AMAZON SALES ANALYSIS

DETAILED PROJECT REPORT

Freeva Dsouza

26th January 2022

Objective:

In this project we will be trying to analyze a food sales dataset with over 65000 records and understand the Sales trend across time, identifying the commodities that are in demand and also giving the stakeholder visual reports to understand all the insights that we gain from the dataset and come up with better decisions to improve the sales model.

Benefits:

Defines key metrics responsible for improving sales.
Provides useful insights to ensure productive sales.

Data Sharing Agreement:

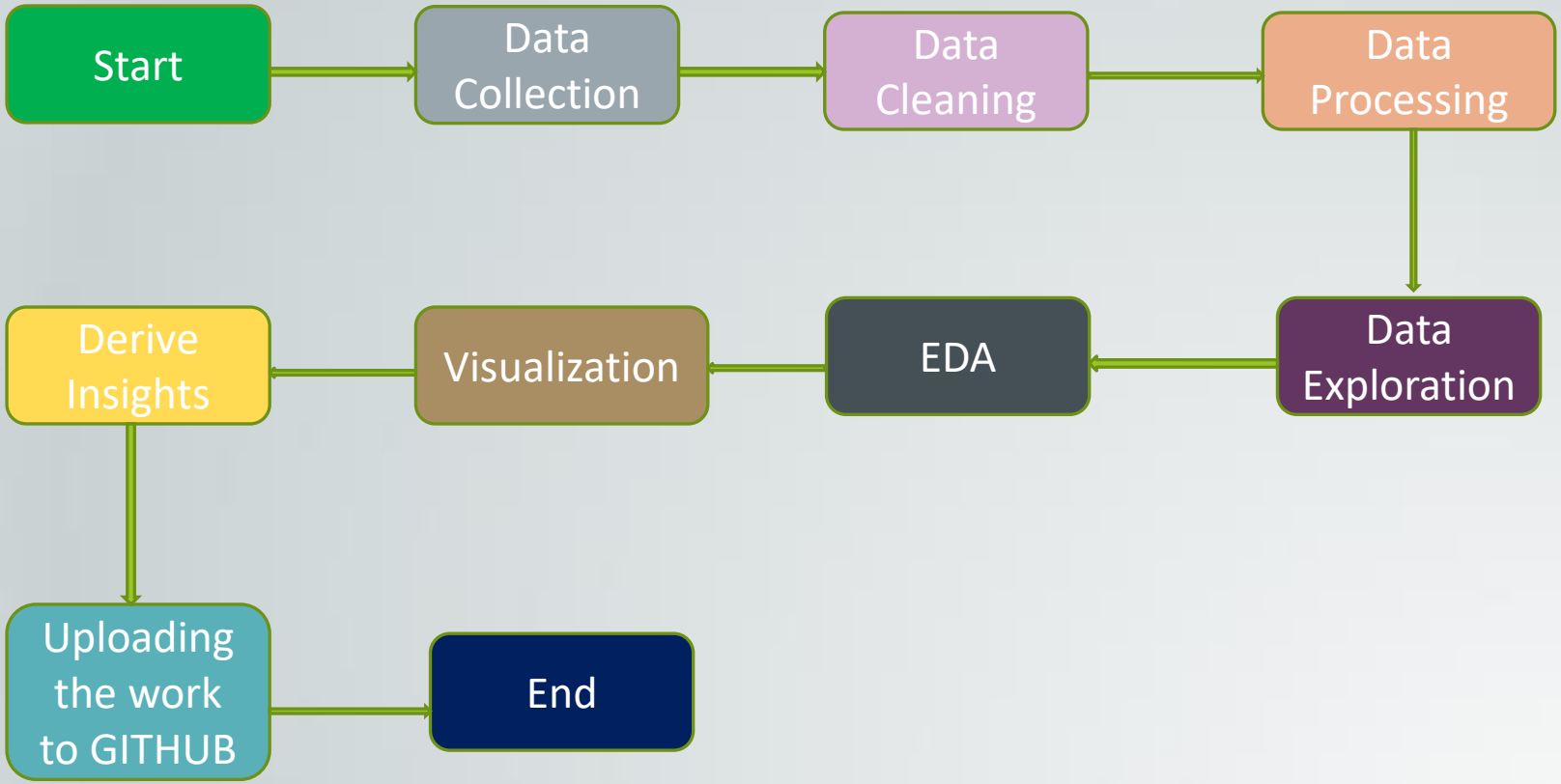
File Name: SALESDATA.csv

No of Columns: 18

Columns Names: CustKey, DateKey, Discount Amount, Invoice Date, Invoice No, Item Class, Item No, Item, Line No, List Price, Order No, Promised Delivery Date, Sales Amount, Sales Amount Based on List Price, Sales Cost Amount, Sales Margin Amount, Sales Price, Sales Quantity, Sales Rep, U/M

Column Data Types: int64, datetime64[ns], float64, datetime64[ns], int64, object, object, object, int64, float64, datetime64[ns], float64, float64, float64, float64, float64, int64, int64

Architecture:



Data Collection:

The dataset was provided by the company in the form of CSV file. The dataset consists of 65282 records of food sales.

The following are the columns in the dataset:

CustKey: Unique ID of all the customers.
Discount Amount: The amount deducted from the selling price of the items.
Invoice Date: The date on which the goods have been billed and the transaction officially recorded.
Invoice Number: A unique sequential code systematically assigned to the invoices.
Item Class: Group of items that share similar properties.
Item: Name of the product.
Line Number: The number that uniquely identifies the item in the sales document.
List Price: The suggested retail price of the product.
Order Number: The number generated against each purchase.

Data Collection:

Promised Delivery date: The date you promised the customer you would deliver the products.
Sales Amount: Total sales generated -> Sales Price*Sales quantity
Sales cost Amount: Total cost price for the quantity of items sold.
Sales Margin Amount: The amount a company makes from a sale of a service or product.
Sales Price: Selling price of each product.
Sales quantity: Number of each products sold
Sales Rep: Unique ID generated for each Sales Representative.
U/M: Units of measure of items sold.

Data Cleaning & Data Processing:

The data cleaning has been carried out on the dataset consisting of 65282 records.

- All the missing values have been removed from the dataset by using two common methods:
  1. Dropping the rows / columns
  2. Imputing them based on other observations i.e. the mean or median

- Eliminated all unnecessary columns which are no use to the analysis.

- Eliminated all duplicate rows and columns to ensure data integrity.

- Transformed all the categorical variables into numerical variables in order to make the dataset more convenient for analysis.

- Extracted independent and dependent variables from the dataset.

EDA:

- In this process various insights and observations from the data were found. Also, KPIs (Key performance metrics) i.e. the factors which are mainly responsible for facilitating efficient sales management were defined.

- A few outliers were detected within the distribution of various attributes and properly examined them at each step.

- Gained better understanding of data aspects like:
  1. Main features of the data
  2. Variables and relationships that hold between them.
  3. Identified which variables are important for the problem.

- We performed Statistical analysis mainly through data manipulation, exploration by using Numpy and Pandas.

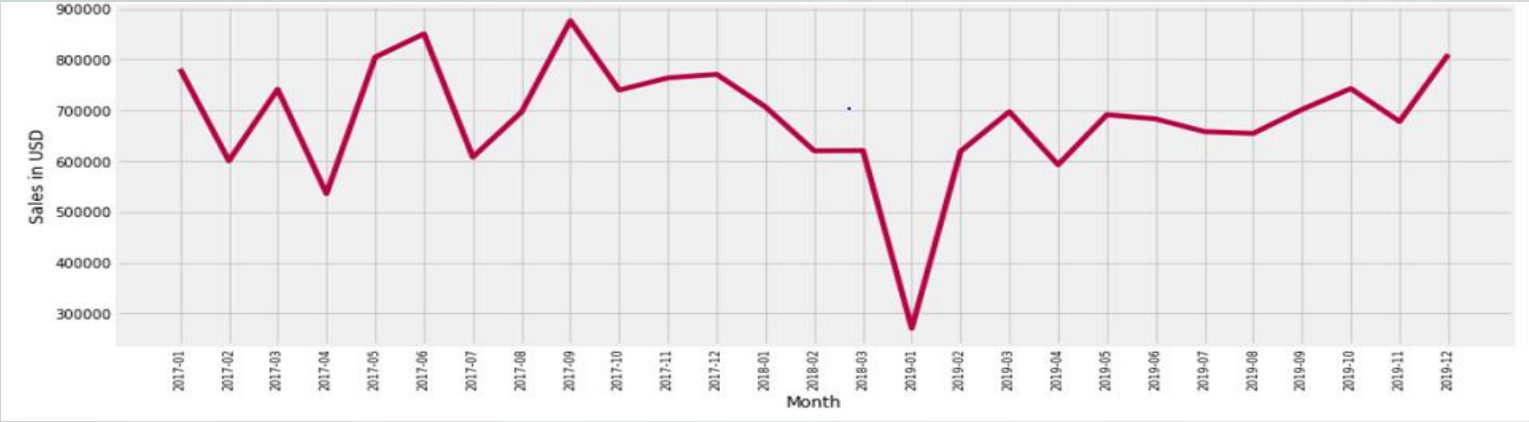
- Established correlations between various attributes present within each dataset and also defined key factors and metrics responsible for boosting up sales.

Visualization:

- Various charts and plots have been prepared based on meaningful data. Using Pythons Matplotlib, Seaborn, Plotly libraries several highly interactive bar charts, trends have been created.

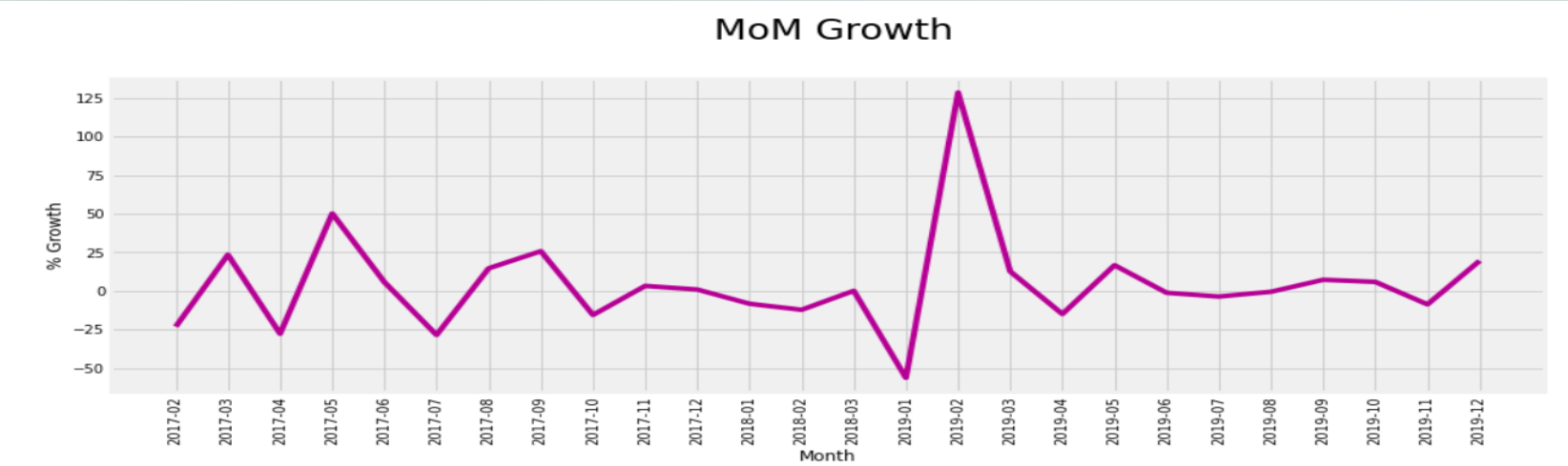
- Defined key metrics by establishing some mathematical relations between numerical variables and leveraged them in order to generate some observations and conclusions

Insights and Observations:



- From the above graph, we can see that the maximum Sales and profit was made in the year 2017.
- September month seems to have made the maximum sales.
- After March 2018 there seems to be significant amount of decline in the sales till December 2018.

Insights and Observations:

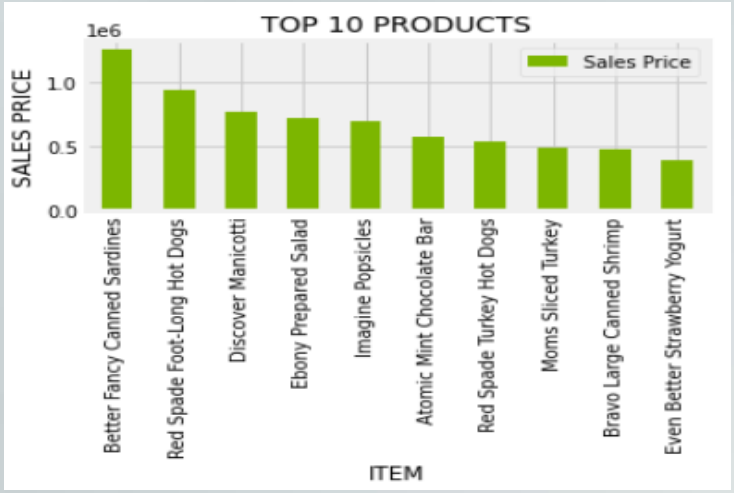


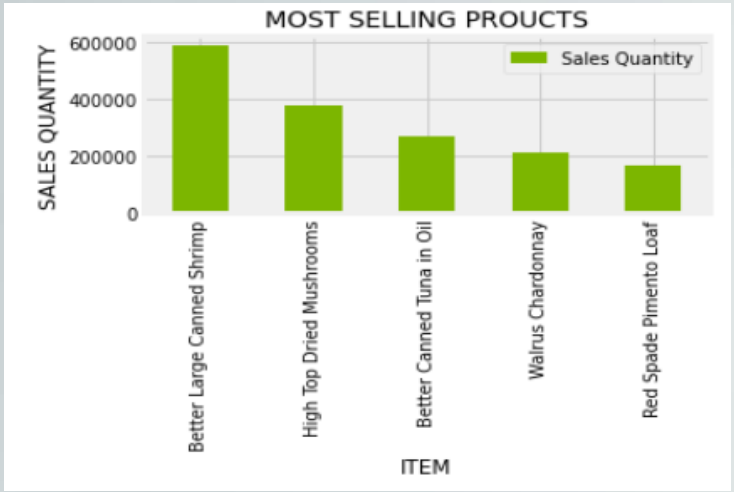
- There seems to be a sudden growth in the sales from 2019 January to 2019 February.
- It tends to drop off really faster between Feb 2019 to March 2019.

Insights and Observations:

- The sales seems to have declined for the whole year from 2018 to 2019.
- The curve tends to increase in the mid year 2019.

Insights and Observations:





The products which are sold the most doesn't seem to making enough sales throughout the years.