**[Streamlit App](https://share.streamlit.io/sfh22/superstore-sales-app/main/Superstore.py) Manual for Superstore CEO**

Designed by: Samer Haidar

**Overview of the Dataset:**

The dataset used for this app contains the transactions which occurred starting from January 2011 to December 2014. The dataset contains 51,290 orders belonging to 1,590 customers. The items bought can belong to one of the following categories: Technology, Furniture, and Office Supplies. The Technology category has the following Sub-Categories: Accessories, Phones, Copiers, and Machines. The Furniture Category has the following Sub-Categories: Chairs, Tables, Bookcases, and Furnishings. Finally, the Office Supplies Category consists of one of the following Sub-Categories: Binders, Supplies, Appliances, Storage, Art, Paper, Envelopes, Fasteners, and Labels. Additionally, the Customers can be one of the following Segments: Consumer, Corporate, and Home Office. The transactions occur in different countries summing up to 147 belonging to several regions worldwide. Moreover, the countries are further divided into states and cities.

**Overview of the Data Analysis Tool:**

The ability to make effective decisions is crucial to an organization’s survival in today’s tumultuous business environment. For firms to evaluate alternatives and make informed choices they must have reliable and timely data upon which to make their decisions. Consequently, the development of effective data management techniques is of central importance to an organization. This Streamlit tool provides on-the-spot insights to decision makers. It assists managers in having an overview of the company’s performance by observing metrics such as Sales, Profits, Quantities Sold, Number of Customers during a specific period. Additionally, the tool includes the ability to filter by the Product Categories, Sub-Categories, Type of Customer, the Location, in addition to time limits. The app also includes machine learning components. Two ML capabilities are provided. The first one is a Sales forecasting model which predicts the Sales for several years in advance. Also, there is a clustering model which segments customers according to an RFM analysis which divides customers according to their overall importance to the business.

1. **Business KPIs:**

The first page of the tool is an overview page which contains the basic metric and KPIs. It includes the sales, profits, number of customers, and the quantities sold. On the sidebar, there is a date input where you can select the date range for which you would like to analyze. Additionally, there is a radio where you can filter by a category of your choice. There are four options: All Categories, Technology, Furniture, and Office Supplies. You can also filter by the corresponding Sub-Categories of the corresponding Categories. On the bottom part of the page, there are four graphs produced using plotly. The graphs show the sales, profits, quantities, and number of customers of the specified Category and corresponding Sub-Category. The metrics are displayed for a specific time period which you can choose.

1. **Profitability Analysis**

Where the first section of the tool provided high-level insights and showcased the performance of the business through a holistic view, this section digs deeper into the business’ numbers and figures by analyzing the company’s metrics by location and knowing the customers who have the highest profits and orders.

* 1. **Profits by Sub-Category Table**

As the company stores its data in transaction-based form, where each entry represents a single item bought by a certain customer during a certain date, it would be beneficial to aggregate the data to be able to observe the sales of the company per month to have a general sense of the performance, and to compare the sales across different months and different years. This table shows the sales of the company per month and per year which enables the managers to compare between the sales of January 2011, January 2012, and January 2013 for example. Additionally, it enables managers to detect trends, seasonality, and unknown random fluctuations in the sales numbers.

* 1. **Most Profitable Sub-Categories**

As a business, you will usually have several products that you sell. These products fall under several categories. However, not all products are always profitable. Some products might endure negative profits and thus does not server the business’ goals. Thus, it is beneficial to keep track of your products’ performance and drop products which don’t have positive profits from your products’ portfolio. This page shows the several sub-categories, their profits, sales, and quantities sold. You can also filter by a date window to see how these products are producing during only a certain period.

* 1. **Most Profitable Customers**

Every business to be able to sustain should seek to optimize their relationship with their customers by better knowing their customers and their preferences. This page shows four tables. The first one shows the customers which have the highest profit return to the business. The second table shows the customers which bought the most as in the largest quantity of products. The third table shows the top bought sub-categories by the customers, and finally, the fourth table shows the most bought products by the customers. For each of the four tables, you can specify the number of tables in the row you want to see. After having a look at these tables, you can have an understanding of your top profitable customers, and the top products and categories your customers are interested in. Thus, you can adjust your sales, marketing, manufacturing strategies accordingly to fit with the insights you gained.

**2.4 Sales by Locations**

Usually, international stores have branches and locations in several countries or states. Thus it is beneficial to understand and know the sales and profits by states and countries to be able to understand the needs of different customers located in different areas. Additionally, some stores in some places may not be profitable for different reasons, and thus perhaps if the store continues to have negative profits for a long period of time, it’s better to close the branches in this location to minimize losses and increase profits as much as possible. The tree map shown in this page shows the sales of the superstore in different states and in their corresponding cities. The manager or decision maker can also filter by Category (All Categories, Technology, Furniture, and Office Supplies), as some categories may have positive profits in a specific location, but other categories might have a negative profit.

1. **Sales Forecasting**

A sales forecast helps every business make better business decisions. It helps in overall business planning, budgeting, and risk management. Sales forecasting allows companies to efficiently allocate resources for future growth and manage its cash flow. Thus, we have created a forecasting too developed using ARIMA which forecasts the sales of the company either as a whole or per category (Furniture, Technology, Office Supplies). You can also specify the number of months in advance that you want to forecast. The four ARIMA models were built after finding the optimal parameters for each of the models. The plot in the top of the page shows the observed sales data in blue and the forecasts in orange. Additionally, at the bottom of the page, a table showing the forecasted sales for each of the months of interest.

1. **RFM Analysis and Customer Segmentation**

The advantage of customer segmentation is that it allows marketers to understand the different needs or purchase patterns of their customers in each subgroup. This will give power to shape the language or promotion which is optimal for success of each campaign.

RFM model is at the core of customer segmentation. RFM studies customers’ behavior and cluster them by using three metrics:

1. Recency (R): measure the number of days since the last purchase to a hypothetical snapshot day.
2. Frequency (F): measure the number of transactions made during the period of study.
3. Monetary Value (M): measure how much money each customer has spent during the period of study.

This customer segmentation tool uses RFM measures as the features of our customers and then uses K-Means Neighbors to segment the customers into three clusters: The Most Important group, the Second Most Important group, and the Third Most Important Group. This clustering is done for each of the three customer types: The Consumer type, the Corporate type, and Home Office type. The Most Important cluster are usually the customers who visit the store the most, have visited the store lately, and spend the much amount at the store, while the least important cluster is the complete opposite. The three clusters are moved into separate data frames which can be downloaded as csv files.