Name: Aqsa Younus

ERP: 18571

**Dataset:**

Dashi Foods Dataset

**Domain Knowledge:**

E-commerce data is the data that is constantly expanding, creating many possibilities for data analysts to see trends, and understand how the business model is working and what we could do to improve business models.

This dataset, in particular, is unique. This dataset dataset belongs to the company Dashi Foods. It is a local Pakistani Company that is dedicated to innovate and produce new food products for convenience with taste. This dataset consists of various sheets which show a database like structure because of the various ID columns connecting all the sheets together.

The sheets are as follows:

1- Customer Ledger: This sheet has all the information regarding the balances of each customer. It gives an idea to how the transactions are being ensued and through what means.

2- Customer: A purely dimensional data, best for machine learning purposes since it has various boolean valued columns.

3- CustomerWise: A sheet rich in KPIs ideal for e-commerce analysis. The sheet can provide answers to potential analysis questions regarding customer buying behavior

4- Item: Similar to Customer sheet, this sheet was also rich in Dimensions by lacking in KPIs, however, had various amount of boolean columns perfect for machine learning algorithms.

5- Purchase: Contains data regarding the purchase of each item, the category and quantity of the product ordered by the customers.

6- DataDump: Though named as datadump, this sheet was the most ideal. It had the right balance of KPIs and dimensions. It also has customer and product data together which would aid in product/Customer behavior analysis

**Thoughts About the Dataset**

* The data had a lot of dimensions by hardly any KPIs, a lot of the KPIs were repeated across multiple sheets, making most of the data redundant
* The data was surprisingly clean though when it came to null values, missing values did not require much imputation especially for the KPIs which is why the analysis is strong for them
* The data was incomplete. I believe more sheets would have given proper insights to the dynamics of the business, therefore would have given even better analysis.
* There was a slight confusion whether the data was entirely e-commerce or not, I believe the ledger report and godown column make it seem like the data partly alludes to physical transactions so I kept that in mind during wrangling and analysis.

**Final Thoughts/Opinions**

I personally would prefer python and Tablaeu over PowerBI. I felt restricted by the limited options that PowerBI provides relative to Python and it is more effort to create charts on PowerBI compared to Tablaeu. Furthermore, the data itself I believe should have had more columns that one could analyze in order to get further insights on the E-commerce data of Dashi Foods.