



E-Commerce Business Analytics Project

Using Excel, SQL, Python & Power BI

1. Introduction

Data analytics plays a critical role in modern business decision-making by transforming raw data into meaningful insights.

This project focuses on analyzing an e-commerce retail dataset to understand sales performance, customer behavior, and product profitability using an end-to-end analytics workflow.

The project demonstrates a complete data pipeline starting from data cleaning to dashboard visualization using multiple industry tools.

2. Project Objective

Analyze overall sales and profit performance.

- Identify high-performing regions and product categories.
- Understand customer purchasing behavior.
- Build an interactive dashboard for business decision support.
- Demonstrate real-world data analytics workflow.

3. Tools & Technology used

Excel	Data cleaning & preprocessing
MySQL	Database creation & querying
Python	Data analysis & visualization
Power BI	Interactive dashboard development

4. Project Workflow

Excel → SQL Database → Python Analysis → Power BI Dashboard

This workflow represents an industry-standard analytics pipeline.



5. Step 1 — Data Cleaning Using Excel

The Superstore dataset was first processed in Microsoft Excel to prepare it for analysis.

Tasks Performed:

- Removed duplicate records
- Checked and handled missing values
- Standardized column names
- Corrected date formats
- Created new columns:
 - Year
 - Month
 - Profit Margin

👉 *Cleaned Excel dataset showing columns and formulas*

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Order_ID	Order_Dat	Ship_Date	Ship_Mod	Customer	Customer_Segment	Country	City	State	Region	Product_ID	Category	Sub_Categ	
2	US-2023-1	1/3/2023	1/7/2023	Standard	CDP-13000	Darren Po	Consumer	United Sta	Houston	Texas	Central	OFF-PA-10	Office Sup Paper	
3	US-2023-1	1/4/2023	1/8/2023	Standard	CPO-19195	Phillina Ol	Home Offi	United Sta	Naperville	Illinois	Central	OFF-BI-10	Office Sup Binders	
4	US-2023-1	1/4/2023	1/8/2023	Standard	CPO-19195	Phillina Ol	Home Offi	United Sta	Naperville	Illinois	Central	OFF-LA-10	Office Sup Labels	
5	US-2023-1	1/4/2023	1/8/2023	Standard	CPO-19195	Phillina Ol	Home Offi	United Sta	Naperville	Illinois	Central	OFF-ST-10	Office Sup Storage	
6	US-2023-1	1/5/2023	#####	Standard	CMB-18085	Mick Brow	Consumer	United Sta	Philadelphia	Pennsylva	East	OFF-AR-10	Office Sup Art	
7	US-2023-1	1/6/2023	#####	Standard	CME-17320	Maria Etez	Home Offi	United Sta	Henderson	Kentucky	South	FUR-CH-10	Furniture Chairs	
8	US-2023-1	1/6/2023	#####	Standard	CME-17320	Maria Etez	Home Offi	United Sta	Henderson	Kentucky	South	OFF-AR-10	Office Sup Art	
9	US-2023-1	1/6/2023	1/7/2023	First Class	JO-15145	Jack O'Bri	Corporate	United Sta	Athens	Georgia	South	OFF-AR-10	Office Sup Art	
10	US-2023-1	1/6/2023	#####	Standard	CME-17320	Maria Etez	Home Offi	United Sta	Henderson	Kentucky	South	OFF-BI-10	Office Sup Binders	
11	US-2023-1	1/6/2023	#####	Standard	CME-17320	Maria Etez	Home Offi	United Sta	Henderson	Kentucky	South	OFF-FA-10	Office Sup Fasteners	
12	US-2023-1	1/6/2023	#####	Standard	CME-17320	Maria Etez	Home Offi	United Sta	Henderson	Kentucky	South	OFF-PA-10	Office Sup Paper	
13	US-2023-1	1/6/2023	1/8/2023	Second Cl	LS-17230	Lycoris Sal	Consumer	United Sta	Los Angeles	California	West	OFF-PA-10	Office Sup Paper	
14	US-2023-1	1/6/2023	#####	Standard	CME-17320	Maria Etez	Home Offi	United Sta	Henderson	Kentucky	South	TEC-PH-10	Technolog Phones	
15	US-2023-1	1/6/2023	#####	Standard	CME-17320	Maria Etez	Home Offi	United Sta	Henderson	Kentucky	South	TEC-PH-10	Technolog Phones	
16	US-2023-1	1/7/2023	#####	Standard	CVS-21820	Vivek Sun	Consumer	United Sta	Huntsville	Texas	Central	FUR-FU-10	Furniture Furnishing	
17	US-2023-1	1/7/2023	#####	Standard	CVS-21820	Vivek Sun	Consumer	United Sta	Huntsville	Texas	Central	OFF-BI-10	Office Sup Binders	
18	US-2023-1	1/8/2023	#####	Standard	CMS-17920	Melanie S	Consumer	United Sta	Laredo	Texas	Central	OFF-AR-10	Office Sup Art	

6. Step 2 — Database Creation Using SQL

The cleaned dataset was imported into MySQL to enable structured querying and efficient analysis.

Steps:

- Created database
- Designed sales table with appropriate data types
- Imported CSV dataset using Import Wizard
- Verified successful data loading

Sample Queries Performed:

- Total sales calculation
- Region-wise sales comparison
- Profit by product sub-category

  MySQL Workbench showing table and query results

```
--Total sales
select round(sum(Sales),3) as Tota_Sales from orders_superstore

--Sales by Region
select Region,round(sum(Sales),3) as total_sale from orders_superstore
group by Region

--Top performing product
select top 5 Sub_Category,round(sum(Profit),3) as Total_Profit from orders_superstore
group by Sub_Category
order by Total_Profit desc

--Total profit
select round(sum(Profit),3) as Overall_Profit from orders_superstore

--most Profitable country
select top 1 Country,sum(Profit) as Overall_profit from orders_superstore
group by country
```



7. Step 3 — Data Analysis Using Python

Python was used for Exploratory Data Analysis (EDA) and feature engineering.

Libraries Used:

- Pandas
- Matplotlib
- Seaborn
- MySQL Connector

Analysis Performed:

- Dataset statistical summary
- Region and category sales analysis
- Customer purchase behavior
- Profit margin calculation
- Visualization using charts

Additional features created:

- Customer total sales
- Profit margin metric

👉 *Jupyter Notebook showing Python code and output graphs*

```
#Statistical summary  
df.describe()
```

	Sales	Quantity	Discount	Profit	year	month	Profit_margin
count	10192.000000	10192.000000	10192.000000	10191.000000	10192.000000	10192.000000	10192.000000
mean	228.233306	3.791601	0.155386	28.720779	2024.720467	7.811911	0.122424
std	619.966119	2.228075	0.206258	232.456479	1.125463	3.295871	0.465096
min	0.444000	1.000000	0.000000	-6599.978027	2023.000000	1.000000	-2.750000
25%	17.219999	2.000000	0.000000	1.762400	2024.000000	5.000000	0.080000
50%	53.890001	3.000000	0.200000	8.690000	2025.000000	9.000000	0.270000
75%	209.500000	5.000000	0.200000	29.293850	2026.000000	11.000000	0.360000
max	22638.480469	14.000000	0.800000	8399.975586	2026.000000	12.000000	0.500000

Figure: Sales by Region (Python Visualization)

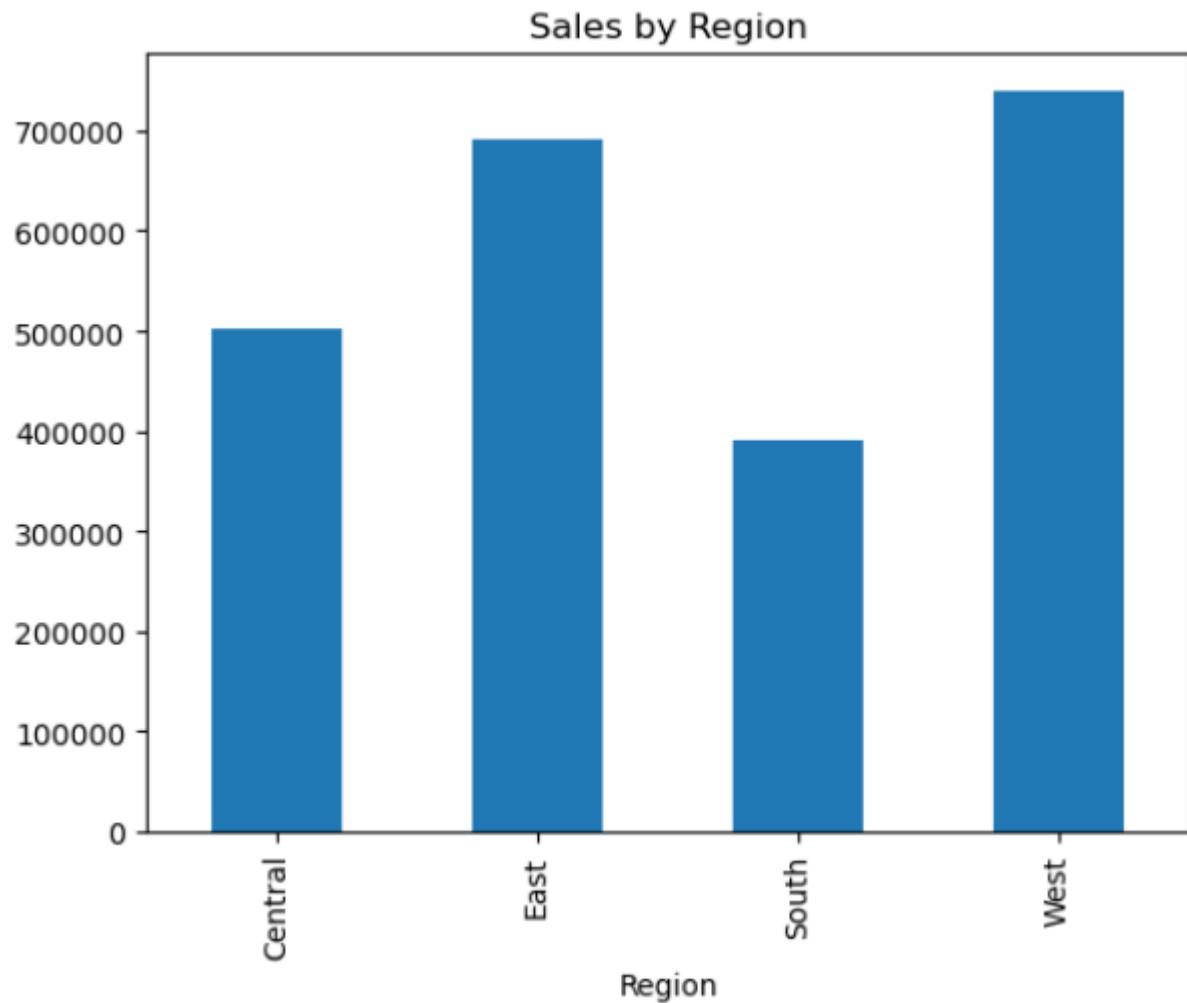
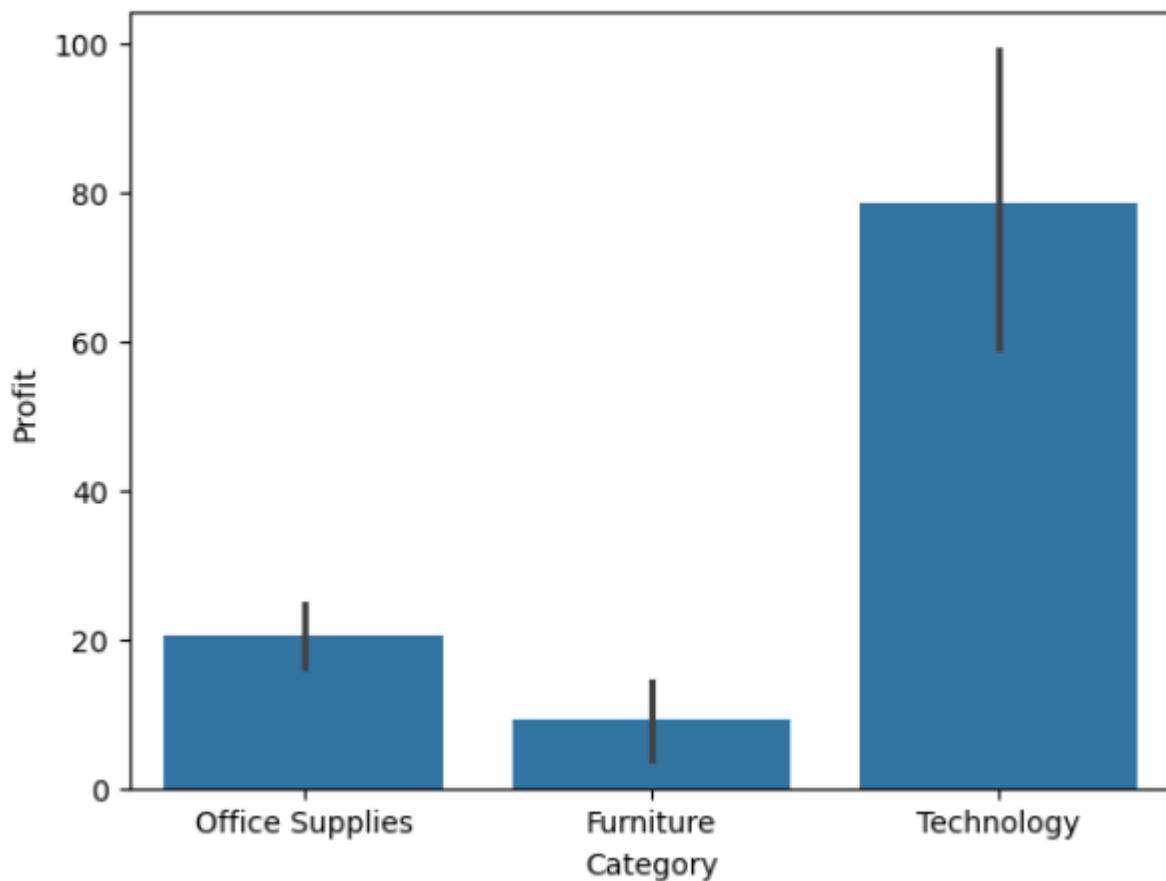


Figure: Profit by Category

<Axes: xlabel='Category', ylabel='Profit'>



8. Step 4 — Power BI Dashboard Development

The final processed dataset was imported into Power BI to create an interactive business dashboard.

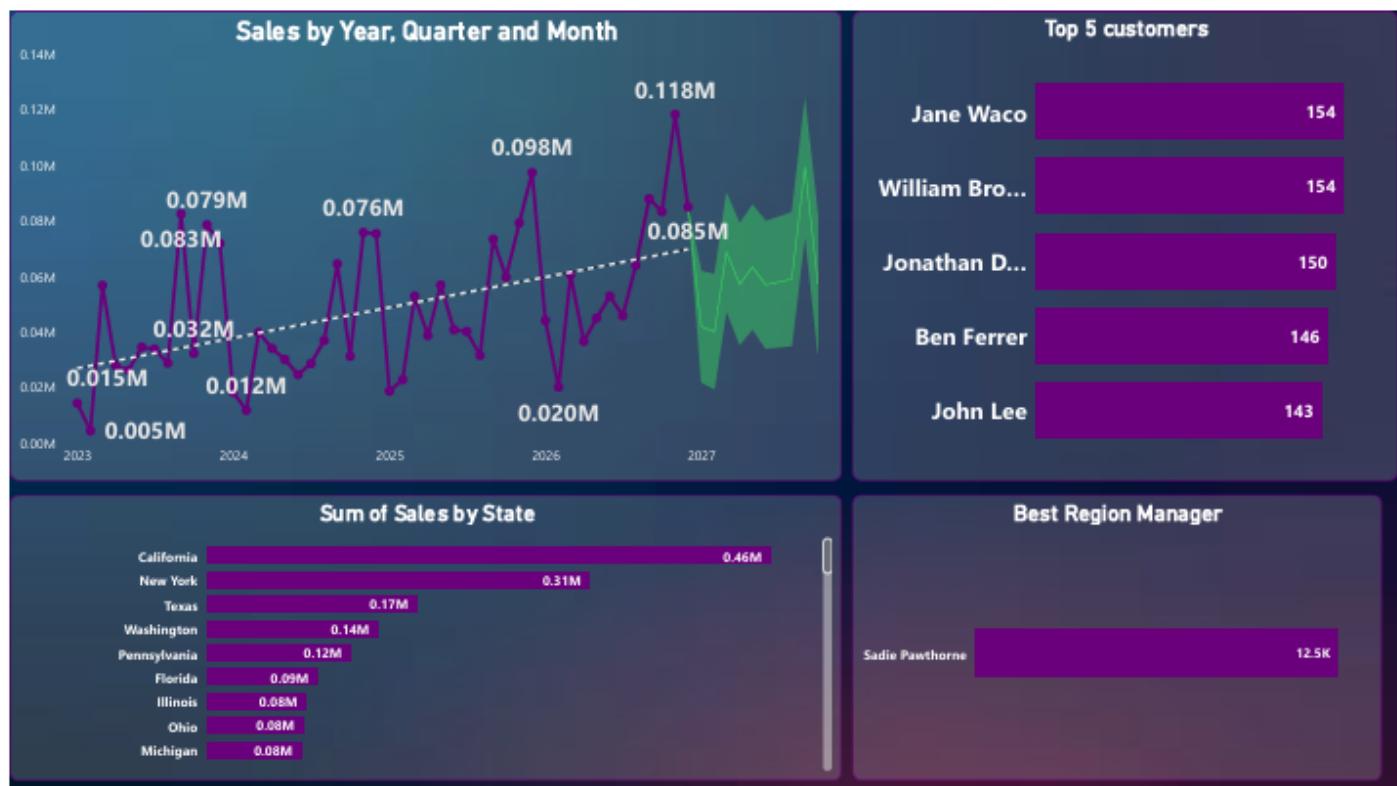
Dashboard Features:

- KPI Cards (Total Sales, Profit, Orders)
- Sales trend over time
- Regional performance analysis
- Category & sub-category comparison
- Top customer analysis
- Interactive filters (Region, Category, Year)

👉 *Main Power BI dashboard overview*



⌚👉 Product performance or regional analysis page





9. Key Performance Indicators

Total Sales: 2.33M+

- **Total Profit:** 292K+
- **Average Delivery Time:** 4 Days
- **Top Performing Region:** West



10. Key Insights

The West region contributed the highest share of revenue.

- The technology category generated strong profitability.
- A small percentage of customers contributed major sales revenue.
- Some sub-categories showed lower profitability, indicating optimization opportunities.



11. Conclusion

- The project analyzed an E-Commerce dataset and achieved a total sales performance of **2.33M** with an overall profit of **292.69K**.
- The dashboard identified **West (31.8%)** and **East (29.7%)** as the highest contributing sales regions.
- The **Consumer segment (50.33%)** generated the maximum share of sales compared to Corporate and Home Office segments.
- Category analysis showed strong performance from **Technology and Furniture**, contributing major revenue to the business.
- Monthly trend analysis indicated steady sales growth with seasonal fluctuations, helping understand demand patterns.
- Overall, the project demonstrates how **Python analysis + Power BI dashboards** can transform raw data into actionable business insights for better decision-making.