**DAB 303 – 004 - Marketing Analytics**

**Project Proposal**

**Superstore Dataset Analysis**

**Name: Arjun Patel**

**Student Id: 0812257**

**Introduction and Motivation:**

I have chosen to work on examining the superstore dataset for my college extend. With developing requests and cut-throat competitions within the showcase, a Superstore dataset is looking for your information in understanding what works best for them. They would like to get it which items, locales, categories and client fragments they should target or maintain a strategic distance from. You'll indeed take this a step advance and attempt and construct a Relapse demonstrate to anticipate Deals or Profit. Go insane with the dataset, but too make beyond any doubt to supply a few trade bits of knowledge to progress.

**Dataset:**

The dataset I will be working with is the superstore dataset sourced from [Tableau](https://community.tableau.com/s/question/0D54T00000CWeX8SAL/sample-superstore-sales-excelxls). The superstore dataset available at the provided tableau link is a comprehensive collection of information related to superstore. It contains a total of 9994 rows and 21 columns, making it a substantial dataset for analysis. It can potentially be used for various analytical purposes, including exploratory data analysis, segmentation, recommendation system development, and trend analysis.

Dataset link: <https://community.tableau.com/s/question/0D54T00000CWeX8SAL/sample-superstore-sales-excelxls>

**Metadata:**

|  |  |
| --- | --- |
| **Column Name** | **Description** |
| **Row ID** | Unique ID for each row. |
| **Order ID** | Unique Order ID for each Customer. |
| **Order Date** | Order Date of the product. |
| **Ship Date** | Shipping Date of the Product. |
| **Ship Mode** | Shipping Mode specified by the Customer. |
| **Customer ID** | Unique ID to identify each Customer. |
| **Customer Name** | It contains name of the customer. |
| **Segment** | The segment where the Customer belongs. |
| **Country** | Country of residence of the Customer. |
| **City** | City of residence of the Customer. |
| **Postal Code** | Postal Code of every Customer. |
| **State** | State of the residence of the Customer. |
| **Region** | Region where the Customer belong. |
| **Product ID** | Unique Id of the product. |
| **Category** | Category of the product ordered. |
| **Sub-Category** | Sub-Category of the product ordered. |
| **Product Name** | Name of the Product. |
| **Sales** | Sales of the Product. |
| **Quantity** | Quantity of the Product. |
| **Discount** | Discount provided on the Product. |
| **Profit** | Profit/Loss incurred. |

**Market Problem:**

After exploring the dataset, I have identified a potential problem related to product recommendations and customer satisfaction. While Superstore offers a wide range of products, customers may struggle to find products that align with their preferences and needs. Additionally, understanding customer sentiments and preferences can be challenging without customer reviews and ratings.

**Problem Approach/Solving the Problem:**

To address this problem, I propose to perform an analysis on the dataset and develop a recommendation system based on product attributes and popularity. This will help customers find relevant products and enhance their shopping experience. Additionally, analysing product attributes and popularity will provide insights into customer preferences and identify areas for improvement also the visualizations also provide the insights of sales and profit of the products so customer and sellers also identifying easily their mistakes and popularity of the products accordingly.

My approach will involve the following steps:

1. Data Cleaning: I will clean the dataset by handling missing values, removing duplicates, and ensuring data consistency, there is not any null values in the dataset so it will be easy for me.
2. Exploratory Data Analysis: I will perform exploratory data analysis to gain insights into the product categories, identify popular products, and understand the distribution of prices also identifying the sales and profit of the products.
3. Attribute Analysis: I will analyse the product attributes such as category, subcategory to understand the preferences of customers, so from this it can be easy to identify the trend of the products.
4. Popularity Analysis: I will analyse the popularity of products based on the number of times they appear in the dataset. This will help identify popular products across different categories and subcategories.
5. Conclusion: I will summarize my findings, provide insights into customer preferences, and propose actionable recommendations for superstore to improve customer satisfaction and product offerings also helps them to identify the trend of the product as well as how can they improve the sales of the product and make some profit from it.

**Tools:**

To conduct this analysis, I will be using the following tools:

* Python: I will use Python programming language along with libraries such as NumPy, Pandas, and Scikit-learn for data cleaning, exploratory data analysis, etc.
* Jupyter Notebook: I will use Jupyter Notebook as my development environment.

**Timeline:**

Managing a project timeline is crucial for effective project management. I estimate that the project has the potential to finish within a time span of four to six weeks.

**Arjun Patel  
0812257**