CITY UNIVERSITY MALAYSIA

Est. 1984

Faculty of Information Technology (FoIT)

Programme	BIT
Subject Code	BIT2053
Subject Name	Fundamental Of Modern Data
Lecturer	Sir Nazmirul Izzad Bin Nassir
Document	Project Report
Due Date	03-Sep-2025
Student Names & IDs	Luqman Siraj - 202309010301 Mazharsolook Arshia – 202309010247
Dashboard Link	https://lookerstudio.google.com/s/vblvOgWGacU
GitHub Repository Link	https://github.com/SirajLuqman/Superstore-Sales-Profit-Analysis/tree/main

Contents

Superstore Sales & Profit Analysis - A Data-Driven Approach

1. Business Scenario & Problem Definition

Modern retail market can be distinguished by the high competition and constantly changing of market conditions. To be successful, companies should abandon previous organizational structures and base on the decisions bases, made on the basis of data. With these obstacles in contention, Superstore is seeking a better understanding of its sales and different profits based on various variables such as the product classifications, geographical location, and time. This project meets that requirement through simulations of an imaginary data-driven analysis that transforms raw (sale) data into useful business knowledge.

This analysis is driven by three core business questions, which will guide the exploration and interpretation of the data.

1.1. Key Business Questions to Address

❖ **Business Question 1: How do sales and profitability compare across different product categories?**

➢ Essentially, we are trying to determine how the three core product lines Tech, Furniture and Office Supplies outweigh. When we examine total sales and profit margins per category, we can identify the largest contributors to revenue, and determine which of the lines should have a strategy rethink to help it ascend up the rankings on the profitability table.

❖ **Business Question 2: Which geographic regions and cities are the top contributors to sales, and which ones are underperforming?**

➢ This is where we are excavating our geography on our sales. Identifying the strong areas (ex-such as West and East) and their principal cities will enable us to identify our treasure markets and implement win-based strategies in other areas. Meanwhile, it is shockingly important to identify falling areas (Central, South, etc.) in order to devise specific growth actions in those locations.

❖ **Business Question 3: What are the underlying trends and patterns in sales over time, and what insights can be gained from seasonal fluctuations?**

➢ This time-based analysis will assist in getting oriented to the business cycle. Moving through the sales figures of 2014 to 2017 allows us to pick up on

long-term patterns, recurring peaks, and the way the seasons mix everything up. He or she uses those insights to make inventory choices, marketing impetuses, and staffing scripts so that we are prepared when demand is suddenly high.

2. Source and Pre-process a Dataset

2.1. Overview of Dataset

The data set in this project is the **Superstore Dataset** which is a public and usable dataset commonly used in business intelligence analysis. This information was obtained on **Kaggle** an excellent data scientist and data analyst site. This data was found in one single file of format.csv. which was converted in Google Sheet to connect with Looker Studio.

2.2. Data Preprocessing and Cleaning

To clean, normalise and be analysis-ready, the following preprocessing functions were performed:

- ❖ **Data Acquisition:** The dataset was downloaded from the Kaggle repository and loaded into a data analysis environment.
- ❖ **Data Inspection:** We have first conducted some initial inspection to learn how the data is structured such as the number of rows and columns, types of data and their presence of anomalies.
- ❖ **Handling Duplicates:** We verified the issue of duplicated records and we eliminated the duplicate rows to ensure integrity of our analysis and avoid over counting of sales.
- ❖ **Missing Value Management:** We identified the presence of any missing or null value in a column. On each discovery we selected a strategy that depends on the context of the data--like imputation in a numerical field or the deletion of rows where rows missing information would spoil the analysis. There was no major missing value in this particular set which needed complicated imputation.
- ❖ **Data Type Conversion:** We had checked the data types of some columns and transformed them when necessary. To illustrate, we made sure that the column of date, namely order date, was formatted correctly, i.e., as a date field and that fields such as sales and profit were expressed using numbers, and not as a string.

- ❖ **Standardization of Text Data:** We examined some text categories i.e. Category and Region, to remove any spelling or capitalization errors. The benefit of this step is that, one category like Technology should not be considered two different categories because of minor variations in spelling.
- ❖ **Feature Engineering (as needed):** Can't say that it is large but we extracted partial features of some fields to assist the analysis. As an example, a field of Profit Margin was calculated dividing Profit by Sales in providing another metric to compare.

These interventions played an essential role in converting the unclean data into normal and structured data that can be used directly in the Business Intelligence (BI) tool.

3. Data Analysis

Through crunching of descriptive statistics and some simple analysis we identified important trends, patterns of distribution and relationships in the Superstore data. We narrowed down the business questions, to unravel valuable insights.

- ❖ **Overall Metrics:** The Superstore incurred RM2,297,204.86 in sales and earned RM286,397.02 in the profit made between **Jan 2, 2014 and Dec 30, 2017**. These macro numbers precondition all the finer-grained digs that further arise.
- ❖ **Sales by Category:** Technology is the leading in the category, with RM836,154.03 followed by Furniture (RM741,999.79) and Office Supplies (RM719,047.06). So, all the categories are important, nevertheless, Tech is the source of primary revenues to the business.
- ❖ **Regional Performance:** The West and East segment is the obviously the largest in terms of revenue. The west contributes 31.6 percent of all sales and the east follows with a close 29.5 percent. Combined, they occupy a greater part of the revenue of the business, more than 60%. Conversely Central and South regions are still behind with 21.4 percent and 17.5 percent respectively and this indicates unexploited growth potential.
- ❖ **Top-Performing Cities:** A further analysis of the city data will reveal that the city of NYC, LA and Seattle are the power brasses as far as sales are concerned. Such target at major metros highlights the importance of such major markets towards the success of the company.

❖ **Sales Over Time:** The time- series data is not a straight line- we can observe distinct peaks and troughs indicating the presence of high seasonality or promotion. The peaks tend to burst at the end of the year perhaps due to the holiday shopping.

4. Visualization Using a BI Tool

4.1. Data Visualization using Google Looker Studio

In order to present the analysed data in a manner that is visualised and relevant to stakeholders, we processed the data through Looker Studio, a powerful and interactive Business Intelligence (BI) tool and a Google product. Clear and user-friendly layout was enhanced into the dashboard to enable you get into the data fast and efficiently.

4.2.   Interactive Dashboard Development

Looker Studio Dashboard: https://lookerstudio.google.com/s/vblvOgWGacU
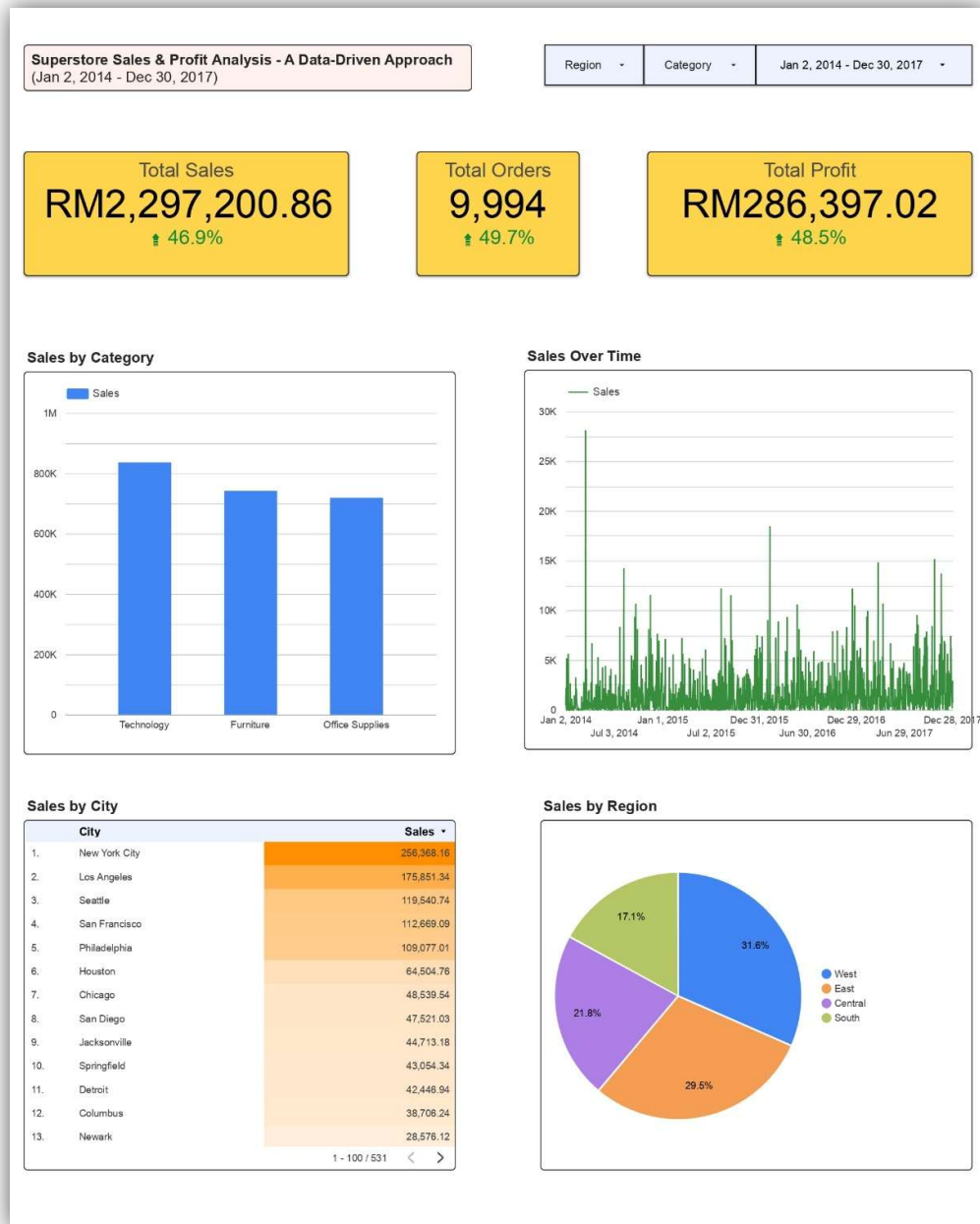

Figure 1: Dashboard

The visualisations and interactions that were stapled to the dashboard are:

Figure 2: KPI Scorecards

❖ **KPI Scorecards:** The dashboard introduces four scorecards that provide a fast at-a-glance representation of most important business metrics: Total Sales (RM2,297, 200.86), Total Profit (RM286,397.02), and Total Orders (9,994).

Figure 3: Bar Chart

❖ **Sales by Category (Bar Chart):** This bar chart represents the sales of every product category Technology, Furniture and Office Supplies. A bar chart where each bar proved easy to understand and visualise the input of sales performance.
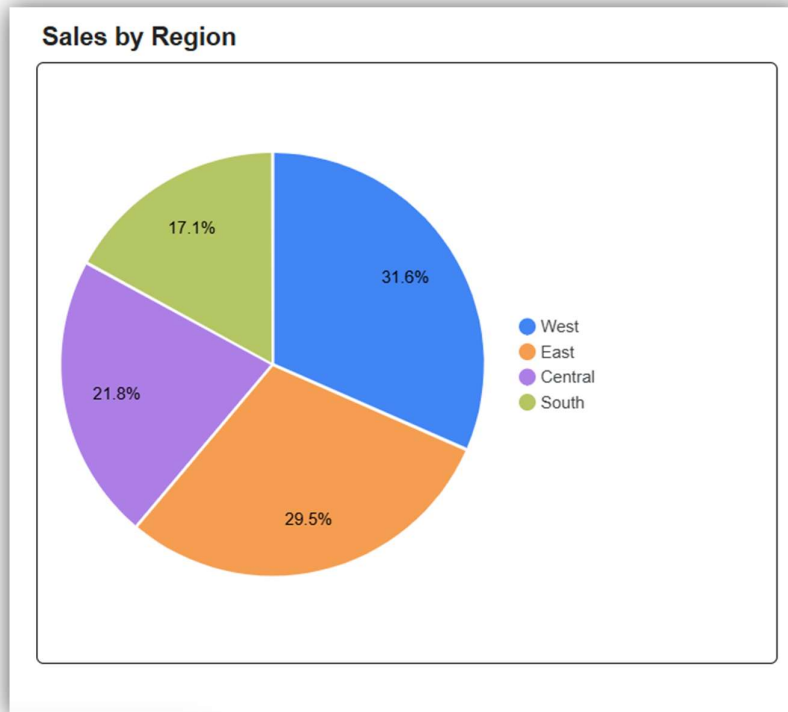
Figure 4: Pie Chart

❖ **Sales by Region (Pie Chart):** Pie chart provides a clean-cut overview of sales contributions by region West, East, Central, and South. It is immediately clear that the dominance of the West and the East is being evident here.
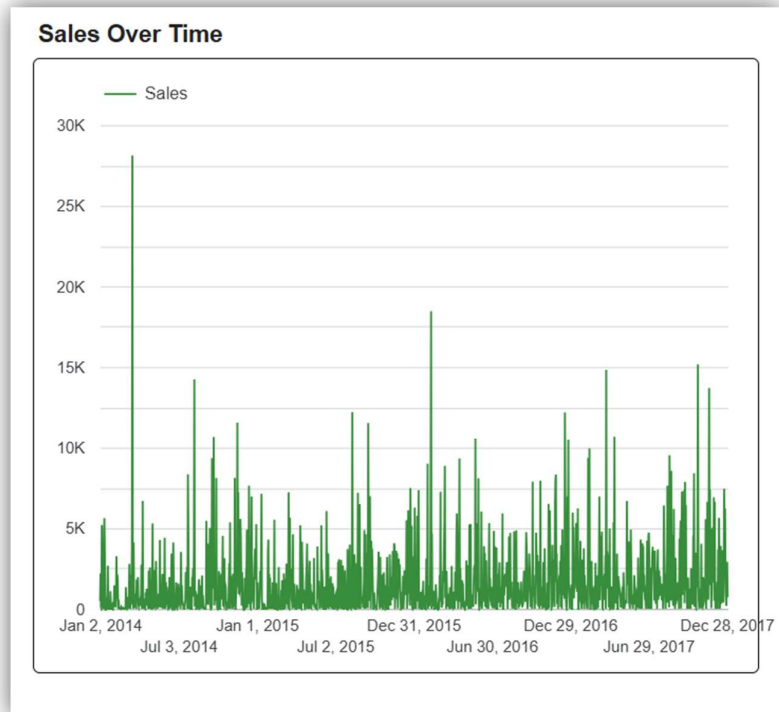
Figure 5: Line Chart

❖ **Sales Over Time (Line Chart):** The line chart is used to monitor sales performance on timeline between 2014 and 2017. It is important to identify trends, seasonal agreement and even major moments to experience high or low sales.

Figure 6: Table

❖ **Sales by City (Table):** Table summarizes by listing all sales by city in detail and order. This will enable users determine the governmental markets that are performing the best and their specific sales data, easily.

Figure 7: Interactive Filters

❖ **Interactive Filters:** Slicers will be conveniently located on the dashboard so that it was possible to dynamically filter according to the points: "Region" and "Category". All the charts on the dashboard are automatically updated by clicking on a single or several regions or categories, creating opportunities to analyze to an extreme depth, allowing stakeholders to investigate the details of certain groups of data independently.

5. Interpret & Recommend

Our analysis of the Superstore sales data has produced some quite solid, data-conscious knowledge that is in fact useful to stimulate business growth. It was based on these insights that actionable recommendations and strategies can proceed.

5.1. Insights and Findings

❖ **Insight 1: The Technology category is the main driver of sales, but its profitability needs a closer look.**

➢ We observe that there is the highest sales volume in Tech products. But the profit margin on this category is not clearly described in the charts we have and therefore we cannot be confident that those high figures are actually benefiting the bottom line. Perhaps the large volume is covering not as high margins as other categories would supply.

❖ **Insight 2: The West and East regions are the core markets, while the Central and South regions are underperforming.**

➢ Analysis of the regional sales reveals a huge imbalance: two regions contribute to more than 60 per cent. of all sales. The implication of that is that although the company enjoys a good market penetration in certain regions, there is a distinct growth potential in others.

❖ **Insight 3: Sales are highly seasonal, with noticeable peaks at the end of each year.**

➢ There is a consistent cyclical trend reflected in the time-series chart but spikes are observed towards the end of the year. The sales pattern seems to wave a lot due to seasonal influences-most probably there was the holiday shopping spurt.

5.2. Recommended Business Strategies

❖ **Recommendation 1: Optimize Product Sourcing and Pricing Strategies for the Technology Category.**

➢ The firm must conduct a profiling profiteering examination of all the Tech products. In case the margins turn out to be low, the firm would consider venturing into new sourcing to reduce costs. Simultaneously, the pricing strategies should be overturned to have a mismatch between remaining competitive and maintaining healthy bottom lines. The move will assist in making the high sales volume translate into a further contribution to the bottom line of the company.

- ❖ **Recommendation 2: Implement Targeted Market Expansion and Marketing Campaigns.**
  - ➢ The Superstore must devise special marketing and sales strategies in the underperforming regions of Central and South. Consider the idea of target ads, special offers, cooperating with local companies to enhance brand presence. The team can use pulling of Western and Eastern data of what is selling and what was successful to draw out a clear map that will be used to boost sales and win a larger market share in those harder regions.
- ❖ **Recommendation 3: Optimize Inventory and Marketing Campaigns Based on Seasonality.**
  - ➢ One way in which such seasonal insights should inform our operations is by letting them do so. Keeping inventory tight means anticipating those peak periods and stocking up so that it is available during those periods, but not taken up during low-need periods. Concurrently, the timing marketing takes advantage of the potential doubling of the effect, and increasing sales, just before the peaks. This foresight approach will reduce holding costs during low business and it will avoid product shortages during high business times.

6. Timeline

- ❖ **Week 1:** Project proposal submission, dataset sourcing, and initial data exploration.
- ❖ **Week 2:** Pre-processing and cleaning of the dataset.
- ❖ **Week 3:** BI tool familiarization, data analysis, and dashboard creation with the specified visualizations.
- ❖ **Week 4:** Interpretation of findings, formulation of business recommendations, and drafting of the final report.
- ❖ **Week 5:** Final review and submission.

7. Reference

- Superstore Dataset. (n.d.). Kaggle. Retrieved from https://www.kaggle.com/datasets/vivek468/superstore-dataset-final