INNOMATICS®
RESEARCH LABS

INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON

Retail Analysis

Presented By: Nikita Nikam

About me

- **Background**

  I have recently completed my Bachelor's Degree in Information Technology from Savitribai Phule Pune University (SPPU).

- **Why you want to learn Data Science**

  I want to learn Data Science because it helps understand large amounts of data and convert it into useful insights. It enables predicting future trends and solving real-world problems using data. Data Science because it combines statistics, programming, and machine learning to analyze complex datasets and build predictive models. It helps uncover hidden patterns, optimize processes, and support data-driven decision-making.

- **LinkedIn and GitHub profile URLs**

  https://www.linkedin.com/in/nikita-nikam-184b39240
  NikitaNikam1 (Nikita Nikam) · GitHub

INNOMATICS RESEARCH LABS

Agenda

- Real World Scenarios
- Objective of the Project
- ER Diagram and schema explanation
- Key analysis questions
- SQL query results with screenshots or summaries
- Final business insights and recommendations
- Conclusion

INNOMATICS RESEARCH LABS

Problem Statement

Retail businesses generate large volumes of transactional data, but without proper analysis, this data cannot be effectively used for decision-making.

The challenge is to transform raw retail data into meaningful insights that help management understand customer behavior, sales performance, product demand, logistics efficiency, and revenue trends.

Objective of the project

- Analyze retail business data using SQL to extract meaningful insights
- Study customer behavior, sales trends, product categories, and seller performance
- Evaluate order processing, payments, reviews, and delivery efficiency
- Support data-driven decisions to improve customer experience, operations, and revenue growth

INNOMATICS RESEARCH LABS

Why a Relational Database Is Needed

Data is spread across multiple entities:

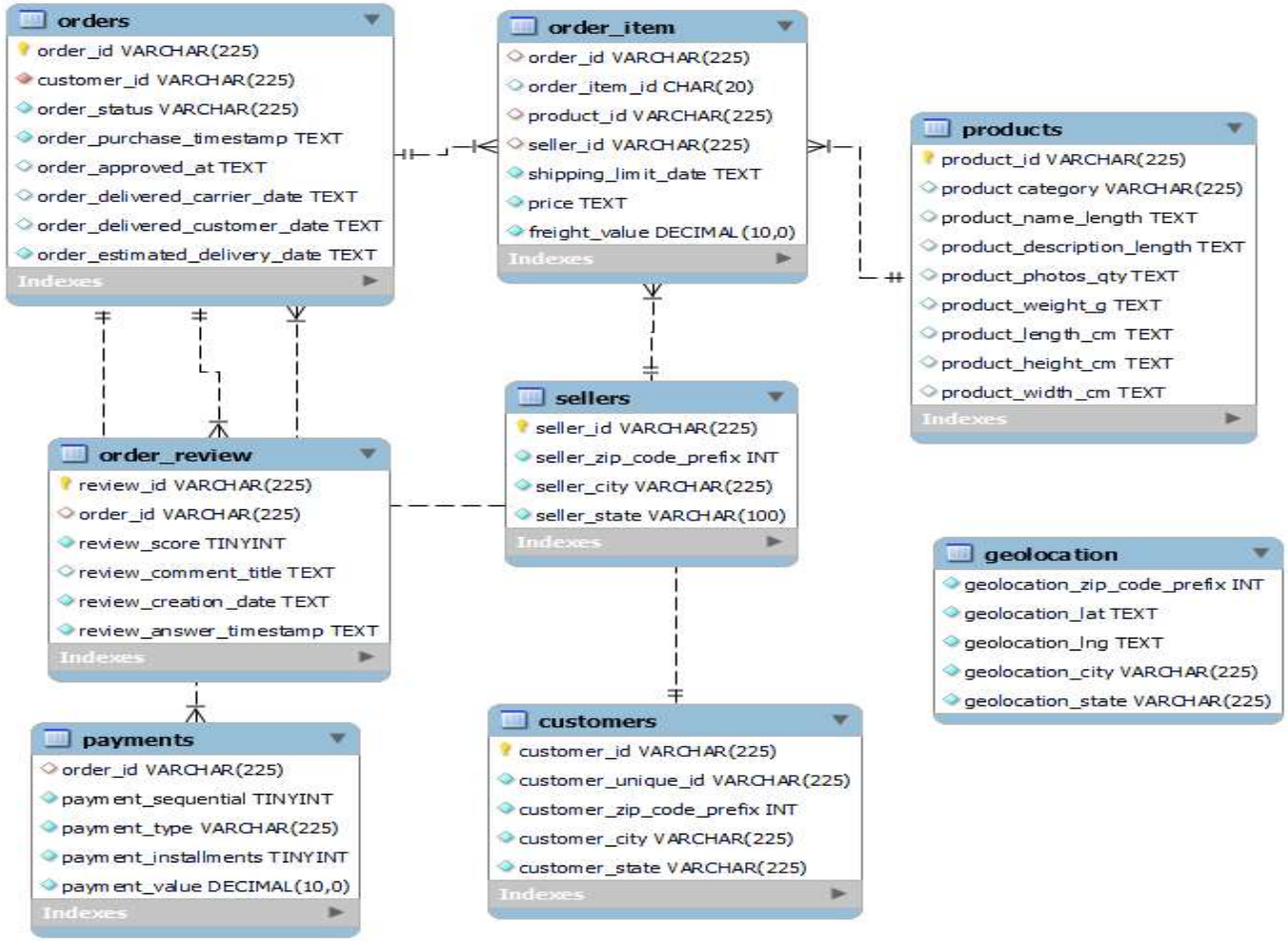
- Customers
- Orders
- Products
- Payments
- Sellers
- Review

Tables are connected using primary and foreign keys.

Enables Data integrity,Efficient joins,Scalable analysis

A relational database is ideal because the data has clear relationships between entities.

ER Diagram and schema explanation



NNOMATICS RESEARCH LABS

SQL query result

1. What is the total number of unique customers?

```
-- 1.  Find the total number of unique customers.
select count(distinct(customer_unique_id)) as total_unique_customers
from customers;  # 96096
```

- This query finds how many different customers are there in total.

result

	total_unique_customers
▶	96096

INNOMATICS RESEARCH LABS

2. Which are the top 5 states with the highest number of customers?

```
-- 2.  Identify the top 5 states with the highest number of customers.

select customer_state, count(*) as customer_count
from customers
group by customer_state
order by customer_count desc
limit 5;
```

result

customer_state	customer_count
SP	41746
RJ	12852
MG	11635
RS	5466
PR	5045

- This analysis highlights top5 states (SP,RJ,MG,RS,PR) with the highest customers.

INNOMATICS RESEARCH LABS

3. What is the customer retention rate (customers who placed more than one order)?

```
select
(count(case when order_count > 1 then 1 end) * 100.0 / count(*)) as retention_rate
from (
    select c.customer_unique_id, count(o.order_id) as order_count
    from customers c
    join orders o on c.customer_id = o.customer_id
    group by c.customer_unique_id
) as customer_orders;
```

- This query measures customer retention rate by identifying how many customers placed more than one order.

result

	retention_rate
▶	3.11876

INNOMATICS RESEARCH LABS

4. What is the total number of delivered and canceled orders?

```
select order_status, COUNT(*) as total_orders
from orders
where order_status in ('delivered', 'canceled')
group by order_status;
```

- This query compares successful deliveries against cancellations to assess operational efficiency.

result

	order_status	total_orders
▶	delivered	96478
	canceled	625

INNOMATICS RESEARCH LABS

5. What is the average delivery time?

```
SELECT
    ROUND(AVG(DATEDIFF(order_delivered_customer_date, order_purchase_timestamp)), 2)
    AS avg_delivery_days
FROM orders
WHERE order_status = 'delivered' AND order_delivered_customer_date is not null;
```

- This query calculates the average delivery time to evaluate logistics performance and understand its impact on customer experience.

result

	avg_delivery_days
▶	12.50

INNOMATICS RESEARCH LABS

6. Find the top 10 most sold product categories.

```
select p.`product category`, count(oi.product_id) as total_sold
from order_item oi
left join products p on oi.product_id = p.product_id
group by p.`product category`
order by total_sold desc
limit 10;
```

- This query identifies the most popular product categories based on total items sold.

result

	product category	total_sold
▶	bed table bath	11115
	HEALTH BEAUTY	9670
	sport leisure	8641
	Furniture Decoration	8334
	computer accessories	7827
	housewares	6964

housewares	6964
Watches present	5991
telephony	4545
Garden tools	4347
automotive	4235

INNOMATICS RESEARCH LABS

7. Find the most common payment type.

```
select * from payments;
select payment_type, count(*) as payment_type_count
from payments
group by payment_type
order by payment_type_count desc
limit 1;
```

- This query identifies the dominant payment methods to help businesses optimize the checkout process and design better promotional offers.

result

payment_type	payment_type_count
credit_card	76795

INNOMATICS RESEARCH LABS

Business Insights

- The platform has a strong customer base, indicating effective market reach.
- Customer demand is concentrated in a few states, enabling targeted regional strategies.
- Low repeat purchases highlight weak customer retention, requiring loyalty and engagement initiatives.
- Order fulfillment is efficient, with deliveries far exceeding cancellations, though further optimization is possible.
- Logistics performance and delivery speed directly impact customer satisfaction and reviews.
- A few product categories drive most sales, making them key priorities for inventory and promotions.
- Credit cards dominate payment methods, presenting opportunities for checkout optimization and card-based offers.

INNOMATICS RESEARCH LABS

Conclusion

This project showcases the use of SQL for real-world retail data analysis, uncovering insights into customer behavior, order performance, product demand, delivery efficiency, and payment preferences. The findings highlight strong customer reach, low retention opportunities, efficient deliveries with scope for improvement, and the dominance of key product categories and payment methods, enabling data-driven business decisions.

INNOMATICS
RESEARCH LABS

THANK YOU

INNOMATICS
RESEARCH LABS