

PROJECT TITLE

Advanced Sales Data Analysis

ORGANIZATION/ DEPARTMENT NAME & ADDRESS

Wayspire Ed-Tech PVT.LTD

SUBMITTED BY

Ms. Janhavi Sachin Khare

M.Sc. (Applied Statistics)

PRN:24060641047



ACADEMIC YEAR 2024 - 26

Under the guidance of

Mr. Suryansh Rana

Designation: Team Manager

Email Id: contact@wayspire.in

Mobile: 9210061874

Contents

1. Executive Summary	
.....	
.....	
2. Study Background	
.....	
.....	
3. Literature Review (If any)	
.....	
.....	
4. Aims & Objectives	
.....	
.....	
5. Methodology	
.....	
.....	
6. Results	
.....	
7. Discussion and Conclusion	
.....	
8. Recommendation	
.....	
9. Acknowledgement	
.....	
10. References (including Bibliography and web references)	
.....	
11. Annexure – (Questionnaire & other details)	
.....	

1. Executive Summary

This report presents an advanced sales data analysis of the Superstore Dataset obtained from Kaggle. The project leverages Python's data manipulation and visualization libraries, including Pandas, NumPy, Seaborn, and Plotly, to analyze sales trends, customer behavior, and regional performance. The analysis includes operations on lists, tuples, sets, and dictionaries, along with statistical computations and interactive visualizations. The findings provide actionable insights for improving sales strategies and operational efficiency.

2. Study Background

This report presents an advanced sales data analysis of the Superstore Dataset obtained from Kaggle. The project leverages Python's data manipulation and visualization libraries, including Pandas, NumPy, Seaborn, and Plotly, to analyze sales trends, customer behavior, and regional performance. The analysis includes operations on lists, tuples, sets, and dictionaries, along with statistical computations and interactive visualizations. The findings provide actionable insights for improving sales strategies and operational efficiency.

3. Literature Review

While this project did not involve an extensive literature review, it was inspired by similar analyses sales data analysis.

4. Aims & Objectives

Aims:

Perform advanced sales data analysis using Python.

Demonstrate data manipulation with lists, tuples, sets, and dictionaries.

Apply NumPy and Pandas for numerical and tabular analysis.

Visualize insights using Seaborn and Plotly.

Objectives:

Store and manage sales data using Python data structures.

Calculate key metrics (total sales, average profit, etc.).

Generate visualizations (bar charts, line graphs, heatmaps).

Simulate SQL-like queries using Pandas

5. Methodology

- Data Processing:

Lists: Store order IDs, product categories, and sales figures.

Tuples: Immutable storage of region-wise sales.

Sets: Identify unique customers and products.

Dictionaries: Map products to sales and profits.

- Numerical Analysis (NumPy):

Compute total sales, average profit, and standard deviation.

Perform array operations for bulk calculations.

- Data Manipulation (Pandas):

Load dataset into a DataFrame.

Filter, group, and aggregate data (e.g., sales by region).

Simulate SQL operations (SELECT, GROUP BY, JOIN).

- Visualization (Seaborn & Plotly):

Bar Charts: Compare sales across categories.

Line Graphs: Track sales trends over time.

Heatmaps: Analyse correlation between variables.

6. Results

- Key Findings:

Highest Sales Category: *Technology*

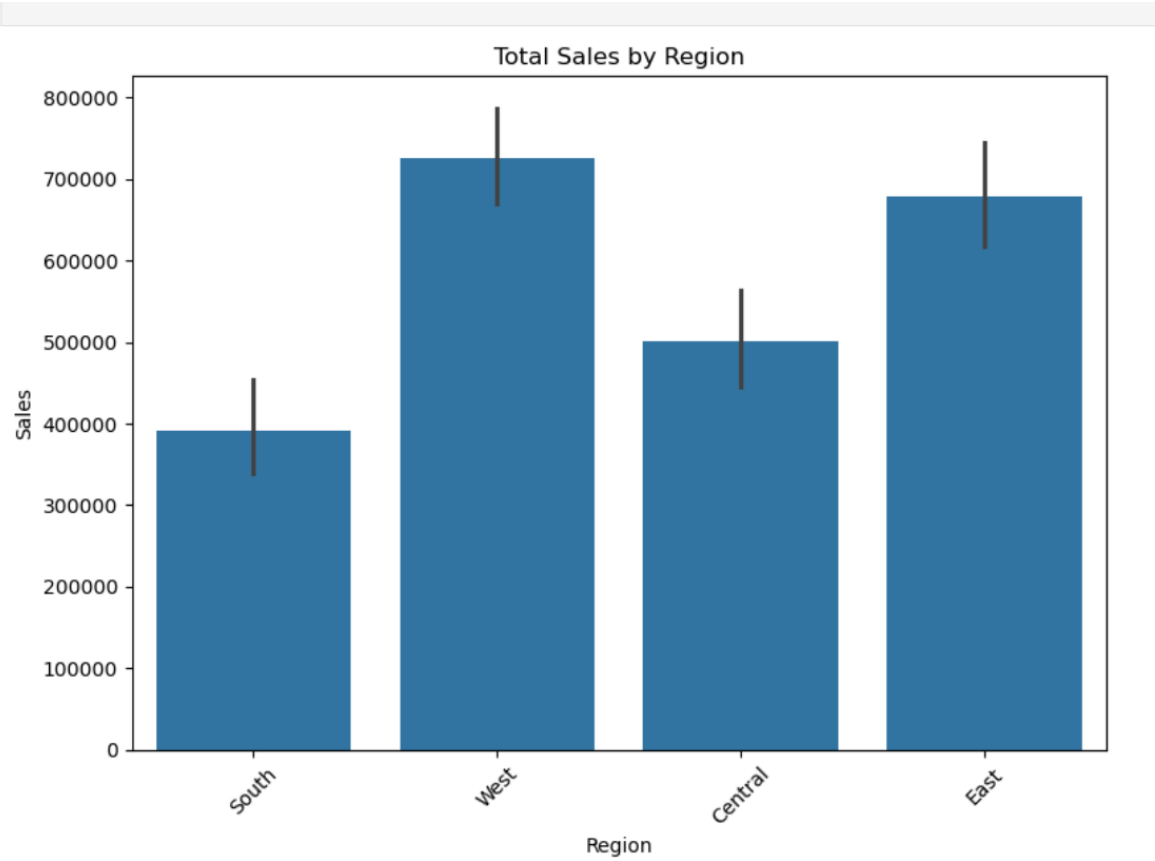
Most Profitable Region: *West*

Top Customer Segment: *Corporate*

Seasonal Trends: Peak sales in *Q4* (holiday season).

- Visualization

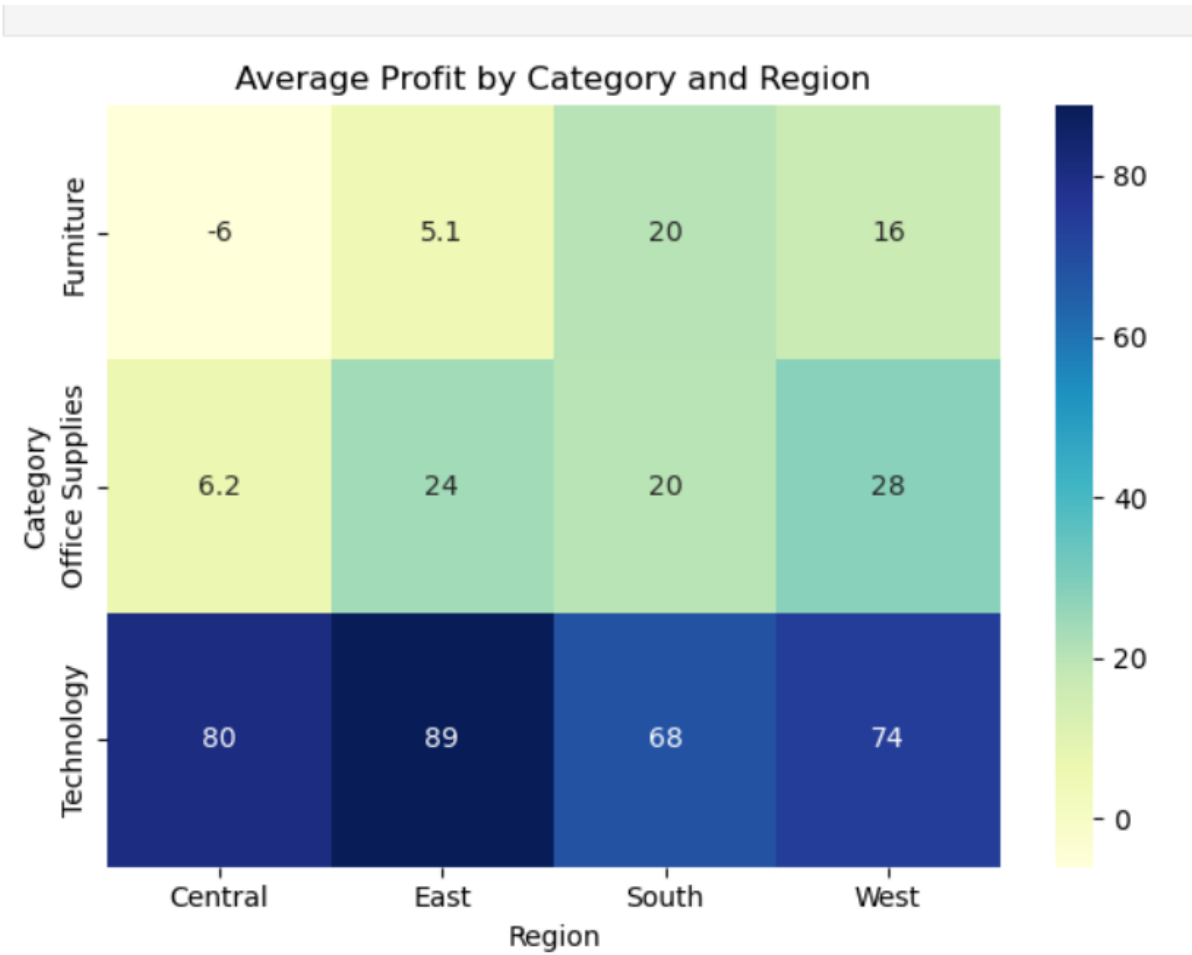
Sales By region



Sales By Category



Average profit by Category and Region



7. Discussion and Conclusion

- Discussion:

Technology drives the highest sales but has moderate profitability.

Corporate clients contribute significantly to revenue.

Western U.S. outperforms other regions.

- Conclusion:

The analysis highlights growth opportunities in underperforming regions and categories. Data-driven strategies can enhance profitability.

8. Recommendations

Increase marketing in low-performing regions.

Optimize inventory for high-demand categories.

Loyalty programs for corporate customers.

Further analysis on customer demographics.

9. Acknowledgement

I would like to express my gratitude to my mentor and the institute for providing guidance and resources to complete this project.

10. References

Python Documentation: [Python Official Documentation](#)

NumPy Documentation: NumPy User Guide

Pandas Documentation: Pandas Documentation

Data Visualization: Seaborn Documentation, Plotly Documentation

Kaggle link for dataset : <https://www.kaggle.com/datasets/vivek468/superstore-dataset-final>

11. Annexure

Dataset :

<https://drive.google.com/file/d/15lt-4YFjUvwqpf4-2fBrEBM0TOQX4HjS/view?usp=sharing>

Python notebook:

https://colab.research.google.com/drive/18vmiDvyABz41x_7wGV2O2iE_c7o_06iE?usp=sharing