Product Analytics on Global Super Store Data Project Proposal

**Overview / Introduction:**

With the development of technology and the strong trend towards e-commerce, it has become necessary for stores that want to maintain their position in the market to keep pace with the development and trend to provide the option of electronic sales through their websites and applications, which help customers to shop and buy all products in the easiest way possible, wherever they are.

We may think that all companies that have adopted the e-commerce approach have succeeded in that, but what we have noticed recently is that some companies have declared bankruptcy and bid farewell to the market, which led me to wonder about the reasons.

The aim of this project was to use data to predict what would be the best selling products in a sample of online stores, in order to develop improvement measures for existing stores and help emerging merchants to identify the best needs and trends

**Question / Need:**

This project originated from one of the open source datasets from Kaggle site, which will help us achieve the goal to answer the following questions:

* What are the most popular products?
* profit margin for each product
* What are the most successful countries and provinces for e-commerce?
* The frequency with which a product is purchased from the same store for the same customer, which indicates satisfaction with the service or product
* Are there some products that are frequently in demand?
* The best-selling products for each town (which shows the needs of the residents of this town and the influencing factors)

This data has been meticulously analyzed using machine learning models, so that shop owners and start-ups can determine their trends based on the results released.

**Data Description:**

This project originated from[**https://www.kaggle.com/apoorvaappz/global-super-store-dataset**](https://www.kaggle.com/apoorvaappz/global-super-store-dataset)that offered by kaggle.com , which contain 51290 purchases by many customers for various products in different countries and With 22 features for each of them, we can consider 16 of them to be very important. which will help us solve the problem and reach the desired results.

The model that I used is Logistic regression, it will deal with many characteristics, most notably are: Order ID, Customer ID , City , Market , Product ID , Quantity , sales , Profit And some other characteristics, where the frequency of demand for the same product and products shared in the same demand and repeated purchase on the same pattern will be studied .

Tools:

- Numpy and Pandas for (EDA) Exploratory Data Analysis

- Scikit-learn for modeling

- Matplotlib and Seaborn for plotting

**Done by:**

Rawan Mohammad Balubaid

[Rawan-balubaid@hotmail.com](mailto:Rawan-balubaid@hotmail.com)

0540386411