zomato

online delivery

Zomato

Data Analysis Using SQL

By Saransh Sharma

Objective :

The objective of this project is to analyze the Zomato dataset to gain insights into user behavior, sales trends, and product popularity. By answering these questions, we can identify patterns and trends that can inform business decisions and improve the overall user experience.

zomato

Schema

goldusers_signup

userid	int
gold_signup_date	date

sales

userid	int
created_date	date
product_id	int

users

userid	int
singup_date	date

product

product_id	int
product_name	text
price	int

All Questions

1. What is the total amount each customer spent on zomato?

2. How many days has each customer visited zomato?

3. What was the first product purchased by each customer?

4. What is the most purchased item on the menu and how many times was purchased by all customers?

5. Which item was the most popular for each customer?

6. Which item was purchased first by the customer after they became a member?

7. Which item was purchased just before the customer became a member?

All Questions

8. What is the total orders and amount spent for each member before they become a member?

9. If buying each product generates points for eg. 5rs = 2 zomoto points and each product has different purchase points for eg. for p1 5rs = 1 zomato point, for p2 10rs = 5 zomato points and p3 5rs = 1 zomato point. Calculate points collected by each customer and for which product most points have been given till now?
(2rs = 1 zomato points)

10. If the first one year after a customer joins the gold program (including their join date) irrespective of what the customer has purchased they earn 5 zomato points for every 10rs spent who earned more 1 or 3 and what was their points earnings in their first year ?
(1 Zomato points = 2RS) = (0.5 points = 1rs)

11. Rank all the transactions of the customers

1. What is the total amount each customer spent on zomato?

```sql
SELECT s.userid, SUM(p.price) as total_amt_spent
FROM sales as s
JOIN product as p
ON s.product_id = p.product_id
GROUP BY userid
ORDER BY userid ASC;
```

	userid integer	total_amt_spent bigint
1	1	5230
2	2	2510
3	3	4570

2. How many days has each customer visited zomato?

```sql
SELECT userid, COUNT(DISTINCT created_date) as distinct_date
FROM sales
GROUP BY userid;
```

	userid integer	distinct_date bigint
1	1	7
2	2	4
3	3	5

3. What was the first product purchased by each customer?

```
SELECT * FROM
(SELECT *, RANK() OVER(PARTITION BY userid ORDER BY created_date) as rnk FROM sales)
WHERE rnk = 1;
```

	userid integer	created_date date	product_id integer	rnk bigint
1	1	2016-03-11	1	1
2	2	2017-09-24	1	1
3	3	2016-11-10	1	1

4. What is the most purchased item on the menu and how many times was purchased by all customers?

```
SELECT userid, COUNT(product_id) as cnt FROM sales
WHERE product_id =
                  (SELECT product_id FROM sales
                   GROUP BY product_id
                   ORDER BY count(product_id) DESC
                   LIMIT 1)
GROUP BY userid;
```

	userid integer	cnt bigint
1	1	3
2	2	1
3	3	3

5. Which item was the most popular for each customer?

```
SELECT * FROM
(SELECT *, RANK() OVER(PARTITION BY userid ORDER BY product_id) as rnk FROM
(SELECT userid, product_id, COUNT(product_id) as cnt FROM sales GROUP BY userid, product_id) AS a) as b
WHERE rnk = 1;
```

	userid integer	product_id integer	cnt bigint	rnk bigint
1	1	1	2	1
2	2	1	1	1
3	3	1	2	1

6. Which item was purchased first by the customer after they became a member?

```
SELECT * FROM
(SELECT a.*, RANK() OVER(PARTITION BY userid ORDER BY created_date ASC) as rnk FROM
(SELECT s.userid, s.created_date, s.product_id, gus.gold_signup_date
FROM sales as s
INNER JOIN goldusers_signup as gus
ON s.userid = gus.userid and created_date >= gold_signup_date) AS a)
WHERE rnk = 1;
```

	userid integer	created_date date	product_id integer	gold_signup_date date	rnk bigint
1	1	2018-03-19	3	2017-09-22	1
2	3	2017-12-07	2	2017-04-21	1

7. Which item was purchased just before the customer became a member?

```
SELECT * FROM
(SELECT a.*, RANK() OVER(PARTITION BY userid ORDER BY created_date DESC) as rnk FROM
(SELECT s.userid, s.created_date, s.product_id, gus.gold_signup_date
FROM sales as s
INNER JOIN goldusers_signup as gus
ON s.userid = gus.userid and created_date <= gold_signup_date) as a)
WHERE rnk = 1;
```

	userid integer	created_date date	product_id integer	gold_signup_date date	rnk bigint
1	1	2017-04-19	2	2017-09-22	1
2	3	2016-12-20	2	2017-04-21	1

8. What is the total orders and amount spent for each member before they become a member?

```
SELECT userid, COUNT(created_date) as order_purchased, SUM(price) as total_smt_spent FROM
(SELECT a.*, p.price FROM
(SELECT s.userid, s.created_date, s.product_id, gus.gold_signup_date
FROM sales as s
INNER JOIN goldusers_signup as gus
ON s.userid = gus.userid and created_date <= gold_signup_date) as A
INNER JOIN product as p
ON a.product_id = p.product_id) as B
GROUP BY userid;
```

	userid integer	order_purchased bigint	total_smt_spent bigint
1	3	3	2720
2	1	5	4030

9. If buying each product generates points for eg. 5rs = 2 zomoto points and each product has different purchase points for eg. for p1 5rs = 1 zomato point, for p2 10rs = 5 zomato points and p3 5rs = 1 zomato point. Calculate points collected by each customer and for which product most points have been given till now? (2rs = 1 zomato points)

```
SELECT d.userid, SUM(total_points) as total_points_earned FROM
(SELECT c.*, amt/points as total_points FROM
(SELECT B.*, CASE
WHEN product_id = 1 THEN 5
WHEN product_id = 2 THEN 2
WHEN product_id = 3 THEN 5
ELSE 0 END AS points FROM
(SELECT a.userid, a.product_id, SUM(price) as amt FROM
(SELECT s.*, p.price
FROM sales AS s
INNER JOIN product as p
ON s.product_id = p.product_id) AS a
GROUP BY userid, product_id) as b) AS C) AS D
GROUP BY userid
ORDER BY userid ASC;
```

First half query

	userid integer	total_points_earned numeric
1	1	1829
2	2	763
3	3	1697

```
SELECT * FROM
(SELECT E.*, RANK() OVER(ORDER BY total_points_earned DESC) rnk FROM
(SELECT d.product_id, SUM(total_points) as total_points_earned FROM
(SELECT c.*, amt/points as total_points FROM
(SELECT B.*, CASE
WHEN product_id = 1 THEN 5
WHEN product_id = 2 THEN 2
WHEN product_id = 3 THEN 5
ELSE 0 END AS points FROM
(SELECT a.userid, a.product_id, SUM(price) as amt FROM
(SELECT s.*, p.price
FROM sales AS s
INNER JOIN product as p
ON s.product_id = p.product_id) AS a
GROUP BY userid, product_id) as b) AS C) AS D
GROUP BY product_id
ORDER BY product_id ASC) AS E)
WHERE rnk = 1;
```

Second half query

	product_id integer	total_points_earned numeric	rnk bigint
1	2	3045	1

10. If the first one year after a customer joins the gold program (including their join date) irrespective of what the customer has purchased they earn 5 zomato points for every 10rs spent who earned more 1 or 3 and what was their points earnings in their first year ?
(1 Zomato points = 2RS) = (0.5 points = 1rs)

```
SELECT b.*, b.price*0.5 as total_points_earned FROM
(SELECT a.*, p.price FROM
(SELECT s.userid, s.created_date, s.product_id, gus.gold_signup_date
FROM sales as s
INNER JOIN goldusers_signup as gus
ON s.userid = gus.userid and created_date >= gold_signup_date and created_date <= gold_signup_date+365) AS a
INNER JOIN product as p
ON a.product_id = p.product_id ORDER BY userid ASC) AS b;
```

	userid integer	created_date date	product_id integer	gold_signup_date date	price integer	total_points_earned numeric
1	1	2018-03-19	3	2017-09-22	330	165.0
2	3	2017-12-07	2	2017-04-21	870	435.0

11. Rank all the transactions of the customers

```
SELECT *, RANK() OVER(PARTITION BY userid ORDER BY created_date ASC) AS rank FROM sales;
```

	userid integer	created_date date	product_id integer	rank bigint
1	1	2016-03-11	1	1
2	1	2016-05-20	3	2
3	1	2016-11-09	1	3
4	1	2017-03-11	2	4
5	1	2017-04-19	2	5
6	1	2018-03-19	3	6
7	1	2019-10-23	2	7
8	2	2017-09-24	1	1
9	2	2017-11-08	2	2
10	2	2018-09-10	3	3
11	2	2020-07-20	3	4
12	3	2016-11-10	1	1
13	3	2016-12-15	2	2
14	3	2016-12-20	2	3
15	3	2017-12-07	2	4
16	3	2019-12-18	1	5

Tools:

This project was completed using Postgre SQL to analyze the dataset and answer the questions posed. The results are presented in a clear and concise manner and summaries to help illustrate key findings.

Conclusion:

This project demonstrates the ability to analyze a dataset using SQL and extract insights that can inform business decisions. The results provide a comprehensive understanding of the Zomato dataset and highlight key trends and patterns that can be used to improve the platform.