	2020-03-07	2020-03-14	11803
7655514	57096	5	2020-03-07	7	2020-03-14	2020-03-22	10991
7655520	57083	5	2020-03-18	1	2020-03-19	2020-04-06	14685
7655524	57120	6	2020-03-21	7	2020-03-30	2020-04-15	15089
7655525	57105	4	2020-03-27	5	2020-04-01	2020-04-07	12892
7655526	57087	4	2020-03-27	5	2020-03-31	2020-04-20	13788
7655528	57127	4	2020-03-31	6	2020-04-01	2020-04-09	14006
7655530	57091	5	2020-04-02	1	2020-04-12	2020-04-15	12918
7655538	57107	4	2020-04-10	3	2020-04-14	2020-05-03	15779
7655539	57113	4	2020-04-10	5	2020-04-20	2020-05-08	11651
7655540	57108	4	2020-04-12	1	2020-04-17	2020-04-25	14498
7655542	57090	3	2020-04-15	4	2020-04-22	2020-05-05	14005
7655544	57123	5	2020-04-17	2	2020-04-23	2020-05-09	13216

The query retrieves details from the ORDERS table where the TOTAL_ORDER_AMOUNT is between 10,000 and 20,000, ordered by ORDERID in ascending order. The resulting table shows multiple orders meeting this criterion, including various order dates and total amounts.

NUMBER OF CUSTOMERS BELONGING TO THE CITY OF EVERY COUNTRY WHERE CUSTOMERS ARE LESS THAN 3

```
SELECT
    COUNT(DISTINCT (CUSTOMERID)) AS CUST_NO, COUNTRY, CITY
FROM
    CUSTOMERS
GROUP BY COUNTRY , CITY
HAVING CUST_NO < 3
ORDER BY CUST_NO DESC;
```

CUST_NO	COUNTRY	CITY
2	Australia	Hobart
2	France	Montpellier
2	Australia	Cairns
2	France	Brest
2	Belgium	Namur
2	France	Lille
2	Australia	Townsville
2	India	Amritsar
2	India	Bhopal
2	Portugal	Amadora
2	Russia	Novosibirsk
1	Australia	Darwin
1	Australia	Geelong
1	Australia	Newcastle
1	Australia	Perth
1	Australia	Sunshine Coast
1	Australia	Sydney
1	Australia	Wollongong

The query counts the distinct CUSTOMERID in the CUSTOMERS table, grouped by COUNTRY and CITY, and filters for cities with fewer than three customers. The result is a table listing such cities, ordered by the customer count in descending order, showing cities from various countries with 1 or 2 customers.

DETAILS OF THE CUSTOMERS WHERE FIRST_NAME STARTS WITH A VOWEL

```
SELECT
    *
FROM
    customers
WHERE
    firstname REGEXP '^[AEIOUaeiou]'
ORDER BY customerid ASC;
```

CustomerID	FirstName	LastName	Date_of_Birth	City	State	Country	PostalCode	Phone	Email	DateEntered
57094	Anthony	James	1957-04-12	Brussels	Brussels-Capital	Belgium	322713	9371469569	Anthony.James@gmail.com	2020-02-07
57099	Andrew	Martinez	1998-03-04	Bruges	West Flanders	Belgium	717216	9289999420	Andrew.Martinez@gmail.com	2020-02-11
57105	Edward	Gonzalez	1972-07-08	Patras	Western Greece	Greece	588452	9234402314	Edward.Gonzalez@gmail.com	2020-02-20
57114	Eric	Rosen	1962-12-20	Amsterdam	North Holland	Netherlands	304934	7552732635	Eric.Rosen@gmail.com	2020-03-01
57124	Anne	Frank	1982-05-09	Belfast	Belfast	Northern Ireland	765669	6694563730	Anne.Frank@gmail.com	2020-03-08
57125	Alexander	Thompson	1983-11-30	Patras	Western Greece	Greece	506559	7083288108	Alexander.Thompson@gmail.com	2020-03-09
57132	Aaron	Paul	1976-10-13	Namur	Namur	Belgium	322042	8668707337	Aaron.Paul@gmail.com	2020-03-17
57134	Adam	Levin	1994-07-06	San Diego	California	United States	529879	7460297856	Adam.Levin@gmail.com	2020-03-19
57142	Ethan	Hunt	1961-06-18	Zurich	Zurich	Switzerland	935721	8581240635	Ethan.Hunt@gmail.com	2020-03-27
57153	Austin	Emerson	1997-11-23	Los Angeles	California	United States	323529	7849411743	Austin.Emerson@gmail.com	2020-04-10
57154	Arthur	Torres	1957-04-30	New York	New York	United States	603298	7244987442	Arthur.Torres@gmail.com	2020-04-10
57163	Albert	Black	1975-03-12	Houston	Texas	United States	531839	8269461310	Albert.Black@gmail.com	2020-04-16
57167	Alan	Walker	1999-05-01	Houston	Texas	United States	425321	7519015025	Alan.Walker@gmail.com	2020-04-20
57173	Eugene	White	1990-11-27	San Francisco	California	United States	333520	6435013361	Eugene.White@gmail.com	2020-04-27
57176	Elijah	Mitchell	1971-06-29	Novosibirsk	Novosibirsk Obl...	Russia	369738	7286676080	Elijah.Mitchell@gmail.com	2020-05-04
57185	Elizabeth	Emerson	1965-04-01	Saint Peter...	Saint Petersburg	Russia	120668	6751710453	Elizabeth.Emerson@gmail.com	2020-05-13
57196	Ashley	White	1998-02-10	New York	New York	United States	803691	9837229493	Ashley.White@gmail.com	2020-06-03
57198	Emily	Young	1987-09-24	Houston	Texas	United States	949975	8408622873	Emily.Young@gmail.com	2020-06-04

This segmentation of customers based on first names starting with vowels reveals unique marketing opportunities. Such a demographic split could be used for personalized marketing campaigns, potentially increasing engagement and conversion rates. This niche focus might also provide insights into specific customer behaviors and preferences.

NUMBER OF ORDERS PLACED THROUGH EACH PAYMENT METHOD

```sql
SELECT
    COUNT(DISTINCT (O.ORDERID)) AS 'NO_OF_ORDERS',
    P.PAYMENTID,
    P.PAYMENTTYPE
FROM
    ORDERS O
        INNER JOIN
    payments P ON P.PAYMENTID = O.PaymentID
GROUP BY P.PAYMENTID , P.PAYMENTTYPE
ORDER BY P.PAYMENTID;
```

NO_OF_ORDERS	PAYMENTID	PAYMENTTYPE
50	1	Debit Card
100	2	POD
1000	3	PayPal
2500	4	Credit Card
700	5	Wallet
649	6	Net banking

The data shows the distribution of orders by payment methods, highlighting customer preferences. Understanding which payment methods are most popular can help streamline payment processes and enhance customer experience. It also aids in identifying trends and potential areas for improvement in payment options.

TOP 10 MOST EXPENSIVE ORDERS BY CUSTOMERS

```sql
SELECT
    C.CUSTOMERID,
    C.FIRSTNAME,
    C.LASTNAME,
    MAX(O.TOTAL_ORDER_AMOUNT) AS AMOUNT
FROM
    ORDERS O
        JOIN
    CUSTOMERS C ON C.CUSTOMERID = O.CUSTOMERID
GROUP BY C.CUSTOMERID , C.FIRSTNAME , C.LASTNAME
ORDER BY AMOUNT DESC
LIMIT 10;
```

CUSTOMERID	FIRSTNAME	LASTNAME	AMOUNT
57212	Amy	Kirsten	156724
57249	Jacqueline	Fernandez	152253
57574	Richard	Corey	149094
57427	Mohammad	Ciaran	146727
57451	Ciaran	Arthur	140382
57115	Jonathan	Moore	138997
57538	Duncan	Elliot	128214
57356	Cooper	Martin	128138
57114	Eric	Rosen	125321
57289	Alfie	Vincent	120979

This slide details the highest-value orders, showcasing key high-spending customers. These orders often come from bulk or luxury purchases, highlighting the importance of catering to premium segments. Focusing on these customers can drive revenue growth and foster loyalty through personalized offers and exclusive deals.

ALL THE DETAILS OF THE CUSTOMER WHO ORDERED ONLY ONCE

```
WITH CUST_COUNT AS
(SELECT CUSTOMERID, COUNT(CUSTOMERID) AS 'COUNT_OF_ORDERS'
FROM ORDERS
GROUP BY CUSTOMERID
HAVING COUNT(CUSTOMERID) = 1)
SELECT C.* FROM CUSTOMERS C
INNER JOIN CUST_COUNT
ON CUST_COUNT.CUSTOMERID=C.CUSTOMERID;
```

CustomerID	FirstName	LastName	Date_of_Birth	City	State	Country	PostalCode	Phone	Email	DateEntered
57165	Gabriel	Thomas	1995-06-10	Saint Petersburg	Saint Petersburg	Russia	501708	7079545540	Gabriel.Thomas@gmail.com	2020-04-18

Identifying single-order customers is crucial for understanding retention challenges. This data helps in devising strategies to convert one-time buyers into repeat customers, such as loyalty programs or follow-up marketing campaigns. Addressing the reasons for low retention can significantly improve customer lifetime value.

THE TOP 3 COUNTRIES WHO HAVE PLACED THE LEAST NUMBER OF THE ORDERS

```
SELECT
    CUSTOMERS.COUNTRY,
    COUNT(DISTINCT (ORDERS.ORDERID)) AS ORDER_COUNT
FROM
    ORDERS
        JOIN
    CUSTOMERS ON CUSTOMERS.CUSTOMERID = ORDERS.CUSTOMERID
GROUP BY CUSTOMERS.COUNTRY
ORDER BY ORDER_COUNT ASC
LIMIT 3;
```

COUNTRY	ORDER_COUNT
South Africa	47
New Zealand	60
Italy	121

Highlighting the countries with the fewest orders provides insight into weaker markets. This information is critical for strategic planning, allowing the company to reassess its approach in these regions. It might indicate a need for localized marketing efforts or adjustments in product offerings to better meet local demand.

THE COUNTRY AND NUMBER OF THE ORDERS

```sql
SELECT
    CUSTOMERS.COUNTRY,
    COUNT(DISTINCT (ORDERS.ORDERID)) AS ORDER_COUNT
FROM
    ORDERS
        JOIN
    CUSTOMERS ON CUSTOMERS.CUSTOMERID = ORDERS.CUSTOMERID
GROUP BY CUSTOMERS.COUNTRY
ORDER BY ORDER_COUNT DESC;
```

COUNTRY	ORDER_COUNT
United States	634
India	425
Portugal	344
Austria	304
Greece	304
France	301
Australia	297
Switzerland	294
Northern Ireland	277
Ireland	259
Germany	251
Belgium	239
Russia	235
Netherlands	233
Romania	207
Poland	167
Italy	121

Breaking down orders by country reveals geographical performance, showing which regions are thriving and which are lagging. This data guides market-specific strategies and resource allocation, ensuring efforts are focused on areas with the highest growth potential. It also helps in understanding cultural and regional preferences.

TOTAL_ORDER_AMOUNT, TOTAL REVENUE, TOTAL PROFIT AND PROFIT MARGIN FOR EACH ORDER ID

```sql
SELECT
    O.ORDERID,
    O.TOTAL_ORDER_AMOUNT,
    SUM(O.TOTAL_ORDER_AMOUNT * OD.Quantity) AS REVENUE,
    (SUM(O.TOTAL_ORDER_AMOUNT * OD.Quantity) - O.TOTAL_ORDER_AMOUNT) AS PROFIT,
    ROUND((SUM(O.TOTAL_ORDER_AMOUNT * OD.Quantity) - O.TOTAL_ORDER_AMOUNT) / SUM(O.TOTAL_ORDER_AMOUNT * OD.Quantity) * 100,
            2) AS PROFIT_MARGIN
FROM
    ORDERS O
        JOIN
    ORDERDETAILS OD ON O.ORDERID = OD.ORDERID
GROUP BY O.ORDERID , O.TOTAL_ORDER_AMOUNT
ORDER BY REVENUE DESC;
```

ORDERID	TOTAL_ORDER_AMOUNT	REVENUE	PROFIT	PROFIT_MARGIN
7656460	152253	23142456	22990203	99.34
7660370	149094	20425878	20276784	99.27
7657624	109499	18176834	18067335	99.40
7656345	156724	17082916	16926192	99.08
7660117	128214	16667820	16539606	99.23
7658555	115128	15427152	15312024	99.25
7656373	125321	14537236	14411915	99.14
7659851	95039	12735226	12640187	99.25
7656803	100679	12584875	12484196	99.20
7658882	138997	12231736	12092739	98.86
7658436	140382	12072852	11932470	98.84
7656811	88539	12041304	11952765	99.26
7659676	146727	11591433	11444706	98.73
7657687	97053	11549307	11452254	99.16
7658773	128138	11532420	11404282	98.89
7656127	120979	11251047	11130068	98.92
7657674	83479	11019228	10935749	99.24
7657056	87526	10678172	10590646	99.18

Detailed financial metrics for each order, including total amount, revenue, profit, and profit margin, offer insights into operational efficiency and profitability. This information is vital for identifying high-margin products and optimizing pricing strategies. It also helps in making informed decisions to enhance overall profitability.

DETAILS OF ALL DIFFERENT PAYMENT METHODS ALONG WITH THE TOTAL AMOUNT OF MONEY TRANSACTED IN YEAR 2020 AND 2021

```sql
SELECT
    (P.PAYMENTTYPE) AS PAYMENT_TYPE,
    P.ALLOWED,
    SUM(CASE
        WHEN YEAR(O.ORDERDATE) = 2021 THEN O.TOTAL_ORDER_AMOUNT
    END) AS '2021 TRANSACTION',
    SUM(CASE
        WHEN YEAR(O.ORDERDATE) = 2020 THEN O.TOTAL_ORDER_AMOUNT
    END) AS '2020 TRANSACTION'
FROM
    PAYMENTS P
        JOIN
    ORDERS O ON P.PAYMENTID = O.PAYMENTID
GROUP BY P.PaymentType , P.Allowed
ORDER BY P.PaymentType;
```

PAYMENT_TYPE	ALLOWED	2021 TRANSACTION	2020 TRANSACTION
Credit Card	Yes	40307060	6540922
Debit Card	Yes	849200	94478
Net banking	Yes	9625900	2148623
PayPal	Yes	15789936	2326674
POD	Yes	1526094	306323
Wallet	Yes	10593886	1810081

Summarizing total transactions by payment methods over two years shows trends in consumer behavior and financial performance. It highlights shifts in payment preferences and helps in forecasting future trends. Understanding these trends is crucial for adapting to changing consumer needs and improving transaction processes.

PRODUCT_ID, QUANTITY AND ORDER DATE FOR DAILY TOP 3 SELLING PRODUCTS BETWEEN JAN 2020 TO MARCH 2020

```sql
WITH PRODUCT_ORDERDATE AS (
SELECT PRODUCTID, SUM(QUANTITY) AS 'TOTAL_Q', ORDERDATE
FROM Orders O
JOIN OrderDetails OD
ON O.OrderID = OD.OrderID
WHERE ORDERDATE BETWEEN '2020-01-01' AND '2020-03-31'
GROUP BY ProductID, OrderDate)
SELECT PRODUCTID, TOTAL_Q, ORDERDATE FROM
(SELECT PRODUCTID, TOTAL_Q, ORDERDATE, DENSE_RANK() OVER (PARTITION BY ORDERDATE ORDER BY TOTAL_Q DESC) AS 'RNK' FROM PRODUCT_ORDERDATE) AS X
WHERE RNK <= 3
ORDER BY ORDERDATE, TOTAL_Q;
```

	PRODUCTID	TOTAL_Q	ORDERDATE
▶	8703	14	2020-01-12
	9268	17	2020-01-12
	6299	18	2020-01-12
	6232	13	2020-01-20
	17335	14	2020-01-20
	13301	19	2020-01-20
	11536	11	2020-02-06
	20138	16	2020-02-06
	12452	19	2020-02-06
	12236	16	2020-02-09
	6315	16	2020-02-09
	17363	16	2020-02-09
	13271	17	2020-02-09
	3971	18	2020-02-09
	10618	16	2020-02-11
	16783	16	2020-02-11
	10768	19	2020-02-11
	8190	19	2020-02-11

Listing the top-selling products for a specific period reveals insights into product popularity and seasonal trends. This data helps in inventory management and planning marketing campaigns around high-demand products. It also provides clues about consumer preferences and potential areas for product development.

THE PRODUCTS WHOSE NAMES CONSIST OF THE WORD 'BABY'. THEN COUNT THE NUMBER FOR EACH CATEGORY AND SUB CATEGORY.

```sql
SELECT
    C.CATEGORYID,
    C.CATEGORYNAME,
    PRODUCTS.SUB_CATEGORY,
    COUNT(C.CATEGORYID) AS PRODUCT_COUNT
FROM
    CATEGORY C
        LEFT JOIN
    PRODUCTS ON C.CategoryID = PRODUCTS.Category_ID
WHERE
    C.CATEGORYNAME LIKE 'BABY%'
GROUP BY C.CategoryID , C.CategoryName , PRODUCTS.Sub_Category
ORDER BY C.CategoryID;
```

CATEGORYID	CATEGORYNAME	SUB_CATEGORY	PRODUCT_COUNT
5005	Baby Care	NULL	1

Analyzing products with "Baby" in their names and categorizing them helps understand the demand for baby-related items. This segmentation can guide targeted marketing and product development efforts. It also shows the company's strength in specific niches, allowing for strategic expansion in similar product lines.

TOP 5 CUSTOMERS WITH HIGHEST SPENDING

```
WITH CUSTOMER_SPENDING AS (
SELECT
C.FIRSTNAME,
C.LASTNAME,
SUM(O.TOTAL_ORDER_AMOUNT) AS TOTAL_SPENDING
FROM CUSTOMERS C JOIN ORDERS O
ON C.CUSTOMERID = O.CUSTOMERID
GROUP BY
C.FIRSTNAME,
C.LASTNAME
)
SELECT
FIRSTNAME,
LASTNAME,
TOTAL_SPENDING
FROM
CUSTOMER_SPENDING
ORDER BY TOTAL_SPENDING DESC
LIMIT 5;
```

FIRSTNAME	LASTNAME	TOTAL_SPENDING
Jacqueline	Fernandez	442545
Toby	Richard	421414
Richard	Corey	408974
Caelan	Lochlan	405883
Cruz	Duncan	401145

Identifying the highest-spending customers highlights key clientele for personalized marketing and loyalty programs. These customers are critical for revenue growth, and understanding their preferences can help in tailoring offers to retain and further engage them. Focusing on high-value customers enhances profitability and customer loyalty.

MOST POPULAR PRODUCT CATEGORY IN TERMS OF THE NUMBER OF ORDERS PLACED.

```sql
WITH PRODUCTSORDER AS (
    SELECT
        P.CATEGORY_ID,
        COUNT(OD.OrderID) AS ORDERCOUNT
    FROM
        PRODUCTS P
        JOIN ORDERDETAILS OD ON P.PRODUCTID = OD.PRODUCTID
    GROUP BY
        P.CATEGORY_ID
)
SELECT
    C.CATEGORYNAME,
    PO.ORDERCOUNT
FROM
    PRODUCTSORDER PO
    JOIN CATEGORY C ON PO.CATEGORY_ID = C.CATEGORYID
ORDER BY
    PO.ORDERCOUNT DESC
LIMIT 1;
```

CATEGORYNAME	ORDERCOUNT
Cleaning & Household	96

Determining the most popular product category by the number of orders provides insights into consumer preferences and market trends. This information helps in optimizing product offerings and marketing strategies. It also indicates areas of strength where the company can invest more resources to maximize sales.

CATEGORY WISE TOTAL SALES

```sql
WITH CATEGORYSALES AS (
    SELECT
        P.CATEGORY_ID,
        SUM(O.TOTAL_ORDER_AMOUNT) AS TOTALSALES
    FROM
        PRODUCTS P
        JOIN ORDERDETAILS OD ON P.PRODUCTID = OD.PRODUCTID
        JOIN ORDERS O ON OD.ORDERID = O.ORDERID
    GROUP BY
        P.CATEGORY_ID
)
SELECT
    C.CATEGORYNAME,
    CS.TOTALSALES
FROM
    CATEGORYSALES CS
    JOIN CATEGORY C ON CS.CATEGORY_ID = C.CATEGORYID
ORDER BY
    CS.TOTALSALES DESC;
```

CATEGORYNAME	TOTALSALES
Cleaning & Household	2071401

Breaking down sales by category shows performance across different product lines, helping in understanding which categories are driving revenue. This data guides strategic decisions on inventory management, marketing, and product development. Focusing on high-performing categories can enhance overall business growth.

NUMBER OF ORDERS AND TOTAL SALES AMOUNT BY EACH CUSTOMER

```sql
SELECT CUSTOMERID, COUNT(ORDERID) AS ORDER_NUM, SUM(TOTAL_ORDER_AMOUNT) AS 'TOTAL_SALES_AMOUNT'
FROM ORDERS
GROUP BY CUSTOMERID
ORDER BY TOTAL_SALES_AMOUNT DESC;
```

CUSTOMERID	ORDER_NUM	TOTAL_SALES_AMOUNT
57249	12	442545
57495	15	421414
57574	14	408974
57455	17	405883
57486	17	401145
57232	13	393376
57213	17	389207
57200	15	388967
57339	14	384503
57102	14	376974
57276	14	375425
57363	16	370318
57114	10	366274
57396	16	363229
57513	16	357263
57373	11	355680
57592	11	355410

Detailing sales data by customer helps in personalizing marketing efforts and improving customer relationship management. Understanding each customer's order history and sales contribution aids in developing targeted strategies to increase their lifetime value. This approach can lead to higher customer satisfaction and loyalty.

THANK YOU