

**AMAZON SALES ANALYSIS**

**Content:**

* Data Description
* BUSINESS UNDERSTANDING
* Data Understanding
* Data Preparation
* Data Visualization in Tableau
* Machine Learning:
* Data Transformation
* Modelling
* Conclusion

**DATA DESCRIPTION OF AMAZON SALES DATASET:**

This dataset provides detailed insights into Amazon sales data, including SKU Code, Design Number, Stock, Category, Size and Colour, to help optimize product profitability.

**Category:** Type of product. (String)

**Size:** Size of the product. (String)

**Date:** Date of the sale. (Date)

**Status:** Status of the sale. (String)

**Fulfilment:** Method of fulfilment. (String)

**Style:** Style of the product. (String)

**SKU:** Stock Keeping Unit. (String)

**ASIN:** Amazon Standard Identification Number. (String)

**Courier Status:** Status of the courier. (String)

**Qty:** Quantity of the product. (Integer)

**Amount:** Amount of the sale. (Float)

**B2B:** Business to business sale. (Boolean)

**Currency:** The currency used for the sale. (String)

**BUSINESS UNDERSTANDING**

This dataset provides a comprehensive overview of e-commerce sales data from different channels covering a variety of products. Using this dataset, retailers and digital marketers can measure the performance of their campaigns more accurately and efficiently.

**Business Analysis:**

1. Sales Analysis

* What is the total revenue generated over a specific time period?
* What are the top 10 products by sales?
* What is the average order value?

1. Customer Analysis

* What are the top cities/states/countries for sales?
* Are there any trends in order cancellations?
* What is the distribution of B2B vs B2C customers?

1. Shipping Analysis

* What is the average shipping time?
* How does the shipping time vary between different shipping service level.
* What percentage of orders are fulfilled by Amazon vs Merchant?

1. Product Analysis

* Which categories of products are most popular?
* Are there specific styles that are more popular than others?
* What is the average quantity of products ordered?

1. Promotional Analysis

* How effective are the promotions?
* Which promotions are most commonly used?

1. Seasonal Analysis

* Are there seasonal trends in sales (e.g., festivals, holidays)?
* How do seasonal trends affect various product categories?

1. Return and Cancellation Analysis

* What is the return and cancellation rate?
* Are there specific products or categories that have a higher return rate?

By Analysing all the factors above a successful business can be set up targeting specific audience, providing discounts sales, improving services etc.

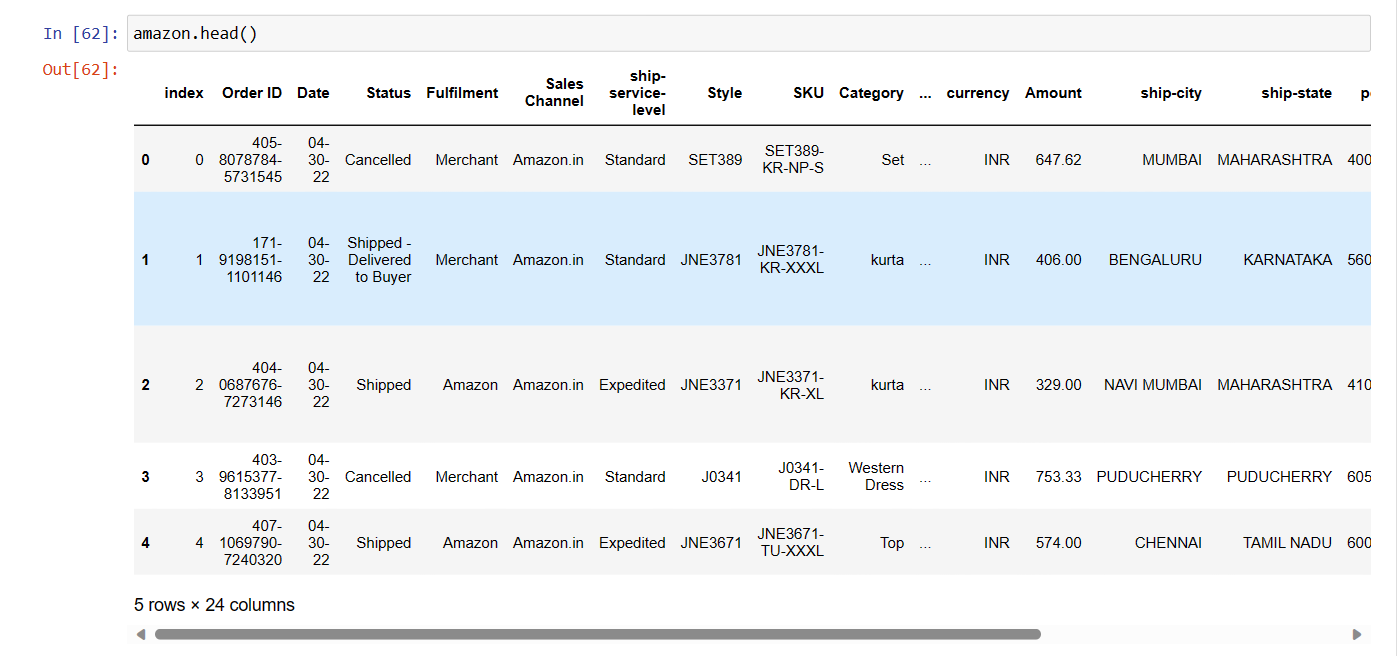
The analyses can be done on SQL, Tableau as well as Python.

**Machine Learning Problem:**

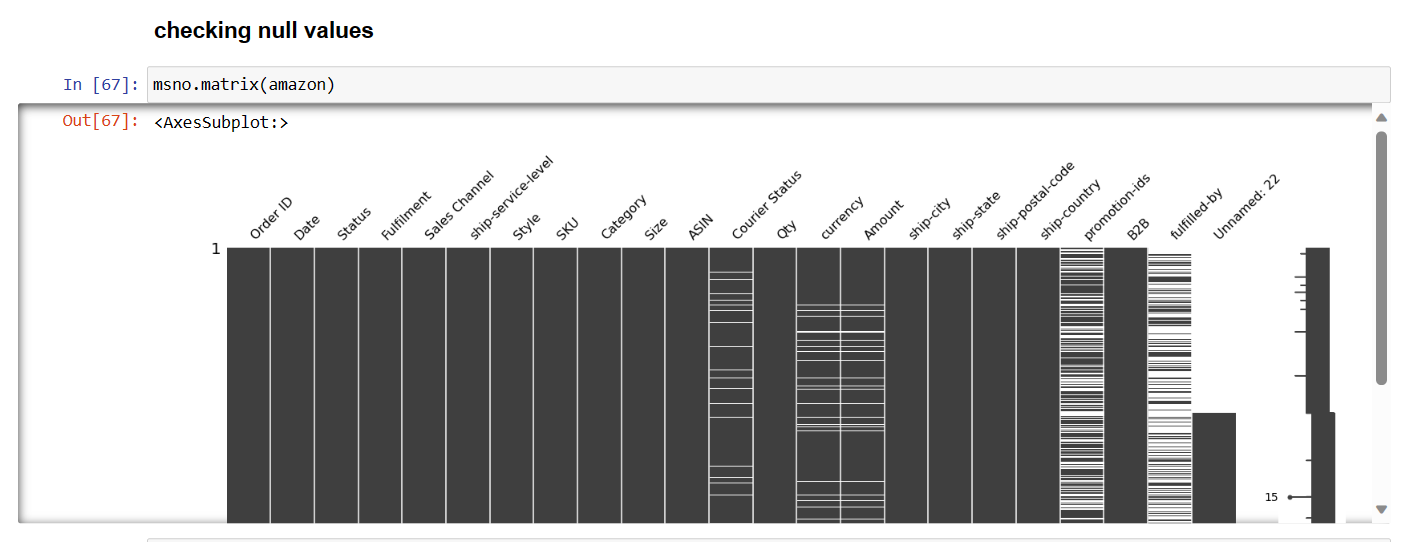
Creating a Clustering model using K-Means Clustering to cluster the Customers and make decisions.

**Data Understanding:**

The Dataset Consists of a total of 23 Columns and 128975 total rows.



Also, we have a lot of Null Values in some of our Columns that we need to handle (the white space in between bars is Null Values)



**Data Preparation:**

**Data Cleaning:**

* The Null Values in our Dataset are mostly in following Columns - 'Unnamed: 22','fulfilled-by','ship-country', 'currency'.
* These Columns are not need by us in dataset as there is not much information. Therefore. We dropped these Columns.
* Also, in few other columns we need to handle the missing values too. Such as, *promotion-ids* null values is filled with no promotion and *Courier Status* is filled with unknown.
* Also, Amount is filled with 0 because if its null its means nothing is purchased by the Customer.
* Lastly, the columns associated with location are also filled with unknown values.

**Removing Duplicate Values:**

We have duplicate values in our dataset and therefore we remove them.

We removed a total of 6 rows in our Dataset.

**Changing Data Type:**

We convert the date column into datetime datatype for better analysis.



Lastly we handled our CustomerType column as for True we replaced it with Business and for False we replaced it with Consumer.

The Data is fully Cleaned and prepared for further Analysis.

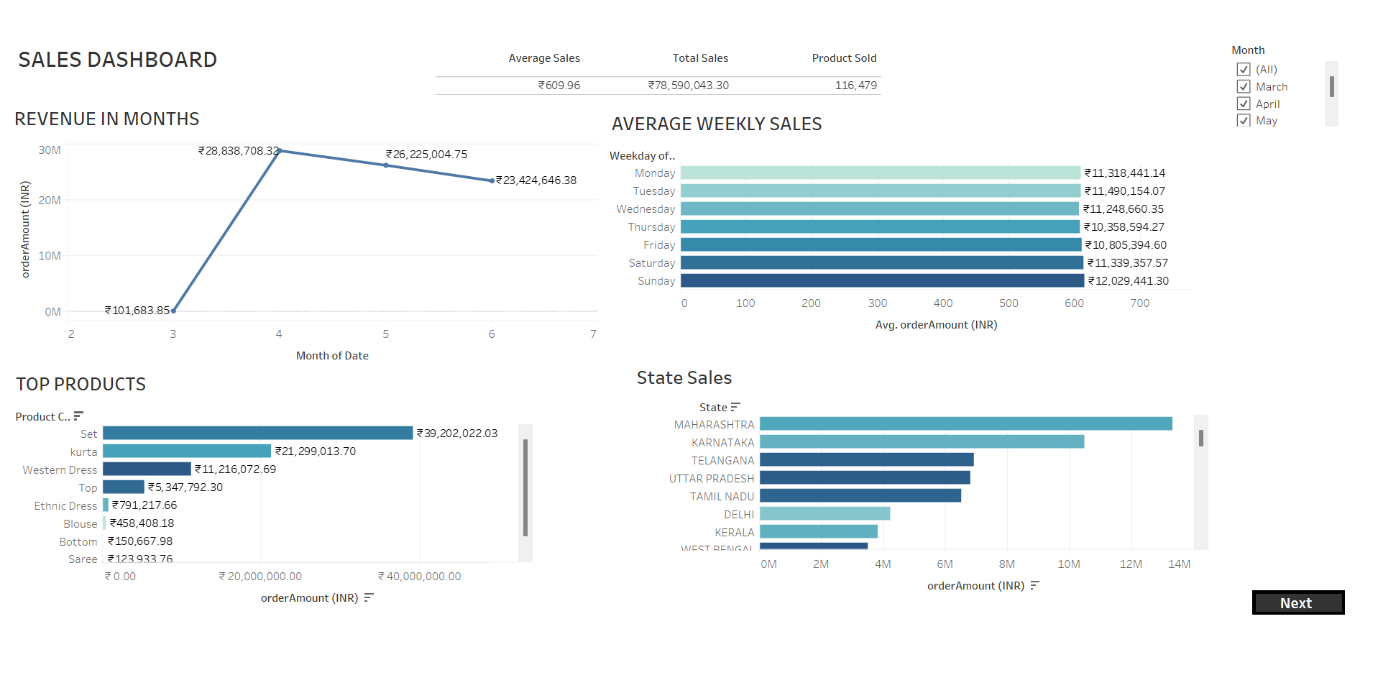
**Data Visualization in Tableau:**

The Cleaned Data is uploaded to Tableau for Business analysis and dashboards were created for Business Decisions.

All the Business Problems Stated Above are solved in the Visualizations.

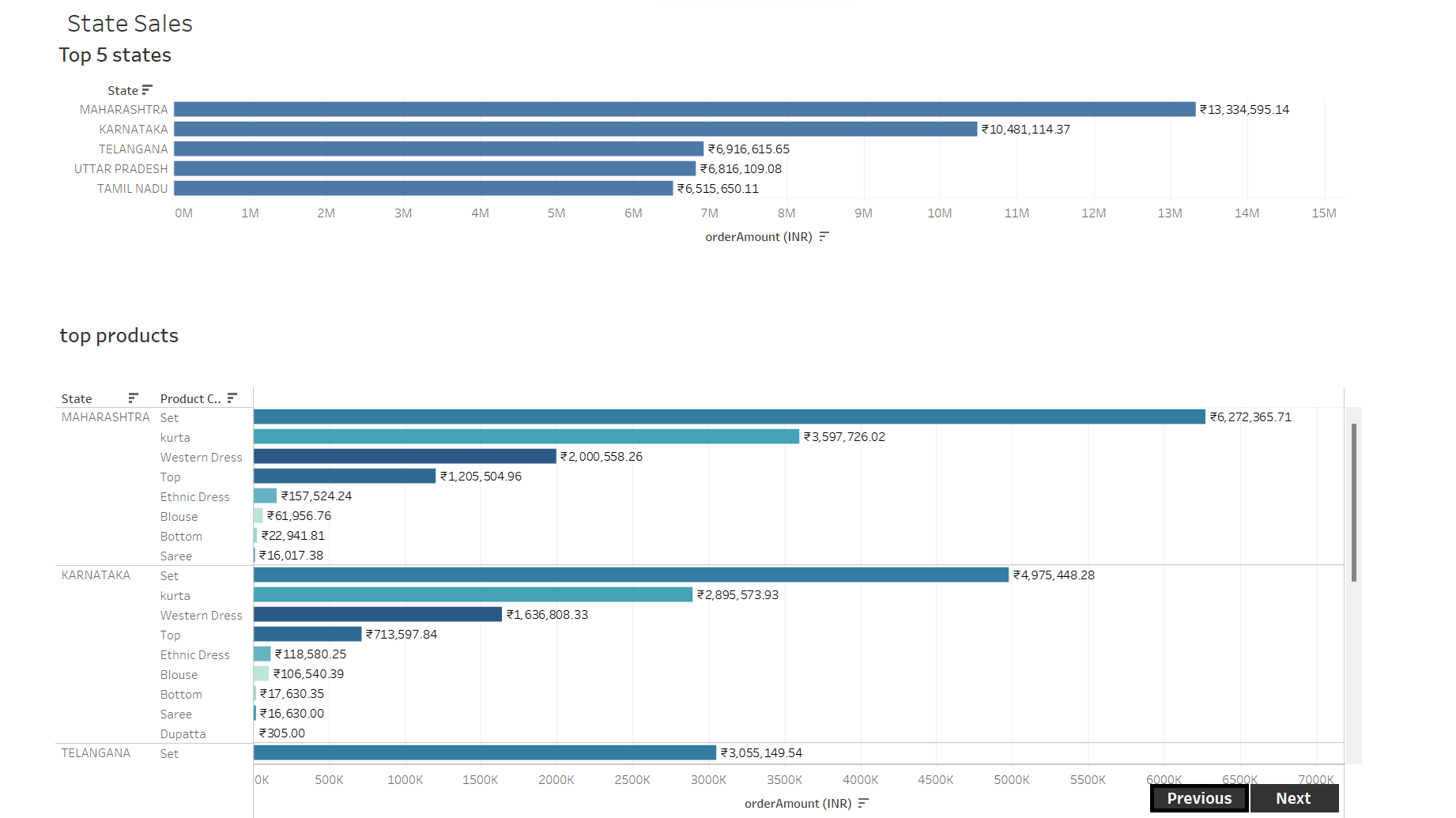
**Sales Dashboard:**

From this Dashboard we get to know that April Month has the highest number of sales with Set being the most purchased product specially on the Weekend days.



**State Sales:**

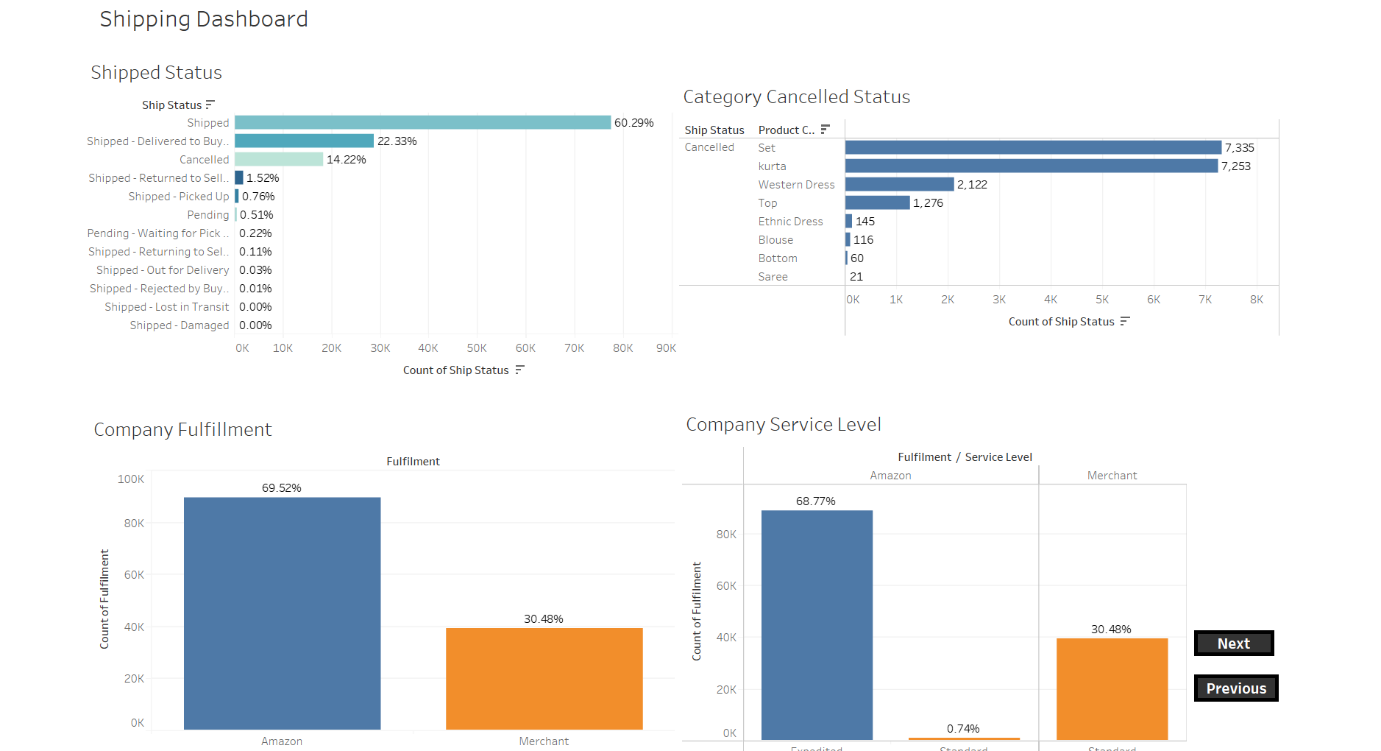
The state of Maharashtra shows the highest number of sales among all the other states.



**Shipping Dashboard:**

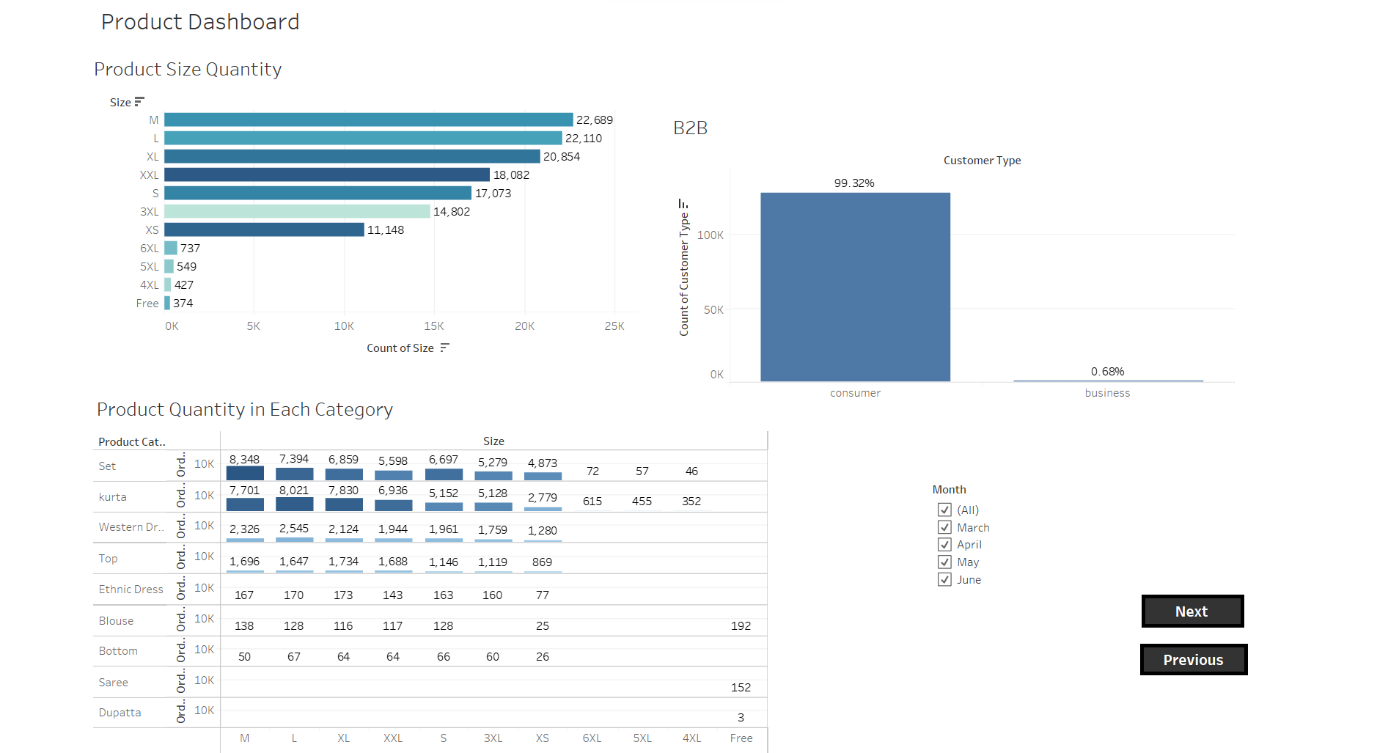
More than half of the total percentage of the products were shipped.

A few were even cancelled that included Set on the top followed by Kurta.

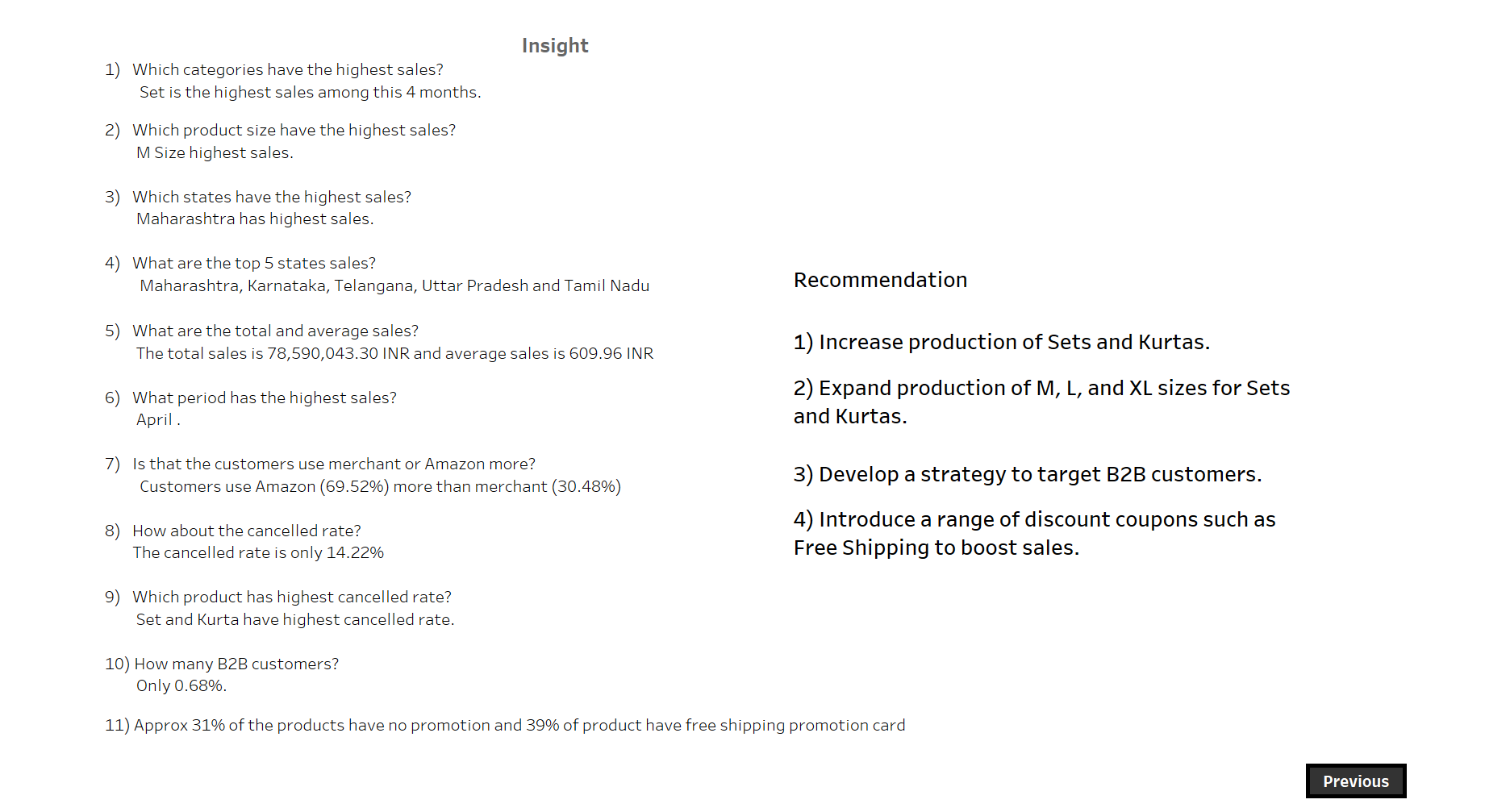


**Product Dashboard:**

Size that was mostly sold was Medium size products. Also most of the products were purchased by direct customers from Amazon rather than merchants.



After looking at the Dashboards we have following insights about our Dataset as well as recommendations.



**Machine Learning:**

The Machine Learning model to be used for this project is Clustering. Here in this project, we use K-Means Clustering Model to Cluster the Customers so that we can make decision based on the Clusters.

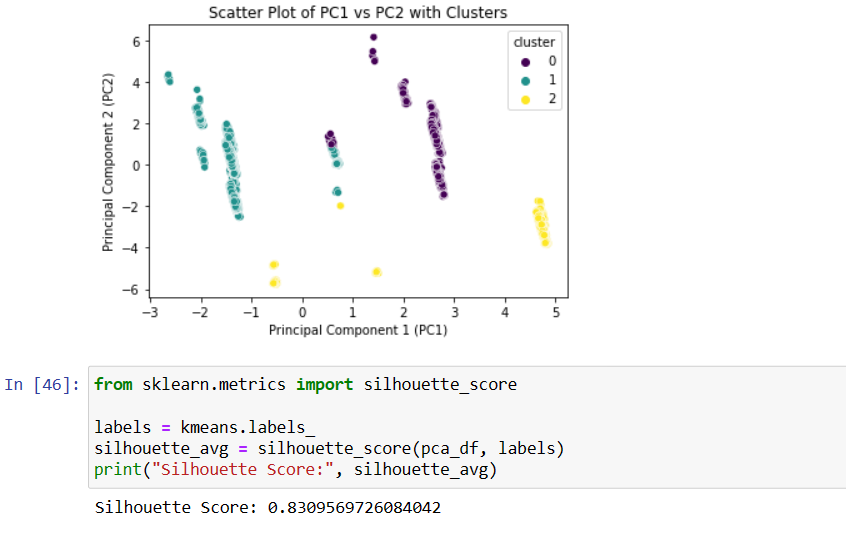
**Data Transformation**

* After Data Cleaning the data needs to be transformed according to the Clustering algorithm. K-means clustering requires transforming the data into Standard Scaler.
* Also, we one hot Encode our Categorical variables as the algorithm only works on the numerical data.
* According to our requirement we also use Principal Component Analysis (PCA) for Multicollinearity and Curse of Dimensionality

**Modelling**

The transformed data is further used in modelling phase. The k-means Clustering requires an optimal value of k, which is derived by plotting an Elbow plot.

After using an optimal value of k, in K-Means Clustering we get a ***Silhouette score of 0.83*** which helps in defining a good Machine Learning Model.



**Conclusion**

Overall we can conclude that Amazon makes a good revenue based on some products and also according to the analysis we can make some recommendations such as increase production of Sets and Kurtas. Also we can expand production of M, L, and XL sizes for Sets and Kurtas. Develop a strategy to target B2B customers. Introduce a range of discount coupons such as Free Shipping to boost sales. Also we can cluster our customer according to our model and target specific Customers.