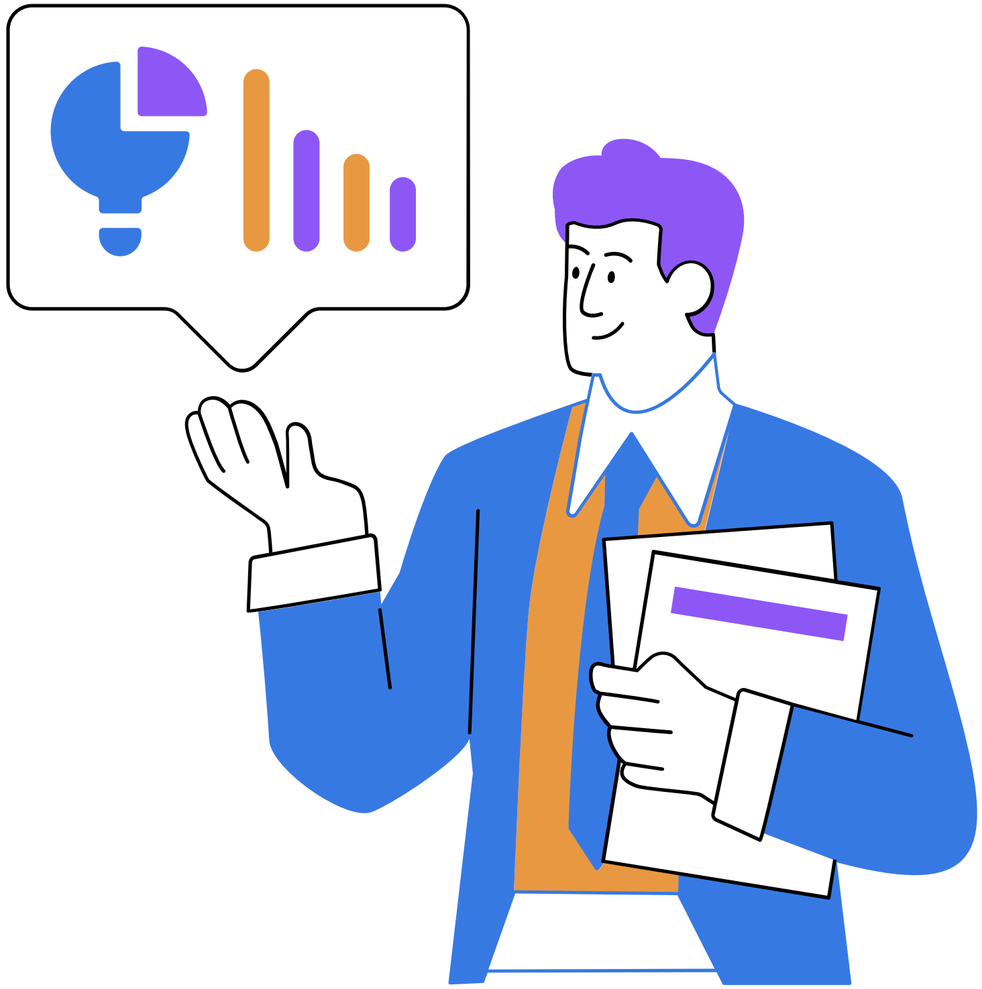
RETAIL SALES ANALYSIS



-By Abhishek Kumar Singh

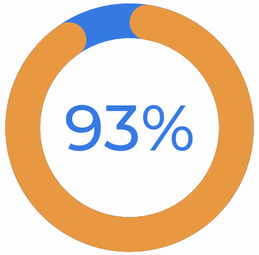
PROJECT INFO

This project involved analyzing a comprehensive retail dataset to derive key insights on sales, profitability, customer behavior, and performance trends across products, regions, and time.

The goal of this project was to help a retail business understand its sales performance, customer behavior, and profitability trends across different dimensions like time, geography, and product category.

Purpose Behind this Analysis

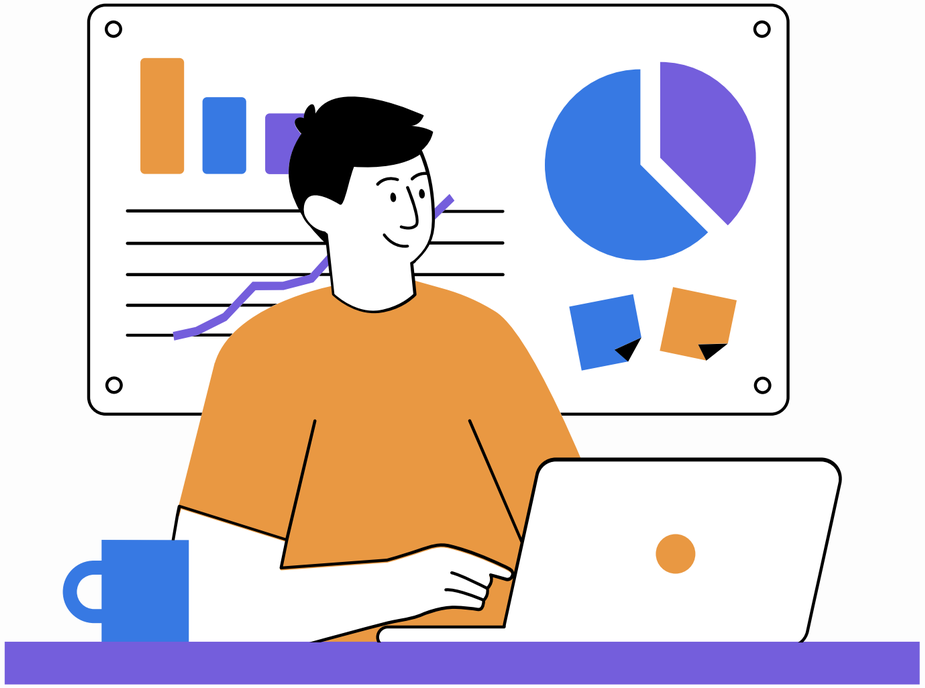
- To understand overall business performance and how much a customer typically spends.
- To identify best-sellers and most profitable products to focus marketing & inventory planning.
- To detect regional demand and optimize supply chain, stocking, and regional promotions.
- Time-Based Trends (Year, Quarter, Month): To uncover seasonality and trends, helping in forecasting and campaign planning.
- To pinpoint what’s growing fast and what’s falling behind. This supports strategic product decisions.
- Shipping & Order Pattern Insights: To assess operational efficiency and customer preferences across shipping modes and categories.



Project Overview

This project analyzes a retail orders dataset containing **9,994** records with **16** initial columns.

The data includes order details, shipping information, product categories, pricing, and geographical information for orders primarily in the United States.



Key Steps Peformed

Data Loading & Initial Exploration

Installed Kaggle API and downloaded the dataset

To begin the analysis, the Kaggle API was installed and configured to enable direct dataset access.

The dataset titled Orders Dataset was downloaded using the Kaggle CLI, specifically targeting the orders_data.csv file, which contains shipment and order-level data for nearly 10,000 transactions across the United States.

Install Required Libraries and Load Data

Library Installation: Installed essential Python libraries for data analysis and database connectivity, including:

- pandas and numpy for data manipulation and numerical operations
- sqlalchemy, mysql-connector-python for interacting with SQL databases

Data Loading: Imported the dataset into a Pandas DataFrame to facilitate initial analysis, cleaning, and transformations.

Explore the Data

- **Objective:** Performed an initial data exploration to understand the structure and distribution of the dataset, check column data types, and identify any potential data quality issues.

Key Steps Peformed

Methods Used:

- .info() to inspect data types and null values
- .describe() to view summary statistics for numerical columns
- .head() to preview the first few rows of data
- .unique() and .nunique() to explore the variety of values in categorical columns

Data Cleaning

- While reading the CSV, specific placeholder values like "Not Available" and "unknown" were explicitly treated as missing (NaN) using the na_values parameter in pandas.read_csv().
- Created a Cleaned DataFrame (df1): A new DataFrame was created with corrected missing values and standardized formats to avoid contaminating the analysis.
- Renamed the following columns for improved understanding and consistency during analysis:
  - 'Profit' → 'Unit_Profit'
  - 'Selling_Price' → 'Unit_Selling_Price'
- Dropped columns that were no longer required after calculations:
  - 'Cost Price', 'List Price', and 'Discount Percent'

Key Steps Peformed

Feature Engineering

- *Calculated Selling Price:* List Price - (List Price * Discount Percent/100)
- *Calculated Unit Profit:* Selling Price - Cost Price
- *Calculated Total Profit per order:* Quantity * Unit Profit
- *Calculated Total Sales:* Quantity * Selling Price
- Converted 'Order Date' to datetime format
- Extracted Month and Quarter from dates

Database Integration

- Connected to MySQL using SQLAlchemy for seamless data handling.
- Exported the cleaned DataFrame (df1) to a table named orders_data in database.
- Automated table creation and data insertion using Python.
- Verified data accuracy by running sample SQL queries to check row count, structure, and sample records.

Key Steps Peformed

SQL Analysis: Complex Queries and Business Problem Solving

- Calculate the total selling price and profits for all orders.
- Identified top 10 highest profit generating products
- Listed all distinct cities
- Calculated total selling price and profits
- Found Technology category orders shipped via Second Class.
- Find the average order value
- Find the city with the highest total quantity of products ordered.
- Rank orders in each region by quantity in descending order
- List all orders placed in the first quarter of any year (January to March), including the total cost for these orders
- Calculate Year and Quarter wise Total Sales
- Find top 3 highest selling products in each region according to quantity
- Find the city with the highest total quantity of products ordered
- Find Total sales in each region
- Find month over month growth comparison for 2022 and 2023 sales.
- calculate % of growth: YoY Growth % = (2023_sales − 2022_sales)/2022_sales × 100
- For each category which month had highest sales
- Which sub category had highest growth by sales in 2023 compare to 2022

Key Steps Peformed

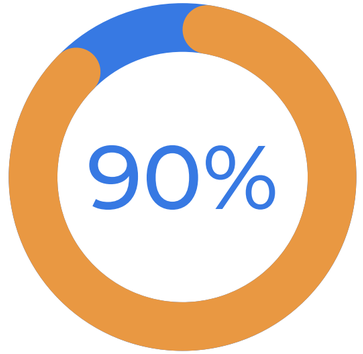
Insights

City and Region-Level Insights

- Distinct Cities: The data spans 531 unique cities, indicating a diverse geographic spread.
- Top City by Quantity Ordered: New York City tops the list with 3,417 products ordered.

Regional Sales:

- West region leads with $3.47M, followed by East ($3.26M), Central ($2.39M), and South ($1.97M).



Sales & Profitability Analysis

- Average Order Value: Approximately $1,108.60 per order
- Highest Profit Orders: Order #2698 generated the highest profit: $21,746.40, with total sales of $130,406.40.

Category & Shipping Mode Trends

- Orders in Technology + Second Class:
  Orders with this combo began as early as January 2022, suggesting frequent use of this ship mode for electronics.

Key Steps Peformed

Insights

Product-Level Insights

Most profitable product: TEC-CO-10004722 ($24,816 profit)

- **Top 3 Highest Selling Products in Each Region:**
  - Products like OFF-BI-10000545, TEC-MA-10000822, and TEC-CO-10001449 dominate across regions.
  - This reflects consistent demand across geographies.

🗓️ **Category-wise Best Months**

Category	Best Month	Sales
Furniture	Aug 2023	$230.5K
Office Supplies	Feb 2023	$287.2K
Technology	Oct 2023	$295.6K

Key Steps Peformed

Quarter-wise Total Sales

Year	Quarter	Total Sales
2022	Q1	$1,275,534.78
2022	Q2	$1,347,755.19
2022	Q3	$1,462,240.58
2022	Q4	$1,497,101.69

Year	Quarter	Total Sales
2023	Q1	$1,559,496.62
2023	Q2	$1,224,460.68
2023	Q3	$1,591,566.40
2023	Q4	$1,574,319.97

Key Steps Peformed

Quarter-wise Total Sales

- Best Quarter Overall: Q3 2023 with $1.59M
- Lowest Quarter: Q2 2023, down to $1.22M
- Yearly Trend:
- 2023 shows a stronger Q1, Q3, and Q4 compared to 2022.
- Q2 2023 underperformed — aligns with June’s big -29.31% drop.

Key Business Takeaways

- Sales go up at certain times of the year, especially in February and during the last few months (Q4).
- Technology and Machines are doing really well. It’s a good idea to focus more marketing on these categories.
- New York City has a lot of orders. Keep giving it priority
- The Machines sub-category grew a lot compared to last year
- Second Class Shipping is popular, especially for electronics. If we improve or speed up this option, customers might be happier.

THANK YOU