SQL for Data Analysis

Introduction

In this task, we explored the power of SQL to analyze data from a sample e-commerce database. By performing various SQL queries, we were able to extract meaningful insights from the dataset. The task aimed to demonstrate proficiency in fundamental SQL concepts, including data filtering, aggregation, and optimization techniques. We used a variety of SQL operations to manipulate and analyze data, allowing for deeper insights into customer behavior, order details, and overall sales performance.

Objective

The objective of this task was to execute SQL queries on a sample e-commerce database to extract valuable insights and showcase proficiency in data analysis using SQL.

Tools Used

- SQLite (via SQLiteOnline)
- SQL Queries: task3_queries.sql

Sample Dataset

- Dataset: ecommerce_sample.sqlite

Key Concepts Utilized

- SQL Queries:
    - SELECT, WHERE, ORDER BY, GROUP BY
    - SQL JOINs: INNER JOIN
    - Subqueries
    - Aggregate Functions: SUM, AVG
    - Views Creation
    - Query Optimization: Using INDEX

Queries Covered

1. Selecting Customers by Country
   Extracted customer details filtered by country.

2. Joining Customer and Order Data
   Combined customer and order data to get a comprehensive view of customer purchases.

3. Aggregating Payments per Customer
   Calculated the total payments made by each customer.

4. Identifying High Spenders with Subqueries
   Used subqueries to find customers with the highest spending.

5. Creating a View of Top Customers
   Created a view to easily access a list of top customers based on their spending.

6. Optimizing Search by Indexing Country
   Applied indexing to optimize search queries by country.

Conclusion

By leveraging SQL queries, this task provided valuable hands-on experience in analyzing a real-world e-commerce dataset. It demonstrates the importance of using SQL's advanced features such as joins, subqueries, and indexing for efficient data retrieval and optimization. These skills are fundamental to conducting comprehensive data analysis in any business domain.

The key analyses performed include:

- **Total Orders per State**

  Identified the volume of orders across different states to understand regional demand and support market segmentation strategies.

	customer_state	total_orders
1	SP	41746
2	RJ	12852
3	MG	11635
4	RS	5466
5	PR	5045
6	SC	3637
7	BA	3380
8	DF	2140
9	ES	2033
10	GO	2020
11	PE	1652
12	CE	1336
13	PA	975
14	MT	907
15	MA	747
16	MS	715
17	PB	536
18	PI	495
19	RN	485
20	AL	413
21	SE	350
22	TO	280
23	RO	253
24	AM	148
25	AC	81
26	AP	68
27	RR	46

- **Monthly Order Trends**

  Analyzed order trends over time to detect seasonal patterns, forecast demand, and assess the effectiveness of marketing campaigns.

	month	orders
1	2016-09	4
2	2016-10	324
3	2016-12	1
4	2017-01	800
5	2017-02	1780
6	2017-03	2682
7	2017-04	2404
8	2017-05	3700
9	2017-06	3245
10	2017-07	4026
11	2017-08	4331
12	2017-09	4285
13	2017-10	4631
14	2017-11	7544
15	2017-12	5673
16	2018-01	7269
17	2018-02	6728
18	2018-03	7211
19	2018-04	6939
20	2018-05	6873
21	2018-06	6167
22	2018-07	6292
23	2018-08	6512
24	2018-09	16
25	2018-10	4

- **Order Summary by State**
  Summarized order metrics such as total sales and item counts per state to evaluate financial performance and regional opportunities.

	customer_state	total_orders	avg_delivery_time
1	AC	80	21
2	AL	397	24
3	AM	145	26
4	AP	67	27
5	BA	3256	19
6	CE	1279	21
7	DF	2080	12
8	ES	1995	15
9	GO	1957	15
10	MA	717	21
11	MG	11355	11
12	MS	701	15
13	MT	886	18
14	PA	946	23
15	PB	517	20
16	PE	1593	18
17	PI	476	19
18	PR	4923	11
19	RJ	12353	15
20	RN	474	19
21	RO	243	19
22	RR	41	29
23	RS	5344	15
24	SC	3547	14
25	SE	335	21
26	SP	40495	8
27	TO	274	17

- **Orders With Delays Over 10 Days**
  Highlighted fulfillment issues by finding orders delayed by more than 10 days, enabling improvements in logistics and customer service.

Results / Messages

	order_id	delivery_days
1	53cdb2fc8bc7dce0b6741e2150273451	14
2	949d5b44dbf5de918fe9c16f97b45f8a	14
3	a4591c265e18cb1dcee52889e2d8acc3	17
4	e69bfb5eb88e0ed6a785585b27e16dbf	18
5	e6ce16cb79ec1d90b1da9085a6118aeb	13
6	82566a660a982b15fb86e904c8d32918	12
7	432aaf21d85167c2c86ec9448c4e42cc	11
8	dcb36b511fcac050b97cd5c05de84dc3	14
9	403b97836b0c04a622354cf531062e5f	18
10	116f0b09343b49556bbad5f35bee0cdf	13
11	83018ec114eee8641c97e08f7b4e926f	13
12	203096f03d82e0dffbc41ebc2e2bcfb7	21
13	95266dbfb7e20354baba07964dac78d5	18
14	fbf9ac61453ac646ce8ad9783d7d0af6	29
15	acce194856392f074dbf9dada14d8d82	12
16	91b2a010e1e45e6ba3d133fa997597be	14
17	ecab90c9933c58908d3d6add7c6f5ae3	30
18	1790eea0b567cf50911c057cf20f90f9	19
19	989225ba6d0ebd5873335f7e01de2ae7	22
20	d887b52c6516beb39e8cd44a5f8b60f7	19
21	8563039e855156e48fccee4d611a3196	31
22	60550084e6b4c0cb89a87df1f3e5ebd9	20
23	434d158e96bdd6972ad6e6d73ddcfd22	17
24	1e7aff52cdbb2451ace09d0f848c3699	12
25	6ea2f835b4556291ffdc53fa0b3b95e8	34
26	6ebaec694d7025e2ad4a05dba887c032	11
27	d17dc4a904426827ca80f2ccb3a6be56	11
28	138849fd84dff2fb4ca70a0a34c4aa1c	13
29	641fb0752bf5b5940c376b3a8bb9dc52	19
30	66e4624ae69e7dc89bd50222b59f581f	25
31	a685d016c8a26f71a0bb67821070e398	24
32	ccbabeb0b02433bd0fcbac46e70339f2	18
33	974c1993ab8024d3ed16229183c2308d	17
34	82bce245b1c9148f8d19a55b9ff70644	20
35	2711a938db643b3f0b62ee2c8a2784aa	18
36	0a4a2fccb27bd83a892fa503987a595b	21
37	e4de6d53ecff736bc68804b0b6e9f635	23
38	6b860b35691d486e45dc98e3514ec5f6	11
39	ec341c54a5ebf8ee0a67a8632aa7579b	13

- **Most Common Order Status**

   Investigated the distribution of order statuses to monitor operational efficiency and uncover potential process bottlenecks.

	order_status	count
1	delivered	96478
2	shipped	1107
3	canceled	625
4	unavailable	609
5	invoiced	314
6	processing	301
7	created	5
8	approved	2

These queries provided a strong foundation for building dashboards, improving data-driven decision-making, and gaining a deeper understanding of the eCommerce business operations.