

## **TASK 2**

# **SQL QUERIES TO EXTRACT AND ANALYZE DATA FROM A DATABASE**

Data Analyst Internship

By : Kajal Prajapati

Project Name: Olist E-Commerce SQL Analysis

Tool Used: MySQL

Dataset: Brazilian E-Commerce Public Dataset (Olist)

Tables Used: 9 Tables including orders, products, payments, order\_items etc.

# OBJECTIVE

- Perform SQL data analysis using JOINS, GROUP BY, Subqueries, Views
- Extract business insights like:
  - Top Selling Products
  - Payment Method Trends
- Apply indexing for performance optimization
- Use views for reusable queries



# DATA LOADING IN MYSQL

- Used LOAD DATA LOCAL INFILE for 9 CSVs
- Enabled local\_infile in MySQL

- Sample SQL:**

```
LOAD DATA LOCAL INFILE "C:\\Users\\Kajal Prajapati\\Documents\\my_sql_dataset\\olist_customers_dataset.csv"  
INTO TABLE customers  
FIELDS TERMINATED BY ','  
ENCLOSED BY ''''  
LINES TERMINATED BY '\\n'  
IGNORE 1 ROWS;
```

## TOP SELLING PRODUCTS

- View Name: top\_selling\_products\_view
- Columns: product\_id, category, total\_sold, total\_revenue

- **Output Table**

```
280 • SELECT * FROM top_selling_products view LIMIT 10;
```

281

	product_category_name	product_id	total_sold	total_revenue
▶	moveis_decoracao	aca2eb7d00ea1a7b8ebd4e68314663af	527	37608.90062713623
	cama_mesa_banho	99a4788cb24856965c36a24e339b6058	488	43025.56069946289
	ferramentas_jardim	422879e10f46682990de24d770e7f83d	484	26577.220558166504
	ferramentas_jardim	389d119b48cf3043d311335e499d9c6b	392	21440.590496063232
	ferramentas_jardim	368c6c730842d78016ad823897a372db	388	21056.80047607422
	ferramentas_jardim	53759a2ecddad2bb87a079a1f1519f73	373	20387.200454711914
	informatica_acessorios	d1c427060a0f73f6b889a5c7c61f2ac4	343	47214.51049041748
	relogios_presentes	53b36df67ebb7c41585e8d54d6772e08	323	37683.42002105713
	beleza_saude	154e7e31ebfa092203795c972e5804a6	281	6325.189935684204
	informatica_acessorios	3dd2a17168ec895c781a9191c1e95ad7	274	41082.59832763672



# PAYMENT METHOD TRENDS

- View Name: payment\_summary\_view
- Columns: payment\_type, transactions, total revenue, avg

- **Output Table**

```
296 • SELECT * FROM payment_summary_view ;
297
```

	payment_type	total_transactions	total_revenue	avg_payment
▶	credit_card	76795	12542084.190177472	163.32
	boleto	19784	2869361.2694940567	145.03
	voucher	5775	379436.87039574794	65.7
	debit_card	1529	217989.7899875641	142.57
	not_defined	3	0	0

# SUBQUERY EXAMPLE

products which sold more than average quantity

```
281 • SELECT *
282     FROM top_selling_products_view
283     WHERE total_sold > (
284         SELECT AVG(total_sold) FROM top_selling_products_view
285     );
286
```

Result Grid	Filter Rows:	Export:	Wrap Cell Content:	Fetch rows:
product_category_name	product_id	total_sold	total_revenue	
moveis_decoracao	aca2eb7d00ea1a7b8ebd4e68314663af	527	37608.90062713623	
cama_mesa_banho	99a4788cb24856965c36a24e339b6058	488	43025.56069946289	
ferramentas_jardim	422879e10f46682990de24d770e7f83d	484	26577.220558166504	
ferramentas_jardim	389d119b48cf3043d311335e499d9c6b	392	21440.590496063232	
ferramentas_jardim	368c6c730842d78016ad823897a372db	388	21056.80047607422	
ferramentas_jardim	53759a2ecddad2bb87a079a1f1519f73	373	20387.200454711914	
informatica_acessorios	d1c427060a0f73f6b889a5c7c61f2ac4	343	47214.51049041748	
relogios_presentes	53b36df67ebb7c41585e8d54d6772e08	323	37683.42002105713	
beleza_saude	154e7e31ebfa092203795c972e5804a6	281	6325.189935684204	
informatica_acessorios	3dd2a17168ec895c781a9191c1e95ad7	274	41082.59832763672	
beleza_saude	2b4609f8948be18874494203496bc318	260	22717.219451904297	



# QUERY OPTIMIZATION

Used Indexes on:

- product\_id, order\_status, payment\_type etc

```
306  -- Optimize Queries with Indexes
307 • CREATE INDEX idx_order_status ON orders(order_status);
308 • CREATE INDEX idx_product_id ON order_items(product_id);
309 • CREATE INDEX idx_payment_type ON order_payments(payment_type);
310
311 • SHOW INDEX FROM orders;
312
```

Result Grid		Filter Rows:	Export:		Wrap Cell Content:										
	Table	Non_unique	Key_name	Seq_in_index	Column_name	Collation	Cardinality	Sub_part	Packed	Null	Index_type	Comment	Index_comment	Visible	Expression
▶	orders	0	PRIMARY	1	order_id	A	0	NULL	NULL		BTREE			YES	NULL
	orders	1	idx_order_status	1	order_status	A	7	NULL	NULL	YES	BTREE			YES	NULL

Result 34 x

Query Completed

## SUMMARY

Loaded & analyzed real-world E-Commerce dataset  
Extracted key insights using SQL (Top Products + Payments)  
Applied subqueries, views, indexing  
Ready for dashboard in Power BI