**SCOPE OF STUDY**

**Topic: *Data Analytics Project on Sales Orders from July 16, 2023 to October 16, 2023***

**RESEARCH OBJECTIVES**

1. To explore related products that are commonly purchased together via Product affinity analysis
2. To explore the potential outcomes of strategic planning via ‘Single scenario’, and ‘multiple scenario’ of ‘What-if’ Scenario analysis relating to ‘Discount rate’, ‘Quantity change (units), and Shipping cost change’.
3. To create customer segments based on the value of Gross Sales (before discount & tax). The segments will be broken down as follows:   
   Major Customers: $1000 - $5000

Important: $426 - $1000

Potential: $75 -$425

Low touch: $0 - $74

1. To create Customer clusters from the correlation of:
2. Discount Amount & Gross sales amount (before Tax & discount)
3. Discount amount & Net Sales Amount
4. To show trends of Net Total Sales, Quantity purchased and Shipping cost over time
5. To create filters for distilling peculiar offers and services of ‘Discount Codes” ‘Accepts Marketing’, ‘Shipping method’ and the ‘Year/Month’
6. To show top & bottom rated Products, Customers and Shipping destinations
7. To show trends of customer patronage, filterable by Location, Customer segments, Clusters & Discount offers

**Research Design and Methodology**

Adopting a mixed research design method of Exploratory Research Design, Descriptive and Causal research design to serve as barometer to view marketing performance for a general understanding of how things are going, dealing with multiple variables and with an intention to inspire an observation of the Functional & Emotional attributes unique to the business. Introspective from the sales activity, explanatory research will be done to identify and understand the extent of the relationship between isolated variables and dependent variables.

**Geographical and Temporal Boundaries**

The dataset contains information of sales orders from 60 provinces in the US and Canada between ‘2023-07-16 08:07:46’ and ‘'2023-10-16 11:27:49’.

**Variables and Measurements**

The raw dataset contains 7829 rows and 75 Columns.

The key variables include ‘Names”, ‘Total’, ‘Line-Item price’, 'Discount Amount', 'Shipping Method', 'Created at', 'Line-item quantity', ‘Shipping address’. Basically, all primary variables indicating the important transaction details of Location, Date, Line-item quantity, Line-Item Price, Discounts, Shipping, Tax, and customers consent to Marketing.

Measures, Aggregates, Averages, and clusters will be created during our analysis and visualizations.

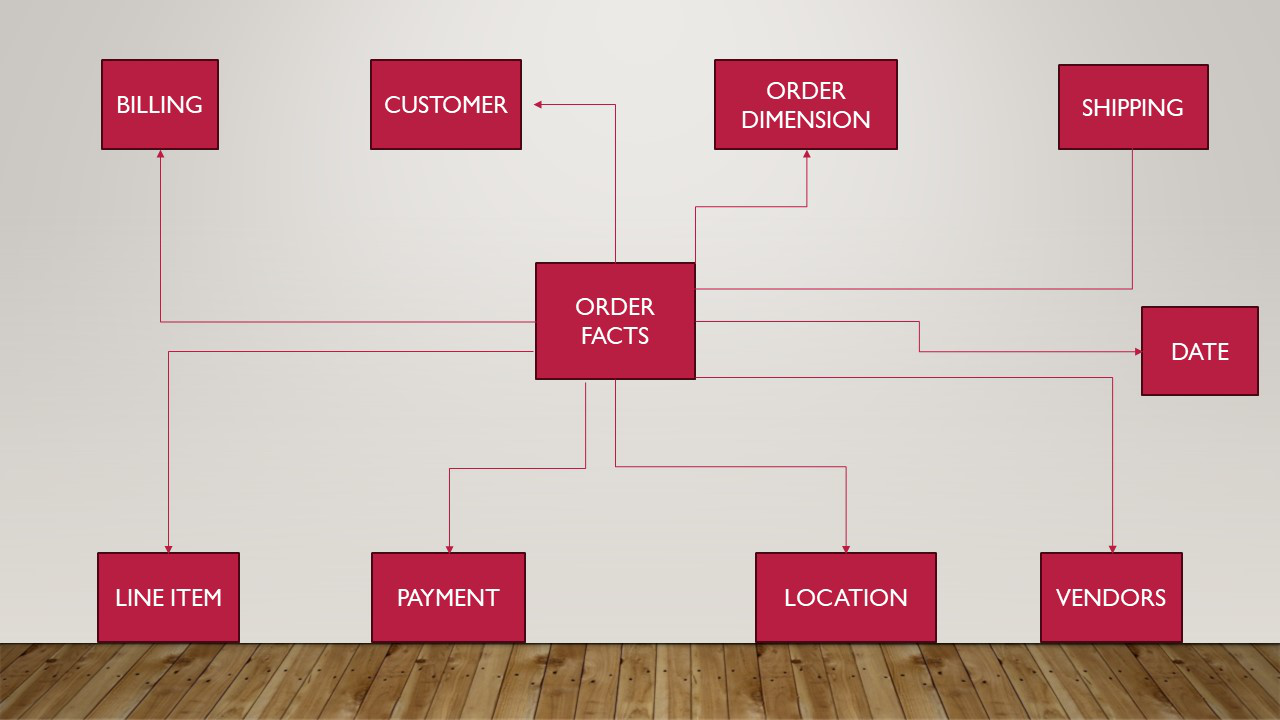
**DATA COLLECTION**

Data was shared on a basis of mutual agreement with the Data Analyst.

**DATA ANALYSIS PROCEDURE**

The project involves refining data through cleaning, modelling, and employing various analytical techniques. These include 'What-If Analysis', 'Market Basket Analysis', 'Dynamic Customer Clustering and Segmentation', examination of trends, utilization of Descriptive and Exploratory Statistics, and Data Visualizations. These methods aim to uncover insights related to day-to-day activities, encompassing Location, Customer behaviour, Product trends, Shipping activities, and Marketing-related patterns.

Image of data Model is attached below:

****

**Significance and Contribution of the Study**

This Data Analytics project will create a template that’d make any similar projects in the nearest future to be churned and ready to show insights much faster.

Analysis that will be done include Market Basket analysis to show products that are usually purchased together, in a suitable user interface.

Scenario analysis to test the sensitivity of Discount rate, Quantity Change (Units), and Shipping cost on Gross Sales, Net sales, Quantity Change, and Shipping price.

Interactive charts to show Product, Customer and Location with filters like Discount Codes, Shipping Method, Shipping Province, and Customer segments/clusters.

Also, visualizations with filters and dynamic clusters will be created. Customers will be segmented in order of the ‘Net Purchases’.

Trend visualizations of store visits will also be created.

**LIMITATIONS**

In the dataset, ‘created at’ i.e. date of entry has 3887 distinct entries, and 49.6% are distinct.

For the totals and discount amount, 50.3% of the raw data has missing values. Some activities didn’t just occur enough to reveal convincing information; however, Machine learning can be used to test possible results in certain instances.

Resource Constraints:Frequency of project meetings and updates.

Changing Conditions: Conditions may change after the analysis is conducted, affecting the relevance of the findings. This is particularly relevant for dynamic environments.

**Assumptions and Simplifications**

Many analyses involve making assumptions or simplifications to model complex situations.

**Ethical Considerations**

Data is not owned by Data Analyst.

To ensure anonymization and de-identification of customers, insights were shown with SKU’s and not ‘Ids’ or personal information. Data Analyst is accountable and responsible, also he seeks ethical review and compliance.