

The Project and Data Management (PDM) Plan

Project Overview

The project plan include the following

- **Project Title:**

Analysis of Superstore Sales Data

- **Summary:**

This project will explore a retail sales dataset from a superstore, focusing on analyzing sales trends, customer segments, product performance, and profitability. The dataset includes transaction-level details such as sales, quantities, discounts, and profit, along with customer and product information. The aim is to extract actionable insights that can help optimize inventory management, sales strategies, and overall business performance.

- **Research Question:**

How do factors like product category, customer segment, and discount rates influence sales and profitability in the Superstore dataset?

- **Project Objectives:**

To explore the relationship between product categories, customer segments, and their corresponding sales and profit performance.

To identify the impact of discounts on sales volume and profitability.

To analyze sales trends over time and by different geographic regions.

To segment customers based on their purchasing behavior and evaluate their contribution to overall profits.

- **Reference List:**

Here are three peer-reviewed papers relevant to the project research question:

Chong, A.Y.L., Chan, F.T.S. & Ponnambalam, S.G. (2020). *The Impact of Big Data Analytics on Supply Chain Management: A Literature Review and Future Research Agenda. Computers & Industrial Engineering*, 149, 106926.

Available at: [link](#)

Davenport, T.H. & Harris, J.G. (2017). *Competing on Analytics: The New Science of Winning. Harvard Business Review Press.*

Available at: [link](#)

Zhao, X. & Xie, J. (2019). *The Effect of Pricing and Discounting on Retail Performance: Evidence from a Large-Scale Data Set.* *Journal of Retailing*, 95(4), 1-15.
Available at: [link](#)

Task List:

Explore and Clean the Dataset:

- Load the Superstore dataset.
- Handle missing values, correct data types, and remove duplicates/outliers.

Define Project Objectives and KPIs:

- Identify key metrics such as total sales, profit, quantity, discount impact, etc.
- Frame questions to guide the analysis (e.g., "Which region is most profitable?").

Perform Sales Analysis:

- Analyze sales by category, sub-category, and region.
- Identify top-performing products and customer segments.

Analyze Profit and Discount Trends:

- Evaluate the effect of discounts on profit.
- Compare profit margins across regions and product lines.

Develop Visualizations

- Create charts (bar, line, heatmaps, etc.) using Python (matplotlib/seaborn) or Power BI.
- Visualize trends in sales, profit, shipping time, and customer demographics.

Start Literature Review:

- Begin reviewing academic and industry sources on sales analytics and data visualization techniques.
- Summarize findings relevant to your approach.

Create Interactive Dashboards:

- Build multi-page dashboards (Power BI or Python with Plotly/Dash).
- Include filters, slicers, and drill-down capabilities.

Draft Initial Sections of the Report:

- Write Introduction, Objectives, and Methodology sections.

- Describe your data sources, tools used, and analysis approach.

Summarize Key Insights:

- Highlight actionable findings and business recommendations.
- Compare them with findings from your literature review.

Prepare for Progress Report:

- Create a PowerPoint with visuals and summary.
- Draft the written progress report for submission.

Complete Final Report:

- Write Literature Review, Results, Discussion, and Conclusion sections.
- Add figures, charts, and reference list.

Proofread and Format the Report:

- Edit for clarity, grammar, and flow.
- Format headings, page numbers, appendices, and citations.

Prepare Final Presentation:

- Build and rehearse a short presentation covering your project journey and findings.

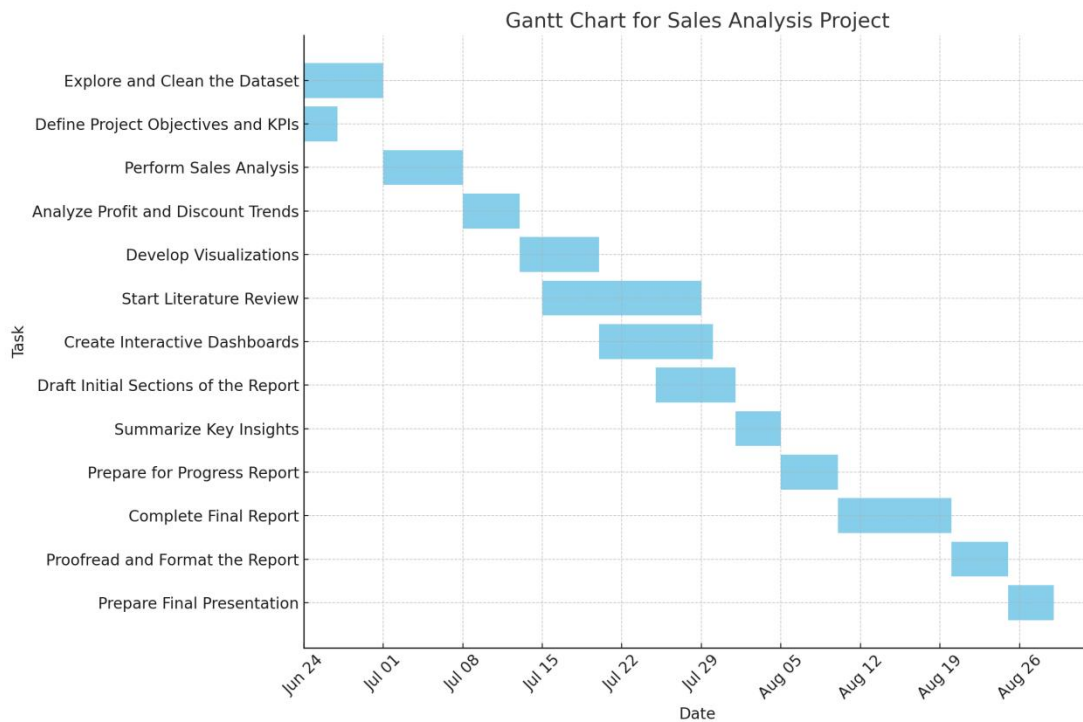
Assessment Milestones:

Progress Report Due: Submit analysis findings and project progress.

Final Presentation Due: Present your insights and visualizations to the class.

Final Report Due: Submit the complete written report with visuals, discussion, and references.

Task List Gantt Chart:



Data Management Plan

Overview of the Dataset

The **Sample - Superstore** dataset is a fictional but commonly used dataset provided by **Kaggle** originally designed for demonstrating data visualization features within the Tableau BI tool. The data simulates sales transactions for a US-based superstore, covering information such as customer details, product categories, order dates, sales, profit, shipping details, and regions.

Country of origin: United States

Original source: Kaggle

Original purpose: Educational/business intelligence tool demo

Data Collection

Source: The dataset is freely available online for public use and academic projects.

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

Collection method: Downloaded as a CSV file manually from the Tableau Community website.

Metadata

Attribute	Details
Format	CSV (Sample - Superstore.csv)
Size	Approx. 51 KB
Records	~9,994 rows (individual orders)
Columns	21 columns including Sales, Profit, Region, Segment, Category, etc.
Code format	Python scripts (.ipynb, .py) and Power BI files (.pbix)
Expected code size	~5–10 MB including visuals and reports

Document Control

GitHub Repository: <https://github.com/UmerCheena/FYP-PDM->

Ethical Requirements

● GDPR Compliance:

The dataset contains no real personal information. It's synthetic, anonymized, and publicly distributed for training and educational use, so **GDPR does not apply**.

● UH Ethical Policies:

The dataset does not involve human participants or sensitive personal data, and conforms with **University of Hertfordshire ethical guidelines** for student projects.

- **Permission to Use:**

The dataset is freely and publicly available. No restrictions are placed on usage for academic or non-commercial purposes. Documentation from Tableau confirms this.

- **Ethical Data Collection:**

The data is synthetic and was created by Tableau for educational/business demo purposes. No real individuals were involved, thus **ethical collection standards were met**.