

COURSERA DATA ANALYTICS CAPSTONE PROJECT

SUPERSTORE SALES ANALYSIS

Name: Annam Nazeer

Course: Introduction to data analytics (Coursera)

Project Type: Capstone Case Study (Personal Project)

Date: October 2025

Tools Used: Microsoft Excel

Dataset: Superstore Dataset

PROJECT OVERVIEW

Objective

To apply the knowledge and practical skills gained from the course “**Introduction to data analytics**” from Coursera to perform a comprehensive business intelligence analysis on retail sales data, showcasing end-to-end data analysis expertise from data cleaning and exploration to generating meaningful business insights and recommendations.

Personal Learning Goals

- Apply data cleaning methods to messy real-world data
- Master advanced Excel analytics features (PivotTables, statistical functions)
- Develop business insight through data-driven insights
- Build professional quality documentation for career portfolio

Dataset Profile

- **Source:** Kaggle
- **Dataset Name:** Superstore Dataset
- **Records:** 9,994 individual transactions
- **Data Structure:** Relational format with customer, product, geographic, and financial dimensions

What I wanted to find out

My Main Questions

1. Which products sell the most?
2. Which areas make the most money?
3. Average Delivery time across regions?
4. Do big discounts hurt our profits?

WHAT I DID TO CLEAN THE DATA

Data Cleaning Steps I Learned

- Fixed dates so they all looked the same
- Removed duplicate orders as I found some repeated data
- Used TRIM function to remove extra spaces
- Made sure numbers looked like numbers and text looked like text
- Created a new column to calculate delivery time

Problems I Solved

- Some dates were showing up wrong, learned about MM/DD vs DD/MM
- Country and City columns had the same information, deleted the extra one
- Some delivery dates seemed impossible figured out the date format issue

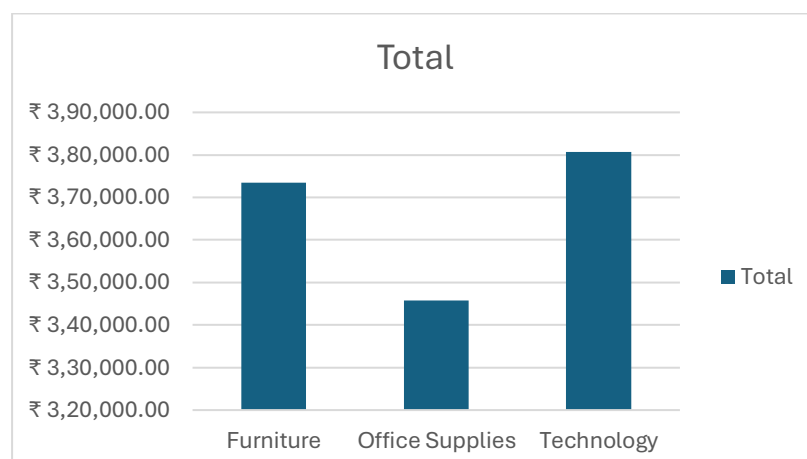
My analysis and what I found

ANALYSIS 1: BEST SELLING PRODUCTS

What I Did: Made a PivotTable and chart

What I Found:

Row Labels	Sum of Sales
Furniture	₹ 3,73,504.72
Office Supplies	₹ 3,45,716.46
Technology	₹ 3,80,640.90
Grand Total	₹ 10,99,862.07



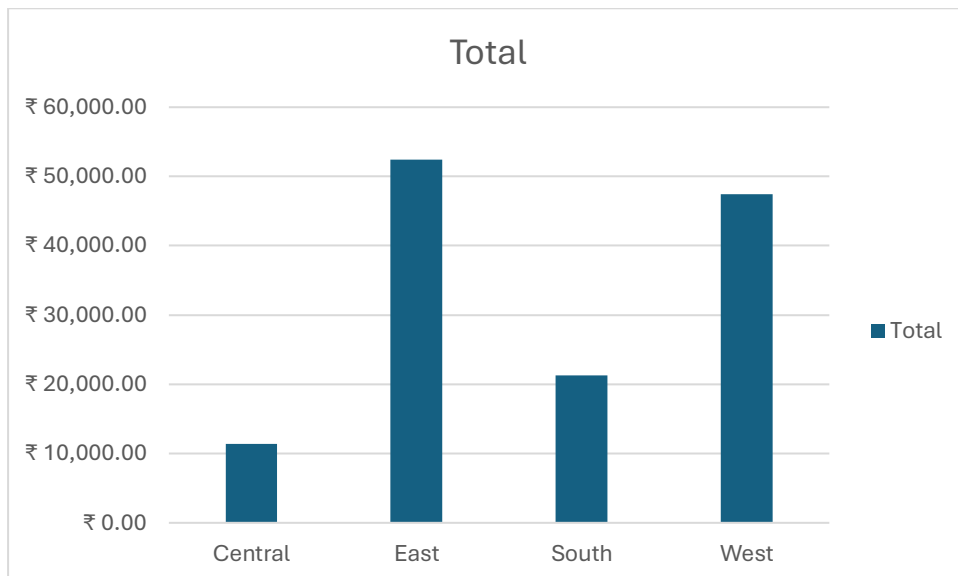
What This Means: The business should focus on these popular products.

ANALYSIS 2: MOST PROFITABLE AREAS

What I Did: Made another PivotTable for regions

What I Found:

Row Labels	Sum of Profit
Central	₹ 11,357.04
East	₹ 52,445.63
South	₹ 21,292.62
West	₹ 47,420.44
Grand Total	₹ 1,32,515.74



What This Means: These areas are doing great while others might need help.

ANALYSIS 3: DELIVERY SPEED

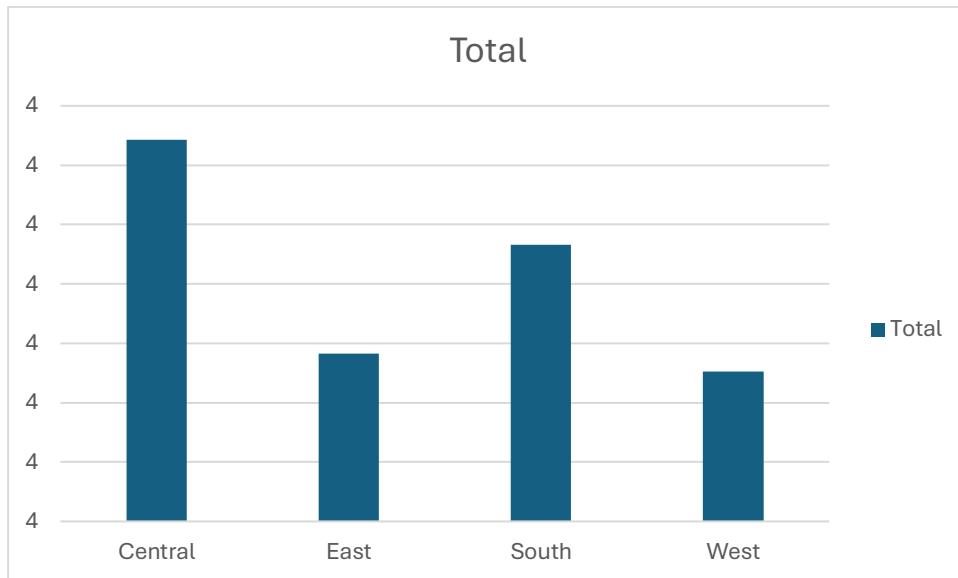
What I Did: Used a formula to calculate delivery days (Ship Date - Order Date)

What I Found: Average delivery is 4 days

Row Labels	Average of Delivery Time
Central	4
East	4
South	4
West	4

Grand
Total

4

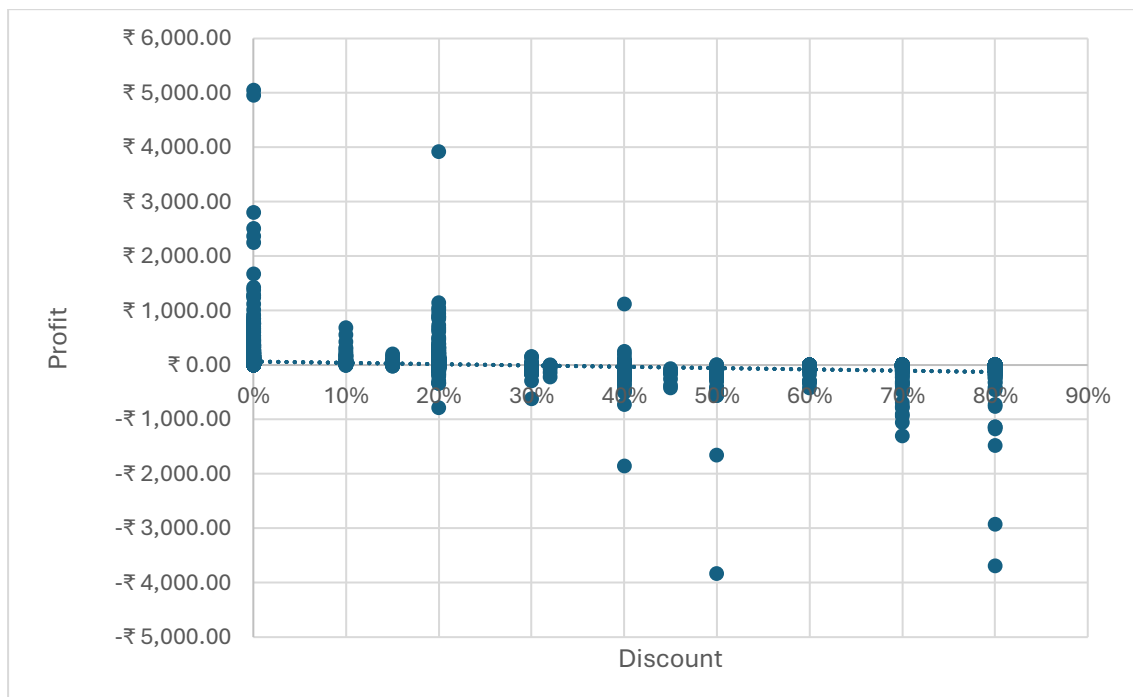


What This Means: Good/bad compared to what customers might expect.

ANALYSIS 4: DISCOUNT PROBLEM

What I Did: Made a scatter plot comparing discounts to profits

What I Found: When discounts are high, the company loses money



What This Means: Big discounts are bad for business

What I learned

Skills I Developed:

- How to clean messy data
- How to make PivotTables
- How to create charts that tell a story
- How to spot business problems in data

Challenging Part:

Fixing the date formats was very difficult but I figured it out eventually.

Resources That Helped Me:

1. How to Clean Data in Excel in 10 Minutes (Free File)- **MyOnlineTrainingHub**
2. 9 Fantastic Functions for Cleaning Dirty Dates in- **Excel-Up4Excel**

Note: - My project demonstrates diagnostic analytics capabilities by identifying root causes behind profitability issues, I applied prescriptive analytics to develop actionable business recommendations, the analysis moved beyond descriptive reporting to strategic insights