# E-Commerce Sales Analysis Project

End-to-End E-Commerce Sales Analysis Using Python & MySQL – From Raw Data to Business Insights with SQL Window Functions, Aggregates & Trends

By: Nishant Kumar Sharma | Aspiring Data Analyst

#### Dataset Source: Amazon E-Commerce Data (inspired by Kaggle)

• Records: 100000+ rows

 Columns: index, Order\_ID, Date, Status, Fulfilment, Sales\_Channel ship\_service\_level, Style, SKU, Category, Size, ASIN, Courier\_Status, Qty, Amount, ship\_city, ship\_state, Total Sales

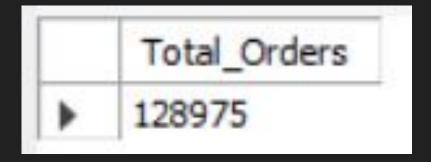
 Objective: Perform SQL-based analysis to extract trends, patterns, and KPIs

1. Total Orders Count

Query: -

Select count(order\_ID) as Total\_Orders

from sales\_data;



#### 2. Monthly Sales

Query: -

Select

CASE

WHEN Month(Date) = 3 THEN 'March'

WHEN Month(Date) = 4 THEN 'April'

WHEN Month(Date) = 5 THEN 'MAY'

WHEN Month(Date) = 6 THEN 'June'

End as Months,

Sum(Total\_sales)

from Sales\_Data

group by Months;

	Months	Sum(Total_sales)
•	April	27847245
	March	98261
	MAY	25326683
	June	22762217

3. Top 3 Category Sales

Query: -

Select Distinct(Category), sum(Total\_Sales) as Total\_Sales

from Sales\_Data

group by Category

LIMIT 3;

	Category	Total_Sales
Þ	Set	37934434
	kurta	20675349
	Western Dress	10707932

Set Category shows highest sales

4. Top 5 City Sales

Query: -

select distinct(ship\_city), sum(Total\_Sales) as Total\_Sales

from Sales\_Data

group by ship\_city

Limit 5;

	ship_dty	Total_Sales	
•	MUMBAI	4173450	
	BENGALURU	7104012	
	NAVI MUMBAI	824415	
	PUDUCHERRY	130507	
	CHENNAI	3500826	

5. Top 5 size sales

Query: -

Select distinct(Size), Sum(Total\_Sales) as Total\_Sales

from Sales\_Data

group by Size

LIMIT 5;

	Size	Total_Sales
•	S	10258811
	3XL	8870309
	XL	12000635
	L	12767020
	XXL	10350376

6. % contribution of each category

```
Query: -
Select Category,
sum(Total sales) as Total Sales,
round(100.0 * sum(Total_Sales) /
(Select Sum(Total Sales) from Sales Data), 2)
as Percentage_Contribution
from Sales Data
group by Category;
```

	Category	Total_Sales	Percentage_Contribution
•	Set	37934434	49.89
	kurta	20675349	27.19
	Western Dress	10707932	14.08
	Тор	5242931	6.90
	Ethnic Dress	762949	1.00
	Bottom	142870	0.19
	Saree	125767	0.17
	Blouse	441259	0.58
	Dupatta	915	0.00

## Techniques & Tools Used

- & Python (Jupyter Notebook) Exploratory Data Analysis (EDA)
- Loaded raw dataset using pandas
- Handled missing values and nulls
- Formatted date column to correct data type
- Created a new 'Total\_Sales' column using (Qty \* Amount)
- Exported cleaned dataset to Excel
- MySQL Data Analysis
- Loaded cleaned Excel into MySQL
- Performed queries using:
- 🔽 SELECT, WHERE, GROUP BY, ORDER BY
- Aggregations: SUM, AVG, COUNT
- Window Functions: RANK(), DENSE\_RANK(), ROW\_NUMBER(), LAG(), MOVING\_AVG
- **EXECUTE:** CASE statements, CTEs, HAVING
- → Performance optimization with INDEX

#### Insights & Observations

- Set category had the highest sales
- Mumbai & Bengaluru contributed most to revenue
- ✓ Sales increased by 28% from March to June
- × ~14% orders were cancelled
- TSize 'M' most frequently ordered
- **2** 50+ duplicate orders found → Possible system issue
- Amazon.in was dominant channel
- Most sales happened on weekends
- Shipped orders = 90% of revenue

#### Challenges & Learnings

- Working with real-world messy data
- Efficient SQL query writing
- Understanding business metrics through data
- Performance optimization using INDEX

# THANK YOU

Email: nish.sh27@gmail.com

GitHub: github.com/Nishantksh277

LinkedIn: linkedin.com/in/nishantksharma

Portfolio: Nishant Sharma

Dataset: <u>Kaggle</u>

Phone: 8799720977