

**BHARATI VIDYAPEETH (DEEMED TO
BE UNIVERSITY) COLLEGE OF
ENGINEERING, PUNE-43**



**Project Based Learning Report on
E-commerce Data Analytics Dashboard for
Big Data Analytics Course**

Submitted By:

Laxita Alizad-2214110125
Sneha Bansal- 2214110132
Amisha Kumari-2214110127
Sanjana Ghadge-2214110147
Sejal Dubal- 2214110305

Under the guidance of
Prof. Sheetal Patil

BHARATI VIDYAPEETH
(DEEMED TO BE UNIVERSITY)
COLLEGE OF ENGINEERING, PUNE – 43

DEPARTMENT OF COMPUTER ENGINEERING



CERTIFICATE

This is to certify that the Project Based Learning report title on E-commerce Data Analytics Dashboard, submitted by **Laxita Alizad (02), Sejal Dubal (47), Amisha Kumari (04), Sneha Bansal (07), Sanjana Ghadge (20)** to the Bharati Vidyapeeth (Deemed to be University), College of Engineering, Pune-43 for the award of the degree of BACHELOR OF TECHNOLOGY in Computer Engineering is a bonafide record of the PBL work done by them under my supervision.

Prof. Sheetal Patil
(Course Coordinator's signature)

Laxita Alizad-2214110125
Sneha Bansal-2214110132
Sejal Dubal-2214110305
Amisha Kumari-2214110127
Sanjana Ghadge-2214110147

ACKNOWLEDGEMENT

We would like to express our heartfelt gratitude to everyone who supported us throughout the journey of our project “E-commerce Data Analytics Dashboard.”

First and foremost, we are deeply thankful to our course coordinator, Prof. Sheetal Patil ma'am, for her valuable guidance, continuous support, and encouragement. Her insights and mentorship played a pivotal role in shaping our understanding and execution of the project.

We also extend our sincere thanks to our Head of Department, Dr. Vanjale Sir, for providing us with this opportunity and for his constant motivation and support.

Our heartfelt appreciation goes to our entire project team — Laxita Alizad, Sejal Dubal, Amisha Kumari, Sneha Bansal, and Sanjana Ghadge. The project was a collective effort, and it was the spirit of collaboration, dedication, and teamwork that made this experience enriching and successful.

We are equally grateful to all the faculty members and individuals who directly or indirectly contributed to the successful completion of our project.

Last but not least, we thank our parents and families for their unwavering support, encouragement, and for being our constant source of strength throughout this academic journey.

Thank you once again to everyone who contributed to making this project a memorable and valuable learning experience for all of us.

ABSTRACT

The E-commerce Data Analytics Dashboard project aims to explore consumer shopping behavior and highlight key trends using data analytics and interactive visualization. With the rapid expansion of the e-commerce sector, businesses generate large volumes of data, making it essential to derive meaningful insights for strategic decision-making. This project leverages the R programming language and the Shiny framework to build a dynamic, user-friendly dashboard that allows intuitive exploration of customer behavior patterns.

The dataset was cleaned and preprocessed using dplyr and readr, while visualizations were developed using Plotly, Leaflet, and DT. The dashboard presents insights such as popular product categories, preferred payment methods, and regional purchase distributions. Designed with simplicity and accessibility in mind, the interface enables non-technical users to interactively filter and interpret data.

This project demonstrates the effectiveness of combining data analytics with interactive tools to support data-driven strategies and enhance business understanding in the e-commerce domain.

INDEX

SR NO.	TITLE	PG.NO
1	Introduction	1
2	Objective	2
3	Methodology	3
4	System Design	4
5	Requirement and Specification	6
6	Key Functionalities of the Application	7
7	Screenshots and Functional Demonstration	8
8	Result Analysis	11
11	Future scope	14
12	Conclusion & Code link	15
13	References	16

LIST OF FIGURES

Figure No.	Figure caption	Page no.
1	Control flow	<u>4</u>
<u>2</u>	Age distribution	<u>8</u>
<u>3</u>	Age & product categories	<u>8</u>
<u>4</u>	Promo code spending	<u>9</u>
<u>5</u>	Popular payment methods	<u>10</u>
<u>6</u>	Shipping type by category	<u>10</u>

INTRODUCTION

In today's data-driven world, businesses rely heavily on analytics to gain actionable insights and make informed decisions. The Shopshere Data Analytics Tool is an interactive dashboard developed using **R Shiny** to analyze and visualize shopping-related data. This project aims to help retailers, analysts, and business stakeholders understand trends related to **sales performance, customer behavior, and inventory management**.

The dashboard is designed with user experience in mind, featuring a clean layout, multiple tabs, and dynamic visualizations. It utilizes powerful R packages such as shiny for web-based interactivity, plotly for engaging graphs, leaflet for mapping, and DT for responsive data tables. The application allows users to explore various metrics through filters and drill-downs, making it easier to identify patterns, peak sales periods, customer preferences, and potential inventory issues.

By integrating real-time interaction and multiple visualization formats, the tool simplifies complex datasets into intuitive insights. Whether used by business owners or data analysts, the tool supports strategic planning and operational improvements. Overall, the Shopshere Data Analytics Tool showcases the potential of combining data science and interactive dashboards to make retail data more accessible, insightful, and useful for decision-making in a competitive marketplace.

This report aims to identify key shopping trends through an in-depth analysis of a dataset containing **3,900 customer entries** across **17 attributes**. These attributes encompass a wide range of information, including **demographic details** (such as age, gender, and location), **purchase data** (like product categories, quantities, and prices), and **payment methods**. The primary objective of this study is to uncover meaningful patterns and relationships within the data that can help explain consumer behavior.

To achieve this, the analysis incorporates a combination of **statistical techniques, interactive data visualizations, and predictive modeling** using R programming. By leveraging tools like **plotly, leaflet, and DT** in the R Shiny framework, the project delivers an intuitive and dynamic dashboard that provides stakeholders with real-time insights.

The findings from this analysis offer actionable intelligence that can support better business decisions, such as optimizing inventory, tailoring marketing strategies, and improving customer engagement. By transforming raw data into comprehensible visual stories, the Shopshere Data Analytics Tool demonstrates the practical value of data science in the retail sector.

OBJECTIVE

➤ **Analyze Consumer Shopping Data**

Extract and interpret key trends from shopping datasets to understand consumer preferences and behavior. By performing data wrangling, summarization, and visualization, we identify trends such as high-performing product categories, preferred payment options, and peak shopping times

➤ **Develop an Interactive Web Application**

Create a user-friendly interface using **Shiny** for dynamic visualization of shopping patterns, enabling stakeholders to explore data insights effectively. This enables stakeholders—including business owners, analysts, and marketers—to gain a clear overview of consumer behavior without requiring deep technical expertise.

➤ **Enhance Decision-Making**

Provide actionable insights that businesses can leverage to optimize marketing strategies, inventory management, and customer engagement. The insights gained through the app can support data-driven decision-making in business contexts.

➤ **Bridge the Gap Between Raw Data and Usability**

Simplify complex data and make it accessible to non-technical users through clear visualizations and summaries. Through clear graphs and tables, users who are not familiar with data science or statistics can still interpret the findings and make informed decisions.

➤ **Lay the Foundation for Future Predictive Analytics**

Establish a base for integrating advanced machine learning techniques for trend forecasting and customer behavior prediction. While the current version of the project focuses on descriptive analysis, the clean and well-structured approach lays the groundwork for future development. Predictive analytics models can be added later to forecast future sales, customer churn, or seasonal trends using machine learning techniques like regression, clustering, or classification.

➤ **Build a Scalable and Extensible Framework**

Design the project in a way that allows future integration of additional datasets, advanced analytics techniques, or deployment to larger-scale platforms. This ensures long-term usability and growth of the system.

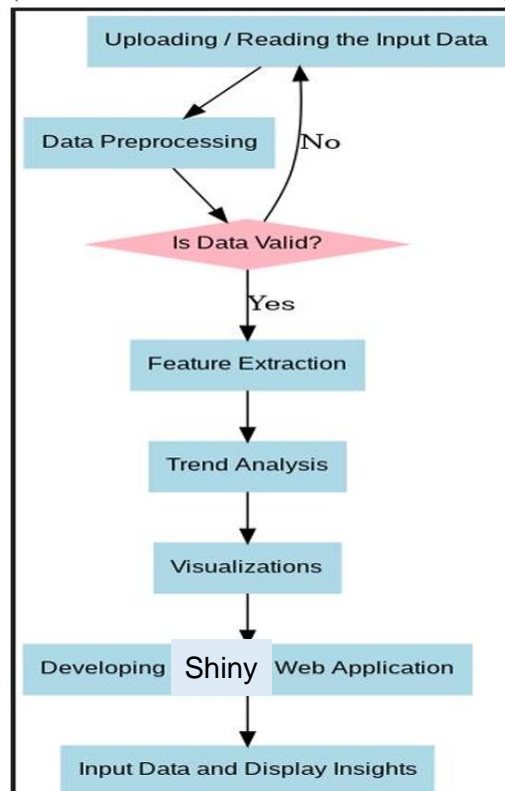
METHODOLOGY

The methodology adopted for this project involves a systematic approach to analyzing customer shopping data using R and Shiny. The project is structured into several key stages to ensure effective data handling, analysis, and visualization.

Workflow Steps

1. Data Collection
 - The dataset containing 3,900 customer records was collected in CSV format.
 - It includes various attributes such as gender, age, category, payment method, purchase amount, and location.
2. Data Preprocessing
 - Using the dplyr and readr libraries, the raw dataset was read and cleaned.
 - Missing values, outliers, and inconsistent entries were handled.
 - Data was transformed and structured into a tidy format suitable for analysis.
3. Exploratory Data Analysis
 - Basic statistical summaries and distributions were generated.
 - The aim was to understand trends such as most used payment methods, gender-wise buying behavior, and age-group preferences.
4. Visualization
 - The plotly and leaflet packages were used to create dynamic, interactive visualizations.
 - Data tables were built using DT, allowing users to explore raw and filtered datasets.
 - Maps were created using leaflet to show geographical trends, if location data was included.
5. Web Application Development
 - The entire analytical dashboard was developed using shiny and shinydashboard.
 - UI elements like filters, tabs, and input fields allow the user to interactively explore insights.
 - All charts, summaries, and maps were integrated into a user-friendly interface.
6. Deployment
 - The app was deployed on ShinyApps.io to make it accessible online.
 - The final product is an end-to-end dashboard that allows businesses and stakeholders to derive insights from shopping data easily.

SYSTEM DESIGN



(Fig no. 1 Control flow)

Explanation of Diagram:

1. Uploading / Reading the Input Data

- **Action:** The app begins by loading customer data from a CSV file using `read_csv()` from the `readr` package.
- **Shiny Functionality:** A `fileInput()` widget can be used in the UI to let users upload their datasets dynamically.

2. Data Preprocessing

- **Action:** Data is cleaned and transformed using the `dplyr` package.
 - Handle missing values, incorrect formats, outliers, etc.
 - Ensure all variables are in the correct data type (e.g., numeric, categorical).
- **Example:** `filter()`, `mutate()`, and `na.omit()` are commonly used here.

3. Is Data Valid?

- **Decision Point:** After preprocessing, the app checks if the data meets certain quality standards.
 - Are there enough rows?
 - Are required columns present and clean?

- **Shiny Functionality:** Conditional messages (`validate()` or `showNotification()`) can be displayed to inform the user.
-

4. Feature Extraction

- **Action:** New features or insights are derived from existing columns.
 - Example: Grouping purchase frequency by gender or age, calculating totals.
 - **Tools:** `group_by()`, `summarise()`, `mutate()`.
-

5. Trend Analysis

- **Action:** Patterns are uncovered using statistical summaries or visual trends.
 - Seasonal trends, preferred payment modes, popular product categories, etc.
 - **Optional:** Basic predictive models or clustering can be introduced.
-

6. Visualizations

- **Action:** Key findings are plotted using libraries like `plotly`, `ggplot2`, and `leaflet`.
 - **Examples:**
 - Pie charts for gender distribution
 - Line graphs for monthly sales
 - Bar plots for product categories
 - Map plots using `leaflet` for geographic distribution
-

7. Developing Shiny Web Application

- **Action:** All analysis and visualizations are wrapped inside a Shiny dashboard layout using `shinydashboard`.
 - Sidebar menu for navigation
 - Dynamic UI components like sliders, dropdowns, etc.
 - Tabs for different types of insights
-

8. Input Data and Display Insights

- **Final Output:** Users interact with the dashboard to:
 - Filter data
 - Select criteria (like gender, city, age group)
 - View visualized insights dynamically
- The UI and server work together to reflect user input in real-time.

Requirement Specification

Hardware Requirements

- **Processor:** Intel i3 or equivalent (sufficient for basic data processing and visualization).
- **RAM:** Minimum 4 GB (to handle small to medium-sized datasets without lag).
- **Storage:** At least 5 GB of free disk space for storing CSV files, analysis outputs, and related resources.
- **Network:** A stable internet connection is required for accessing R packages, libraries, and deploying the Shiny application online (if hosted).

3.2.2 Software Requirements

Operating System

- Compatible with Windows, macOS, or Linux based on the user's preference.

Programming Language and Libraries

- **R (version 4.x or above)** – Used for data analysis, visualization, and developing the Shiny dashboard.
- **Required R Packages:**
 - shiny – For building the interactive web application.
 - shinydashboard – For designing dashboard layout and UI components.
 - plotly – For creating interactive charts and graphs.
 - dplyr – For efficient data manipulation and transformation.
 - readr – For reading and processing CSV files.
 - leaflet – For generating interactive maps based on location data.
 - DT – For rendering dynamic and searchable data tables.

Tools and Environment

- **RStudio IDE** – For writing, testing, and running R scripts and Shiny applications.
- **Shiny Server** (or **shinyapps.io**) – For deploying and sharing the web application.
- **GitHub (Optional)** – For version control and sharing the source code repository.

KEY FUNCTIONALITIES OF APPLICATION

Dataset Details

The dataset used for this project was downloaded from the Kaggle website. It provides detailed information about customers, their purchases, and inventory-related data, making it well-suited for developing a shopping analytics dashboard.

This dataset contains a total of 3,900 entries (rows) and 18 columns. Each row represents a unique transaction or customer entry, and the columns capture various aspects such as personal details, product information, and sales metrics.

User-Friendly Interface with Custom Styling

- Customized UI using shinydashboard, enhanced with CSS for modern aesthetics.
- Gradient background, shadowed boxes, and responsive fonts ensure a visually appealing experience.
- Custom audio controls and styling embedded via JavaScript and CSS.

Interactive Geospatial Analysis (Map Tab)

- Implemented using the **Leaflet** library.
- Displays total purchase amounts across various U.S. states using circle markers.
- Hovering over markers shows location-specific spending insights.

Speech-Enabled Accessibility Features

- Integrated **Speech Synthesis API** via JavaScript to read out:
 - Welcome message
 - About section
 - Conclusion section
- Accessible buttons to **play** and **stop** the audio for each section.

Full Dataset Viewing (Dataset Tab)

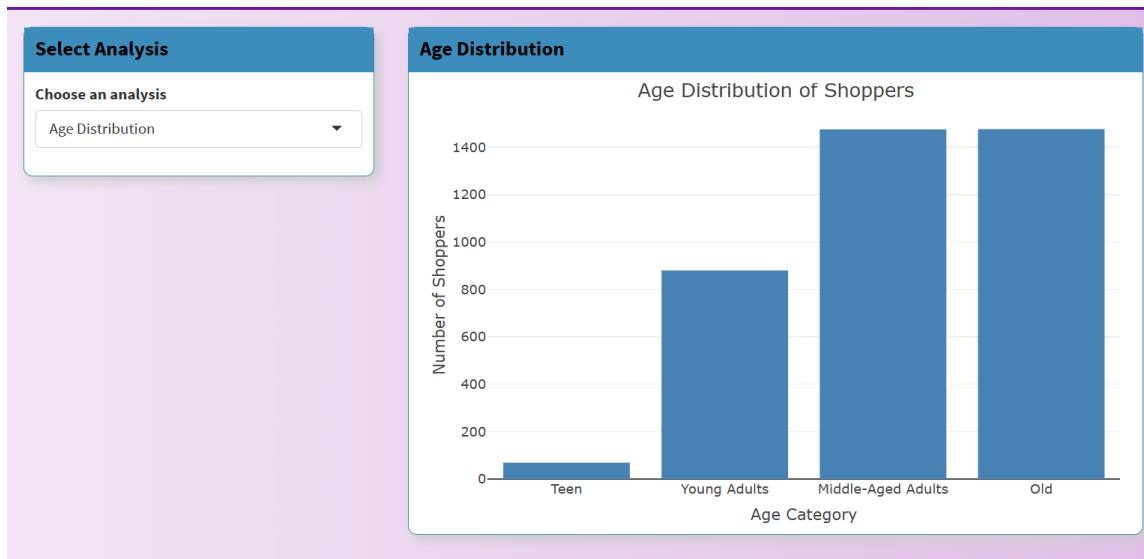
- Displayed using the **DT** package with features like:
 - Search
 - Sorting
 - Pagination
 - Copy/Download table options

Multiple Analysis Tabs

- Includes additional focused tabs for deep-dives:
 - Sales Performance
 - Customer Trends
 - Inventory Management
 - Marketing Insights
 - Store Operations

SCREENSHOTS & FUNCTIONAL DEMONSTRATION

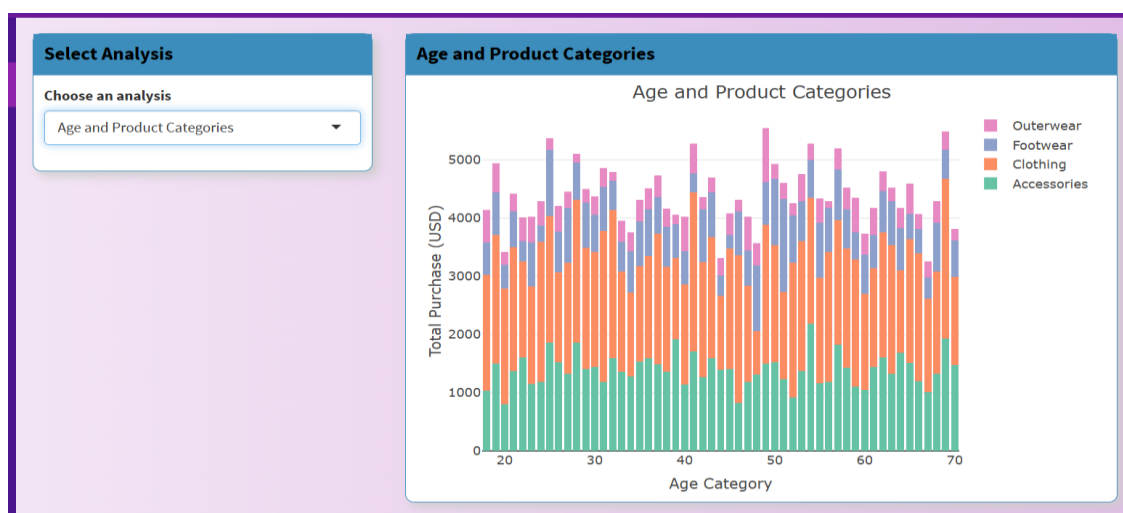
1) Distribution of Customer Ages :



(fig 2:Age distribution)

- The dataset reveals that the majority of customers fall into the Middle-Aged Adults and Older Adults categories, with each group accounting for over 1,500 entries.
- Insight: The analysis suggests that shopping activities are predominantly driven by middle-aged and older adults, likely due to their greater purchasing power, financial stability, and well-defined consumer needs. Businesses targeting these demographics may achieve higher engagement and conversion rates.

2) Age and Product Categories

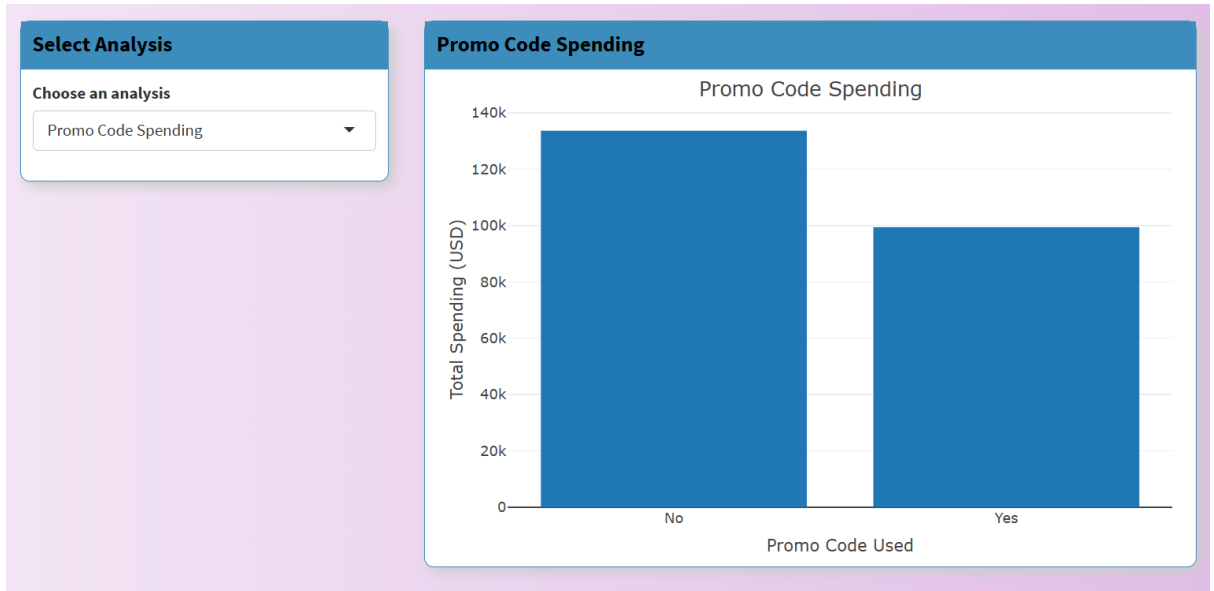


(fig 3:Age & product categories)

Insight: The visualization indicates that **middle-aged customers are the most active buyers** across a diverse range of product categories. **Clothing and Accessories** are

essential across all age groups, making them key drivers of retail revenue. Retailers can use this data to **strategically target age-based segments** with tailored product recommendations.

3) Promo Code Usage



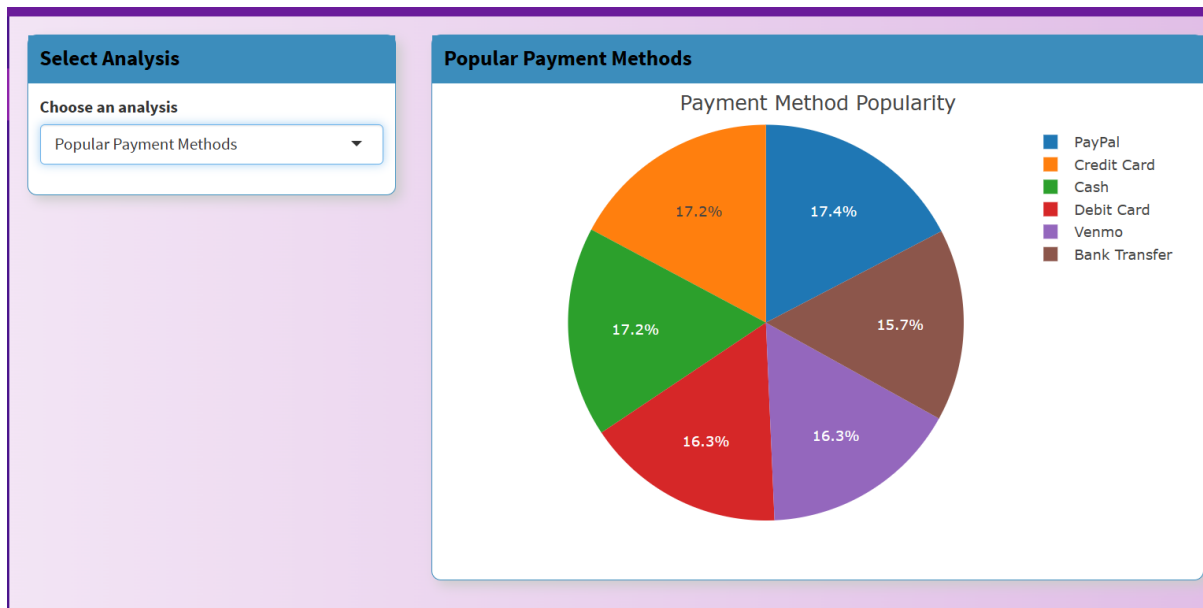
(fig 4:Promo code spending)

The analysis compares transactions **with** and **without promo codes**, focusing on their influence on total spending.

Insight: While **promo codes attract a significant volume of buyers**, they also **reduce the profit margin** per transaction. This suggests that **promo strategies** should be **used selectively**—perhaps during off-peak seasons or for customer acquisition—rather than as a regular practice.

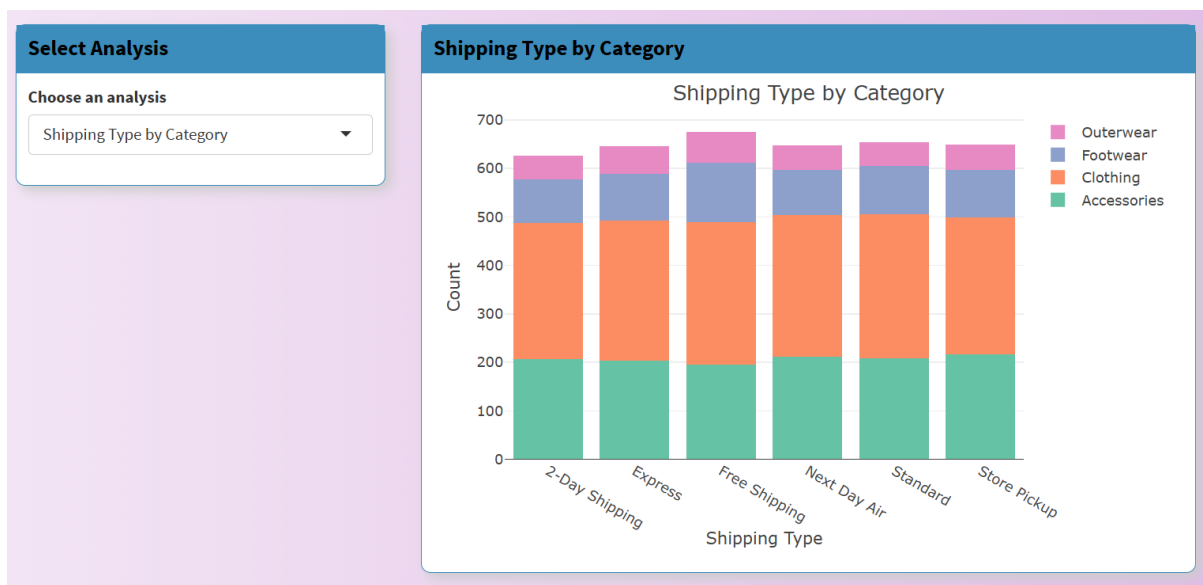
4) Popular Payment Methods

- The distribution of payment methods provides insights into customer preferences and transaction convenience.
- **Breakdown of Payment Method Usage:**
 - **17%** of transactions were completed using **credit cards, PayPal, and cash**, indicating a strong preference for **traditional and widely accepted** payment modes.
 - **16%** of payments were processed through **Venmo and debit cards**, showing that **digital wallet adoption** is on the rise, especially among younger customers.
- **Insight:** The data reveals a **diverse range of payment preferences** among customers. Businesses should ensure support for **multiple payment gateways** to maximize convenience and conversion rates.



(fig 5:Popular payment methods)

5) Shipping Type by Product Category



(fig 6:Shipping type by category)

Clothing consistently dominates across all shipping types, indicating its high demand regardless of delivery speed.

Accessories follow closely behind, especially in **Store Pickup** and **2-Day Shipping**, suggesting that customers prefer faster or convenient delivery for these items.

Insight:

Businesses should ensure **flexibility in shipping options** for clothing and accessories due to their high demand. Offering **expedited or free shipping** can attract more customers, especially for frequently purchased categories.

RESULT ANALYSIS

Distribution of Customer Ages

- The highest number of customers belong to the Old and Middle-Aged Adults categories, with over 1500 entries for each group.
- Young Adults make up a moderate portion, while Teens are significantly underrepresented.

Insight:

The majority of shopping activities are driven by older and middle-aged customers, likely due to their higher purchasing power and specific needs.

Purchases by Gender

- Males contribute to a significantly higher total purchase amount compared to females, with a difference exceeding 50%.

Insight:

Marketing campaigns can target male shoppers for high-ticket items, while focusing on increasing female engagement and loyalty.

Category-Wise Purchases

- Clothing is the most frequently purchased category, followed by Accessories.
 - Popular clothing items: Blouses, Shirts, Pants
 - Popular accessories: Jewelry, Belts, Sunglasses
- Footwear and Outerwear show relatively lower purchase counts.

Insight:

Retailers can focus on expanding clothing collections and promoting trending accessories to cater to majority demand.

Effect of Discounts on Purchases

- When discounts were not applied, the purchase amount was \$130,000 USD.
- When discounts were applied, the amount was \$100,000 USD.

Insight:

Discounts significantly influence customer behavior, encouraging more purchases during sale periods.

Popular Payment Methods

- Most commonly used payment methods include:
 - Credit Card (17%)
 - PayPal and Cash
 - Venmo and Debit Card (16%)
 - Bank Transfer (15%)

Insight:

Ensuring smooth transaction experiences for these payment methods can enhance customer satisfaction.

Promo Code Usage

- Purchases without promo codes amounted to **\$130,000 USD**.
- Purchases with promo codes totaled **\$100,000 USD**.

Insight:

Promo codes increase engagement but may reduce revenue per purchase. A balance between profitability and promotions is essential.

Shipping Preferences

- Shipping options include:
 - 2-Day Shipping, Express, Free Shipping, Next-Day Air, Standard, Store Pickup
- Categories like Clothing, Accessories, Footwear, and Outerwear are evenly distributed across shipping types.

Insight:

Offering diverse shipping methods caters to varying customer needs and improves shopping convenience.

Popular Colors Chosen

- Frequently chosen colors include:
 - Olive, Yellow, Silver, Teal, Green, Black, Cyan, Violet

Insight:

Highlighting these colors in product recommendations or marketing campaigns can align with customer preferences.

Concluding Statements

Age Demographics

- Older adults and middle-aged customers are the primary shoppers.
- Tailored promotions for these age groups (e.g., discounts on premium products, loyalty programs) can maximize sales.

Gender Influence

- Males dominate in terms of spending.
- Personalized recommendations for men's products or bundles targeting their preferences may boost sales.

Category Preferences

- Clothing is the core segment for customer retention.
- Accessories complement clothing purchases, suggesting cross-selling strategies (e.g., bundling jewelry with apparel).

Strategic Opportunities

- Underrepresented groups (Teens and Females) show growth potential.
- Campaigns focused on trendy and affordable items could engage these segments.
- Lower-performing categories like Footwear and Outerwear may need reevaluation in terms of variety, quality, or targeted advertising.

Result Analysis & Identifying Trends

- Use insights from age demographics, gender influence, and category preferences to identify emerging trends.
- Focus on strategic marketing and inventory management to address gaps and drive business growth effectively.

FUTURE SCOPE

While the **Shopping Trend Analysis** project already provides valuable insights into customer purchasing behavior, there are several potential areas for future improvement and expansion. These enhancements could significantly increase the model's effectiveness, accuracy, and real-world applicability.

1. Data Expansion

- **Real-Time Data Integration:**

Currently, the dashboard relies on static CSV files. Integrating real-time data from online shopping platforms (via APIs) could enhance the dashboard's relevance by reflecting the most up-to-date trends.

- **Larger and Diverse Datasets:**

Expanding the dataset with more entries and incorporating diverse data sources (e.g., online reviews, customer feedback, or external economic indicators) could provide deeper insights and better predictive power.

2. Advanced Predictive Modeling

- **Machine Learning Algorithms:**

Incorporating machine learning techniques such as clustering, regression, or classification can add intelligent features like customer segmentation, demand forecasting, and targeted marketing.

- **Time Series Analysis:**

Using models like ARIMA or LSTM (Long Short-Term Memory networks) would allow for forecasting future sales trends based on historical patterns and seasonality.

3. Personalization Features

- **Customer Segmentation:**

Grouping customers based on demographics and shopping behavior could enable personalized recommendations, enhancing user experience.

- **Recommendation System:**

Implementing a collaborative or content-based recommendation system can guide customers toward relevant products, increasing user engagement and potential sales.

4. Data Privacy and Security

- **Secure Data Handling:**

As customer data privacy is critical, future versions of the application should implement secure data handling practices including encryption, secure authentication, and compliance with data protection regulations.

5. Scalability

- **Cloud Integration:**

Hosting the application on cloud platforms like AWS or Google Cloud can handle larger datasets and improve system scalability and performance.

- **Database Integration:**

Transitioning from static CSV files to dynamic databases such as MongoDB or MySQL would allow more efficient data storage, retrieval, and management.

6. Enhanced Visualizations

- **Interactive Features:**

Adding more user controls such as filters, drill-downs, dynamic charts, and real-time updates would make the dashboard more engaging and insightful for users.

CONCLUSION

The **Shopping Trend Analysis** project successfully demonstrates how data visualization and interactive dashboards can provide actionable insights into customer purchasing behavior. Developed using **R** with libraries such as shiny, shinydashboard, plotly, dplyr, and others, the application allows users to intuitively explore patterns in age, gender, location, and product preferences.

By leveraging a dataset of 3,900 entries and 18 attributes sourced from **Kaggle**, this project enables meaningful analysis through interactive charts, searchable data tables, and a geographic map. The user-friendly dashboard empowers decision-makers—whether technical or non-technical—to understand customer behavior and adapt strategies in marketing, inventory, and product design.

This project lays a strong foundation for future expansion, such as incorporating machine learning models, real-time data integration, customer segmentation, and cloud deployment. It demonstrates not only the importance of **data-driven decision-making** but also the potential of interactive tools to democratize access to powerful insights.

In conclusion, **Shopping Trend Analysis** is more than just a data dashboard—it's a step toward smarter, more informed retail decisions backed by technology and analytics.

Deployment & Source Code

Live Dashboard Link:

<https://sanjanaahello.shinyapps.io/FinalApp/>

The project has been deployed using ShinyApps.io allowing users to interact with the dashboard online without any local setup.

GitHub Repository:

https://github.com/sanjanaGhadge/BDA_PBL/

This repository contains the full source code, including the R scripts, UI/server logic, and the dataset used.

REFERENCES

- IdeaScale Blog. *What is Trend Analysis?* Retrieved from: <https://ideascale.com/blog/what-is-trend-analysis>
- Creswell, J. W. (2013). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (4th ed.).
- Merriam, S. B., & Tisdell, E. J. (2015). *Qualitative Research: A Guide to Design and Implementation* (4th ed.).
- Booth, W. C., Colomb, G. G., & Williams, J. M. (2008). *The Craft of Research* (3rd ed.).
- Trochim, W. M. K., & Donnelly, J. P. (2008). *The Research Methods Knowledge Base* (3rd ed.).
- *Analysis of Shopping Trends Employing E-Commerce Applications: A Comparative Case Study.*
- Kaggle Dataset: Islam, M. M. (n.d.). *E-Commerce Data Analysis*. Retrieved from: <https://www.kaggle.com/datasets/mmohaiminulislam/ecommerce-data-analysis>