Instacart E-Commerce SQL Analysis Project

Project Overview

This project focuses on analyzing **Instacart e-commerce data** using SQL to understand customer behavior, product demand, repeat purchases, and loyalty patterns. The goal is to practice **real-world SQL analysis** similar to what data analysts do in product-based companies.

Dataset Description (Simple Explanation)

- **instacart_orders** – Contains order-level information such as user, order sequence, and order timing
- **order_products** – Contains products included in each order and whether they were reordered
- **products** – Product names and their category mappings
- **departments** – High-level product categories (produce, dairy, beverages, etc.)

Together, these tables help analyze **what customers buy, how often they buy, and how loyal they are**.

Key Business Questions & SQL Analysis

1. Department with Highest Sales

Business Question: Which department sells the most products?

```sql
SELECT d.department, COUNT(*) AS total_sales
FROM order_products op
JOIN products p ON op.product_id = p.product_id
JOIN departments d ON p.department_id = d.department_id
GROUP BY d.department
ORDER BY total_sales DESC;
```

Explanation: - Joins orders → products → departments - Counts total product sales per department - Helps identify **top-performing departments**

---

2. Percentage of Repeat Customers

Business Question: How many customers place more than one order?

```sql
SELECT
    (COUNT(DISTINCT user_id) * 100.0 /
```

```
        (SELECT COUNT(DISTINCT user_id) FROM instacart_orders)
    ) AS repeat_percentage
FROM instacart_orders
WHERE order_number > 1;
```

Explanation: - `order_number > 1` means repeat order - Calculates repeat customers as a percentage - Shows **customer retention strength**

---

3. Loyal Customers (Ordered More Than 15 Times)

Business Question: Who are the most loyal customers?

```
SELECT user_id, COUNT(*) AS loyal_customers
FROM instacart_orders
GROUP BY user_id
HAVING loyal_customers > 15
ORDER BY loyal_customers DESC;
```

Explanation: - Groups orders by user - Filters customers with more than 15 orders - Identifies **high-value customers**

---

4. Products with Reorder Rate Above 70%

Business Question: Which products customers buy again and again?

```
SELECT p.product_name,
       AVG(op.reordered) AS reorder_rate
FROM order_products op
JOIN products p ON op.product_id = p.product_id
GROUP BY p.product_name
HAVING reorder_rate > 0.7
ORDER BY reorder_rate DESC;
```

Explanation: - `reordered = 1` means product was repurchased - Average gives reorder rate - Highlights **customer-favorite products**

---

5. Rank Customers by Total Orders

Business Question: Who are the top customers by order volume?

```sql
SELECT user_id,
       COUNT(order_id) AS total_orders,
       RANK() OVER (ORDER BY COUNT(order_id) DESC) AS customer_rank
FROM instacart_orders
GROUP BY user_id
ORDER BY total_orders DESC;
```

Explanation: - Uses `RANK()` window function - Orders customers based on total orders - Helps in **customer segmentation**

---

6. First Product Ordered by Each User

Business Question: What was the first product each customer bought?

```sql
SELECT user_id, product_id
FROM order_products op
JOIN instacart_orders io ON op.order_id = io.order_id
WHERE io.order_number = 1;
```

Explanation: - First order represents onboarding behavior - Helps understand **initial customer preferences**

Additional Useful Queries

Average Orders per Customer

```sql
SELECT AVG(order_count) AS avg_orders
FROM (
    SELECT user_id, COUNT(*) AS order_count
    FROM instacart_orders
    GROUP BY user_id
) t;
```

Most Frequently Ordered Products

```sql
SELECT p.product_name, COUNT(*) AS order_count
FROM order_products op
JOIN products p ON op.product_id = p.product_id
GROUP BY p.product_name
```

```
ORDER BY order_count DESC
LIMIT 10;
```

Skills Demonstrated in This Project

- SQL Joins (INNER JOIN)
- Aggregations (`COUNT`, `AVG`)
- Window Functions (`RANK()`)
- Subqueries
- Business-oriented data analysis
- E-commerce customer behavior analysis

Conclusion

This project demonstrates practical SQL usage on a **real e-commerce dataset**, focusing on **customer loyalty, repeat behavior, and product demand**. The analysis reflects real business questions commonly asked in retail and product-based companies.