

**MCSD1123- 01 BIG DATA MANAGEMENT**

**Assignment 1A Report**

**Lecturer:**

PM DR. Mohd Shahizan Othman

| **Members** | **Matric NO.** |
| --- | --- |
| Hazem Taher Omran Fenneer | MCS231019 |
| Shivanesh | MCS231014 |
| Mustafa | MCS212012 |
| Nur Shahirah | MEC233005 |

**INTRODUCTION**

The case study focuses on the data processing and visualization of a dataset named dataset2.txt. This dataset contains information on the sales performance in two years, focusing on customers, products, sales personals, sales region, target sales, total sales, sales years, sales months and sales quarters. The primary objectives of this case study are to transform and analyze the data in Google Sheets and create a visualization dashboard to gain insights into the dataset.

**Data Entry**

We imported that data from the .text file into google sheet using the import function. To import the .txt file into Google Sheet, first choose the ‘File’ tab and click ‘Import’. Then, it will appear as the figure 1 below. Choose ‘Import data’ and the data will automatically be arranged with the column header in Dataset2 sheet.

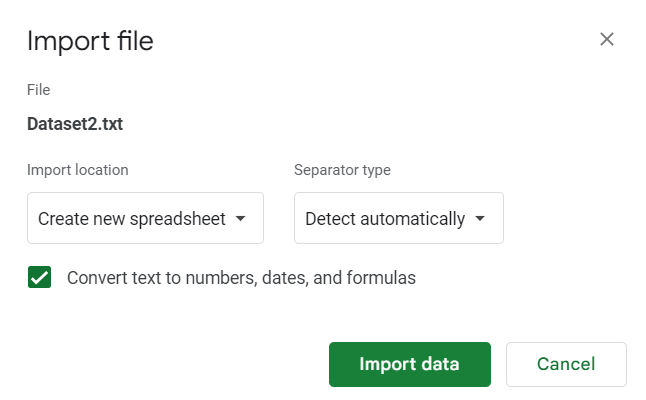


Figure 1: Import data

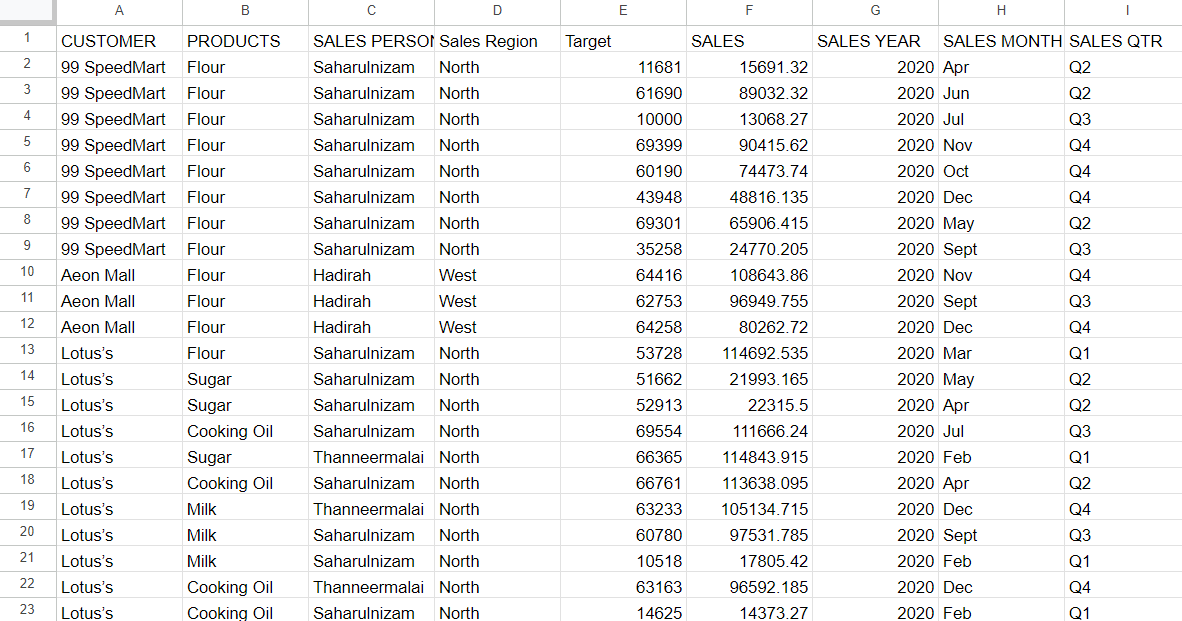


Figure 2: Result of import data function

As shown in figure 2, the data was processed and clear, but we have inspected it to ensure that it is clean without white spaces or interference. After we ensured that the data was cleared, we started to visualize it. Create another sheet.

Data Processing

In order to gain a comprehensive understanding of data visualization and processing, it is imperative to first establish the objectives of the task and delineate the essential components within the dataset. This dataset pertains to the sales performance of Rising Star across distinct regions for the years 2020 and 2021. The dataset encompasses the following attributes:

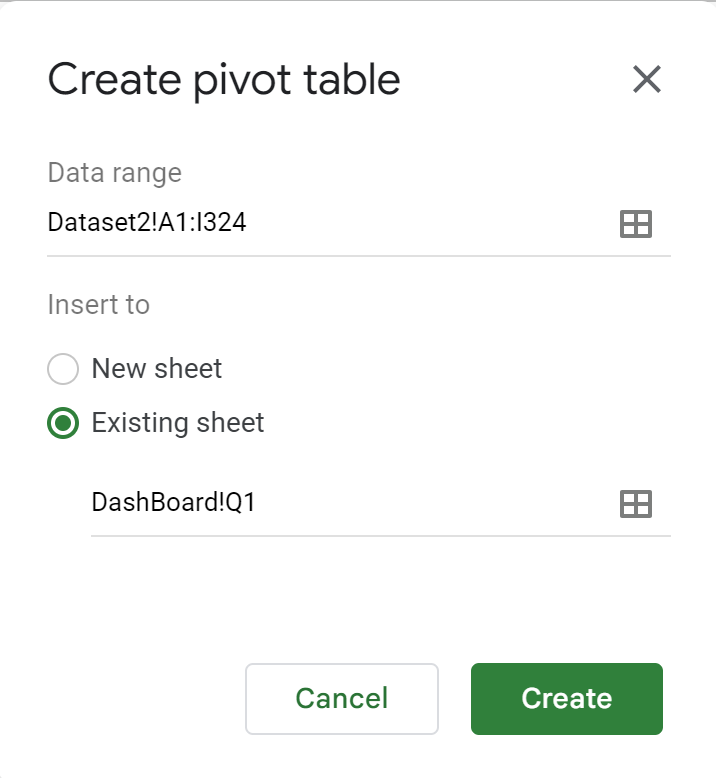
* Customers
* Products
* Sales Personnel
* Sales Region
* Target
* Sales
* Sales Year
* Sales Month
* Sales Quarter

The primary aim of this project is to develop an interactive dashboard that incorporates the following key elements:

* Sales Year
* Sales Region
* Products
* Sales Personnel
* Customers

Each element within this dashboard contains valuable information that can be seamlessly harnessed to manipulate graphs and diagrams, ultimately providing the insights and visualization capabilities required by Rising Star.

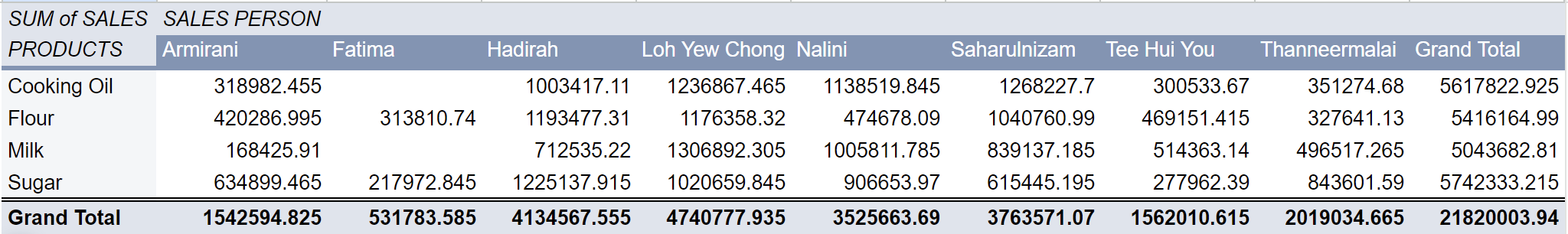
To effectively process the data, we have employed pivot tables for each chart. This strategic approach ensures that the data is accurately represented. The process involved creating a dedicated sheet to host all the pivot tables, achieved through the "Insert" menu and selecting "Pivot Table (into an existing sheet)." Once the data accuracy was confirmed, we embarked on the visualization phase.

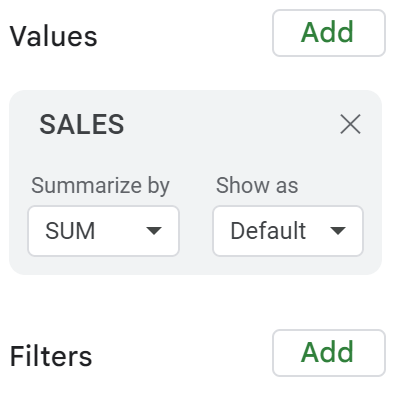
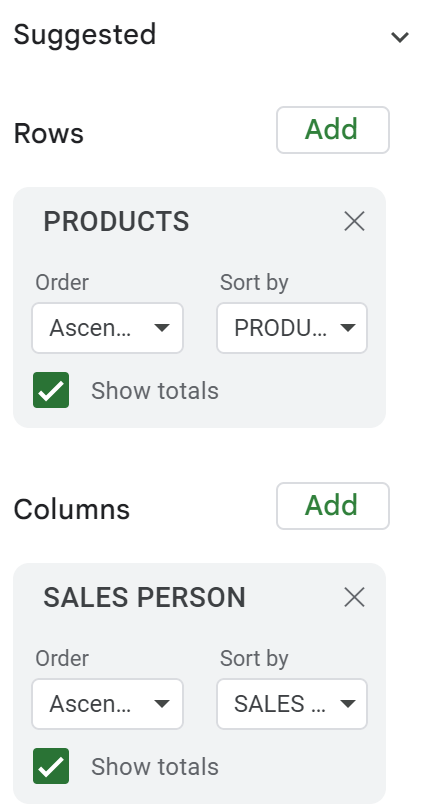


(figure 3: creating a pivot table)

Within the pivot table framework, we meticulously constructed five distinct tables, each tailored to provide indispensable data for chart creation. These tables encompass the following critical dimensions:

* Products Enriched with Seller and Sales Insights: This table amalgamates product data with associated seller details and pertinent sales information, fostering a comprehensive understanding of the product sales landscape.
* Sales Region Aggregated with Total Sales: Here, we judiciously summarize the sales figures for each region, affording a clear overview of regional sales performance.
* Customers' Cumulative Sales: This table succinctly compiles and presents the sum of sales for each customer, facilitating the identification of high-value clients and market trends.
* Sales Month: Sum of Sales and Target Sales: This table provides a detailed breakdown of sales performance, showcasing the total sales achieved during each month and comparing it against the set target sales.
* Sales Year and Month: Total Sales Summation: In this table, we meticulously aggregate sales data, providing insights into both the yearly and monthly sales figures, thereby offering a comprehensive view of the sales trajectory.
* These tables serve as the cornerstone of our data visualization initiative, underpinning the creation of informative and insightful charts and graphs.





(figure 4: the created pivot table and its values)

**Data visualization**

After the creation of the pivot tables, we proceeded with the data visualization. We methodically ensured that each table contained the requisite inputs for accurate data representation, all while considering the most suitable chart types for the task at hand.

The visualization process involved selection of chart types:

1. Pie Charts for Region and Customer:

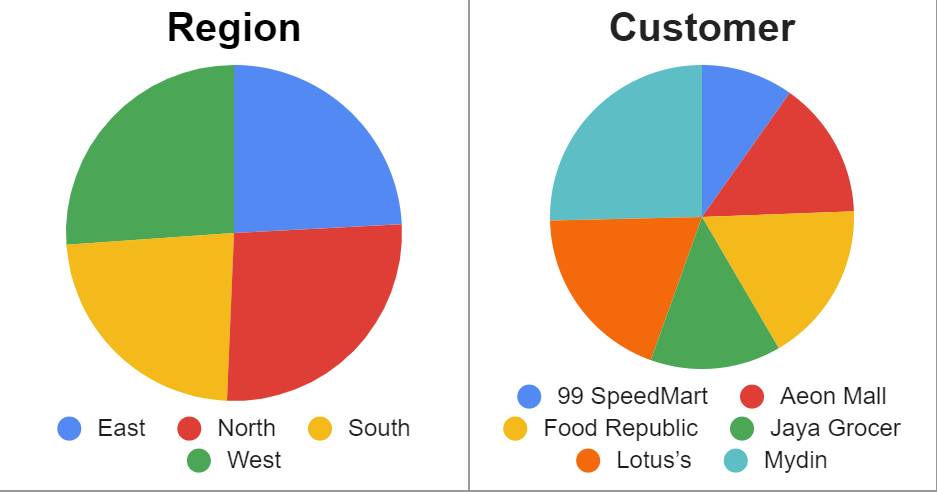


Figure 5: Illustrate regional and customer-related insights, we used the simplicity of pie charts. With only two key columns necessary for representation, these charts effectively conveyed the distribution of sales within regions and among customers.

1. Stacked Charts for Product Sales:

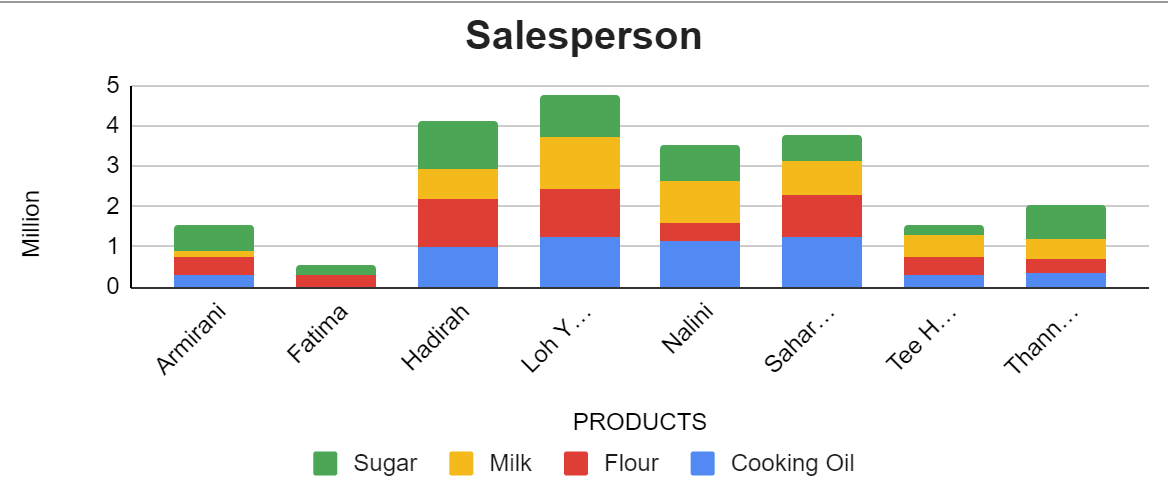


Figure 6: We employed stacked charts to delve into the intricate web of product sales, providing a holistic view of which products were sold by specific sellers and the corresponding sales volumes. The stacking mechanism enabled a clear breakdown, considering factors such as who is the seller, what products that are offerings, and the sales volumes.

1. Line Charts for Sales Trend Analysis:

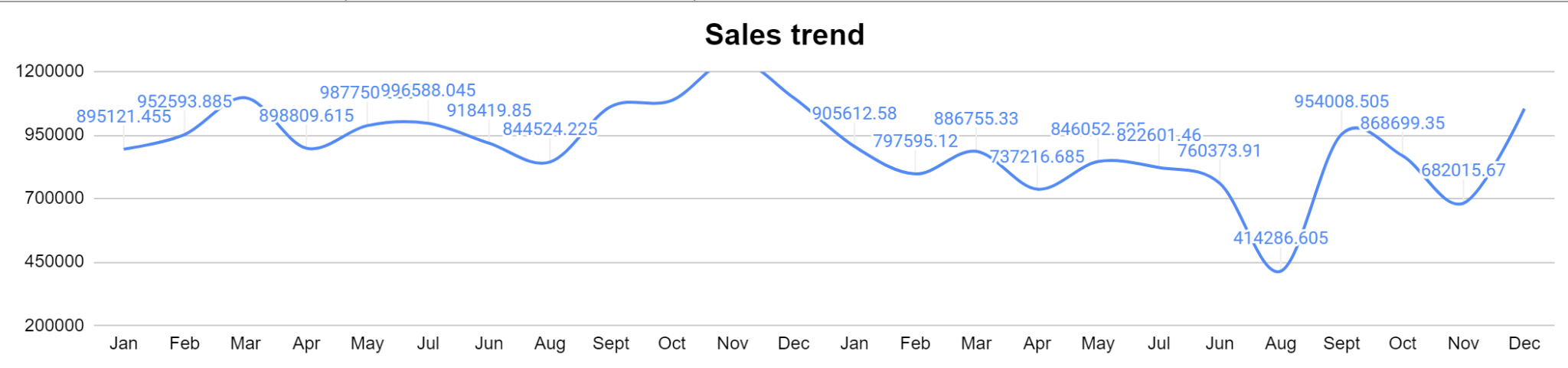


Figure 7: For comprehensive trend analysis of sales data, line charts were chosen. These charts encapsulated sales dynamics over time, revealing patterns and fluctuations.

1. Comparison Chart for Sales vs. Target:

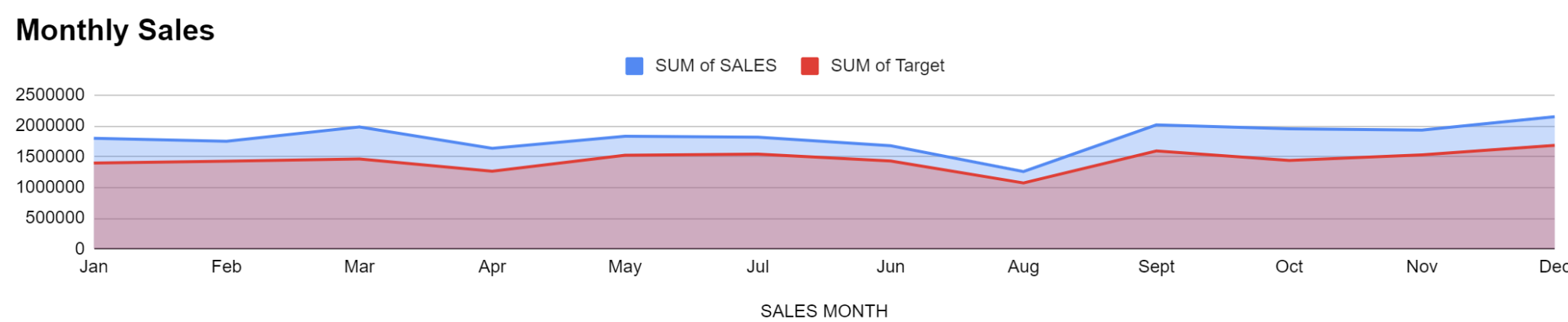
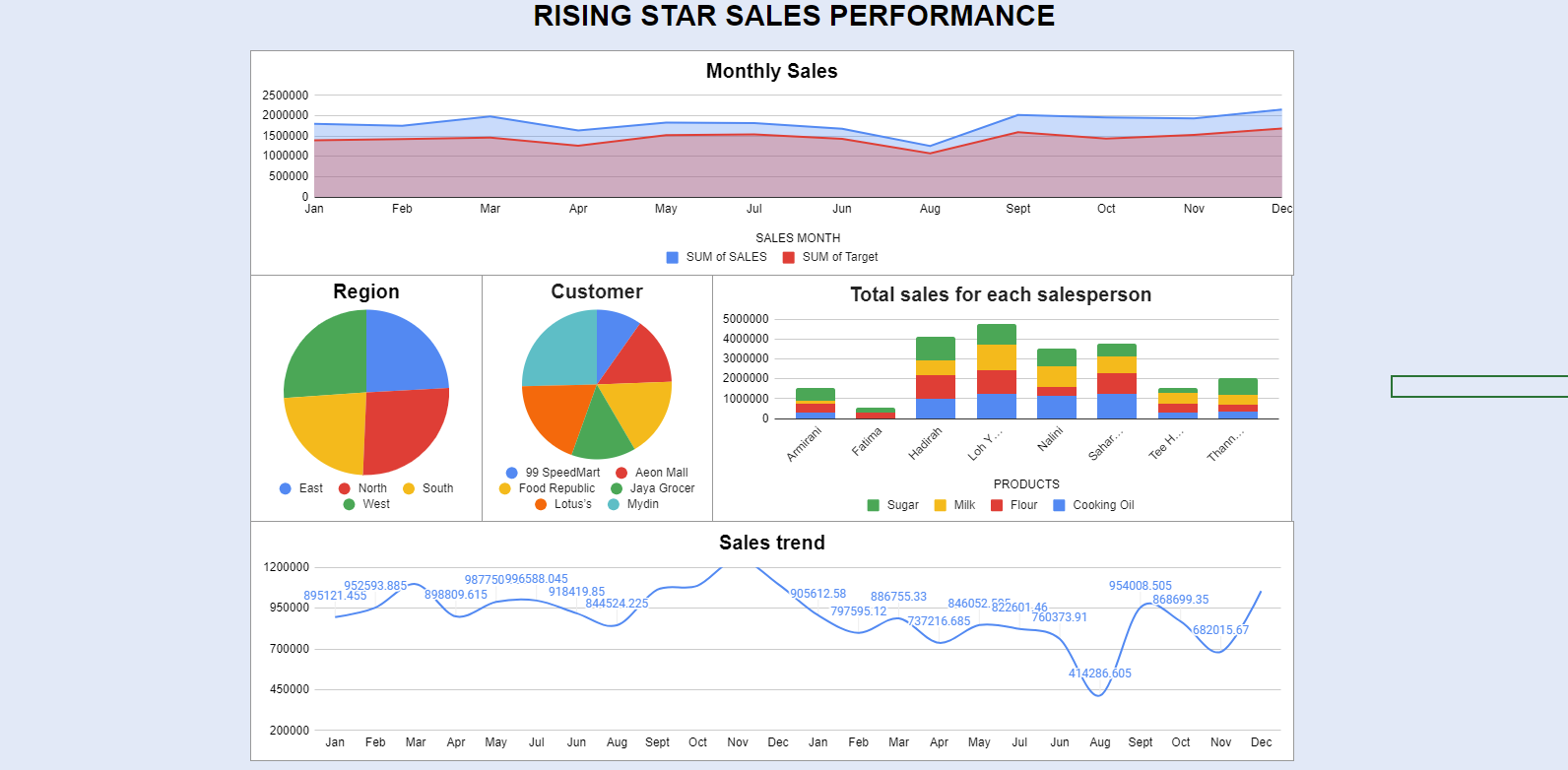


Figure 8: In tandem with the line chart, a comparison chart was designed to provide a clear contrast between actual sales and target sales. This visual representation allowed for immediate assessments of goal attainment and performance gaps.

These figures are used as visual anchors to the dashboard, ensuring that the data was not only correctly visualized but also delivered with impact and interpretability. The choice of each chart type was driven by its aptness in effectively conveying the specific data dimensions and relationships under scrutiny.

Following the chart creation phase, we embarked on the development of the interactive dashboard, which effectively encapsulated the visual representation of our data.



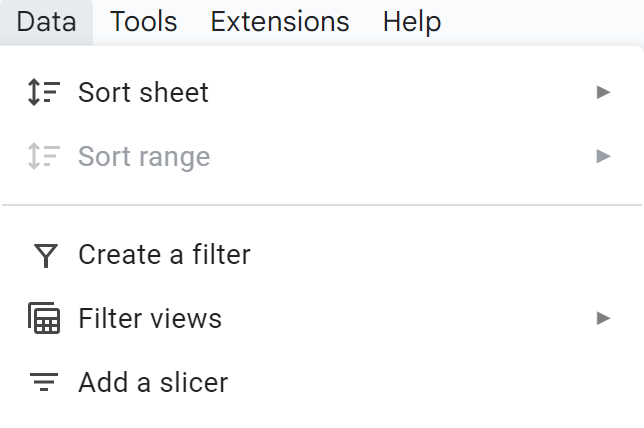
(Figure 9: an overview of the dashboard layout)

With the charts integrated into the dashboard, we sought to enhance user engagement and data exploration by implementing five slice lists, further optimizing the dashboard's interactivity and user-friendliness.

The process involved the following steps:

1. Slice List Creation:

Within Google Sheets, we harnessed the "Data" tab to initiate the slice list creation process.

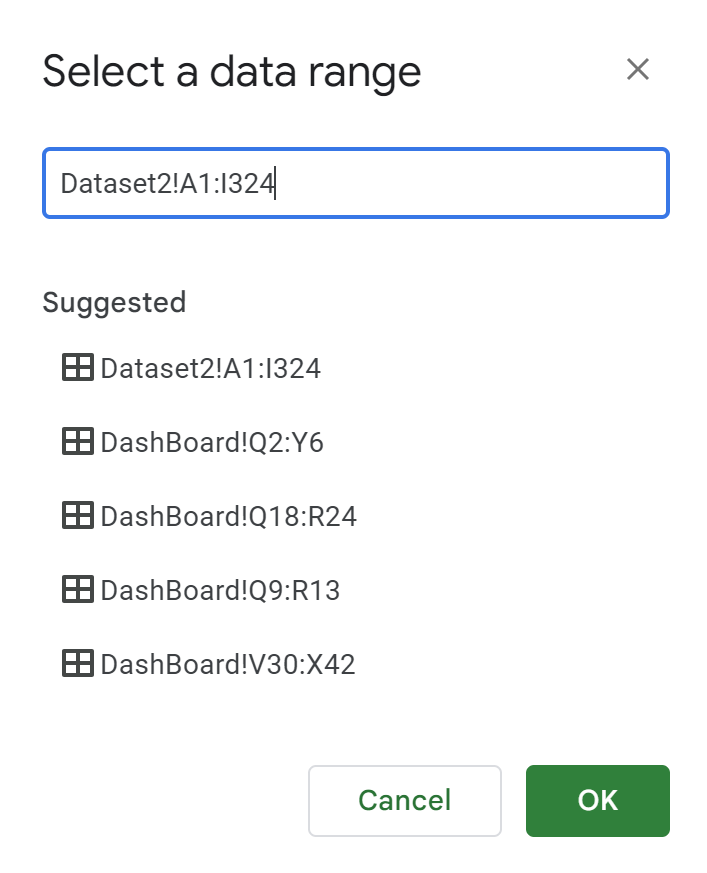


(figure 10: adding a slicer)

Adding a slicer allowed us to generate user-friendly filters that would facilitate data manipulation within the dashboard.

1. Slicer Range Selection:

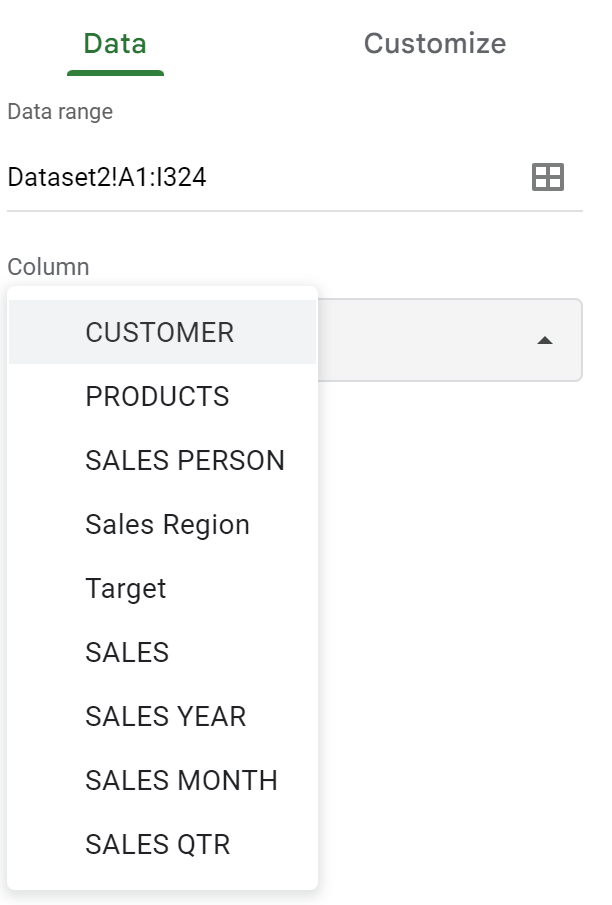
In a strategic decision, we opted to encompass the entire dataset within the slicer's selectable range. This choice ensured that any adjustments made through the slicer would universally affect all linked figures within the dashboard. This comprehensive approach harmonized data changes across all visuals.



(figure 11: setting the slicer data range)

1. Column Selection for Slicing:

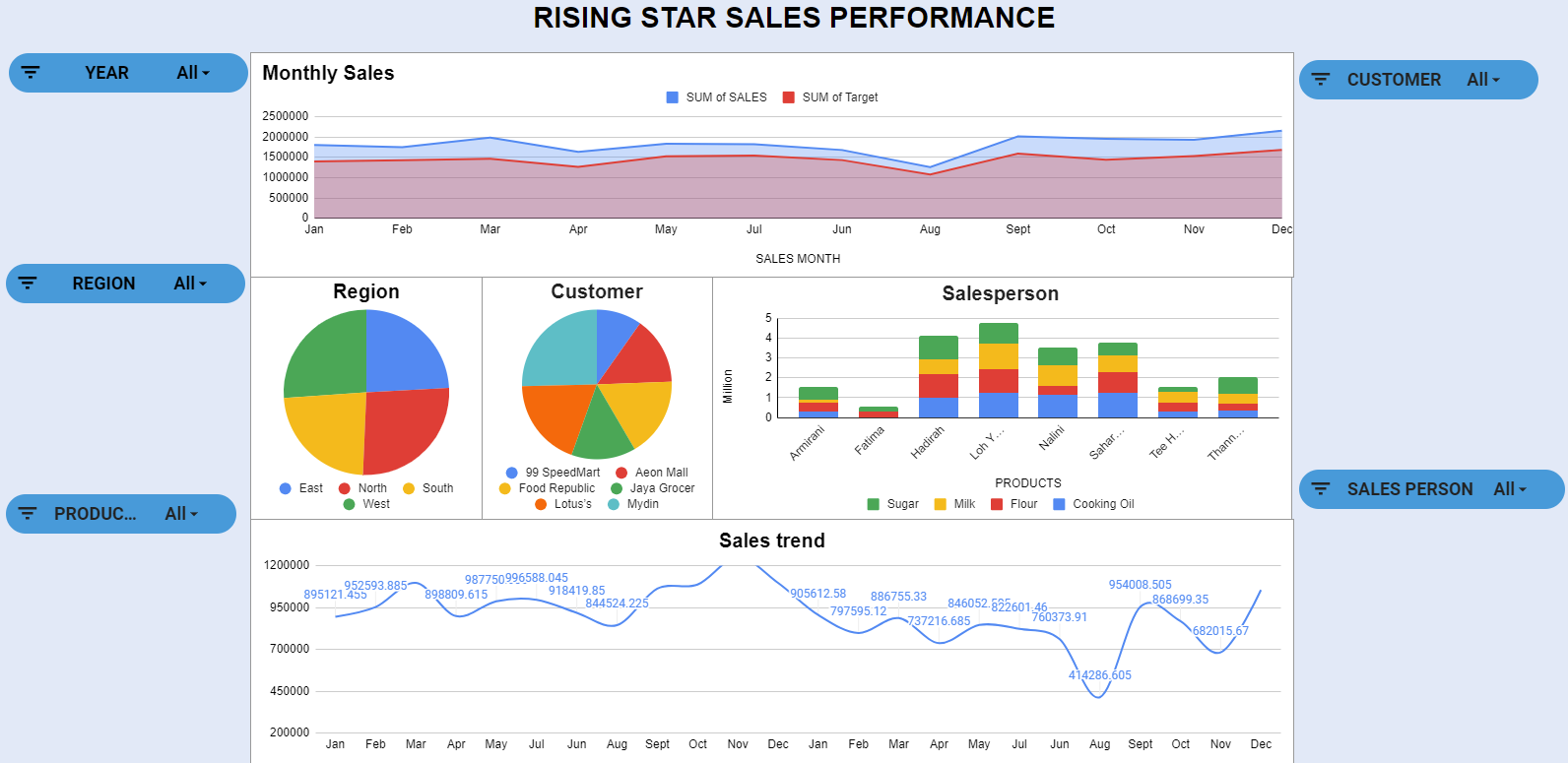
By pinpointing the specific column for slicing, we defined the parameter by which the slice list would influence data presentation. This provided a fine-tuned control mechanism for users.



(figure 12: available columns for selecting the parameter)

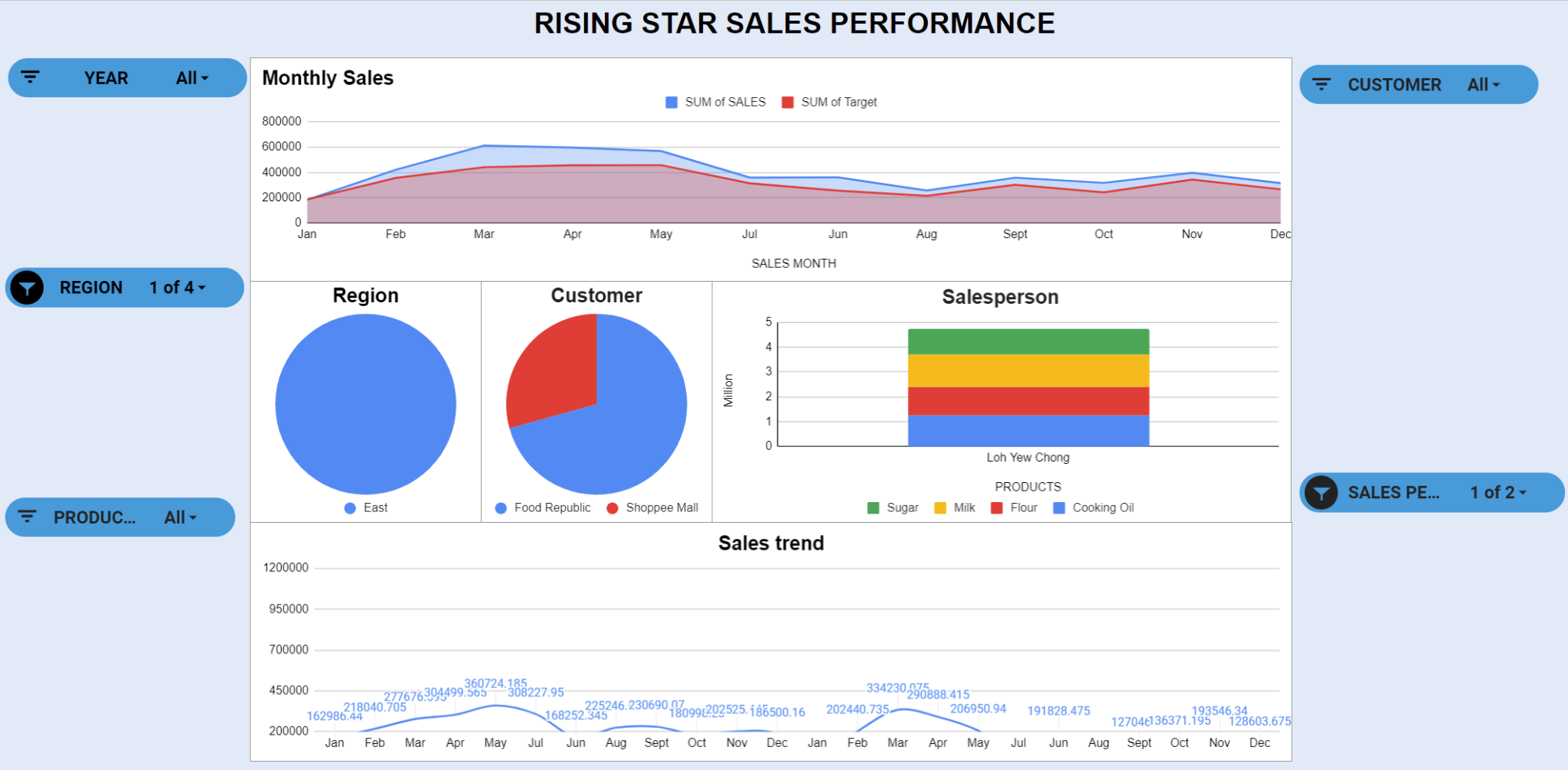
1. Integration into the Dashboard:

With the slice list fine-tuned and tested for functionality, we seamlessly integrated it into the dashboard, ensuring that users could leverage this feature to refine their data exploration experience.



(figure 13: the entire dashboard)

User-centric design considerations, such as linking the slicer's range to the entire dataset, created a harmonized and intuitive experience for interacting with the data visualizations, ultimately enhancing the utility and practicality of the dashboard.

To test the dashboard, figure 14 has select seller Loh Yew Chong in the region east:  


(figure 14: testing the date)

Conclusion

In conclusion, we have processed and visualized the dataset focusing on sales performance. We created an interactive dashboard using Google Sheets, highlighting key elements like sales year, region, products, personnel, and customers. We used various chart types, such as pie charts, stacked charts, line charts, and comparison charts to represent the data effectively.

The dashboard included slicer lists for creating an interactive dashboard, enhancing the overall experience.