AMAZON SALES ANALYSIS AND PREDICTION

Table of Contents

List of Figures

List of Abbreviations

Abbreviation	Description
ML	Machine Learning
EDA	Exploratory Data Analysis
MSE	Mean squared error
RMSE	Root mean squared error
MAE	Mean absolute error
B2B	Business to Business
SKU	Stock keeping unit

Abstract

In E-Commerce industry machine learning applications directly affect customer service and business growth. It is very important to generate valuable knowledge about customer behavior and sales trends which bringing benefits for both customers and sellers. Accurate sales forecasting helps the organizations to make informed business decisions and enhance the long-term and short-term business requirements. In this scenario, the purpose of this project is to analyze comprehensive review of various factors that affects the Amazon sales data and apply machine learning methods to predict the future sales of the products. In this project we have taken amazon sales analysis dataset and made detailed analysis of the data and by applying various ML techniques and metrics to evaluate the performance. After evaluating, the model is deployed using Flask application framework.

1. Problem Definition

1.1 Overview

The project aims to analyze sales data from Amazon to predict future sales using machine learning techniques. With the exponential growth of E-Commerce, understanding sales patterns and predicting future trends is crucial for businesses to optimize inventory management, marketing strategies and overall business performance.

E-commerce platforms like Amazon generate vast amounts of sales data, including product sales, customer demographics, and purchase behaviors. Analyzing this data can provide valuable insights into customer preferences, market trends and sales dynamics.

1.2 Problem Statement

The problem is to develop a predictive model that utilizes monthly sales data from Amazon to forecast future sales accurately. This involves identifying relevant features, selecting appropriate ML algorithms and evaluating model's performance to ensure reliable predictions.

1.2.1 Objectives

- Analyzing the amazon sales data and visualizing the different variables to determine the various trends and patterns.
- Develop a ML model capable of predicting future sales based on historical sales data.
- Evaluate the model's accuracy and effectiveness in predicting sales trends and deploy the ML model using Flask framework.

2. Introduction

In the world of rising new technology and innovation, E-Commerce industry is advancing with the help of data analysis. Data analysis can help them to understand their business in a quiet different manner and helps to improve the quality of the service by identifying the weak areas of the business.

Amazon sales data analysis focuses on the process of analyzing customer behavior, sales and several other attributes in order to make improved, data driven decisions. Analysis of the sales data is the main factor that contributes to sellers improving their business and increasing their revenue.

Sales forecasting helps the business to enhance the growth and make improved business decisions. Sales prediction involves using statistical and ML technique to forecast future sales and the predictive models can helps to make more accurate sales forecasts.

This project demonstrates how different analysis help to make better business decisions and help analyze customer trends and satisfaction, which can lead to new and better products and services.

3. Literature Survey

Analysis of sales data of a company or retailer has been becoming a widely discussed topic. The sales data can consist of many records, and filtering of sales data to find meaningful intuitions are common techniques in sales analysis. In financial year 2022, Amazon Marketplace reported the highest revenue of 214.62 billion Indian rupees. This was an increase compared to the previous year. Amazon Internet reported a revenue of over 89 billion Indian rupees in during the same year.

MAINI,KAMAL : STUDY OF CUSTOMER ATTITUDE AND PERCEPTION ON SHOPPING WITH AMAZON INDIA LTD (May-2017)

The e-commerce sector has seen unprecedented growth in 2014. The growth was driven by rapid technology adoption led by the increasing use of devices such as smartphones and tablets, and access to the internet through broadband, 3G, etc. which led to an increased online consumer base. Consumers are increasingly adopting electronic channels for purchasing the products of their daily need. The development of information technology has enhanced customer awareness and created a situation where people prefer shopping online rather than physically visiting the retail stores since online shopping provides quality products as well as saves time Methodology: The purpose of this study is to analyse factors affecting on online shopping behaviour of consumers by designing a self-closed ended questionnaire to collect the data. The dimensions included in the study are web site design, reliability, responsiveness, trust and personalization. A research model is developed for understanding the perceptions of rational customers regarding online shopping. The model proposed that e-service quality dimensions are casually linked to the two performance measures of overall service quality and customer satisfaction, and in turn influence customer purchase intentions. We specifically focused on the site Amazon India ltd and studied customer perception and attitude towards shopping with them.

SATYENDRA KUMAR SHARMA, SWAPNAJIT CHAKRABORTI, TANAYA JHA : ANALYSIS OF BOOK SALES PREDICTION AT AMAZON MARKETPLACE IN INDIA: A MACHINE LEARNING APPROACH (Information Systems and e-Business Management 17, 261-284, 2019)

Prediction of customer demand is an important part of Supply Chain Management, as it helps to avoid over or under production and reduces delivery

time. In the context of e-commerce, accurate prediction of customer demand, typically captured by sales volume, requires careful analysis of multiple factors, namely, type of product, country of purchase, price, discount rate, free delivery option, online review sentiment etc., and their interactions. For e-tailers such as, Amazon, this kind of prediction capability is also extremely important in order to manage the supply chain efficiently as well as ensure customer satisfaction. This study investigates the efficacy of various modelling techniques, namely, regression analysis, decision-tree analysis and artificial neural network, for predicting the sales of books at amazon.in, using various relevant factors and their interactions as predictor variables. Sentiment analysis is carried out to measure the polarity of online reviews, which are included as predictors in these models. The importance of each independent predictor variable, such as discount rate, review sentiment etc., is analysed based on the outcome of each model to determine top significant predictors which can be controlled by the marketer to influence sales. In terms of accuracy of prediction, the artificial neural network model is found to perform better than the decision-tree based model. In addition, the regression analysis, with and without sentiment and interaction factors, generates comparable results. The comparative analysis of these models reveals several significant findings. Firstly, all three models confirm that review volume is the most important and significant predictor of sales of books at amazon.in. Secondly, discount rate, discount amount and average ratings have minimal or insignificant effect on sales prediction. Thirdly, both negative sentiment and positive sentiment of the reviews are individually significant predictors as per regression and decision-tree model, but they are not significant at all as per neural network model. This observation from the neural network model is contrary to the extant research which claims that both negative and positive sentiment are significant with the former having more influence in predicting sales. Finally, the interaction effects of review volume with negative and positive sentiment are also found to be significant predictors as per all three models. Hence, overall, out of various factors used for sales prediction of books, review volume, negative sentiment, positive sentiment and their interactions are found to be the most significant ones across all models. The results of this study can be utilized by online sellers to accurately predict the sales volume by adjusting these significant factors, thereby managing the supply chain effectively

4. Result

4.1 Data collection

The amazon sales analysis dataset has been collected from www.Kaggle.com and the dataset contains 128975 rows and 24 columns.

4.2 Data Visualization

Data Visualization is the method of representing the data in a graphical and pictorial way, data scientists depict a story by the results they derive from analyzing and visualizing the data. The best tool used is Tableau which has many features to play around with data and fetch wonderful results.

4.2.1. Monthly sales Analysis

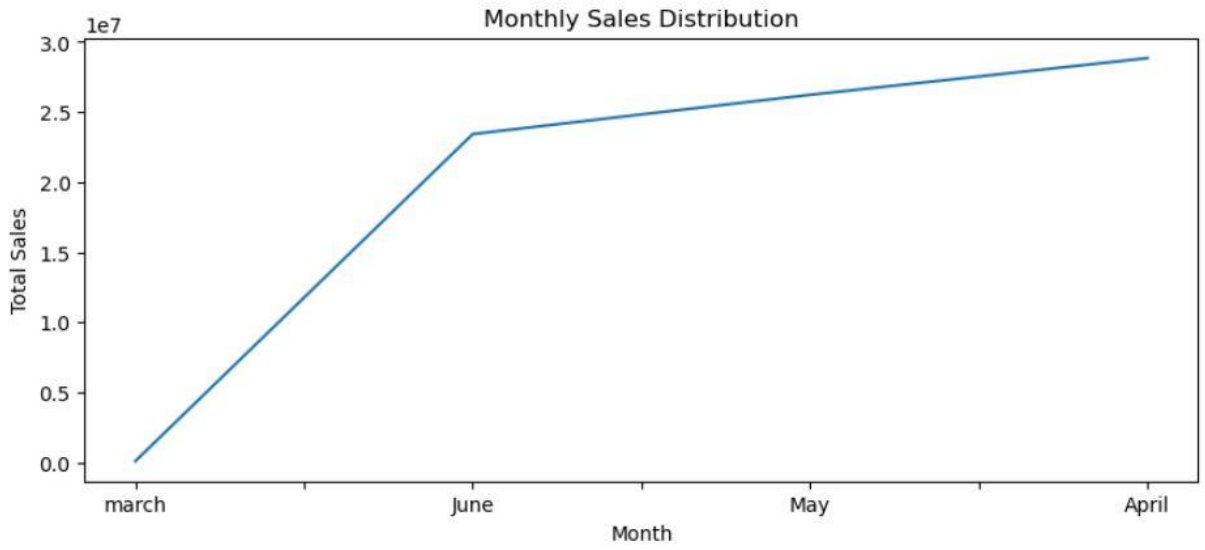

Fig. 1: Monthly sales vs total sales

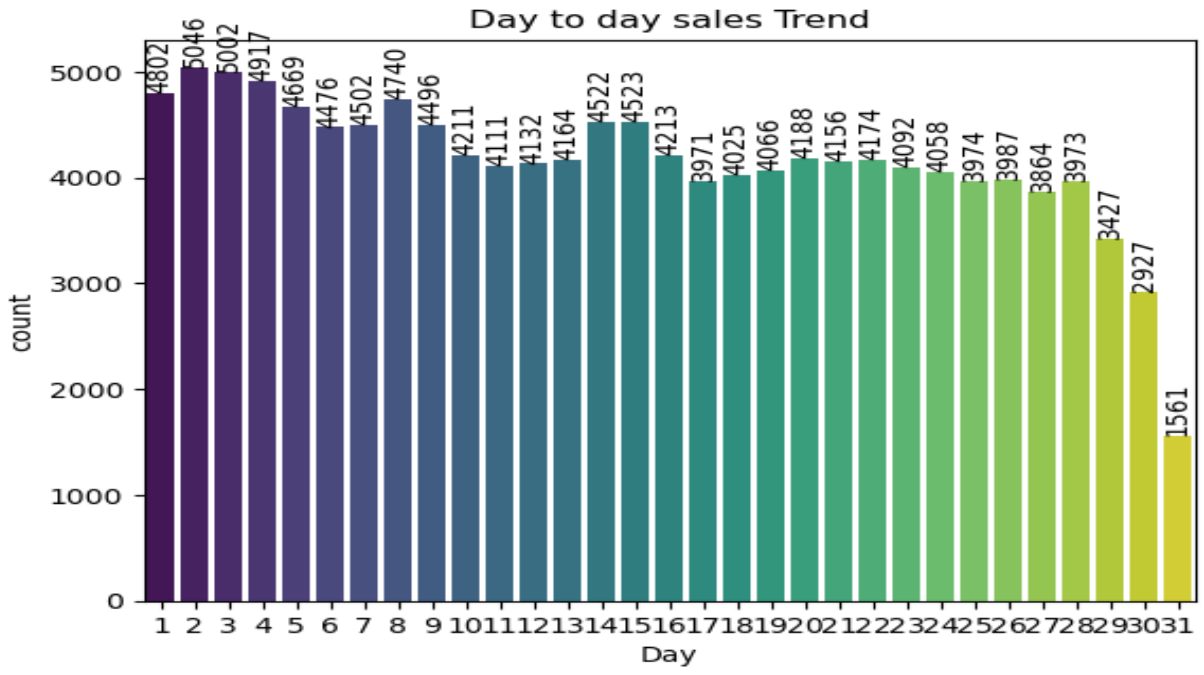

Fig. 2: Day to day sales trend

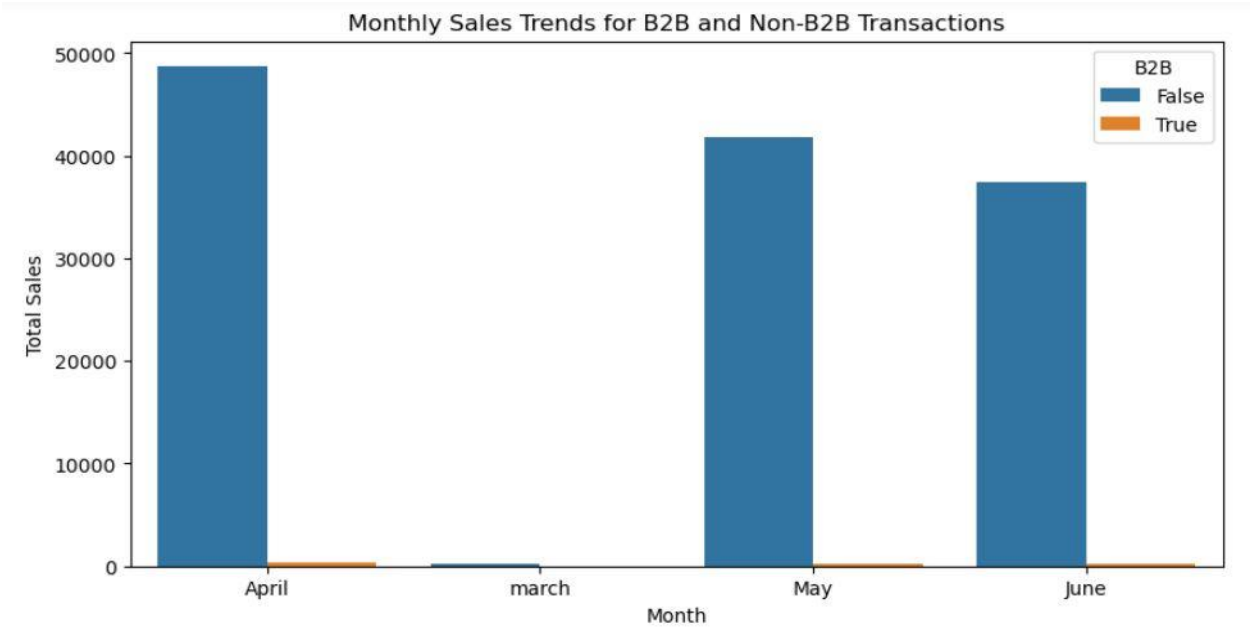

Fig. 3: Sales trend-B2B VS Non B2B

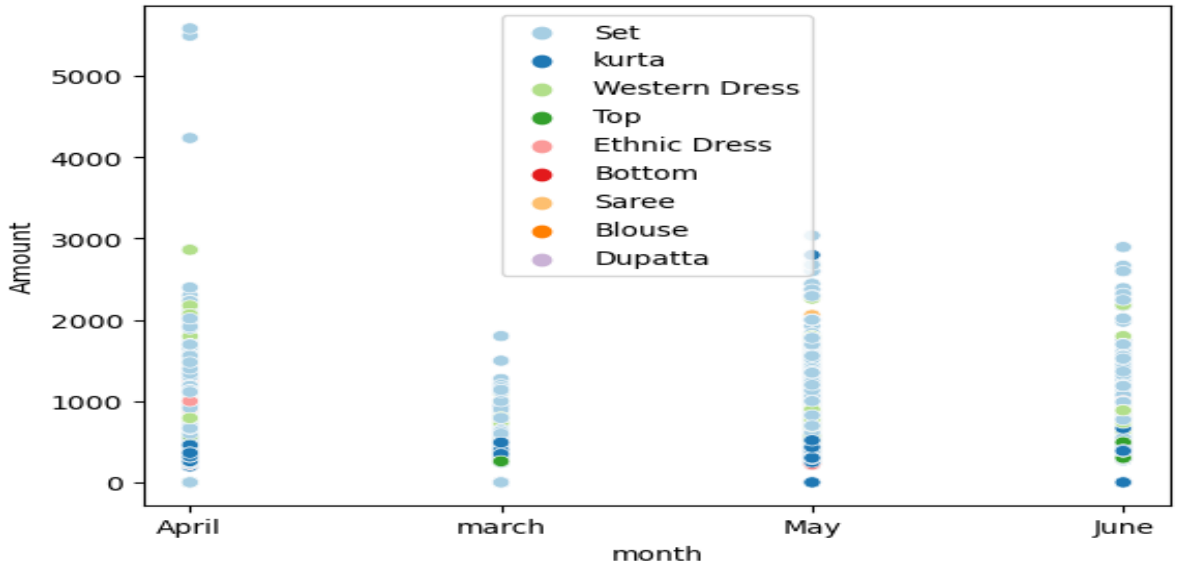

Fig. 4: Monthly sales by category

4.2.2 Product based Analysis

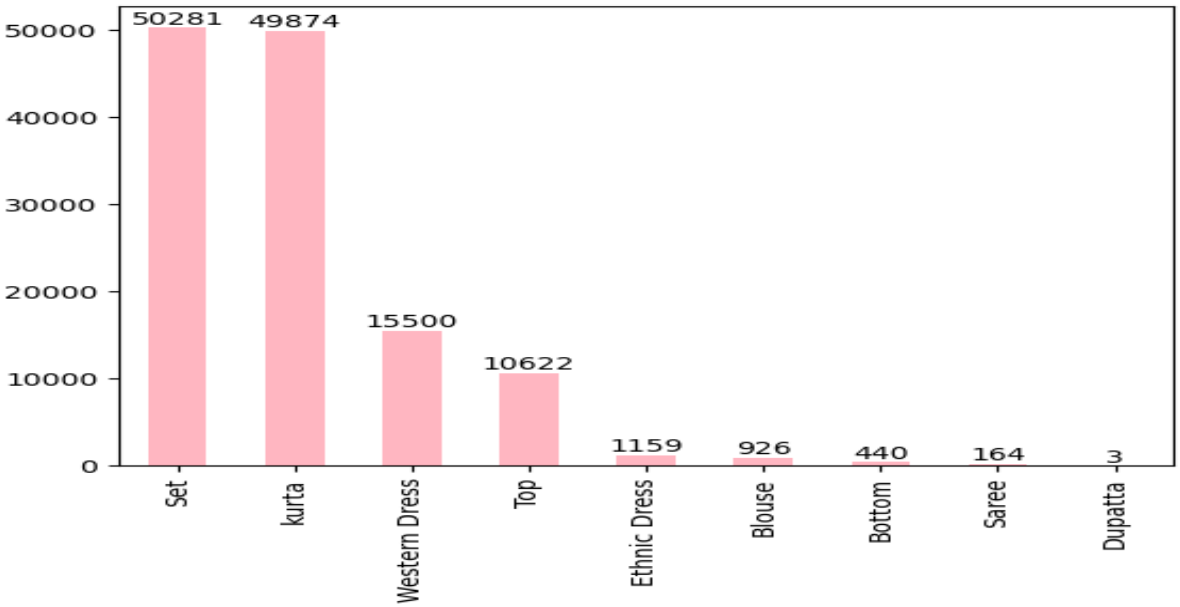

Fig. 5: Highest sold product category

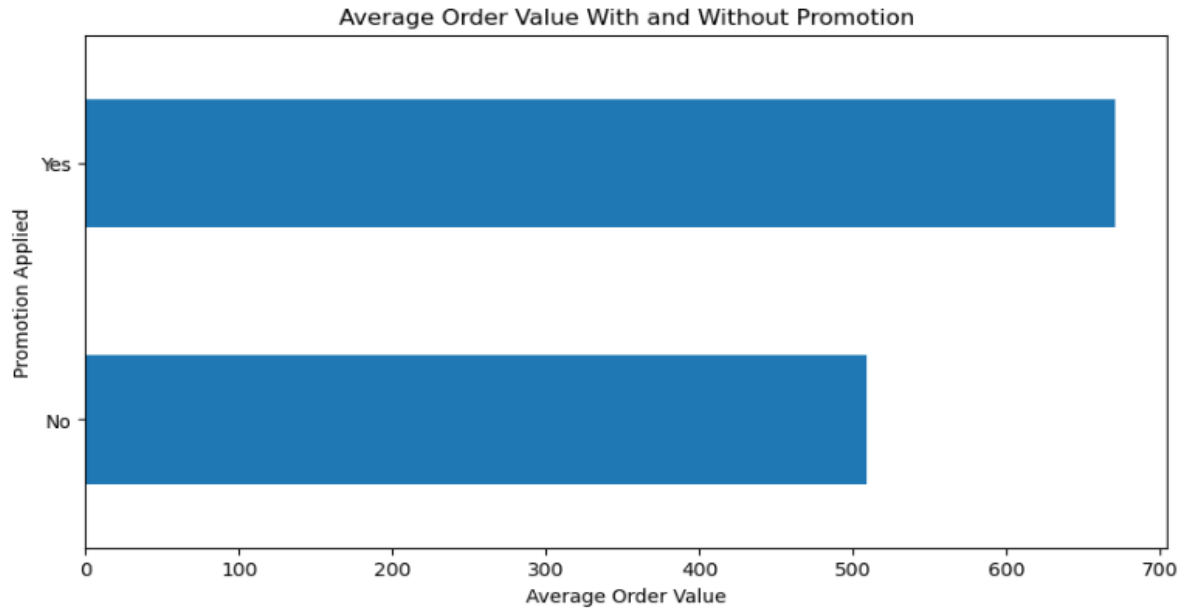

Fig. 6: Average order value with and without promotion

4.2.3 Customer Location Analysis

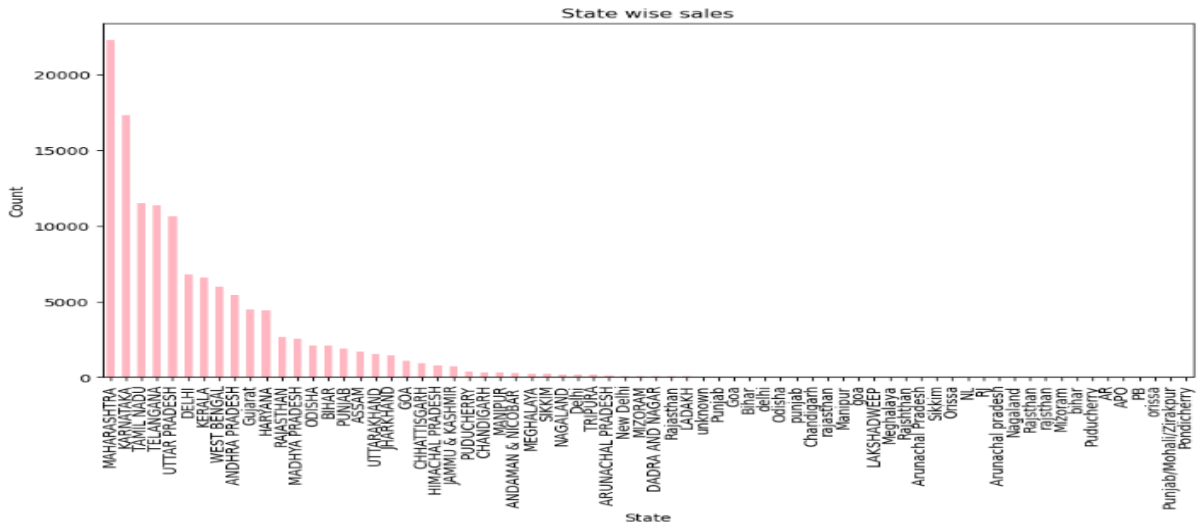

Fig. 7: State wise sales analysis

4.2.4 SKU Analysis

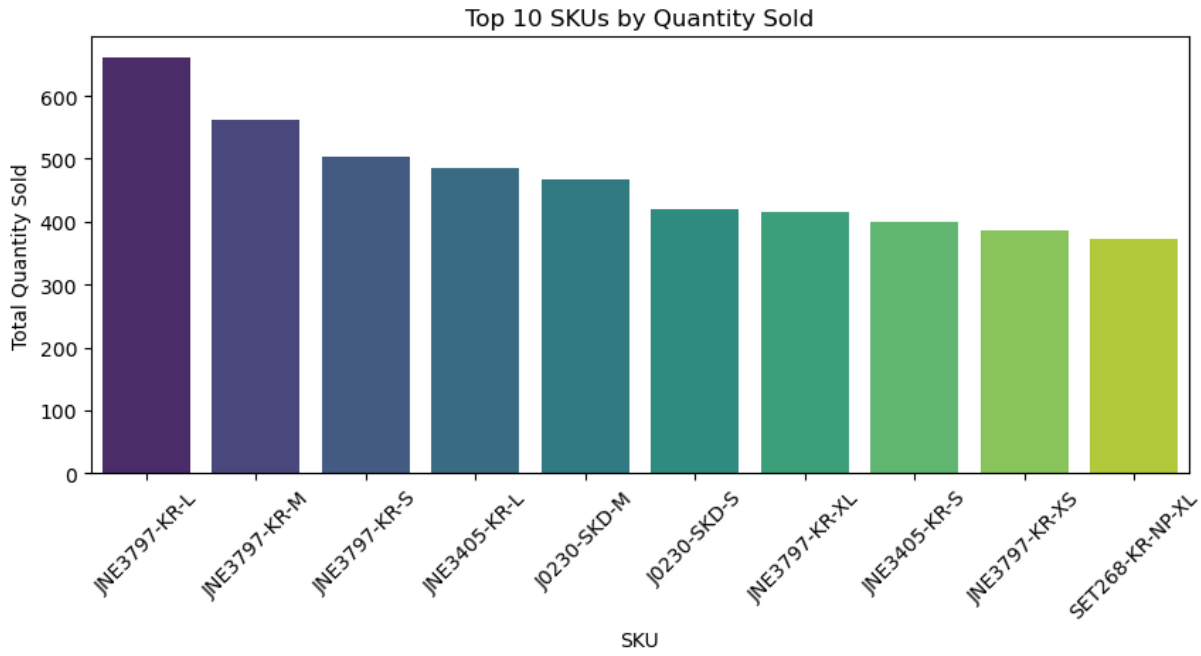

Fig. 8: Top 10 SKUs

4.3 Data preprocessing

Pre-processing is the important step in data mining which involves:

- Data Cleaning: The collected data may contain missing values, outliers and Errors. Therefor it is essential to clean the data by imputing missing values, removing outliers etc.
- Data preparation: After data cleaning, it is important to prepare for analysis which includes scaling, normalization and encoding of categorical values.
- Train-test split: This involves splitting the dataset into training and test data. The training set is used to build the model, while testing set is used to evaluate the model's performance.

```
x = data[["month","Category"]]
y = data["Qty"]
```

```
x_train, x_test, y_train, y_test = train_test_split(x, y,
                                                    test_size=0.2,
                                                    random_state=42)
print(x_train.shape, x_test.shape, y_train.shape, y_test.shape)
```

```
(100638, 2) (25160, 2) (100638,) (25160,)
```

```
from sklearn.preprocessing import StandardScaler
# Scale the input features

scaler = StandardScaler()
x_train_scaled = scaler.fit_transform(x_train)
x_test_scaled = scaler.transform(x_test)
```

Fig. 9: Data preprocessing

4.4 Evaluating the model

By using Decision tree regression algorithm, the model is evaluated to asses its performance. The performance metrics used for evaluation includes, mean squared error (MSE), Root mean squared error (RMSE) or R-squared.

```
from sklearn.tree import DecisionTreeRegressor
model = DecisionTreeRegressor()
model.fit(x_train_scaled, y_train)
```

```
▾ DecisionTreeRegressor
DecisionTreeRegressor()
```

```python
# Evaluate the model
from sklearn.metrics import mean_absolute_error #MAE

mae = mean_absolute_error(y_test, y_pred)
print(f'Mean Absolute Error: {mae}')

# Evaluate the model
from sklearn.metrics import mean_squared_error
mse = mean_squared_error(y_test, y_pred)
print(f'Mean Squared Error: {mse}')

# Assuming y_pred and y_test are NumPy arrays or similar data structures
rmse = np.sqrt(mean_squared_error(y_test, y_pred))

print(f'RMSE: {rmse:.2f}')

from sklearn.metrics import r2_score
# Assuming y_pred and y_test are NumPy arrays or similar data structures
r2 = r2_score(y_test, y_pred)

print(f'R^2 Score: {r2:.2f}')
```

```
Mean Absolute Error: 0.03512392424192033
Mean Squared Error: 0.020717685972782104
RMSE: 0.14
R^2 Score: 0.78
```

Fig. 10: Model Evaluation

4.5 Prediction and Model Deployment

Once the model is trained and evaluated, it can be used to make predictions on new data. Here we use Flask application framework to deploy the model to predict sales for a future period based on historical sales data.

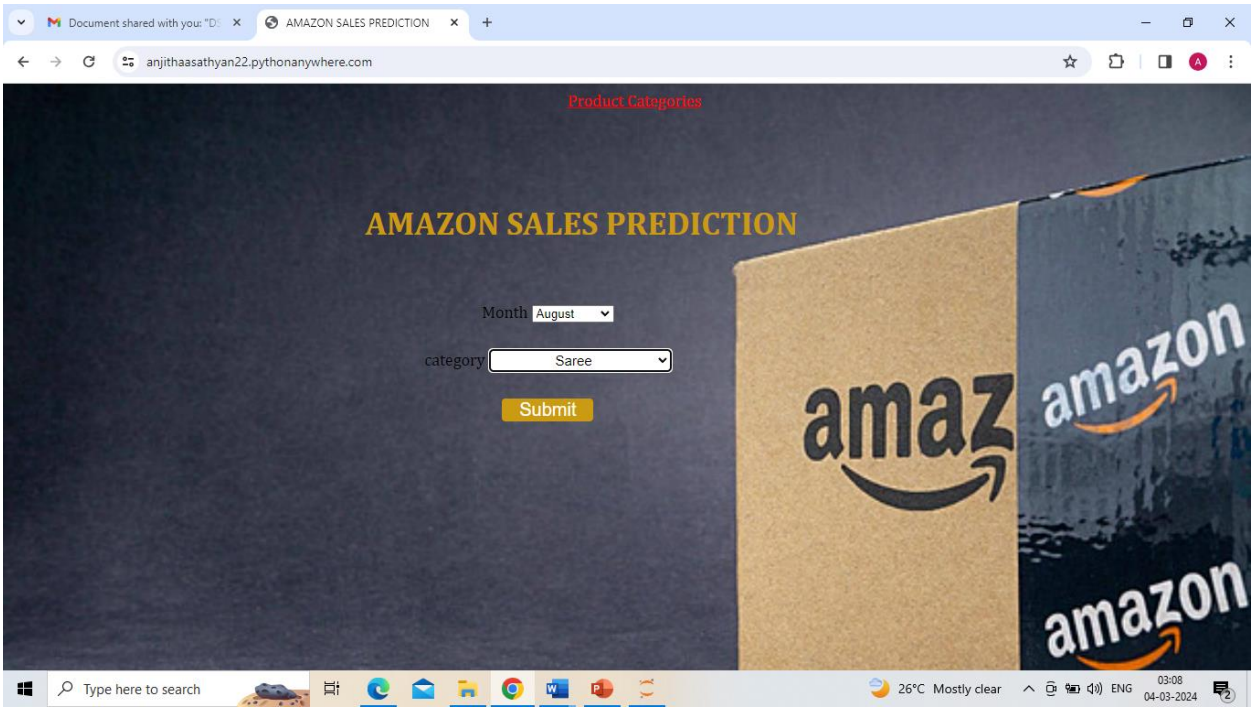

Fig. 11: Sales prediction Home page

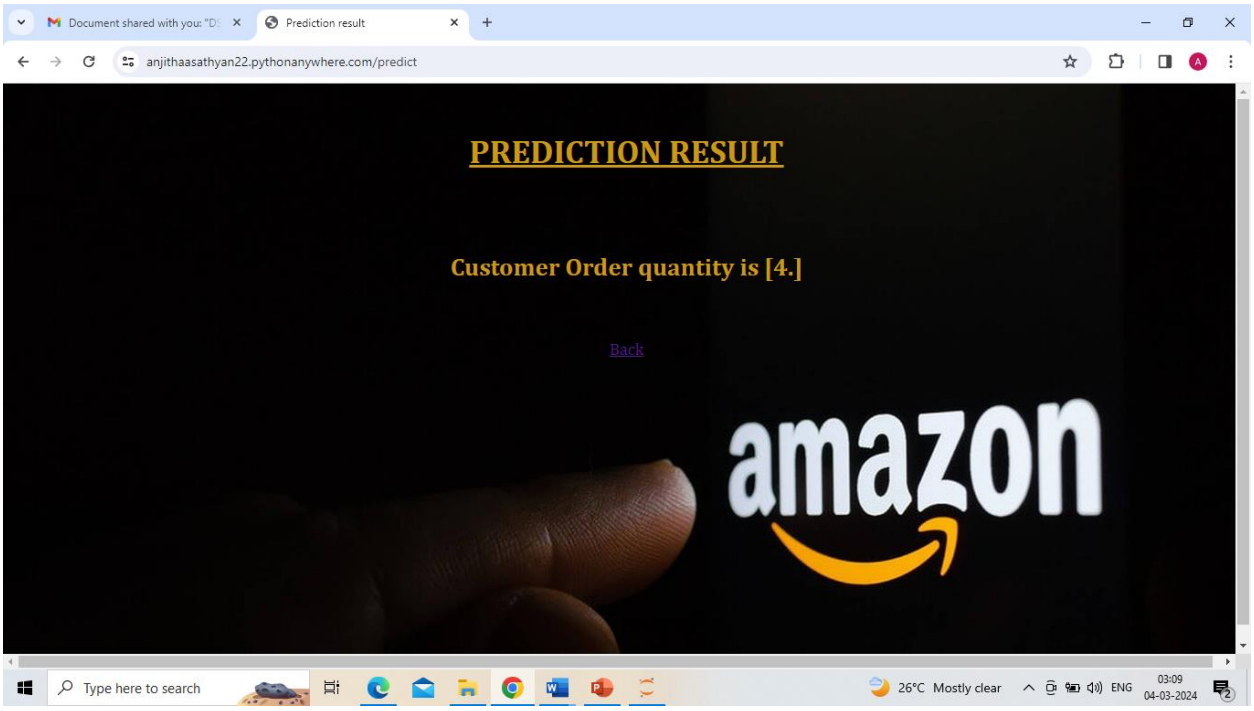

Fig. 12: Prediction Result

5. Conclusion

With the demand of E-commerce applications, sales forecasting is needed for business organizations for making business decisions and predicting the future sales. Accurate forecasting will help companies to enhance the sales and market growth. In the Amazon sales analysis project, we build a machine learning model to predict the future sales quantity using historical sales data of Amazon. This project involves analyzing various factors such as month, category, size, SKU, B2B etc. to determine their impact on sales. By predicting the purchase quantity, the company can apply some marketing strategies for certain sections of customers and able to make profits.

Overall, the amazon sales analysis and prediction project have significant potential to help Amazon to make informed decisions and improve profitability of sales.

6.References

[1] Avinash kumar, Neha Gopal & Jatin Rajput(2020). An Intelligent Model For Predicting the Sales of a Product.

[2] Sunitha Cheriyan, Shaniba Ibrahim, Saju Mohanan & Susan Treesa (2018) Intelligent Sales Prediction Using Machine Learning Techniques.

[3]. Jaccard, J., & Turrisi, R. (2018). Interaction effect in multiple regression second edition. Sage Publications,
Thousand Oaks CA.

[4]. Sullivan, J. (2019). Data cleaning with r and the tidyverse: Detecting missing values. https : / /towardsdatascience . com / data - cleaning -with - r - and - the - tidyverse - detecting-missing-valuesea23c519bc