

# Detailed Project Report

AMAZON SALES DATA ANALYSIS

SIDDHARTH S SRIVASTAVA

## 1. Problem Statement:

Sales management has gained importance to meet increasing competition and the need for improved methods of distribution to reduce cost and to increase profit. Sales management today is the most important function in a commercial and business enterprise.

## 2. Objectives:

- The goal of this project is to perform ETL, create a visually appealing dashboard, and to perform analysis on the same to find meaningful trends and patterns.
- In order to deduce important metrics and patterns in the dataset, this project will use the provided data to perform ETL and data analysis.
- Additionally, several visualizations and reports are created to represent significant linkages.

### 3. Benefits

- Help in making wiser business decisions.
- Aid in customer satisfaction and trend monitoring, which can serve current consumers and attract new ones.

### Dataset information

**CustKey:** Primary key for customer dataset

**Discount Amount:** Discount given for each product

**Invoice date:** The date on which the item was ordered.

**Invoice number:** Unique invoice number for each customer.

**Item:** Item name

**Line number:** Unique line number

**List Price:** Price listed for each item

**Order Number:** Unique Order Number.

**Promised Delivery Date:** Estimated delivery date

**Sales Quantity:** Number of units ordered

**Sales Amount based on list Price:** [List Price] \* [Sales Quantity]

**Sales Amount:** [Sales Amount Based on List Price] -  
[Discount Amount]

**Sales Cost Amount:** Cost of Goods Sold (Expenses related  
to production, storage and delivery of goods)

**Sales Margin Amount:** [Sales Amount] - [Sales Cost  
Amount]

**Sales Price:** [Sales Amount] / [Sales Quantity]

**Sales Rep:** Sales Representative

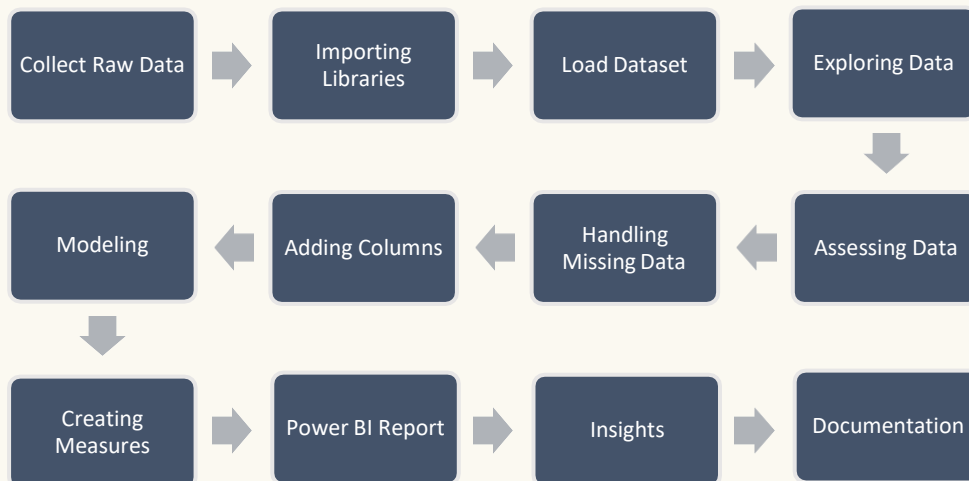
**U/M:** Unit of measure (EA - Each, SI - Some SI unit like kgs or  
galloons, PR - Pair)

**Year:** Year of order

**Month:** Month of order

**Quarter:** Quarter of a year (Q1- Jan to Mar, Q2 - Apr to  
Jun, Q3 - Jul to Sep, Q4 - Oct to Dec)

## 4. Architecture



- 1. Collect Raw Data** - This step involves extracting the data from different sources relevant to the problem statement or obtaining data from the client
- 2. Importing Libraries** – Import analysis related python libraries example – Pandas, Numpy, Plotly, datetime etc
- 3. Data Wrangling** – Contains following steps gathering data, assessing data, handling missing data and adding columns

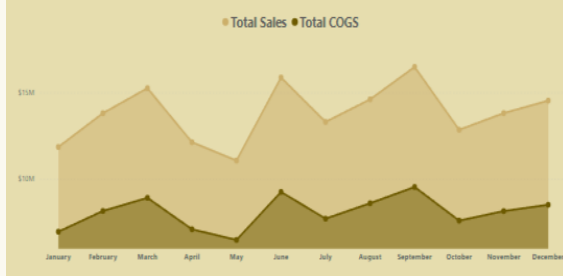
**4. Exploring Data** – Once the data is loaded and pre-processed, we perform data analysis using python libraries and Business Intelligence tools like Power BI

**5. Data Modelling** - Data Modelling is one of the features used to connect multiple data sources in BI tool using a relationship.

A relationship defines how data sources are connected with each other and you can create interesting data visualizations on multiple data sources

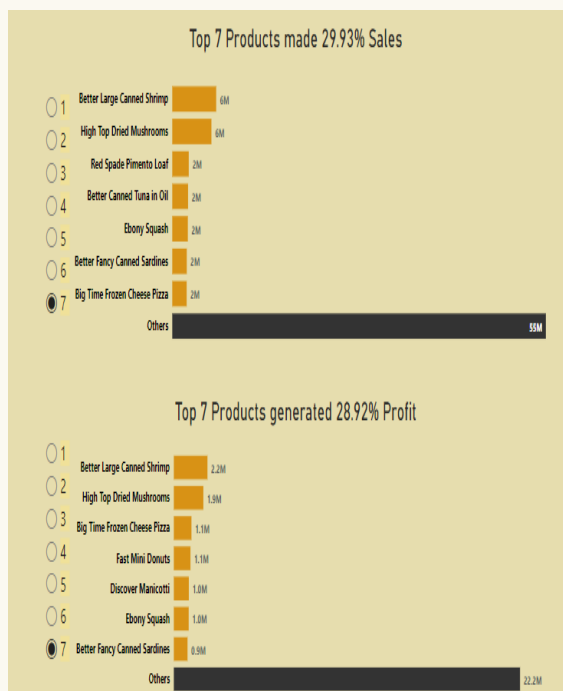
## 6. Insights

### 1. Sales by Month



- According to the following chart, we can conclude that March, June and September seem to be the best months for Sales, whereas there is a significant drop in Sales in January and May.

### 2. Sales by Product



- The general trend seems to be that the top 7 products account for around 30% of Sales as well as Profit. This trend seems to be prevalent across different years.
- The best Product in terms of Sales and Profit is "Better large Canned Shrimp". It accounted for around 6%, 9% and 6% of profit in the years 2017, 2018 and 2019 respectively.



## 7. Q & A

Q1) What's the source of data?

- The Dataset was taken from iNeuron's Provided Project Description Document

Q2) What was the type of data?

- The data was the combination of numerical and Categorical values

Q 3) What's the complete flow you followed in this Project?

- Refer page 5 for better Understandings

Q4) What techniques were you using for data?

- Removing unwanted attributes
- Cleaning data by removing column with missing values
- Converting Numerical data into Categorical values

Q 6) What were the libraries that you used in Python?

- I used Pandas, NumPy, Matplotlib, Seaborn and Plotly libraries