EDA REPORT

Sudharsan Vanamali

Summary:

Project Link: https://github.com/Astrasv/Business_Analytics.git

This Business Analytics Project leveraging Power BI and SQL, designed to analyze and cleanse the **AdventureWorks** database. AdventureWorks is an open-source database provided by Microsoft, and this project harnesses its rich dataset to extract meaningful insights and create compelling visualizations.

The initial data cleansing and transformations were done in SQL which solves the major problem of **TIME CONSUMPTION** of python. While the EDAs are done using **python** and Data analytics frameworks

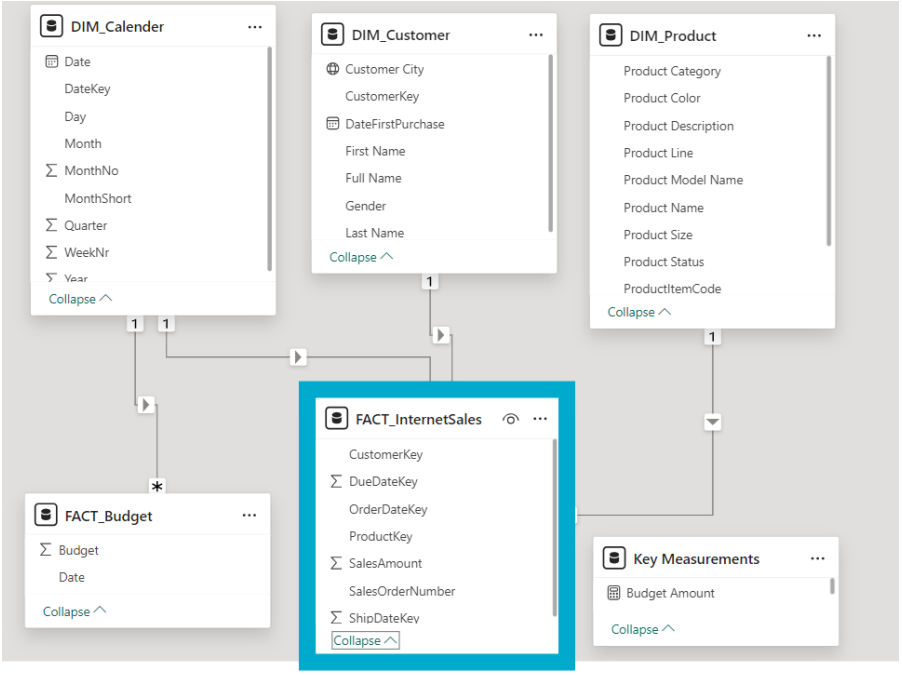
Python Frameworks used in EDA:

- Pandas
- Numpy
- Matplotlib
- Seaborn
- Sklearn

Data set description:

The data from AdventureWorks Database is cleansed

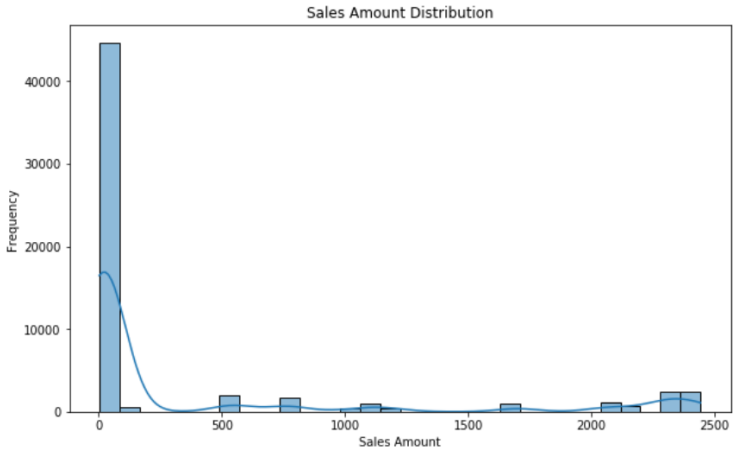
The cleansed dataset contains the following relationships and attributes



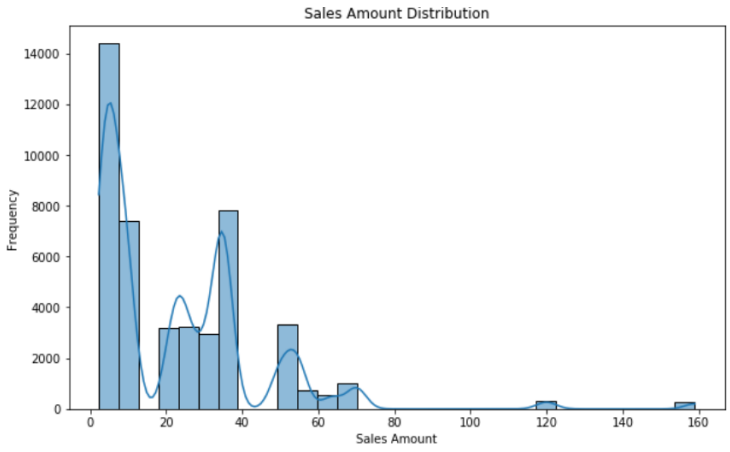
Out of the 3 dimension tables and 2 fact tables, I have picked the FACT_InternetSales Data to provide business insights (Dataset is highlighted in Blue)

This EDA solves a smaller sub-problem of the Internet Sales insights over the broad scope of the project

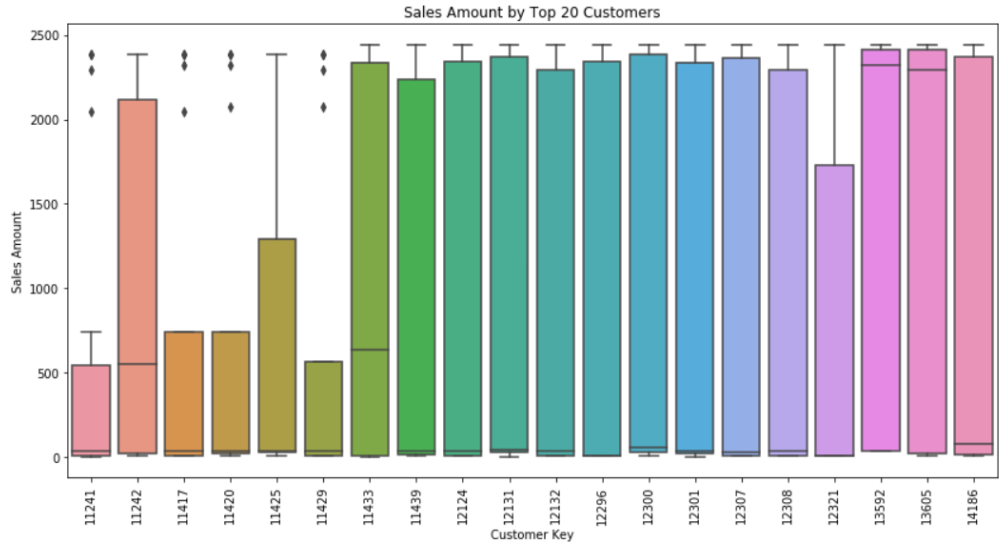
Patterns, Predictions and Suggestions to the Company:



We can conclude that company make lesser income from costlier products than the least expensive ones. Lets take a look into the cheaper ones in detail

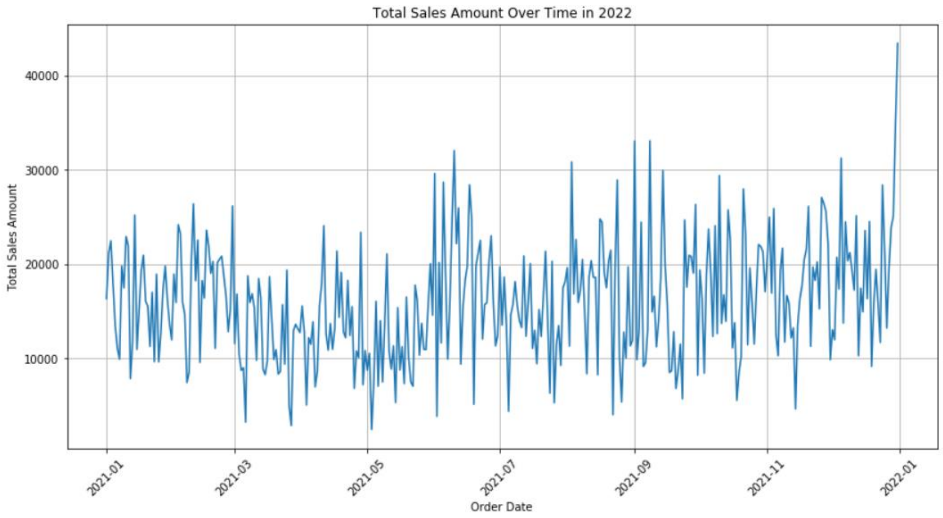


We can clearly get a inference that the company has more amount of sales for low priced products in the range of 1-10 Dollars



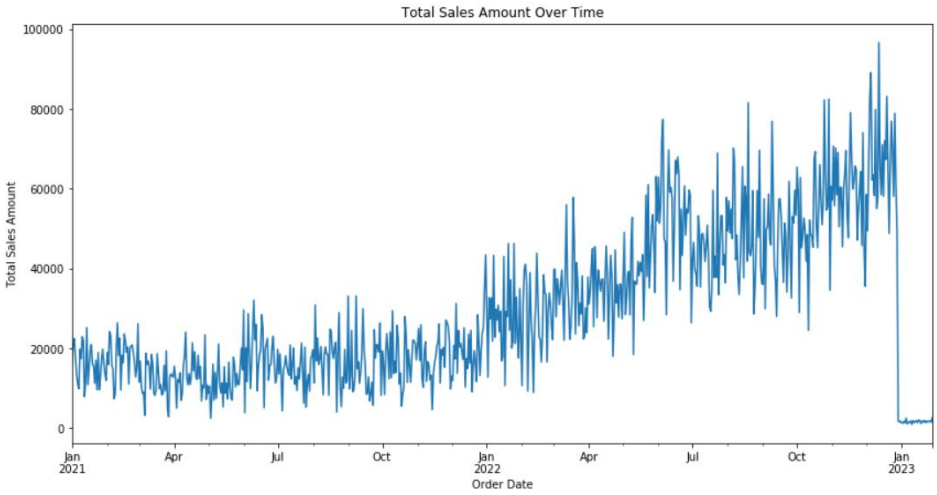
Our analysis indicates that our top customers purchase products across a wide range of prices. While a small subset of customers predominantly buy higher-priced items and contribute significantly to the company's revenue, the majority engage in varied purchasing behavior.

Based on these insights, we recommend that the company continue to offer a diverse range of products at different price points to cater to the broader customer base and maximize overall revenue.



It is evident that the company's profit has been increasing each year. However, we observe a dip in purchases during May and a significant increase as we approach the New Year and Christmas, particularly around November.

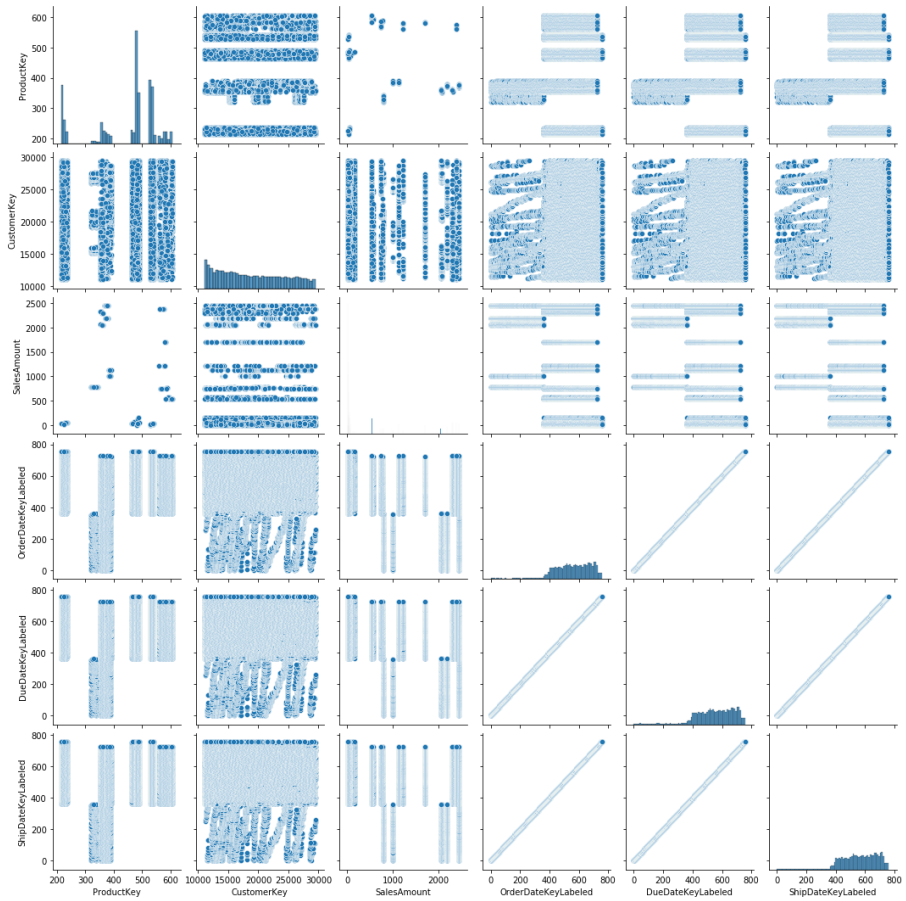
To capitalize on these trends, we recommend importing more goods couple of months before the festive seasons to meet the heightened demand and further boost sales.

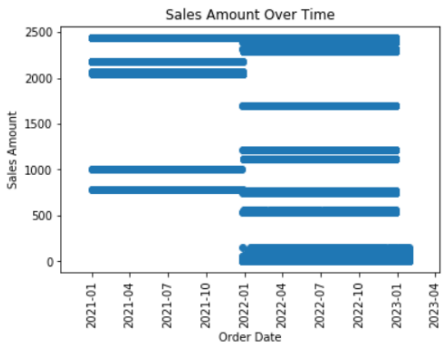


It is evident that the company's sales and advertising teams are performing well, as there has been a consistent increase in sales over the years. Based on this trend, we can predict a similar rise in future sales if the sales and advertising teams continue to receive support and encouragement for their efforts.

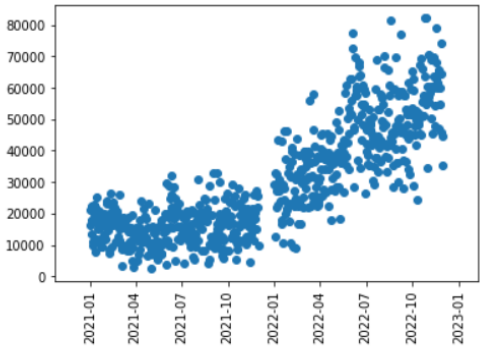
NOTE: The dip in Jan 2023 is due to lack of data

We can see lots of horizontal/vertical strips in the scatter plot, meaning the data is categorical and some behaviours has gaps





Since we can see strips and its categorical, we need to have some aggregations on overall Sales amount on one particular day



Since we are having graphs like this, we build a any regression model to predict the future sales amount

Models Selected:

This scatter plot appears to show a relationship between two variables with a non-linear pattern. There are several types of models that could potentially fit this kind of data:

1. **Polynomial Regression**: This model can capture the curvature in the data by including polynomial terms of the predictor variables.
2. **Logarithmic or Exponential Models**: Depending on the pattern, a logarithmic or exponential model might fit the data well.
3. **Generalized Additive Models (GAMs)**: These models allow for flexible fitting of the data by using non-linear functions.
4. **Piecewise Linear Regression**: This model fits multiple linear segments to different parts of the data.