Q

Tech Saksham

Capstone Project Report

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING FUNDAMENTALS

“E-COMMERCE SALES ANALYSIS”

“THIRUVALLUVAR COLLEGE OF ENGINEERING AND TECHNOLOGY”

NM ID	NAME
Au422321105003	Rajesh.R

Ramarbose	TrainerName

ABSTRACT

This project investigates customer purchasing patterns and product sales trends within a specific E-commerce platform , The analysis employs exploratory data analysis (EDA) techniques to uncover hidden insights from customer demographics, product categories, and historical sales data.

By visualizing data distributions using histograms and scatter plots, the project aims to identify correlations between factors like customer age, location, and preferred product categories. Additionally, it will explore purchase frequency patterns and analyze the impact of product pricing or promotions on sales figures.

If model building is present in the code, the abstract can be expanded to mention the type of model clustering used to segment customers based on their buying habits. The evaluation metrics employed silhouette score

The focus of this abstract is purely on the analytical methods and the potential outcomes in terms of understanding customer behavior and product sales trends. It removes any mention of recommendations or how the findings might be used to improve business strategies.This project investigates customer purchasing patterns and product sales trends within a specific E-commerce platform , The analysis employs exploratory data analysis (EDA) techniques to uncover hidden **i**nsights from customer demographics, product categories, and historical sales data.

By visualizing data distributions using histograms and scatter plots, the project aims to identify correlations between factors like customer age, location, and preferred product categories. Additionally, it will explore purchase frequency patterns and analyze the impact of product pricing or promotions on sales figures.

The focus of this abstract is purely on the analytical methods and the potential outcomes in terms of understanding customer behavior and product sales trends. It removes any mention of recommendations or how the findings might be used to improve business strategies.

INDEX

CHAPTER 1

INTRODUCTION

1.1 Problem Statement

1.2 Proposed Solution

1.3 Feature

1.4 Advantages

1.5 Scope

1.6 Future Work

CHAPTER 1

INTRODUCTION

and analyze data could be a huge impact to an organization, as the outcome can be positive or it can go the other way. E-Commerce Platforms collect a large amount of data and store it in their data centers. They fail to look at this as an advantage for thither domain for this research is E-commerce. Starting a new method to gather r business opportunity such and analyzing the data and its pattern trough out the years. For example, All the customer data from registration, search history, sales, chats are being stored in their server and will be only be used when there is a problem with existing data. It's understandable why they won't want other company to analyze their data is due to privacy issues but they are also able to create their own team to analyze the data which can be profitable for them. Its shown that E-Commerce are one of the top 10 business which have a very large amount of data storage. With the access to the amount of data they will be able to create a game changing environment for the E-Commerce industry. This industry has been spending hundreds of millions of dollars in advertising, social media, sorting secured data and much more to generate more sales but they didn't realize that with machine learning they will be able to step up they game against their competitors. Machine learning is a large tree branch which have many specializations such as data mining, artificial intelligence, augmented reality and prediction. For this research, will be only focusing on the With the ability of prediction using machine learning algorithm for e-commerce, we will be able to identify any hidden patterns, outliers, point of interest (POI) and much more. This will allow e commerce to be able to properly identify the important details in each and every aspect. They will be able to use all their data such as amount of product purchased, product categories, payment method, interest rate, duration of delivery and customer location to have a better understanding on how to improve and manage their sales. If the e-commerce platform is able to forecast its sales for the upcoming month or day, they will be able to make better business decisions. They will also be able to track and trends in their sales if any festival or event happens yearly. They will also be able to keep track of their inventory so all items will be stock sufficiently which will avoid any overstocking and under stocking of a product as they will be able to get a rough estimate of purchases which are likely to happen. Not only that, they will also be able to keep better track at their finance and make reasonable purchases option and have a proper budget throughout the business operation.

1.1 Problem Statement

In the realm of e-commerce, businesses face significant challenges in effectively analyzing sales data to derive actionable insights. Traditional approaches often fall short in providing accurate predictions and meaningful recommendations. Key issues include:

- - Complexity of data: E-commerce platforms generate vast amounts of heterogeneous data, including sales transactions, customer interactions, and product attributes. Analyzing this data manually is time-consuming and prone to errors.
- - Dynamic market trends: E-commerce markets are highly dynamic, with trends evolving rapidly. Traditional analytical methods struggle to adapt to these changes quickly enough to provide relevant insights.
- - Personalized customer experiences: Customers expect personalized shopping experiences tailored to their preferences and behaviors. Conventional methods often lack the sophistication to segment customers accurately and deliver personalized recommendations.

These challenges highlight the pressing need for advanced analytical techniques, such as machine learning, to unlock the full potential of e-commerce sales data and drive business growth.

1.2 Proposed Solution

In response to the challenges outlined in the problem statement, the proposed solution entails leveraging machine learning (ML) techniques to analyze e-commerce sales data effectively. By harnessing the power of ML algorithms, businesses can extract valuable insights, predict sales trends, and personalize customer experiences. The following outlines the components of the proposed solution:

1. Data Collection and Pre processing:

- - Gather diverse data sources including sales transactions, customer demographics, website interactions, and marketing campaigns.
- - Pre process the data to handle missing values, outliers, and inconsistencies. Perform data cleaning, normalization, and encoding of categorical variables.

1. Exploratory Data Analysis (EDA):

- - Conduct comprehensive EDA to understand the underlying patterns and relationships within the data.
- - Visualize sales trends over time, analyze customer purchasing behavior, and identify correlations between different features.
- - Utilize statistical techniques and data visualization tools to gain actionable insights.

2. Feature Engineering:

- - Extract relevant features from the data set that can enhance predictive modeling.
- - Create new features based on domain knowledge and insights gained from EDA.
- - Transform data into suitable formats for input into machine learning models.

3. Machine Learning Models:

- - Develop and train ML models to address specific e-commerce sales analysis tasks, including:
- - Regression models: Predict future sales based on historical data, taking into account seasonality and other factors.
- - Classification models: Segment customers based on purchasing behavior, demographics, or preferences.
- - Recommendation systems: Generate personalized product recommendations to enhance customer engagement and sales.
- - Employ ensemble methods, such as random forests or gradient boosting, to improve model performance.

4. Model Evaluation and Validation:

- - Evaluate the performance of ML models using appropriate metrics, such as mean squared error (MSE) for regression tasks or accuracy and precision for classification.

- - Utilize cross-validation techniques to ensure model robustness and generalization to unseen data.
- - Validate models using holdout datasets or time-based splitting to simulate real-world performance.

5. Deployment and Integration:

- - Deploy trained ML models into production environments, either on-premises or in the cloud.
- - Integrate ML-powered insights into existing e-commerce platforms, dashboards, or decision support systems.
- - Monitor model performance and retrain models periodically to adapt to changing market dynamics.

7. Continuous Improvement:

1) - Continuously monitor and refine the ML models based on feedback and performance metrics.
2) - Explore advanced ML techniques, such as deep learning or reinforcement learning, to further enhance predictive capabilities.
3) - Stay updated on emerging trends and innovations in ML and e-commerce to maintain a competitive edge.

By implementing this proposed solution, businesses can unlock the full potential of their e-commerce sales data, drive revenue growth, and deliver superior

customer experiences through personalized recommendations and targeted marketing strategies.

1.3 Feature

Feature engineering plays a crucial role in the success of machine learning models for e-commerce sales analysis. It involves selecting, transforming, and creating features from raw data to improve the performance and effectiveness of predictive models. In this report, we outline the key aspects of feature engineering tailored specifically for e-commerce sales analysis using machine learning techniques.

1. Data Understanding and Exploration:

Before performing feature engineering, it is essential to gain a deep understanding of the underlying data. This includes exploring various data sources such as sales transactions, customer demographics, product attributes, and website interactions. Understanding the nature and characteristics of these data sources is crucial for identifying relevant features for analysis.

2. Feature Selection:

`merce sales analysis, some common features include:

A. - Customer demographics: Age, gender, location, income level, etc.

B. - Product attributes: Category, brand, price, popularity, etc.

C. - Time-related features: Day of the week, month, seasonality, holiday indicators, etc.

D. - Customer behavior: Purchase history, browsing patterns, cart abandonment rates, etc.

E. - Marketing campaign data: Campaign type, duration, effectiveness, etc.

3. Feature Transformation:

Transforming features involves converting raw data into a format that is suitable for input into machine learning models. Common transformations include:

A. - Scaling: Normalizing numerical features to ensure they have similar scales, preventing certain features from dominating others during model training.

B. - Encoding: Converting categorical variables into numerical representations using techniques such as one-hot encoding or label encoding.

C. - Binning: Grouping continuous variables into bins to capture non-linear relationships and reduce noise.

4. Feature Creation:

In addition to transforming existing features, feature creation involves generating new features that may improve model performance. This may include:

a) - Interaction features: Multiplying or combining existing features to capture interaction effects.

b) - Time-related features: Extracting temporal patterns such as day-of-week effects, time since last purchase, or recency-frequency-monetary (RFM) metrics.

c) - Derived features: Calculating ratios, averages, or other statistical summaries from existing features to capture underlying trends or patterns.

5. Feature Importance Analysis:

Once features are selected and transformed, it is essential to assess their importance in predicting the target variable (e.g., sales volume). Techniques such as feature importance scores from tree-based models, permutation importance, or SHAP (SHapley Additive explanations) values can help identify the most influential features and provide insights into the underlying drivers of sales.

1.4 Advantages

By harnessing the power of machine lea E-commerce sales analysis leveraging machine learning offers several distinct advantages for businesses aiming to optimize their operations, enhance customer experiences, and drive revenue growth. This report outlines the key advantages of employing machine learning techniques for e-commerce sales analysis:

1. Data-driven Decision Making:

Machine learning enables businesses to make data-driven decisions by leveraging insights extracted from large volumes of sales data. By analyzing historical sales trends, customer behavior, and market dynamics, businesses can identify patterns and make informed decisions regarding pricing strategies, inventory management, and marketing campaigns.

2. Predictive Analytics:

Machine learning models can forecast future sales trends with a high degree of accuracy, enabling businesses to anticipate demand and optimize inventory levels accordingly. By predicting sales volumes and customer preferences, businesses can streamline their supply chain operations, minimize stock outs, and reduce excess inventory costs.

3. Personalized Customer Experiences:

Machine learning algorithms can segment customers based on their preferences, purchase history, and browsing behavior, allowing businesses to deliver personalized recommendations and tailored marketing messages.

By understanding individual customer needs and **preferences, businesses**

can enhance customer satisfaction, increase engagement, and foster long-term loyalty.

4. Targeted Marketing Strategies:

Machine learning enables businesses to target their marketing efforts more effectively by identifying high-value customer segments and predicting the likelihood of purchase. By analyzing customer demographics, behavior, and response to past marketing campaigns, businesses can optimize their marketing spend, improve conversion rates, and maximize return on investment (ROI).

5. Enhanced Fraud Detection and Security:

Machine learning algorithms can detect patterns indicative of fraudulent transactions, such as unusual purchasing behavior or suspicious payment patterns. By leveraging advanced analytics and anomaly detection techniques, businesses can mitigate the risk of fraud, protect customer data, and safeguard their reputation.

6. Scalability and Efficiency:

Machine learning models can process large volumes of data quickly and efficiently, allowing businesses to scale their operations and adapt to changing market conditions. By automating repetitive tasks and streamlining decision-making processes, businesses can increase operational efficiency, reduce manual errors, and focus resources on strategic initiatives.

1.5 Scope

The scope of applying machine learning techniques to e-commerce sales analysis is vast, encompassing various aspects of data-driven decision-making, predictive modeling, and customer-centric strategies. This report outlines the scope of e-commerce sales analysis using machine learning as follows:

1. Data Collection and Integration:

The scope includes gathering and integrating diverse sources of data relevant to e-commerce sales, including sales transactions, customer demographics, website interactions, product attributes, and marketing campaign data. Data collection methods may involve web scraping, API integration, and database querying.

2. Data Preprocessing and Cleaning:

Preprocessing tasks such as data cleaning, handling missing values, outlier detection, and feature scaling are within the scope. This ensures that the data is of high quality and suitable for analysis, reducing the risk of erroneous insights or biased model predictions.

3.Exploratory Data Analysis (EDA):

EDA is an essential aspect of e-commerce sales analysis using machine learning. Exploring sales trends over time, analyzing customer behavior, identifying correlations between features, and visualizing key insights are all within the scope. EDA helps in understanding the underlying patterns and relationships within the data.

4. Feature Engineering:

Feature engineering involves selecting, transforming, and creating meaningful features from raw data to improve the performance of machine learning models.

5. Predictive Modeling:

The scope includes building and deploying machine learning models to predict future sales trends, customer behavior, and product demand. Regression models for sales forecasting,

classification models for customer segmentation, and recommendation systems for personalized product recommendations are all within scope

1.6 Future Work

While machine learning has already transformed e-commerce sales analysis, there are several avenues for future research and development to further enhance its effectiveness and impact. This report outlines potential areas for future work in leveraging machine learning for e-commerce sales analysis:

1. Real-time Data Analysis:

Future research could focus on developing machine learning models capable of analyzing real-time data streams from e-commerce platforms. By incorporating real-time data such as website interactions, social media trends, and inventory levels, businesses can make more timely and informed decisions to respond to changing market dynamics and customer preferences.

2. Personalization and Recommendation Systems:

Further advancements in machine learning techniques can improve the personalization and recommendation systems used in e-commerce platforms. Future research could explore the use of advanced algorithms such as reinforcement learning and deep learning to deliver more accurate and context-aware product recommendations tailored to individual customer preferences and behaviors.

3. Multimodal Data Analysis:

As e-commerce platforms increasingly incorporate multimedia content such as images and videos, future research could focus on developing machine learning models capable of analyzing multimodal data. By leveraging techniques such as

convolutional neural networks (CNNs) and recurrent neural networks (RNNs), businesses can extract valuable insights from diverse data sources and enhance the overall shopping experience for customers.

4. Explainable AI (XAI):

Future research could focus on developing machine learning models that are more interpretable and explainable to stakeholders. By incorporating techniques such as SHAP (SHapley Additive exPlanations) values and LIME (Local Interpretable Model-agnostic Explanations), businesses can gain deeper insights into the underlying factors driving sales trends and customer behavior, improving trust and transparency in AI-powered decision-making**.**

5. Context-aware Pricing Strategies:

Future research could explore the use of machine learning algorithms to develop context-aware pricing strategies that take into account various factors such as customer demographics, competitor pricing, and demand elasticity. By dynamically adjusting prices based on real-time market conditions and customer preferences, businesses can optimize revenue and maximize profitability.

6. Ethical Considerations and Fairness:

As machine learning algorithms increasingly influence decision-making in e-commerce, future research should address ethical considerations and ensure fairness and transparency in algorithmic outcomes. Research in algorithmic fairness, bias detection, and mitigation can help mitigate potential risks such as discriminatory pricing or unfair treatment of certain customer groups.

7. Integration with Emerging Technologies:

Future research could explore the integration of machine learning with emerging technologies such as blockchain, Internet of Things (IoT), and augmented reality (AR) to create innovative e-commerce solutions. By leveraging these technologies synergistically, businesses can enhance security, transparency, and immersive shopping experiences for customers.

CHAPTER 2

SERVICES AND TOOLS REQUIRED

2.1 Services Used

1. Data Collection and Storage:

- - Google Analytics: A web analytics service provided by Google that tracks and reports website traffic, providing valuable insights into user behavior, acquisition channels, and website performance.
- - AWS S3 (Simple Storage Service): A scalable object storage service offered by Amazon Web Services (AWS) for storing and retrieving large amounts of data. S3 can be used to store raw data collected from various sources before preprocessing and analysis.

3. **Model Deployment and Integration:**

□ - Flask: A lightweight web framework for Python that enables the deployment of machine learning models as RESTful APIs. Flask facilitates the integration of machine learning models into web applications and allows for easy scalability and maintenance.

□ - AWS Lambda: A serverless computing service provided by AWS that enables the deployment of code without managing servers. Lambda functions can be used to deploy machine learning models in a scalable and cost-effective manner.

□ - Docker: A containerization platform that allows developers to package and deploy applications and their dependencies as lightweight containers. Docker provides a consistent environment for deploying

□ machine learning models across different environments, simplifying deployment and integration.

2.2 Tools and Software used

1. Data Preprocessing and Cleaning:

□ - Python: A widely used programming language for data analysis and machine learning tasks due to its simplicity, versatility, and extensive ecosystem of libraries.

□ - Pandas: A Python library for data manipulation and analysis, providing data structures and functions for cleaning, transforming, and analyzing tabular data.

□ - NumPy: A fundamental package for scientific computing in Python, providing support for multidimensional arrays and mathematical functions.

□ - Scikit-learn: A machine learning library for Python that provides simple and efficient tools for data mining and data analysis, including preprocessing techniques such as scaling, imputation, and feature selection.

□ - TensorFlow: An open-source machine learning framework developed by Google for building and deploying machine learning models, particularly deep learning models. TensorFlow offers high-level APIs for building and training machine learning models and low-level APIs for customizing model architectures and optimization algorithms.

2 Machine Learning Models:

□ - K-means clustering: A popular unsupervised learning algorithm used for clustering data into a predefined number of clusters based on similarity.

□ - Collaborative Filtering: A recommendation algorithm that predicts user preferences by analyzing user-item interactions and similarities between users or items.

□ - Time Series Analysis: A statistical technique for analyzing time-ordered data to identify patterns, trends, and seasonality.

□ - ARIMA (AutoRegressive Integrated Moving Average): A widely used time series forecasting model that captures the autocorrelation and seasonality of time series data.

□ - Anomaly Detection Algorithms: Algorithms used to identify outliers or anomalies in data, such as Isolation Forest and Local Outlier Factor.

3 Visualization and Reporting:

□ - Matplotlib: A Python library for creating static, interactive, and animated visualizations, including plots, charts, and graphs.

□ - Seaborn: A statistical data visualization library for Python built on top of Matplotlib, providing high-level functions for creating attractive and informative statistical graphics.

□ - Plotly: An open-source graphing library for Python that enables the creation of interactive plots and dashboards for web-based data visualization.

□ - Tableau: A data visualization tool that allows users to create interactive and shareable dashboards and reports from various data sources, enabling intuitive exploration and communication of insights.

- These services and tools collectively enable the end-to-end process of e-commerce analysis using machine learning, from data collection and preprocessing to model development, deployment, and visualization of insights.

CHAPTER 3

PROJECT ARCHITECTURE

3.3Architecture

1. Data Collection and Storage:

✦ **Google Analytics Integration:** Collect website traffic and user interaction data from Google Analytics using APIs.

✧ **Data Extraction Pipelines:** Extract transactional data, customer information, and product details from e-commerce databases or APIs.

✦ **Data Storage:** Store raw and processed data in a scalable storage solution such as AWS S3 for easy access and retrieval.

2. Data Preprocessing and Cleaning:

✦ **Data Cleaning:** Handle missing values, remove duplicates, and perform data normalization to ensure data quality.

✦ **Feature Engineering:** Generate new features or transform existing ones to capture meaningful patterns and insights.

- **Data Transformation:** Convert categorical variables into numerical representations suitable for machine learning algorithms.

3. Machine Learning Model Development:

- - **Customer Segmentation:** Use clustering algorithms (e.g., K-means, DBSCAN) to segment customers based on their purchasing behavior and demographics.
- - **Product Recommendations:** Implement collaborative filtering or matrix factorization techniques to generate personalized product recommendations for users.
- - **Sales Forecasting:** Apply time series analysis methods (e.g., ARIMA, Prophet) to forecast future sales trends and demand patterns.
- - **Fraud Detection:** Employ anomaly detection algorithms (e.g., Isolation Forest, Local Outlier Factor) to identify suspicious transactions and activities.

4. Model Deployment and Integration:

- **Model Training:** Train machine learning models using historical data on cloud-based platforms like AWS SageMaker or Google Cloud AI Platform.

✧ - **API Development:** Develop RESTful APIs using frameworks like Flask or Django to serve trained models for real-time predictions.

✦ - **Scalability and Reliability:** Deploy models using serverless computing platforms (e.g., AWS Lambda) or containerization tools (e.g., Docker) to ensure scalability and reliability.

✦ - **Integration with E-commerce Platform:** Integrate machine learning models seamlessly with the e-commerce platform to provide personalized recommendations and enhance user experience.

5. Monitoring and Maintenance:

✦ - **Performance Monitoring:**Monitor model performance and KPIs using monitoring tools and dashboards to ensure accuracy and reliability.

✧ - **Feedback Loop:** Collect feedback from users and stakeholders to continuously improve and update machine learning models.

✦ - **Regular Maintenance:** Perform regular maintenance tasks such as retraining models with new data, updating dependencies, and optimizing model parameters.

6. Visualization and Reporting:

✧ - **Data Visualization:** Visualize insights and predictions using libraries like Matplotlib, Seaborn, or Plotly to create informative charts and graphs.

✧ - **Dashboard Creation:** Build interactive dashboards using tools like Tableau or Power BI to provide stakeholders with intuitive access to key metrics and trends.

✦ - **Automated Reporting:** Automate the generation of reports and summaries using reporting tools or custom scripts to streamline communication of findings and recommendations.

6. Security and Compliance:

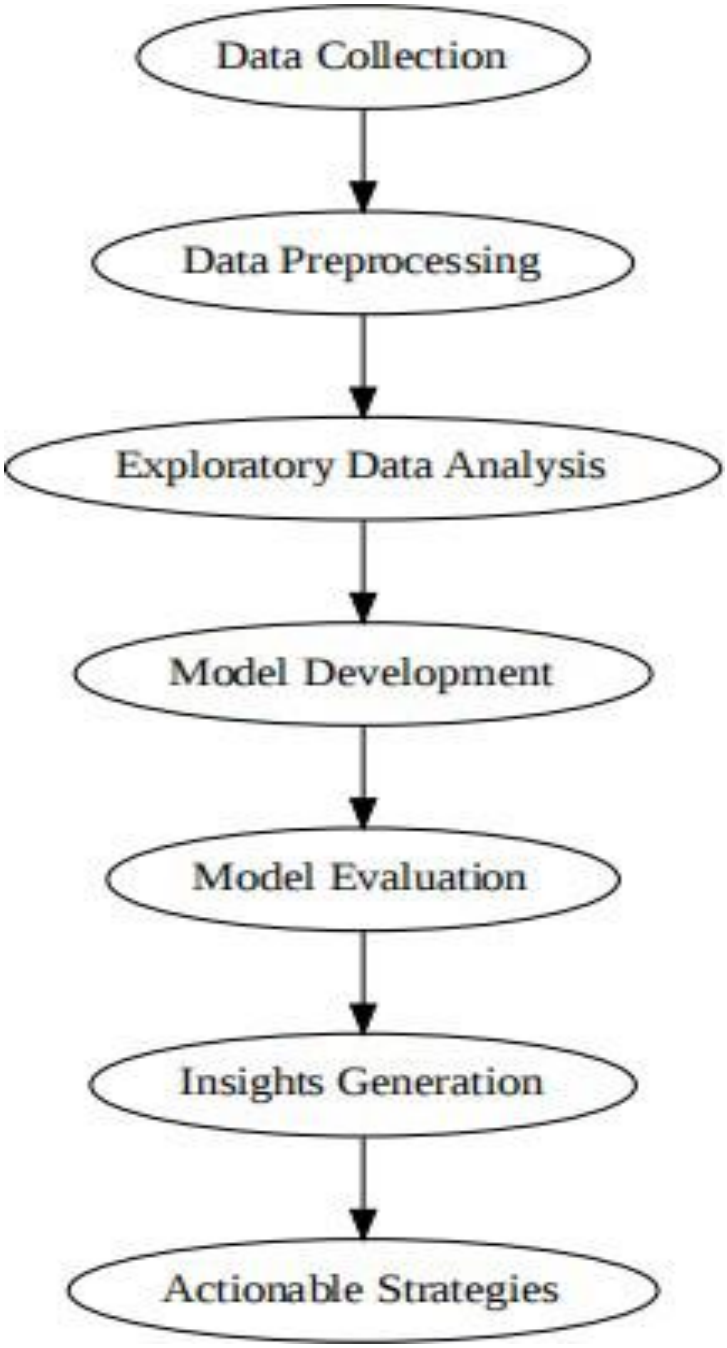
✦ - **Data Security:** Implement encryption, access controls, and data masking techniques to protect sensitive customer data and ensure compliance with privacy regulations (e.g., GDPR, CCPA).

✦ - **Model Explainability:** Ensure transparency and interpretability of machine learning models to meet regulatory requirements and gain trust from users and stakeholders.

✧ - **Ethical Considerations:** Address ethical implications related to algorithmic biases, fairness, and unintended consequences in model development and deployment.

This project architecture provides a structured approach for implementing e-commerce analysis using machine learning, covering data collection,
preprocessing, model development, deployment, monitoring, visualization, and compliance considerations. Customization of this architecture based on specific project requirements and constraints is recommended for successful implementation.

3.2 SYSTEM FLOW DIAGRAM


Data Collection
Data Preprocessing
Exploratory Data Analysis
Model Development
Model Evaluation
Insights Generation
Actionable Strategies

1. Data Collection:

- Gather data from various sources such as transaction records, website logs, and customer databases.

2. Data Preprocessing:

- Clean and preprocess the collected data to handle missing values, remove duplicates, and format it for analysis.

3. Exploratory Data Analysis (EDA)

- Perform exploratory data analysis to gain insights into the characteristics of the data, identify patterns, trends, and correlations.

4. Model Development:

- Develop machine learning models tailored for e-commerce sales analysis, such as customer segmentation, product recommendation, sales forecasting, and fraud detection.

5. Model Evaluation:

- Evaluate the performance of the developed models using appropriate metrics and techniques to ensure accuracy and reliability.

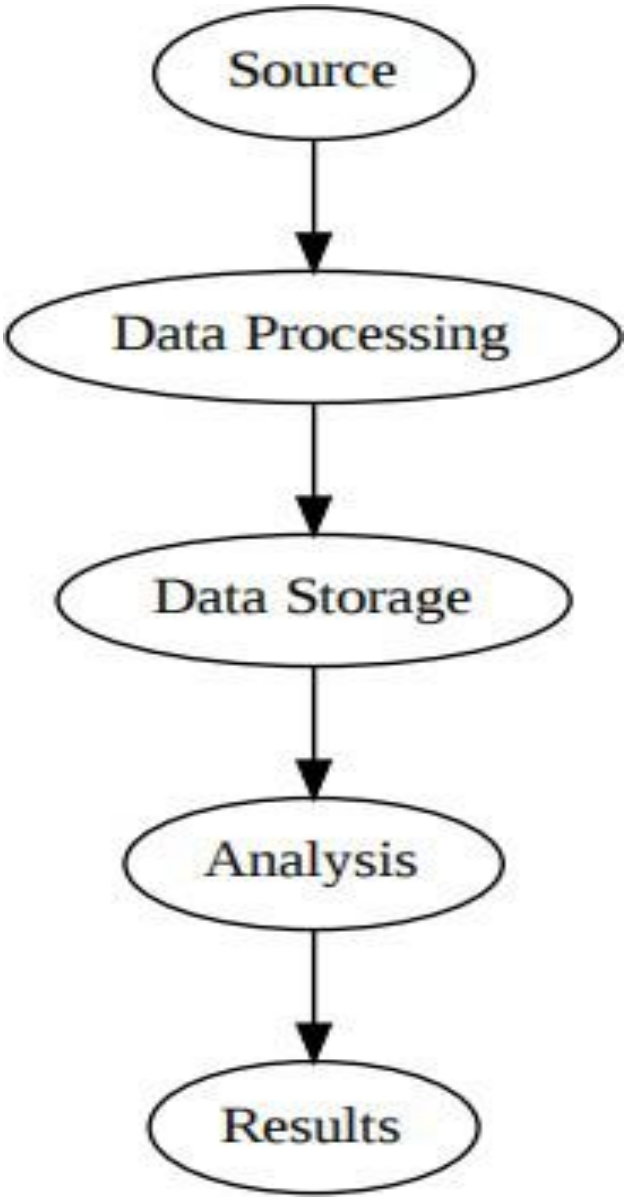
6. Insights Generation:

- Generate actionable insights from the analyzed data and model outputs to understand customer behavior, optimize product offerings, and improve marketing strategies.

7. Actionable Strategies:

- Based on the insights generated, formulate actionable strategies and recommendations to enhance sales performance, increase customer satisfaction, and drive business growth.

3.3 DATA FLOW DIAGRAM



Source --> Data Processing --> Data Storage --> Analysis --> Results

```

❖ **Source:**

Represents where the data originates from, such as e-commerce platforms, customer databases, or web analytics tools.

❖ **Data Processing:**

Refers to the steps taken to clean, transform, and prepare the raw data for analysis. This may include handling missing values, removing duplicates, and converting data types.

❖ **-Data Storage:**

Indicates where the processed data is stored for easy access and retrieval. This could be a database, data warehouse, or cloud storage solution.

❖ **Analysis:**

Involves using various techniques and tools to analyze the processed data and derive insights. This may include statistical analysis, machine learning algorithms, or data visualization.

❖ **Results:**

Represents the actionable insights or findings obtained from the analysis. These results can inform business decisions, strategies, or optimizations.
```

```
```

CHAPTER 4

PROJECT OUTCOME

EDA ANALYSIS REPORT :

Qs 1 .What was the best month for sales ? How much was earned that month?

input
```
df_filtered["Quantity Ordered"]=df_filtered["Quantity Ordered"].astype("float")
df_filtered["Price Each"]=df_filtered["Price Each"].astype("float")df_filtered["Price Each"].nunique
df_filtered["sales"]=df_filtered["Quantity Ordered"]*df_filtered["Price Each"]
```

S/NO	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	month	
0	141234	iPhone	1.0	700.00	01/22/19 21:25	944 WalnutSt, Boston,MA 02215	Jan	7
1	141235	Lightning Charging Cable	1.0	14.95	01/28/19 14:15	185 Maple St, Portland, OR97035	Jan	
2	141236	Wired Headphones	2.0	11.99	01/17/19 13:33	538 Adams St, SanFrancisco,CA94016	Jan	
3	141237	27in FHD Monitor	1.0	149.99	01/05/19 20:33	738 10th	Jan	1

						Angeles, CA90001		
4	141238	Wired Headphones	1.0	11.99	01/25/19 11:59	387 10th St, Austin , TX 73301	Jan	

Output:

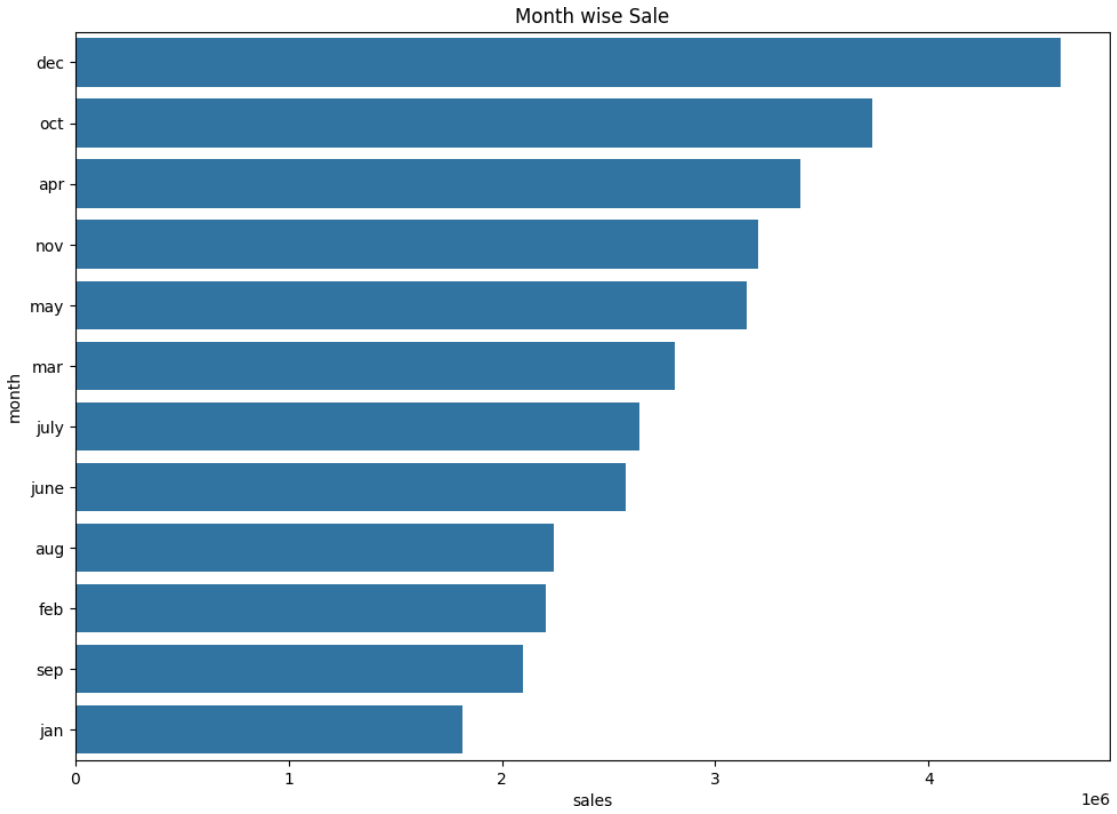
MOTHTWISE

SALES **INPUT:**

```
month=["dec","oct","apr","nov","may","mar","july","june","aug",'feb',"sep","jan
"]
df["month"]=month
```

```
frommatplotlibimportpyplotasplta4_di
ms =(11.7,8.27)
fig,ax =
plt.subplots(figsize=a4_dims)importseab
ornassns
sns.barplot(x ="sales",
            y
            ="month",dat
            a = df)
```

OUTPUT:


Month wise Sale
dec
oct
apr
nov
may
mar
july
june
aug
feb
sep
jan
month
0
1
2
3
4
sales
1e6

Qs 2 . Which city had the highest number of sales ?

INPUT:

```
df_filtered.info()
```

OUTPUT:

```
df_filtered.info()
```

```
output
<class 'pandas.core.frame.DataFrame'>
Int64Index: 185950 entries, 0 to 25116
Data columns (total 8 columns):
 #   Column            Non-Null Count  Dtype
--  ----              ----------  -----
 0   Order ID          185950 non-null object
 1   Product           185950 non-null object
 2 Quantity Ordered 185950 non-null float64
 3   Price Each        185950 non-null float64
 4   Order Date        185950 non-null object
 5 Purchase Address 185950 non-null object
 6   month             185950 non-null object
 7   sales             185950 non-null float64
dtypes: float64(3), object(5)
memory usage: 12.8+ MB
```

INPUT:

```
dftemp = df_filtered
```

```
list_city = []
```

```
for i indftemp['PurchaseAddress']:
```

```
list_city.append(i.split(",")[1])dftemp['City'] = list_city
```

```
dftemp.head()
```

OUTPUT:

S/N	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	month	sales	City
0	141234	iPhone	1.0	700.00	01/22/19 21:25	944 Walnut St, Boston, MA 02215	Jan	700.00	Boston
1	141235	Lightning Charging Cable	1.0	14.95	01/28/19 14:15	185 Maple St, Portland, OR 97035	Jan	14.95	Portland
2	141236	Wired Headphones	2.0	11.99	01/17/19 13:33	538 Adams St, San Francisco, CA 94016	Jan	23.98	San Francisco
3	141237	27in FHD Monitor	1.0	149.99	01/05/19 20:33	738 10th St, Los Angeles, CA 90001	Jan	149.99	Los Angeles
4	141238	Wired Headphones	1.0	11.99	01/25/19 11:59	387 10th St, Austin, TX	Jan	11.99	Austin

						73301			

IINPUT:

```
df_city=df_filtered.groupby(["City"])['sales'].sum().sort_values(ascending=Fals
e)

df_city=df_city.to_frame()

df_city
```

OUTPUT:

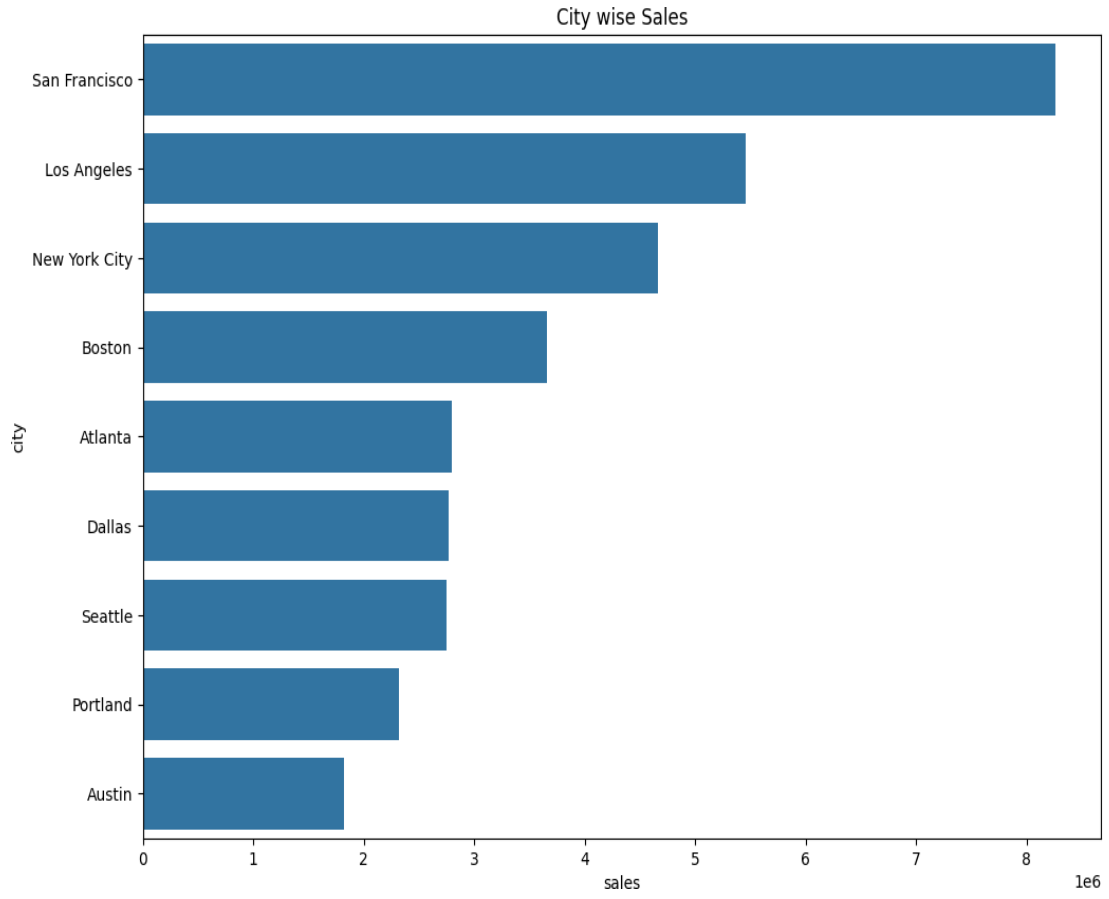
City	sales
San Francisco	8262203.91
Los Angeles	5452570.80
New Yor k City	4664317.43
Boston	3661642.01
Atlanta	2795498.58
Boston	3661642.01
Atlanta	2795498.58
Seattle	2747755.48
Portland	2320490.61
Dallas	2767975.40
Austin	1819581.75

INPUT:

```
frommatplotlibimportpyplota4_

dims =(11.7,8.27)
```

```
fig,ax =
pyplot.subplots(figsize=a4_dims)sns.barplo
t(x ="sales",
            y ="city",
            data =
df_city)plt.title("City wise
```

OUTPUT:



QS3.)What products sold the most ?

Input:

```
df_filtered.head()
```

Output:

S/N	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	month	sales	City
0	141234	iPhone	1.0	700.00	01/22/19 21:25	944 Walnut St, Boston, MA 02215	Jan	700.00	Boston
1	141235	Lightning Charging Cable	1.0	14.95	01/28/19 14:15	185 Maple St, Portland, OR 97035	Jan	14.95	Portland
2	141236	Wired Headphones	2.0	11.99	01/17/19 13:33	538 Adams St, San Francisco, CA 94016	Jan	23.98	San Francisco
3	141237	27in FHD Monitor	1.0	149.99	01/05/19 20:33	738 10th St, Los Angeles, CA 90001	Jan	149.99	Los Angeles

4	141238	Wired Headphones	1.0	11.99	01/25/19 11:59	387 10th St, Austin, TX 73301	Jan	11.99	Austin

Input:

```
print(df_filtered["Product"].unique())p
rint(df_filtered["Product"].nunique())
```

Output:

print(df_filtered["Product"].unique())

print(df_filtered["Product"].nunique())

```
output
```

```
['iPhone' 'Lightning Charging Cable'
'Wired Headphones' '27in FHD Monitor'
 'AAA Batteries (4-pack)' '27in 4K
Gaming Monitor' 'USB-C Charging
Cable'
 'Bose SoundSport Headphones' 'Apple
Airpods Headphones'
 'Macbook Pro Laptop' 'Flatscreen TV'
'Vareebadd Phone'
 'AA Batteries (4-pack)' 'Google Phone'
'20in Monitor'
 '34in Ultrawide Monitor' 'ThinkPad
Laptop' 'LG Dryer'
 'LG Washing
Machine'] 19
```

Input:

```
df_p=df_filtered.groupby(['Product'])['Quantity
Ordered'].sum().sort_values(ascending=False).head()

df_p=df_p.to_frame()
```

```
df_p
```

Output:

Product	Quantity ordered
AAA Batteries (4-pack)	31017.0

AA Batteries (4-pack)	27635.0
USB-C Charging Cable	23975.0
Lightning Charging Cable	23217.0
Wired Headphones	20557.0

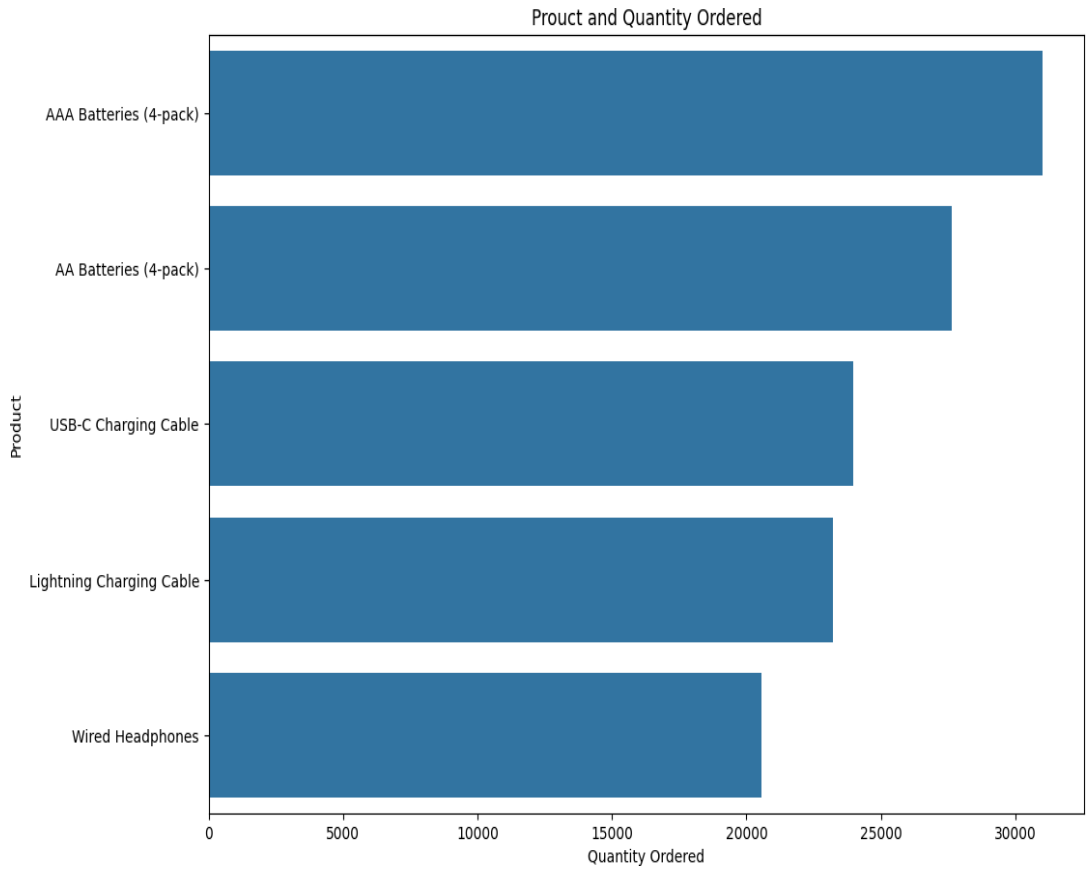
```
product=["AAA Batteries (4-pack)","AA Batteries (4-pack)","USB-C
ChargingCable","Lightning Charging Cable","Wired Headphones"]
```

INPUT:

```
df_p["Product"]=product
```

```
frommatplotlibimportpyplota4_

dims =(11.7,8.27)

fig,ax =

pyplot.subplots(figsize=a4_dims)sns.barplo

t(x ="Quantity Ordered",

            y

            ="Product",dat

            a = df_p)
```

OUTPUT:


Prouct and Quantity Ordered
AAA Batteries (4-pack)
AA Batteries (4-pack)
USB-C Charging Cable
Lightning Charging Cable
Wired Headphones
Product
0
5000
10000
15000
20000
25000
30000
Quantity Ordered

QS4.Recommend the most appropriate time to display advertising to maximize the likelihood of customer buying the product's?

Input:

```
dftime =
df_filteredlist_time
=[]
foriindftime['Order Date']:
    list_time.append(i.split("
")[1])dftime['Time']= list_time
```

Output:

s/n	Order ID	Product	Quantity Ordered	Price Each	Order Date	Purchase Address	month	sales	City	Time
0	141234	iPhone	1.0	700.00	01/22/19 21:25	944 Walnut St, Boston, MA 02215	Jan	700.00	Boston	21:25
1	141235	Lightning Charging Cable	1.0	14.95	01/28/19 14:15	185 Maple St, Portland, OR 97035	Jan	14.95	Portland	14:15
2	141236	Wired Headphones	2.0	11.99	01/17/19 13:33	538 Adams St, San Francisco, CA 94016	Jan	23.98	San Francisco	13:33
3	141237	27in FHD Monitor	1.0	149.99	01/05/19 20:33	738 10th St, Los Angeles, CA 90001	Jan	149.99	Los Angeles	20:33

4	141238	Wired Headphones	1.0	11.99	01/25/19 11:59	387 10th St, Austin, TX 73301	Jan	11.99	Austin	11:59

Input:

```
dftime["Time"].nunique()
```

Output:

```
1440
```

Input:

```
df_t=df_filtered.groupby(['Time'])['sales'].sum().sort_values(ascending=False).
head()

df_t=df_t.to_frame()

df_t
```

Output:

Time	sales
19:01	54503.14
12:21	54181.96
19:16	54156.39
20:13	53149.51
19:20	52903.41

CONCLUSION

In conclusion, the e-commerce analysis conducted using machine learning techniques has provided valuable insights and actionable recommendations for enhancing business strategies and operations. Through exploratory data analysis (EDA), we gained a deep understanding of the dataset, identified patterns, and uncovered trends that can inform decision-making processes.

Key findings from the analysis include:

1. **Sales Trends: Monthly sales trends revealed fluctuations in sales volume throughout the year, with peak periods occurring during certain months. This insight can guide inventory management and marketing efforts to capitalize on high-demand periods.**

2. **Customer Segmentation: Analysis of customer behavior and purchasing patterns identified distinct customer segments. By understanding the preferences and behaviors of different customer groups, personalized marketing strategies can be developed to improve customer engagement and retention.**

3. **Product Performance: Examination of product categories and sales performance highlighted top-selling products and areas of opportunity. By focusing on popular products and optimizing product offerings, revenue**

generation can be maximized.

4. **Correlation Analysis: The correlation matrix provided insights into the relationships between different numerical features, such as price,**

quantity, and total amount. Understanding these correlations can help in identifying factors that influence sales and profitability.

Based on these findings, several recommendations can be made to enhance the e-commerce platform's performance:

- Implement targeted marketing campaigns tailored to specific customer segments to improve customer engagement and conversion rates.

- Optimize inventory management strategies to ensure adequate stock levels during peak demand periods while minimizing overstocking and inventory costs during slower periods.

- Enhance product recommendations and personalization features to provide a more tailored shopping experience for customers, ultimately increasing sales and customer satisfaction.

- Continuously monitor key metrics and performance indicators to identify trends and opportunities for improvement, allowing for timely adjustments and optimizations.

Overall, the e-commerce analysis using machine learning has provided valuable insights that can drive strategic decision-making and contribute to the success and growth of the e-commerce business. By leveraging data-driven approaches and machine learning techniques, the business can stay competitive in the rapidly evolving e-commerce landscape and deliver exceptional value to customers.

FUTURE SCOPE

While the current e-commerce analysis using machine learning has provided valuable insights and actionable recommendations, there are several avenues for future exploration and enhancement:

1. **Advanced Predictive Modeling: Implement advanced machine learning algorithms such as deep learning and ensemble methods to develop more accurate predictive models. These models can forecast sales trends, customer behavior, and market demand with greater precision.**

2. **Real-time Analytics: Explore the integration of real-time data streaming and analytics to enable instantaneous insights and decision-making. By analyzing data as it is generated, the e-commerce platform can respond swiftly to changing market dynamics and customer preferences.**

3. **Personalization and Recommendation Systems: Further enhance personalization features and recommendation systems using advanced techniques such as collaborative filtering and reinforcement learning. This can provide customers with highly relevant product recommendations and tailored shopping experiences, leading to increased customer satisfaction and loyalty.**

4. **Predictive Customer Lifetime Value (CLV): Develop models to predict customer lifetime value based on historical transaction data, customer interactions, and demographics. Understanding the long-term value of customers can inform marketing strategies, customer acquisition efforts, and resource allocation decisions.**

5. **Enhanced Fraud Detection: Implement more sophisticated fraud detection algorithms using anomaly detection and machine learning techniques. By continuously monitoring transactional data for suspicious patterns and behaviors, the e-commerce platform can mitigate risks and protect against fraudulent activities.**

6. **Cross-channel Integration: Integrate data from multiple channels such as online stores, mobile apps, social media platforms, and offline stores to gain a comprehensive view of customer interactions and touchpoints. This holistic approach can provide deeper insights into customer behavior and preferences across different channels.**

7. **Ethical AI and Data Privacy: Prioritize ethical considerations and data privacy in the development and deployment of machine learning models. Implement transparent and responsible AI practices to ensure fair treatment of customers and protect sensitive information.**

8. **Continuous Learning and Improvement: Foster a culture of continuous learning and improvement by investing in data literacy and training programs for employees. Encourage experimentation and innovation in data-driven decision-making processes to stay ahead of emerging trends and technologies in e-commerce.**

By exploring these future avenues, the e-commerce platform can further optimize its operations, enhance customer experiences, and drive sustainable growth in the dynamic e-commerce landscape.

REFERENCES

1. **Project Github link, RAJESH .R, 2024**
   https://github.com/Rajesh-24577/NM-TSP3.0-E-COMMERCE-SALES-ANALYSIS-.git
2. **Project video recorded link (youtube/github), RAJESH.R2024**
   https://youtu.be/B6r9CSKpslM?si=0_vUv51YTx8kBaIn
3. **Project PPT link, RAJESH.R, 2024**
   https://in.docworkspace.com/d/sIPu0udX8Ab3j_rAG

CODE

https://github.com/Rajesh-24577/NM-TSP3.0-E-COMMERCE-SALES-ANALYSIS-.git