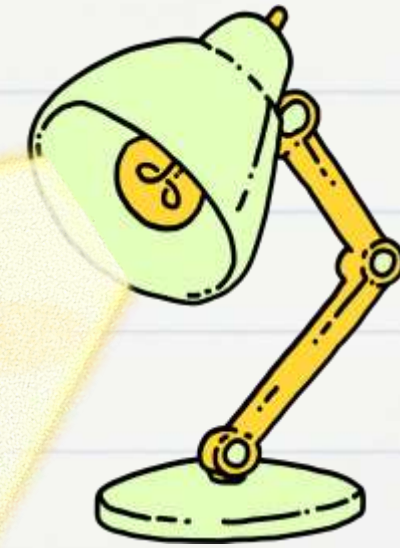




CSA1668-Data warehousing and data mining for data analysis

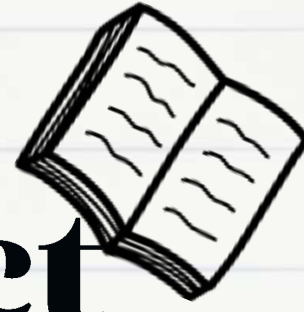


Customer Behavior Analysis and Prediction



Guided by:
Akshaya K
(Course Faculty)
Data warehouse and data mining
SSE,SIMATS

Project by:
M.Ram Jeyanth,S.Hariharan
(192211167,192211110)
Data warehouse and data mining
SSE,SIMATS



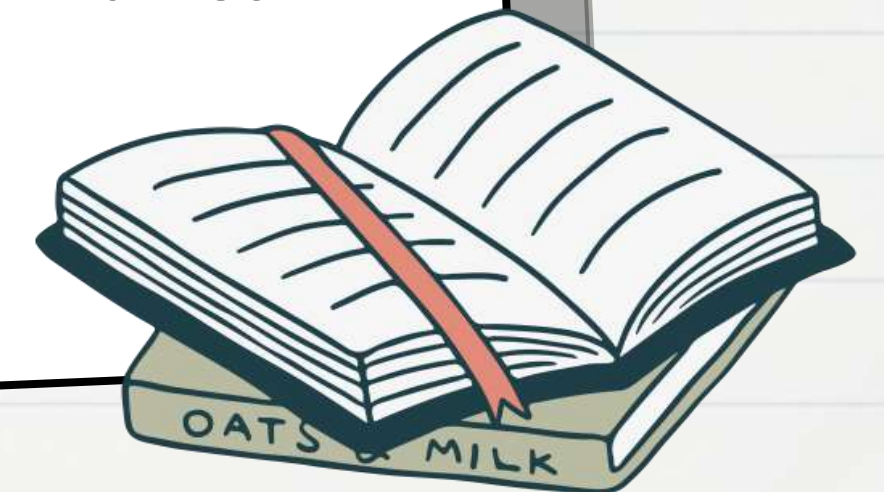
Abstract

Customer behavior analysis and prediction involve studying and interpreting customer data to understand purchasing patterns, preferences, and trends. This field utilizes various data sources, including transactional records, social media interactions, and demographic information, to build comprehensive customer profiles. Advanced analytical techniques such as machine learning, statistical analysis, and data mining are employed to predict future behaviors and trends. By identifying patterns and insights, businesses can tailor marketing strategies, enhance customer experiences, and improve overall operational efficiency. Ultimately, effective customer behavior analysis and prediction enable companies to make informed decisions, foster customer loyalty, and drive business growth.

Keywords: Data mining,trends,Customer data,Marketing strategies,Customer Experiences and Machine Learning.

Introduction

Customer behavior analysis and prediction is a crucial aspect of modern business strategy, aiming to decipher and anticipate the actions and preferences of consumers. By leveraging vast amounts of data from various sources such as purchase histories, social media activity, and demographic details, businesses can gain deep insights into customer needs and behaviors. Advanced methodologies including machine learning, predictive analytics, and data mining are employed to create accurate forecasts of future buying patterns and trends. This process not only aids in personalizing marketing efforts and enhancing customer experiences but also plays a significant role in optimizing inventory management, improving customer retention, and driving overall business growth. As a result, companies can make data-driven decisions that are critical for staying competitive in today's dynamic market landscape.





Hardware and software requirements

Hardware

- 1) Processor: 11th Gen Intel(R) Core(TM) i5-11400H @ 2.70GHz 2.69 GHz 64-bit operating system, x64-based processor
- 2) Storage: 512 GB
- 3) Memory: 8 GB
- 4) Graphic card: In-built Radeon Graphics

Software

- 1) OS: Windows 10 or 11
- 2) Processor: Intel core(i5)
- 3) Integrated development environment (IDE): Visual Studio

Packages

- 1) Import Pandas
- 2) `import plotly.express as px`
- 3) `import plotly.graph_objects as go`

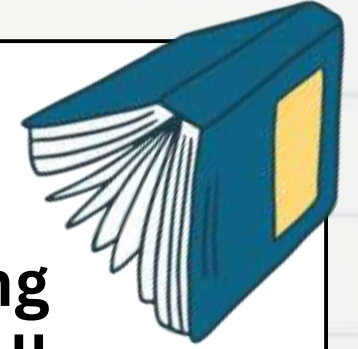




Literature survey

A literature survey on customer behavior analysis and prediction reveals a rich body of research spanning various methodologies and applications. Early studies primarily relied on statistical techniques to analyze purchasing patterns and customer segmentation. With the advent of big data and advanced machine learning algorithms, more recent research has focused on predictive analytics and real-time data processing. Techniques such as clustering, decision trees, and neural networks are extensively explored to enhance accuracy in predicting customer behaviors. Additionally, the integration of social media data and sentiment analysis has provided deeper insights into customer preferences and sentiments. Overall, the literature highlights the evolving nature of customer behavior analysis, emphasizing the increasing importance of personalized marketing and customer-centric business strategies.

Benefits



- **Improved Customer Experience:** Personalize interactions and offerings, leading to more satisfying and engaging customer experiences, which enhance overall satisfaction and loyalty.
- **Enhanced Marketing Strategies:** Identify target audiences more accurately and optimize marketing campaigns, creating more effective and relevant content, leading to higher conversion rates.
- **Increased Sales and Revenue:** Anticipate demand and manage inventory more efficiently, reducing stockouts and overstock situations, thus maximizing sales opportunities.
- **Customer Retention and Loyalty:** Identify at-risk customers and implement personalized retention strategies, such as offers and loyalty programs, to reduce churn rates.
- **Informed Product Development:** Guide product development by understanding customer needs and preferences, resulting in products and services that align with market demands.





Advantages & Disadvantages



Advantages

- **Enhanced Personalization:** Tailors products, services, and communications to individual preferences, leading to increased customer satisfaction and loyalty.
- **Proactive Decision-Making:** Allows businesses to anticipate customer behavior and market trends.
- **Competitive Advantage:** Provides insights that help businesses stay ahead of competitors by understanding customer choices.

Disadvantages

- **Privacy Concerns:** Raises issues about the extent of data collection and usage, potentially leading to trust issues and legal challenges if privacy regulations are not followed.
- **High Costs and Resources:** Requires significant investment in technology, data management, and skilled personnel, making it costly and resource-intensive.
- **Data Quality and Accuracy:** Relies on high-quality, accurate data; poor data can lead to incorrect insights and poor business decisions.



Existing System

1. Google Analytics:

Use Case: An e-commerce site uses Google Analytics to track user interactions, identify popular products, and optimize the user journey for higher conversion rates.

2. Salesforce Einstein Analytics

- **Use Case:** A retail company integrates Einstein Analytics with its CRM to forecast customer purchases, personalize marketing campaigns, and increase sales through targeted promotions.

3. IBM Watson Customer Experience Analytics:

- **Use Case:** A travel agency uses IBM Watson to analyze customer feedback, map out user journeys, and predict customer satisfaction levels to enhance the overall travel experience.

4. Adobe Analytics

- **Use Case:** A media company employs Adobe Analytics to understand audience engagement across multiple channels, optimize content strategies, and increase subscriber retention.





Proposed System



- Implement priority-based scheduling algorithms to ensure timely processing of multimedia data.
- Utilize efficient memory management techniques to handle large volumes of multimedia data.
- Incorporate real-time task synchronization mechanisms for seamless coordination among various system components.
- Provide robust error handling and recovery mechanisms to maintain system stability and reliability.
- Support hardware acceleration for multimedia processing tasks to enhance system performance and efficiency.



Program:

```
project.py - C:/Users/M.Ram.jeyanth/Desktop/DWDM/project.py (3.12.0)
File Edit Format Run Options Window Help

import pandas as pd
import plotly.express as px
import plotly.graph_objects as go

data = pd.read_csv("ecommerce_customer_data.csv")
print(data.head())
# Summary statistics for numeric columns
numeric_summary = data.describe()
print(numeric_summary)
# Summary for non-numeric columns
categorical_summary = data.describe(include='object')
print(categorical_summary)
# Histogram for 'Age'
fig = px.histogram(data, x='Age', title='Distribution of Age')
fig.show()
# Bar chart for 'Gender'
gender_counts = data['Gender'].value_counts().reset_index()
gender_counts.columns = ['Gender', 'Count']
fig = px.bar(gender_counts, x='Gender',
             y='Count',
             title='Gender Distribution')
fig.show()
# 'Product Browsing Time' vs 'Total Pages Viewed'
fig = px.scatter(data, x='Product_Browsing_Time', y='Total_Pages_Viewed',
                 title='Product Browsing Time vs. Total Pages Viewed',
                 trendline='ols')
fig.show()
# Grouped Analysis
gender_grouped = data.groupby('Gender')['Total_Pages_Viewed'].mean().reset_index()
gender_grouped.columns = ['Gender', 'Average_Total_Pages_Viewed']
fig = px.bar(gender_grouped, x='Gender', y='Average_Total_Pages_Viewed',
             title='Average Total Pages Viewed by Gender')
fig.show()
devices_grouped = data.groupby('Device_Type')['Total_Pages_Viewed'].mean().reset_index()
devices_grouped.columns = ['Device_Type', 'Average_Total_Pages_Viewed']
fig = px.bar(devices_grouped, x='Device_Type', y='Average_Total_Pages_Viewed',
             title='Average Total Pages Viewed by Devices')
fig.show()
data['CLV'] = (data['Total_Purchases'] * data['Total_Pages_Viewed']) / data['Age']

data['Segment'] = pd.cut(data['CLV'], bins=[1, 2.5, 5, float('inf')],
                        labels=['Low Value', 'Medium Value', 'High Value'])

segment_counts = data['Segment'].value_counts().reset_index()
segment_counts.columns = ['Segment', 'Count']
```

```
# Create a bar chart to visualize the customer segments
fig = px.bar(segment_counts, x='Segment', y='Count',
             title='Customer Segmentation by CLV')
fig.update_xaxes(title='Segment')
fig.update_yaxes(title='Number of Customers')
fig.show()
# Funnel analysis
funnel_data = data[['Product_Browsing_Time', 'Items_Added_to_Cart', 'Total_Purchases']]
funnel_data = funnel_data.groupby(['Product_Browsing_Time', 'Items_Added_to_Cart']).sum().reset_index()

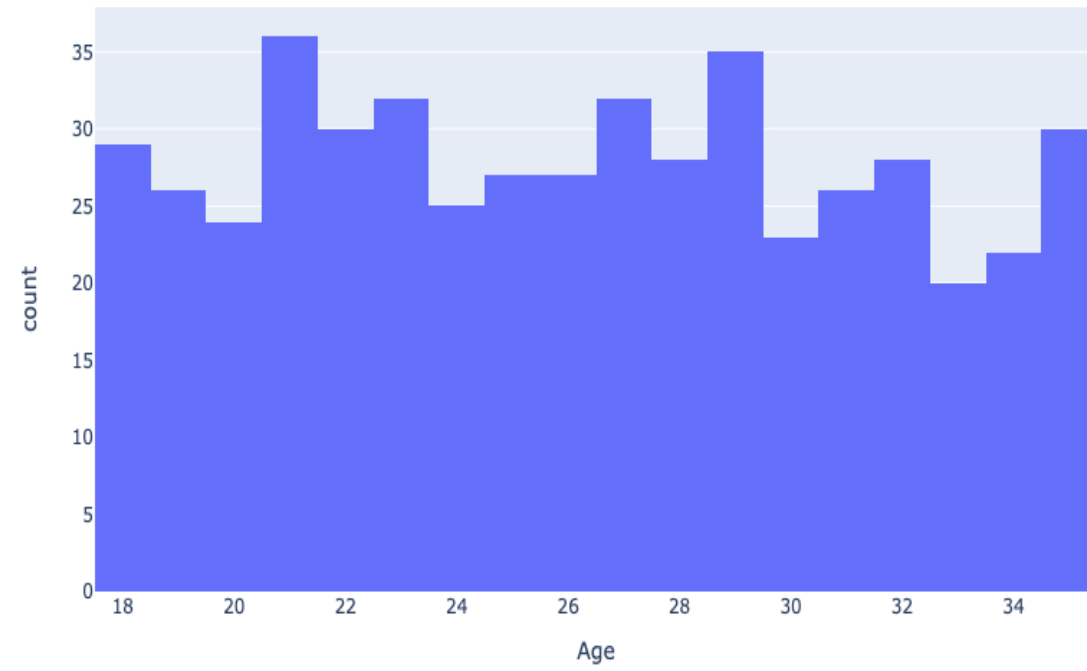
fig = px.funnel(funnel_data, x='Product_Browsing_Time', y='Items_Added_to_Cart', title='Conversion Funnel')
fig.show()
# Calculate churn rate
data['Churned'] = data['Total_Purchases'] == 0

churn_rate = data['Churned'].mean()
print(churn_rate)
```

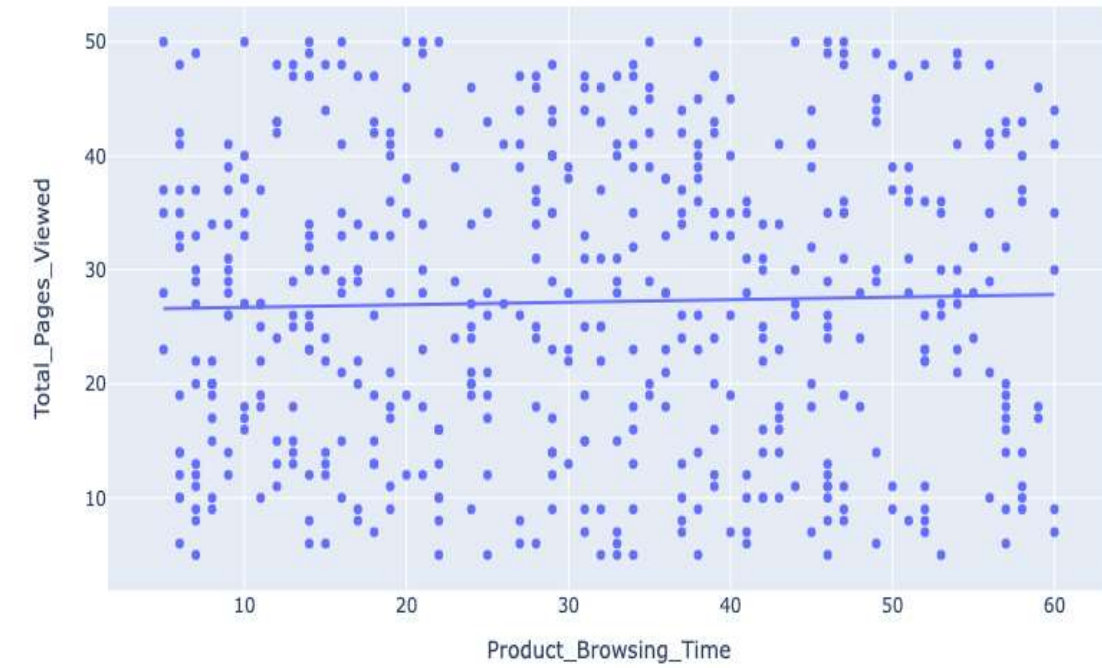
Customer Behavior Analysis is a valuable process that empowers businesses to make data-driven decisions, enhance customer experiences, and remain competitive in a dynamic market.

OUTPUT:

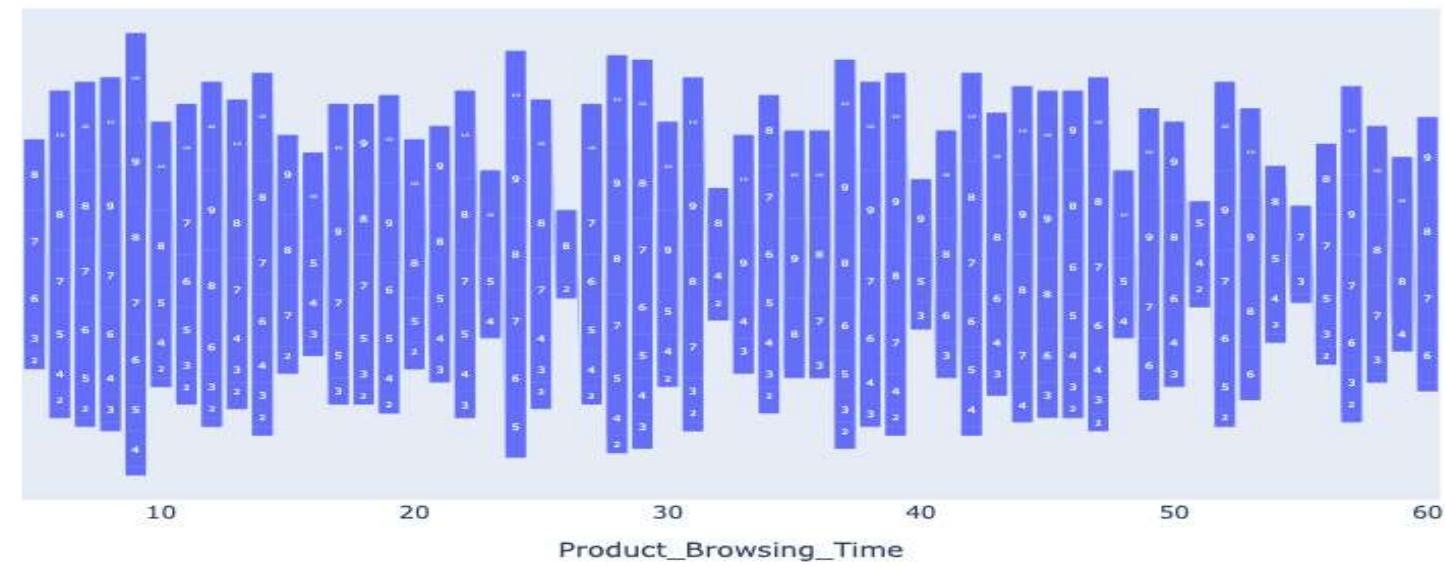
Distribution of Age



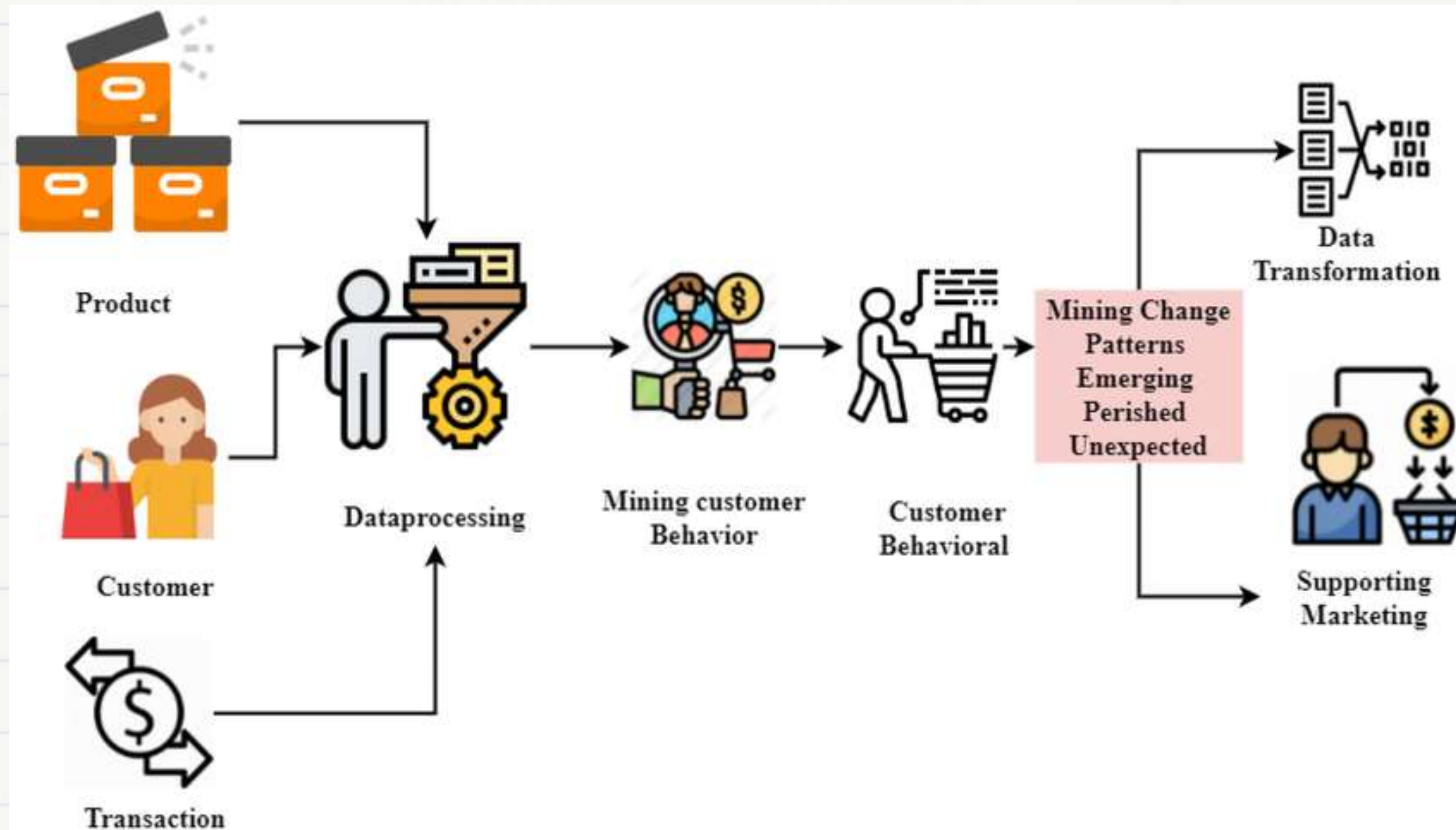
Product Browsing Time vs. Total Pages Viewed

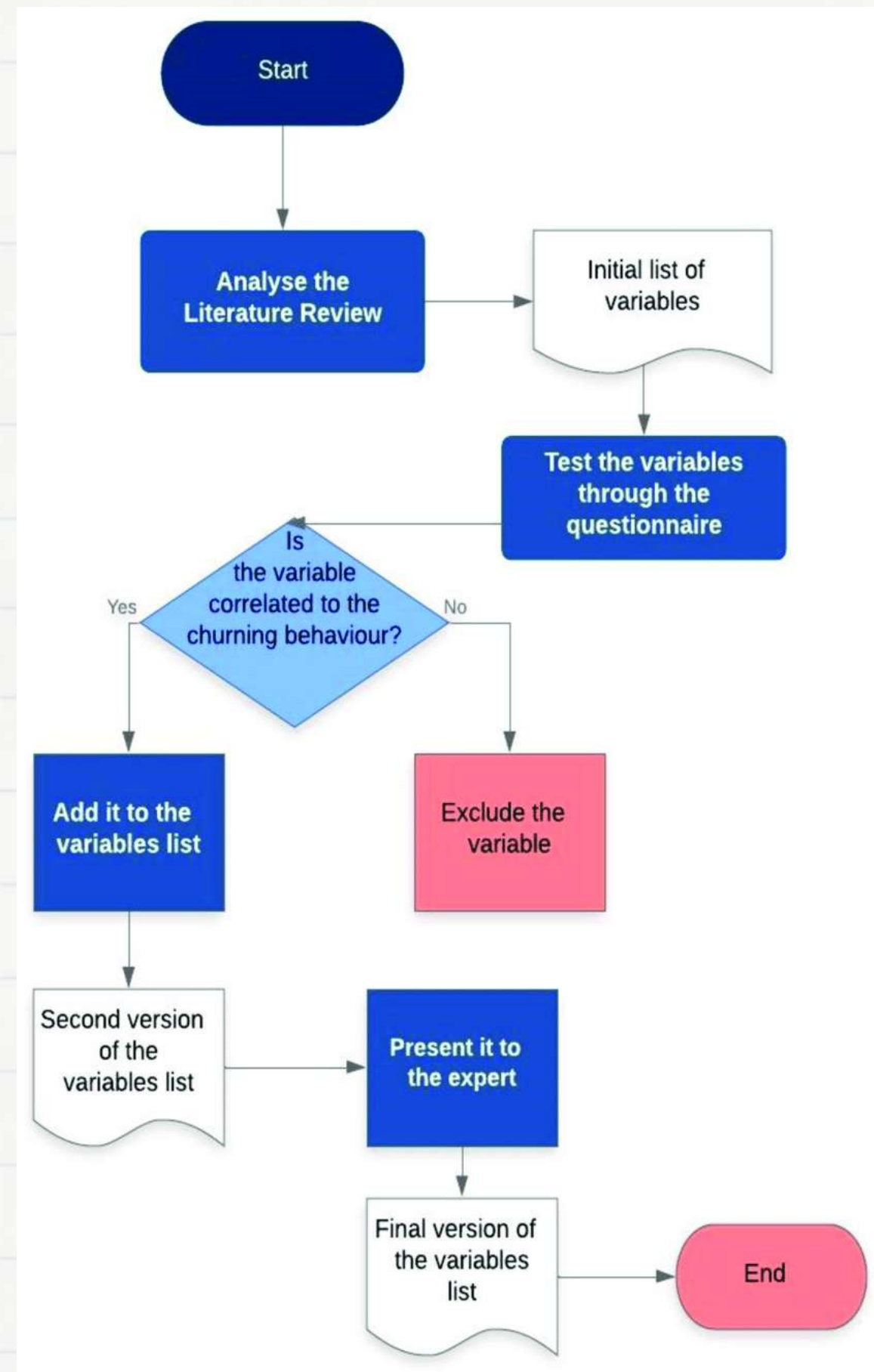


Conversion Funnel



Design







Implementation



Data Collection: Gather data from various sources such as transaction records and customer interactions.

Data Preprocessing: Clean and prepare the data to ensure accuracy and consistency.

Analytics Application: Use machine learning algorithms to identify patterns and predict future behaviors.

Real-Time Analytics: Implement real-time processing for up-to-date insights.

Visualization and Dashboards: Create visual tools to interpret results and support decision-making.

Strategy Enhancement: Leverage insights to improve customer engagement and drive more effective strategies.



Future Scope

- AI will refine personalization techniques, creating highly customized marketing strategies based on individual customer data and preferences.
- Enhanced analytics will enable businesses to monitor and respond to customer behavior instantly, improving engagement and decision-making.
- The expansion of Internet of Things devices will provide additional behavioral data, offering more detailed insights into customer habits and preferences.
- Advances in natural language processing will allow for a deeper understanding of customer emotions and feedback, leading to more effective communication and service.
- More sophisticated predictive models will enable businesses to better forecast customer lifetime value, optimizing acquisition and retention efforts.
- Increased focus on privacy and ethical data use will ensure that customer information is handled responsibly, building trust and complying with regulations.



Conclusion

In conclusion, analyzing and predicting customer behavior is crucial for businesses aiming to enhance their competitive edge. By leveraging data-driven insights, companies can tailor their strategies to meet evolving customer needs, improve customer satisfaction, and drive growth. Predictive models enable businesses to anticipate trends and make informed decisions, ultimately fostering stronger customer relationships and achieving long-term success. Embracing these analytical techniques ensures that organizations remain agile and responsive in a dynamic market landscape.





Reference

- Surendro, Kridanto. "Predictive analytics for predicting customer behavior." *2019 International Conference of Artificial Intelligence and Information Technology (ICAIIIT)*. IEEE, 2019.
- Kalaivani, D., and P. Sumathi. "Factor based prediction model for customer behavior analysis." *International Journal of System Assurance Engineering and Management* 10.4 (2019): 519-524.
- Khade, Anindita A. "Performing customer behavior analysis using big data analytics." *Procedia computer science* 79 (2016): 986-992.
- Li, Jing, Shuxiao Pan, and Lei Huang. "A machine learning based method for customer behavior prediction." *Tehnički vjesnik* 26.6 (2019): 1670-1676.
- Raju, Siva Subramanian, and Prabha Dhandayudam. "Prediction of customer behaviour analysis using classification algorithms." *AIP conference proceedings*. Vol. 1952. No. 1. AIP Publishing, 2018.





thank you
so much

Team members:

M.Ram Jeyanth(192211167)

S.Hariharan(192211110)

