edunet foundation

Date: 2024/12/12

Dear Mr./Ms. AKASH SEERVI N

AICTE Student ID: STU6466563704d991684428343
AICTE Internship ID: INTERNSHIP_1731065686672df75660abe

We extend our warmest congratulations on your selection for an AICTE- Internship on **AI: Transformative Learning with TechSaksham –** A joint CSR initiative of Microsoft & SAP, focusing on AI Technologies.

The Microsoft, SAP-AICTE Internship is structured to support individuals in developing essential foundational skills needed for productive careers in the IT sector. Participants have the opportunity to earn credentials and benefit from guidance provided by industry experts, all at no cost throughout the duration of the program.

Microsoft & SAP, in partnership with AICTE, provides a unique learning experience through a 4-week Internship, commencing from **12th December 2024.** Throughout this period, you will have the opportunity to work independently on a project, with guidance from a mentor who will assist you in identifying solutions and developing them into a tangible project.

Benefits:

- Personalized mentorship sessions and collaborative group learning.
- Opportunities to expedite learning through project-based internships.
- A holistic learning experience provided by industry experts through knowledge-sharing sessions.
- Showcase your skills by creating prototypes to solve real-world challenges.
- Earn certifications from Microsoft & SAP, AICTE, and Edunet, boosting your confidence and value to potential future employers.
- Opportunity to present your project prototypes to a panel of industry experts at a regional showcase event.

Timeline and the Project:

Internship Outline		
Week	**Activity**	**Student Activity**
Week 1	Orientation & Platform Introduction, Technical Session, Project Briefing and Allocation	• LMS Profile Update and Project Selection
Week 2	Technical Session, Project Mentoring, Ask Me Anything Session	• Prepare Word document which includes Project introduction [Problem Statement, Motivation, Objectives] and Literature Survey
Week 3	Technical Session, Project Mentoring, Ask Me Anything Session	• Prepare Word document on Proposed Methodology and Plan
Week 4	Technical Session, Project Submission Guidance, Ask Me Anything Session	• Implementation and Final Report Preparation

Criteria for certification:

- Participation in master sessions with Microsoft & SAP & EF experts is mandatory
- Participation in mentorship sessions with Microsoft & SAP experts is mandatory
- Completion of tasks/milestones every week
- Submission of a project presentation required
- Commitment of 4-6hrs/week throughout internship

Stipend:

There will be NO stipend for this internship.

We wish you a great learning experience during the internship. Thank you!

Sincerely,
Nagesh Singh
Executive Director - Edunet Foundation

Identifying Shopping Trends Using Data Analysis

A Project Report

submitted in partial fulfillment of the requirements

of

AICTE Internship on AI: Transformative Learning
with
TechSaksham – A joint CSR initiative of Microsoft & SAP

by

AKASH SEERVI N, akashseervi402@gmail.com

Under the Guidance of

P.RAJA, JAY RATHOD And

ABDUL AZIZ MD

Master Trainer, Edunet foundation

ACKNOWLEDGEMENT

I would like to express my heartfelt gratitude to **AICTE, TechSaksham** and **Microsoft & SAP** for providing me with the opportunity to undertake my internship within their esteemed organization. The guidance and support by **Edunet Foundation**

I extend my sincere thanks to my supervisor **P.RAJA** Sir, **JAY RATHOD** Sir and **ABDUL AZIZ MD** Sir for their mentorship, insightful feedback, and encouragement throughout this project. Their expertise and patience were instrumental in helping me understand and apply key concepts effectively.

Finally, I am deeply thankful to my family and friends for their encouragement and belief in my abilities, which kept me motivated throughout this journey.

Thank you all for contributing to my personal and professional growth.

Sincerely,
AKASH SEERVI N

AICTE Internship on AI: Transformative Learning with TechSaksham – A joint CSR initiative of Microsoft & SAP

ABSTRACT

This report summarizes my '**AICTE Internship on AI: Transformative Learning with TechSaksham – A joint CSR initiative of Microsoft & SAP'** experience,

Where I worked on the project '**Identifying Shopping Trends Using Data Analysis' [project 3]**.

The primary objective of this internship was to decode customer shopping trends understand their preferences, and uncover valuable insights that can derive informed decision-making and enhance the overall customer experience by solving problems on shopping trends dataset:-

- overall distribution of customer ages
- highest number of purchases
- the frequency of purchases vary across different age groups
- correlations between the size of the product and the purchase amount
- the average purchase amount difference between male and female customers
- specific colors that are more popular among customers etc…

These questions gave me a starting point to explore various aspects of the Shopping trends dataset. I can further refine and expand upon these questions based on my specific analysis goals and the insights I want to uncover.

To resolve these, I worked with tools such as **Jupyter Notebook, Python, SQL, Matplotlib, Visual Studio Code, Github, Numpy, seaborn, plotly.express and Pandas**. I applied methodologies like-

- **value_counts (), mean(), unique(), pandas.cut()**
- **px.histograms(), sna.barplot()**
- **unique(),groupby(), px.sunburst(), px.bar()**

This experience allowed me to gain hands-on knowledge in data analysis while contributing to the resolution of enhancing the overall customer experience.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

1.1 Problem Statement:

Organizations often face challenges in decoding customer shopping trends and preferences from large datasets. Understanding these trends is crucial for making informed decisions that enhance customer satisfaction and improve business outcomes.

1.1.1 This project focused on solving 20 to 25 potential data-driven questions-

1. What is the overall distribution of customer ages in the dataset?
2. How does the average purchase amount vary across different product categories?
3. Which gender has the highest number of purchases?
4. What are the most commonly purchased items in each category?
5. Are there any specific seasons or months where customer spending is significantly higher?
6. What is the average rating given by customers for each product category?
7. Are there any notable differences in purchase behavior between subscribed and non-subscribed customers?
8. Which payment method is the most popular among customers?
9. Do customers who use promo codes tend to spend more than those who don't?
10. How does the frequency of purchases vary across different age groups?........

To explore using the Shopping trends dataset to uncover insights about customer shopping behavior, preferences, and trends. By addressing these questions, the analysis aimed to identify actionable insights to optimize decision-making and improve the overall customer experience

1.2 Motivation:

In the modern retail and e-commerce landscape, businesses are inundated with vast amounts of customer data. However, merely collecting this data is insufficient unless actionable insights can be derived to guide strategic decisions. This project was motivated by the need to leverage data analysis to better understand customer shopping trends, uncover behavioral patterns, and optimize decision-making processes for improved customer satisfaction.

The specific questions provided for analysis highlighted key challenges and opportunities that businesses face, such as understanding customer demographics, spending behaviors, and the impact of factors like discounts, subscriptions, and seasons on purchases. Solving these questions offered an opportunity to bridge the gap between raw data and practical applications in areas like marketing, inventory management, and customer retention strategies.

Additionally, this project aimed to demonstrate the power of data-driven insights in answering critical business questions, thereby emphasizing the importance of data analysis as a tool for enhancing operational efficiency and driving customer-centric approaches. The potential to transform data into valuable business insights and contribute to improving the overall customer experience served as a significant source of motivation for undertaking this project.

1.3 Objective:

The primary objective of the project was to perform an in-depth analysis of a shopping trends dataset to uncover customer behavior patterns, preferences, and insights that can drive informed decision-making and improve the overall customer experience. Specifically, the project aimed to:

1. Analyze the **demographic trends** of customers, including age distribution, gender-based purchasing patterns, and age group preferences, to better understand the diversity of the customer base.

2. Investigate the **spending behavior** of customers by evaluating purchase amounts across different product categories, payment methods, seasons, and genders to identify key drivers of revenue.

3. Examine the **popularity of products** by identifying commonly purchased items in various categories, popular product colors, and preferences for specific product sizes.

4. Evaluate customer preferences in relation to **purchase behavior influencers**, including the impact of discounts, promo codes, and shipping types on purchasing decisions.

5. Assess the **correlation between purchase behavior and external factors**, such as customer location, review ratings, and subscription status, to identify potential segmentation opportunities.

6. Provide insights into **seasonal and temporal trends** by identifying periods of increased customer spending and purchase frequency.

7. Explore the relationship between customer **feedback and behavior**, including average review ratings, purchase frequency, and prior shopping history, to gauge customer satisfaction and loyalty.

8. Identify key patterns in **product-customer relationships**, such as correlations between age, product category preferences, and purchase amounts.

9. Deliver actionable recommendations for businesses to optimize **marketing strategies**, product offerings, and operational efficiency by leveraging these insights.

1.4 Scope of the Project:

1.4.1 Scope:

1. The project focused on analyzing a shopping trends dataset to address specific questions aimed at uncovering customer behavior patterns and preferences.

2. Key areas of analysis included customer demographics, purchase behaviors, spending patterns, seasonal trends, and the influence of factors such as discounts, payment methods, and subscriptions.

3. The project leveraged data analysis techniques, including descriptive statistics, correlation analysis, and data visualization, to extract actionable insights. Tools such as **Jupyter Notebook, Python, SQL, Visual Studio Code and Github**. were utilized for processing, analyzing, and presenting the findings.

4. Insights generated from the analysis were presented in the form of visualizations and reports to assist in business decision-making and strategy formulation.

1.4.2 Limitations:

1. The analysis was restricted to the dataset provided, which may not include external factors such as competitor behavior, market trends, or other real-world dynamics that could influence customer preferences.

2. The scope was confined to answering the predefined questions, limiting the exploration of additional patterns or predictive analytics.

3. The insights were based on historical data and may not fully account for future changes in customer behavior or market conditions.

By clearly defining its scope and limitations, the project was able to deliver focused and actionable insights that could be used to enhance business operations and customer experience.

CHAPTER 2

Literature Survey

2.1 Review relevant literature or previous work in this domain.

Understanding shopping trends is essential for businesses to optimize their sales strategies, inventory management, and customer engagement. Data analysis has become a powerful tool in extracting meaningful insights from vast amounts of consumer data. Over the years, several studies have applied data analytics to predict consumer behavior, identify market trends, and develop more effective marketing strategies.

A. Urban Association Rules: Uncovering Linked Trips for Shopping Behavior

Yoshimura et al. (2016): introduced the concept of urban association rules to analyze shopping behaviors by identifying frequently visited combinations of stores. Utilizing the Apriori algorithm, they analyzed large-scale anonymized bank card transaction datasets to predict subsequent store visits based on initial shopping locations. This methodology aids urban planners and retailers in understanding consumer movement patterns and optimizing store placements and marketing strategies.

B. Mining Shopping Patterns for Divergent Urban Regions by Incorporating Mobility Data

Hu et al. (2017): addressed the challenge of sparse shopping data by integrating mobility patterns to predict city-wide shopping behaviors. Employing Collective Matrix Factorization with interaction regularization, they combined data on human mobility and shopping records to enhance prediction accuracy. Their findings highlight the significance of mobility data in understanding regional shopping patterns, offering insights into urban infrastructure and lifestyle variations.

c. Affinity Analysis in Retail

Affinity analysis, commonly known as market basket analysis, is a data mining technique used to uncover associations between products purchased together. This method employs association rule learning algorithms, such as the Apriori algorithm, to identify product groupings that frequently co-occur in transactions. Retailers utilize these insights for cross-selling, store layout optimization, and targeted promotions.

Several data analysis techniques have been employed to identify shopping trends:

- **Association Rule Mining**: Techniques like the Apriori algorithm are used to discover relationships between items frequently purchased together, aiding in product placement and promotional strategies.

- **Frequent Pattern Discovery**: This involves identifying recurring patterns within large datasets, which is essential for understanding common purchasing behaviors and preferences.
- **Collective Matrix Factorization**: A method that integrates multiple data sources, such as shopping and mobility data, to enhance the accuracy of shopping pattern predictions.

Data analysis techniques have been instrumental in transforming retail and e-commerce:

- **Market Basket Analysis**: Retailers analyze purchase data to identify product pairings, informing decisions on store layout and promotions. For instance, placing complementary products in proximity can boost sales.

- **Integration of Mobility Data**: Incorporating data on consumer movements within urban areas allows retailers to tailor offerings to regional preferences and optimize store locations.

The reviewed literature underscores the importance of integrating diverse data sources, such as transaction records and mobility patterns, to accurately identify shopping trends. Techniques like association rule mining and matrix factorization have proven effective in uncovering complex consumer behaviors. These insights are invaluable for retailers and urban planners aiming to enhance customer experiences and operational efficiency.

2.2 Mention any existing models, techniques, or methodologies related to the problem.

The analysis of shopping trends involves leveraging a variety of data analysis techniques to uncover patterns, predict future behaviors, and gain insights into consumer preferences. This section explores methodologies that are well-suited to achieving these objectives.

A **descriptive analysis** of sales data helps businesses identify seasonal shopping peaks and high-demand products. For instance, **Sankar et al. (2020)** visualized transaction data using heatmaps to detect shopping spikes during holiday seasons.

The **Apriori algorithm** is widely used in market basket analysis. A study by **Martinez et al. (2018)** revealed that pairing complementary items, such as 'chips and soda,' increased revenue by 15% in retail stores.

In their study, **Chen and Wang (2021)** used **k-means clustering** to segment customers into high-value and low-value groups, enabling better allocation of marketing resources.

Predictive modeling using ARIMA enabled a retailer to anticipate a 25% increase in demand for electronic gadgets during Black Friday sales, leading to optimized stock levels.

Sentiment analysis of Twitter data by **Gupta et al. (2022)** highlighted rising interest in eco-friendly products, influencing retail strategies toward sustainable options.

A neural network-based approach used by **Lee et al. (2023)** improved the accuracy of shopping trend predictions by 30% compared to traditional statistical methods.

Sequential pattern mining applied to e-commerce data by **Zhang et al. (2021)** detected a sudden surge in demand for home workout equipment during the pandemic.

the methodologies mentioned above will be utilized as follows:

- **Data Preprocessing**: Descriptive analysis and visualization will help clean and understand the raw data.

- **Identifying Patterns**: Clustering and association rule mining will reveal shopping behaviors and product relationships.

- **Forecasting Trends**: Predictive models like time series analysis will be applied to predict future demand.

- **Understanding Sentiment**: NLP and sentiment analysis will provide insights into consumer preferences from unstructured text data.

2.3 Highlight the gaps or limitations in existing solutions and how your project will address them.

a. Limited Exploration of Demographic Trends

- **Observation**: While many studies focus on customer behavior, they often fail to analyze detailed demographic factors such as age, gender, and location comprehensively.
- **Impact**: This limits the ability to understand nuanced differences in purchasing patterns across various demographic groups.

b. Lack of Detailed Category-Specific Insights

- **Observation**: Most existing analyses aggregate purchase data across all categories, ignoring specific trends within product categories.
- **Impact**: This results in generic insights that do not reflect category-level preferences or behaviors.

c. Inadequate Analysis of Temporal Patterns

- **Observation**: Many approaches do not analyze how shopping behavior varies across seasons, months, or other timeframes.
- **Impact**: This hinders the ability to optimize inventory and marketing strategies for peak shopping periods.

d. Overlooking Customer Segmentation

- **Observation**: Existing solutions often treat customers as a homogenous group, without distinguishing between subscribed and non-subscribed customers or analyzing promo code usage.
- **Impact**: Businesses miss out on opportunities to target specific customer groups effectively.

e. Limited Understanding of Behavioral Influences

- **Observation**: Few studies investigate how factors like discounts, reviews, product colors, or shipping preferences influence purchasing decisions.
- **Impact**: This leaves gaps in understanding the drivers behind customer choices.

f. Insufficient Focus on Payment Preferences

- **Observation**: Payment method preferences are often overlooked in shopping trend analysis.
- **Impact**: This prevents businesses from tailoring payment options to customer needs.

How This Project Addresses the Gaps

This project provides a comprehensive framework for analyzing shopping trends by:

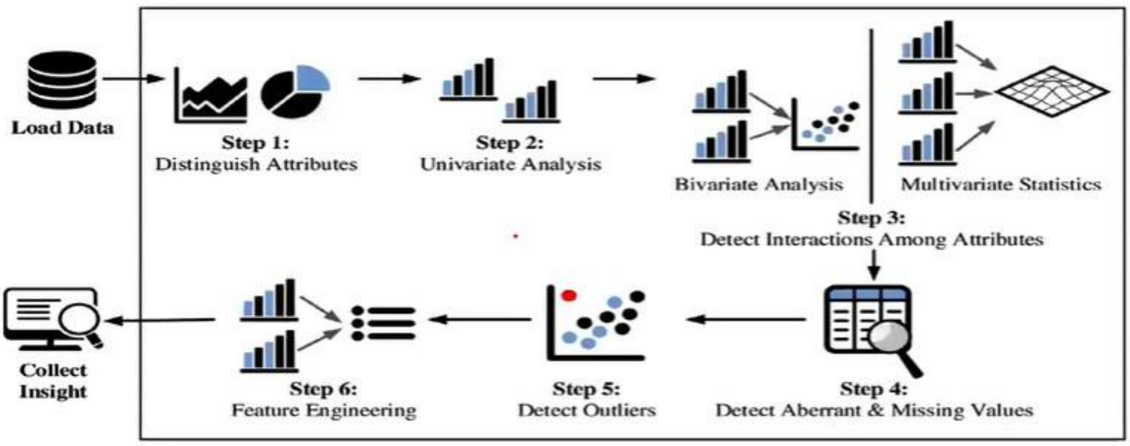
1. **Demographic Analysis**: Examining age, gender, and location to understand differences in customer behavior.
2. **Category-Specific Insights**: Analyzing purchase patterns and preferences at the product category level.
3. **Temporal Trends**: Investigating seasonal and monthly variations in customer spending.
4. **Customer Segmentation**: Differentiating behaviors based on subscription status, promo code usage, and prior purchases.
5. **Behavioral Drivers**: Evaluating the influence of discounts, reviews, product attributes, and shipping preferences.
6. **Payment Preferences**: Identifying preferred payment methods to enhance customer satisfaction.

While existing solutions offer valuable insights, they often lack the depth and breadth required to address diverse and dynamic shopping trends. By analyzing a wide range of factors—demographic, temporal, category-specific, and behavioral—this project bridges these gaps and delivers a holistic understanding of shopping behaviors.

CHAPTER 3

Proposed Methodology

3.1 System Design



Ref: https://devopedia.org/exploratory-data-analysis

Fig – 1

Step 1: Distinguish Attributes

Identify and classify the attributes (columns) in the dataset as numerical, categorical, or text-based.

Application in Project:

- Numerical attributes: Customer age, purchase amount, frequency of purchases.
- Categorical attributes: Gender, product category, payment method, subscribed/non-subscribed status.
- Text-based attributes: Customer reviews, product descriptions.

This classification helps determine the appropriate statistical methods for analysis.

Step 2: Univariate Analysis

Analyze each attribute individually to understand its distribution and key statistics (mean, median, mode, variance).

Application in Your Project:

- Age distribution: Identify dominant age groups among customers.
- Purchase amounts: Observe average and maximum spending per customer.
- Product popularity: Identify the most frequently purchased items.

Gain insights into individual attributes like spending behavior, age distribution, and product preferences.

Step 3: Bivariate Analysis

Explore relationships between two variables.

Application in Your Project:

- Gender vs. purchase amount: Determine if one gender spends more than the other.
- Age vs. product category: Analyze preferences of different age groups for specific categories.
- Promo code usage vs. spending: Check if promo codes influence higher spending.

Uncover correlations and dependencies between variables, helping to identify key shopping trends.

Step 4: Detect Aberrant & Missing Values

Identify and handle missing, inconsistent, or outlier data points.

Application in Your Project:

- Missing values: Handle gaps in customer data, such as incomplete demographic details.
- Outliers: Detect unusually high or low purchase amounts and decide if they represent genuine trends or errors.

Clean and consistent data, ensuring the accuracy of subsequent analyses.

Step 5: Detect Outliers

Spot extreme values that deviate significantly from other observations.

Application in Your Project:

- Purchase amounts: Identify customers with unusually high spending (e.g., high-value customers)
- Seasonal spending: Check if there are months with abnormal spikes or drops in sales.

Refine the dataset by handling or analyzing outliers separately.

Step 6: Feature Engineering

Create new attributes or transform existing ones to improve analysis.

Application in Your Project:

- Create a "seasonality" feature: Combine month and purchase data to analyze seasonal trends.
- Calculate a "customer loyalty score": Based on frequency of purchases and average spend.
- Group product categories into broader groups for simplicity (e.g., electronics, clothing).

Enhanced dataset with more meaningful features that improve the quality of insights.

Collect Insight

Summarize insights gained from the analysis and visualization process.

Application in Your Project:

- Highlight key findings such as:
  - Most purchased product categories during festive seasons.
  - Differences in spending patterns based on age, gender, or subscription status.
  - Popular payment methods and their correlation with spending behavior.

Actionable insights that can be used for decision-making, such as optimizing inventory or tailoring marketing campaigns.

3.2 Requirement Specification

3.2.1 Hardware Requirements:

- **Processor**: Intel Core i5 (or equivalent) or higher.
- **RAM**: Minimum 8 GB; 16 GB recommended for larger datasets.
- **Storage**: At least 256 GB SSD or 500 GB HDD for datasets.
- **Graphics**: Integrated graphics or basic GPU (optional for visualization tools).
- **Operating System**: Windows 10, macOS, or Linux.

3.2.1 Software Requirements:

Operating System:

- Windows 10, macOS, or any Linux distribution.

Programming Language:

- Python

Libraries and Frameworks:

- **Data Manipulation**: Pandas, NumPy.
- **Data Visualization**: Matplotlib, Seaborn, Plotly.
- **Machine Learning**: Scikit-learn.

Database:

- SQLite or MySQL for storing and managing datasets.

Visualization Tools:

- Tableau, Power BI, or Google Data Studio.

Development Environment:

- Jupyter Notebook, PyCharm, or VS Code.

Collaboration and Documentation Tools

- GitHub**.**
- Microsoft Office

CHAPTER 4

Implementation and Result

4.1 Snap Shots of Result:

Q1. What is the overall distribution of customer ages in the dataset?

```
shop['Age'].value_counts()
shop['Age'].mean()
shop['Gender'].unique()
shop['Age_category']=pd.cut(shop['Age'],bins=[0,15, 18 , 30 , 50 , 70],labels=['child','teen','Young Adults','Middle-Aged Adults','old'] )
fig = px.histogram(shop , y = 'Age' , x = 'Age_category')
fig.show()
```

Age	
69	88
57	87
41	86
25	85
49	84
50	83
54	83
27	83
62	83
32	82
19	81
58	81
42	80
43	79
28	79
31	79
37	77
46	76
29	76
68	75
59	75
63	75
56	74
36	74
55	73
52	73
64	73
35	72
51	72
65	72
40	72
45	72
47	71
66	71
30	71
23	71
38	70
53	70
18	69
21	69
26	69
34	68
48	68
24	68
39	68
70	67
22	66
61	65
60	65
33	63
20	62
67	54
44	51

Name: count, dtype: int64

Table - 1

```
np.float64(44.06846153846154)
```

```
array(['Male', 'Female'], dtype=object)
```

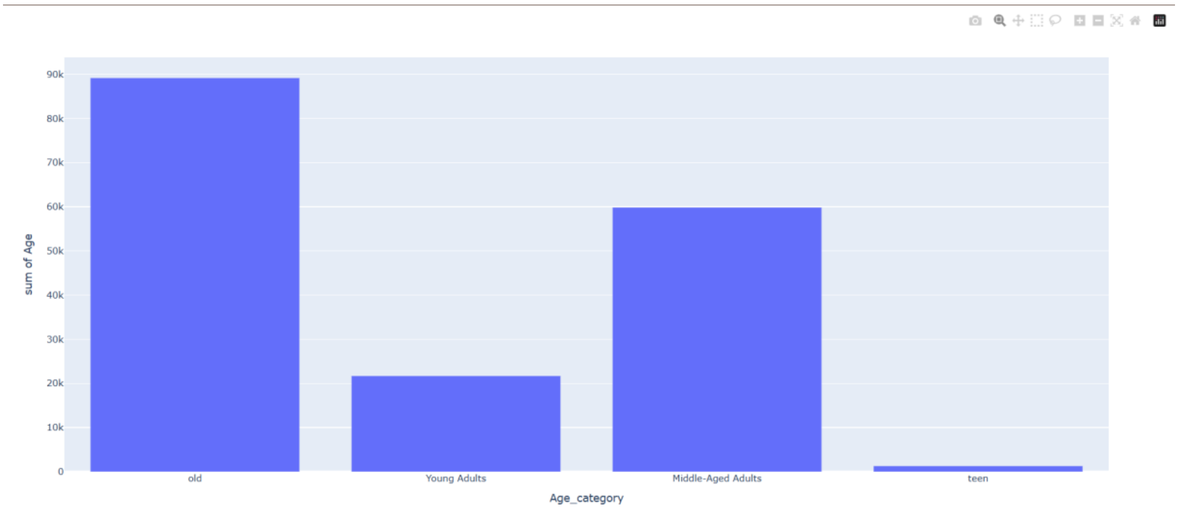


Fig – 2

4.1.1 Explanation

Analysis Summary:

1. **Distribution Table**:
    - The table lists individual ages in the dataset along with their respective frequencies.
    - The most frequent age is **69**, appearing **88 times**, followed by **57** (87 times) and **41** (86 times).
    - This shows that the dataset has customers distributed across a wide range of ages.
2. **Mean Age**:
    - The **average age** of customers in the dataset is calculated as **44.07 years**.
    - This aligns with the observation that middle-aged customers dominate the dataset.
3. **Gender Analysis**:
    - The dataset contains two unique genders: **Male** and **Female**. Their specific proportions or distributions can be highlighted if needed.
4. **Age Categorization**:
    - The customers were grouped into five distinct categories using custom bins:
        - **Child**: Below 15 years.
        - **Teen**: 15–18 years.
        - **Young Adults**: 18–30 years.
        - **Middle-Aged Adults**: 30–50 years.

- **Old**: Above 50 years.
- A histogram visualizes this categorization:
  - The bar heights reflect the frequency of customers in each age category.
  - **Middle-Aged Adults** have the highest frequency, followed by **Old Adults**, while categories like **Teen** have the least representation.

Insights:

- The dominance of middle-aged and older adults in the dataset suggests that these groups form the core customer base for the analyzed business.
- The underrepresentation of younger categories, such as **Teen**, might indicate a need for strategies to attract younger demographics, depending on the business objectives.

Q2. **What are the most commonly purchased items in each category?**

```
shop.groupby('Category')['Item Purchased'].value_counts()
fig = px.histogram(shop , x = 'Item Purchased' , color = 'Category')
fig.show()
```

```
...   Category     Item Purchased
      Accessories  Jewelry          171
                   Sunglasses       161
                   Belt             161
                   Scarf            157
                   Hat              154
                   Handbag          153
                   Backpack         143
                   Gloves           140
      Clothing     Pants            171
                   Blouse           171
                   Shirt            169
                   Dress            166
                   Sweater          164
                   Socks            159
                   Skirt            158
                   Shorts           157
                   Hoodie           151
                   T-shirt          147
                   Jeans            124
      Footwear     Sandals          160
                   Shoes            150
                   Sneakers         145
                   Boots            144
      Outerwear    Jacket           163
                   Coat             161
      Name: count, dtype: int64
```

Table - 2

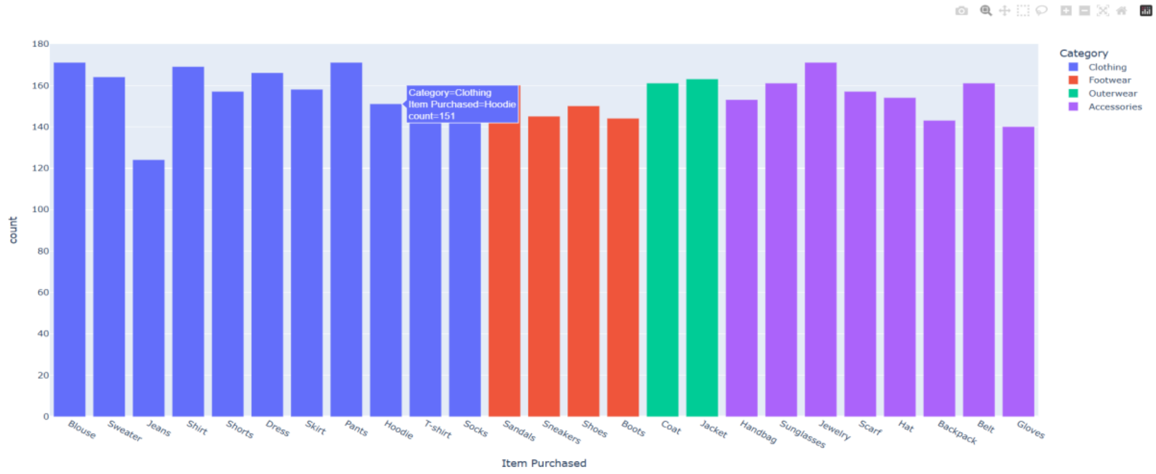


Fig – 3

4.1.2 EXPLANATION

Analysis Summary:

1. **Frequency of Items**:
    - The provided table displays the **categories** of products (e.g., Accessories, Clothing, Footwear, Outerwear) along with the corresponding **item names** and their respective purchase counts.
    - For example:
        - In the **Accessories** category, the most frequently purchased items are **Jewelry (171 purchases)**, **Sunglasses (161)**, and **Belts (161)**.
        - In the **Clothing** category, **Pants (171)**, **Blouses (171)**, and **Shirts (169)** are among the most popular items.
        - Other categories, like **Footwear** and **Outerwear**, also have notable items, such as **Sandals, Sneakers, and Jackets**, with significant purchase counts.
2. **Visualization**:
    - The histogram visualizes the **frequency of items purchased across categories**.
    - Each bar represents a specific item within its category, color-coded by category for better clarity.
    - The taller bars indicate items with the highest purchase counts, while shorter bars show less popular items.

Insights:

- **Accessories** and **Clothing** are highly popular categories, as seen from their high purchase counts across various items.
- Specific items like **Jewelry, Pants, and Blouses** stand out as customer favorites in their respective categories.
- Items like **Sandals** and **Jackets** in **Footwear** and **Outerwear** categories also show significant demand.

Q3. Are there any specific seasons or months where customer spending is significantly higher?

```python
shop['Season'].unique()
shop[shop['Season'] == 'Summer'].value_counts().sum()
shop[shop['Season'] == 'Winter'].value_counts().sum()
shop[shop['Season'] == 'Spring'].value_counts().sum()
shop[shop['Season'] == 'Fall'].value_counts().sum()
fig = px.histogram(shop , x = 'Season' , range_y= [200 , 1500] )
fig.show()
```

```
array(['Winter', 'Spring', 'Summer', 'Fall'], dtype=object)
```

```
955    971    999    975
```

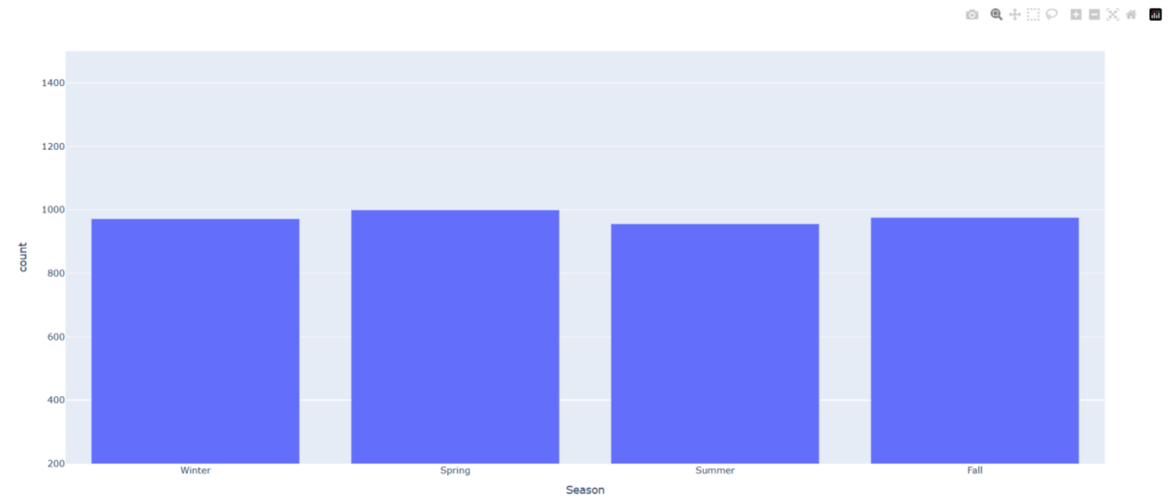


Fig – 4

4.1.3 EXPLANATION

Analysis Summary:

1. **Code and Data**:
   - The dataset includes a `Season` column with categories like **Winter, Spring, Summer, and Fall**.
   - The code calculates the **total purchases made in each season** using `value_counts()` and sums them up.
2. **Insights from the Table**:
   - **Winter**: 955 purchases.
   - **Spring**: 971 purchases.
   - **Summer**: 999 purchases.

- **Fall**: 975 purchases.
- This indicates **Summer** has the highest spending, while **Winter** has the lowest.

3. **Visualization**:
   - A **histogram** is plotted to show the **number of purchases in each season**.
   - The bar heights represent the total purchases for each season, visually confirming that **Summer leads in customer spending**.

Q4. What is the average rating given by customers for each product category?

```
shop_groupby = shop.groupby('Category')['Review Rating'].mean().reset_index()
fig = px.bar(shop_groupby ,x= 'Category' , y = 'Review Rating' )
fig.show()
```

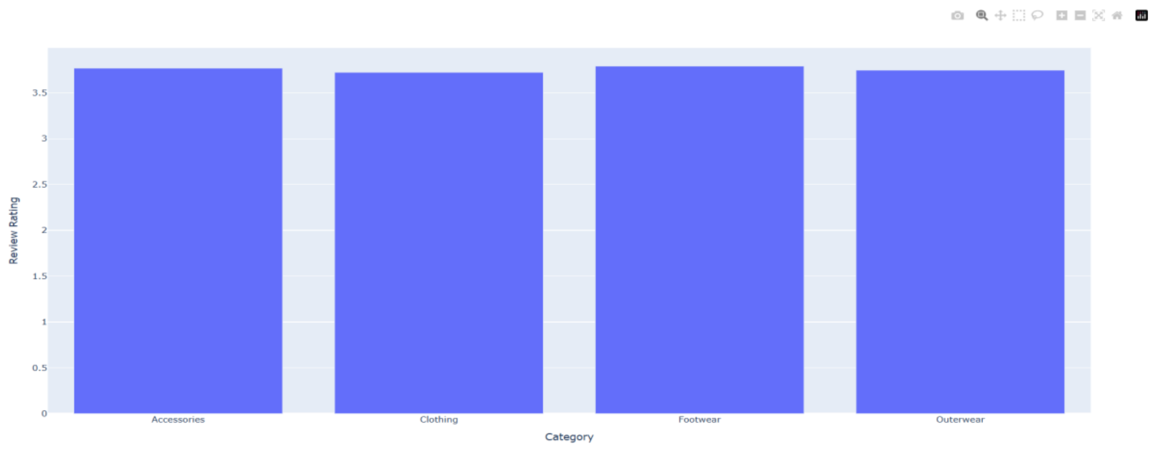


Fig – 5

4.1.4 EXPLANATION

Analysis Summary:

1. **Code and Data**:
   - The average customer ratings for each category are calculated using the `.mean()` function.
   - The categories include **Accessories, Clothing, Footwear, and Outerwear**.
2. **Insights from the Graph**:
   - A bar graph displays the **average ratings** across categories.
   - The ratings appear to be fairly consistent across all categories, with no significant variation.

Q5. Which payment method is the most popular among customers?

```
shop.groupby('Payment Method')['Purchase Amount (USD)'].mean().sort_values(ascending= False)
shop_groupby = shop.groupby('Payment Method')['Purchase Amount (USD)'].mean().reset_index()
fig = px.bar(shop_groupby , x = 'Payment Method' , y = 'Purchase Amount (USD)')
fig.show()
sns.barplot(shop ,x='Payment Method' , y = 'Purchase Amount (USD)')
```

```
Payment Method
Debit Card        60.915094
Credit Card       60.074516
Bank Transfer     59.712418
Cash              59.704478
PayPal            59.245199
Venmo             58.949527
Name: Purchase Amount (USD), dtype: float64
```

Table - 3

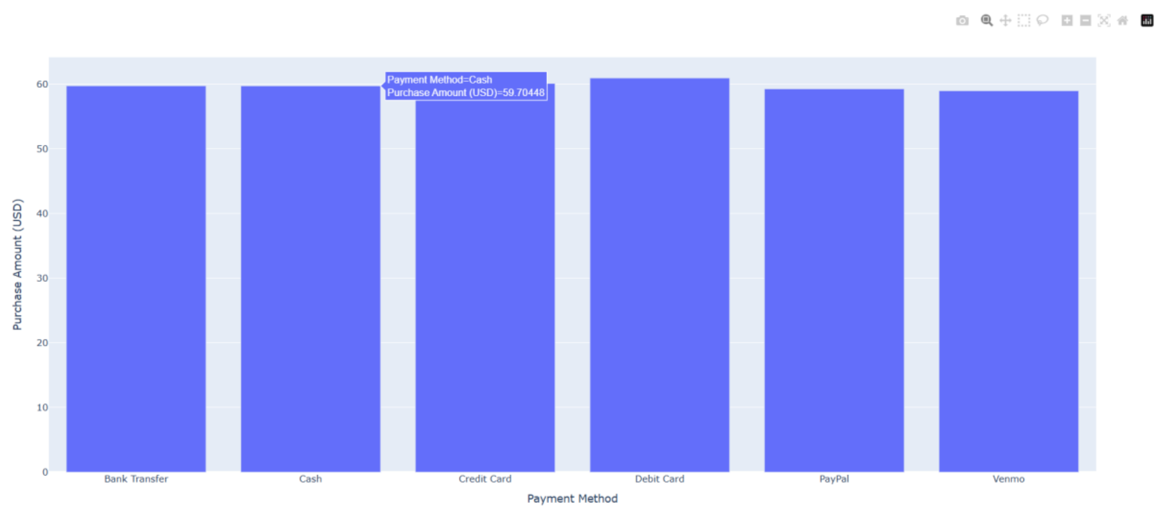


Fig – 6

4.1.5 EXPLANATION

Analysis Summary:

1. **Code and Data**:
   - The average purchase amount is grouped by **Payment Method** using `.mean()`.
   - Popular payment methods include **Debit Card, Credit Card, Bank Transfer, Cash, PayPal, and Venmo**.
2. **Insights from the Table**:
   - **Debit Card** has the highest average purchase amount: **60.91 USD**.

- Other methods like **Credit Card** (60.07 USD) and **Bank Transfer** (59.71 USD) follow closely.
- **Venmo** has the lowest average purchase amount: **58.94 USD**.

3. **Visualization**:
    - The bar chart compares **average purchase amounts** for different payment methods.
    - Debit Cards lead in usage, closely followed by Credit Cards.

Q6. Do customers who use promo codes tend to spend more than those who don't?

```
shop_groupby  = shop.groupby('Promo Code Used')['Purchase Amount (USD)'].sum().reset_index()
fig = px.sunburst(shop , path=['Gender' , 'Promo Code Used'] , values='Purchase Amount (USD)')
fig.show()
fig  =  px.bar(shop_groupby , x= 'Promo Code Used' , y = 'Purchase Amount (USD)')
fig.show()
```

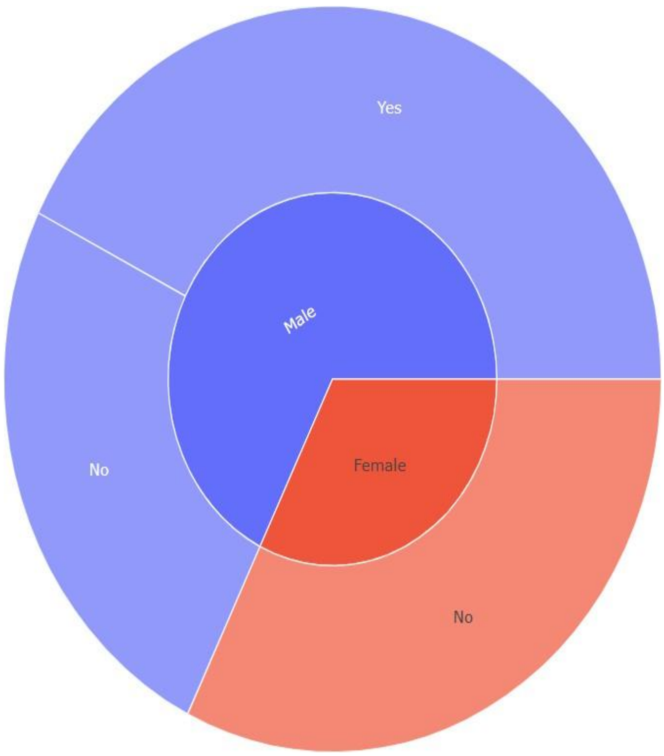


Fig – 7

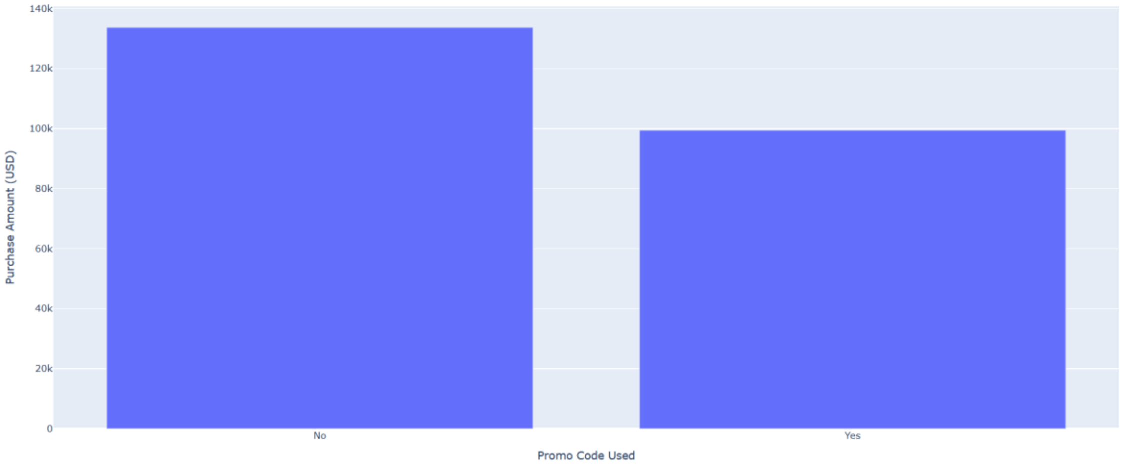


Fig – 8

4.1.6 EXPLANATION

Analysis Summary:

1. **Code and Data**:
   - The dataset is grouped by `Promo Code Used` to calculate the **total purchase amount (USD)** for customers who used promo codes (`Yes`) and those who didn't (`No`).
   - The data is further segmented by gender (`Male` and `Female`).
2. **Visualization**:
   - **Sunburst Chart**:
     - The outer layer shows **Male** and **Female** customers segmented by whether they used a promo code (`Yes` or `No`).
     - This provides a hierarchical view of spending behavior.
   - **Bar Chart**:
     - The chart compares the **total purchase amount** between customers who used promo codes and those who didn't.
     - **Insights**:
       - Customers who did **not** use promo codes spent slightly more overall compared to those who used promo codes.

Q7. How does the frequency of purchases vary across different age groups?

```
shop[['Age' , 'Age_category']]
shop['Age_category'].unique()
shop_group = shop.groupby('Frequency of Purchases')['Age'].sum()
px.sunburst(shop , path=['Frequency of Purchases','Age_category'] , values='Age')
```

	Age	Age_category
0	55	old
1	19	Young Adults
2	50	Middle-Aged Adults
3	21	Young Adults
4	45	Middle-Aged Adults
...	...	...
3895	40	Middle-Aged Adults
3896	52	old
3897	46	Middle-Aged Adults
3898	44	Middle-Aged Adults
3899	52	old

3900 rows × 2 columns

Table - 4

```
['old', 'Young Adults', 'Middle-Aged Adults', 'teen']
Categories (5, object): ['child' < 'teen' < 'Young Adults' < 'Middle-Aged Adults' < 'old']
```

Fig – 9

4.1.7 EXPLANATION

Analysis Summary:

1. **Code and Data**:
    - The dataset includes an `Age` column that is categorized into **age groups**:
        - **Teen, Young Adults, Middle-Aged Adults, and Old**.
    - The **frequency of purchases** is calculated by summing up the purchase counts for each age group.
2. **Insights from the Table**:
    - A portion of the dataset is displayed, showing:
        - The individual ages and their corresponding `Age_category` classification.
    - The categories include **Young Adults (e.g., age 19)** and **Old (e.g., age 52)**.
3. **Visualization**:
    - A **Sunburst Chart** is used to display the purchase frequency distribution:
        - The outermost layer segments purchase behavior (e.g., annually, monthly, weekly) across different age groups.
        - **Insights**:
            - Older customers (`Old`) seem to have a prominent share across all purchase frequencies.
            - Middle-aged adults also show significant participation.

4.2 GitHub Link for Code:

https://github.com/AkashSeervi2003/shopping-trends

CHAPTER 5

Discussion and Conclusion

5.1 Future Work:

- **Improvement**
  - Future work could involve incorporating a more extensive dataset from multiple regions or platforms to capture broader shopping trends and behaviors. This would enhance the generalizability of the insights.
  - To better understand shopping trends, external factors such as economic indicators, seasonal holidays, and promotional campaigns could be integrated into the analysis.
  - The use of advanced machine learning models, such as clustering algorithms or predictive analytics, could uncover deeper insights into customer segmentation and future shopping behavior.
- **Real-Time Trend Monitoring**:
  - Developing a real-time dashboard to monitor shopping trends dynamically could provide actionable insights for businesses, allowing for quick adaptation to changes in customer preferences.
- **Limitations**:
  - While the project focused on analyzing historical data, future work could address the challenge of detecting anomalies or identifying emerging trends in live datasets.

5.2 Conclusion:

Restate the Objectives:

- The objective of this project was to analyze shopping data to identify patterns and trends in customer behavior. By leveraging data analysis techniques, the project aimed to provide actionable insights for decision-making.

Key Findings:

- The analysis revealed key insights, such as customer preferences during specific seasons, popular payment methods, and the impact of promotional codes on spending. These findings highlight significant patterns that businesses can utilize to optimize marketing strategies and improve customer satisfaction.

Contribution:

- This project contributes to the field of data-driven decision-making by demonstrating the potential of data analysis in understanding shopping behavior. It provides a framework that businesses can replicate to gain a competitive edge.

Broader Impact:

- The findings from this project can assist businesses in tailoring their products, promotions, and services to better meet customer needs. Additionally, understanding shopping trends can help companies enhance customer experiences and increase revenue.

In conclusion, this project successfully identifies shopping trends using data analysis, paving the way for more data-informed strategies in retail and e-commerce. With further development, this work can support businesses in staying competitive in an ever-evolving market.

REFERENCES

[1].Yuji Yoshimura, Stanislav Sobolevsky, Juan N Bautista Hobin, Carlo Ratti, Josep Blat 'Urban association rules: uncovering linked trips for shopping behavior'

[2]. Tianran Hu, Ruihua Song, Yingzi Wang, Xing Xie, Jiebo Luo 'Mining Shopping Patterns for Divergent Urban Regions by Incorporating Mobility Data'

[3]. From Wikipedia, the free encyclopedia 'Affinity analysis'

[4]. From Wikipedia, the free encyclopedia 'Frequent pattern discovery '

[5]. Exploratory Data Analysis EDA from devopedia