Customer Segmentation using DataScience

TEAM MEMBER: Sakthibala A

NM ID : aut22leai03

Customer Segmentation:

It is the process of grouping customers according to how and why they are buy products.

Problem Definition:

The problem is to implement data science techniques to segment customers based on their behavior, preferences, and demographic attributes. The goal is to enable businesses to personalize marketing strategies and enhance customer satisfaction. This project involves data collection, data preprocessing, feature engineering, clustering algorithms, visualization, and interpretation of results.

Main Objectives:

- The main goal for the customer segmentation using data science is to divide the customer base into distinct groups based on similar characteristics.
- This segment will helpful for many Business purpose.

Project Phases:

1. Data Collection
2. Data Preprocessing
3. Feature Engineering

4. Visualization
5. Interpretation of Results

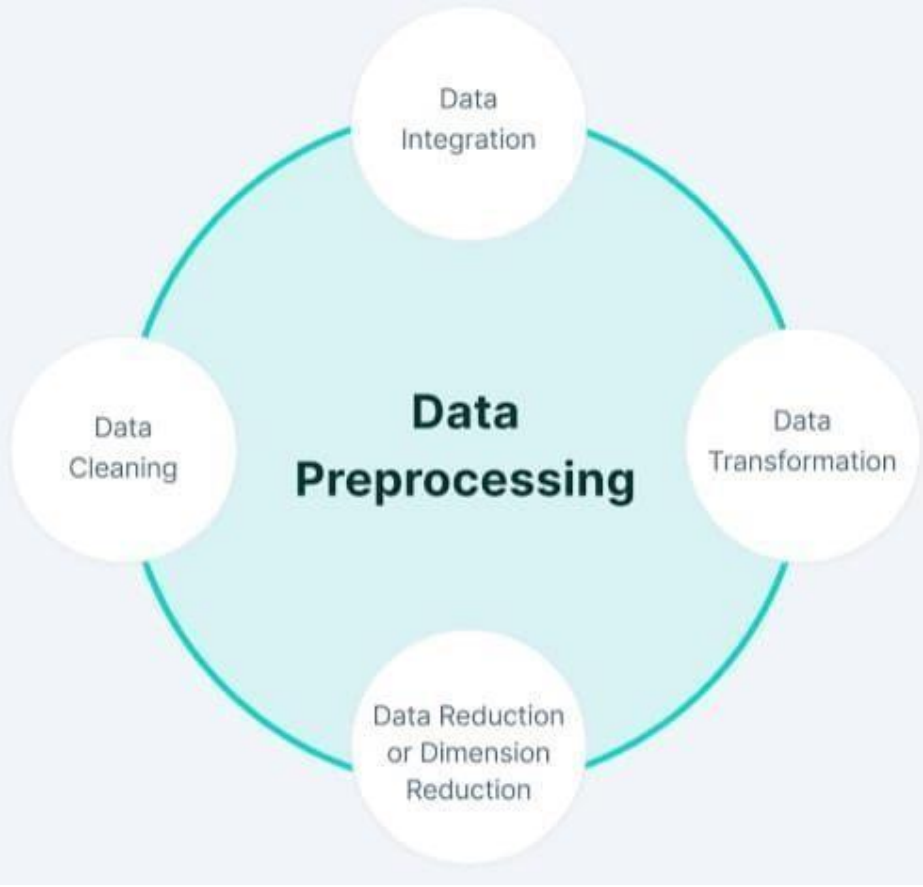
Phase 1: Data Collection

- The main goal for the data collection phase is gathering the necessary data sources for customer segmentation.
- Collecting the customer transaction data are should be include here.
- This phase also collecting the data about the behavior of the customer.
- Huge volumes of data are needed for analysis.
- Example : Considering our given Mall dataset, we have collect the each customer's annual income for predicting the spending time.

Phase 2: Data Preprocessing

- After collecting the data that should be well prepared and clean for analysis.
- Handling the missing values, outliers, and data inconsistencies are should be including here.
- It transform the data through scaling, encoding categorical variables, and feature engineering.
- It integrate the data from different sources.
- Data cleaning, Data integration, Data transformation and Dimension reduction are the important factors in Data Preprocessing.

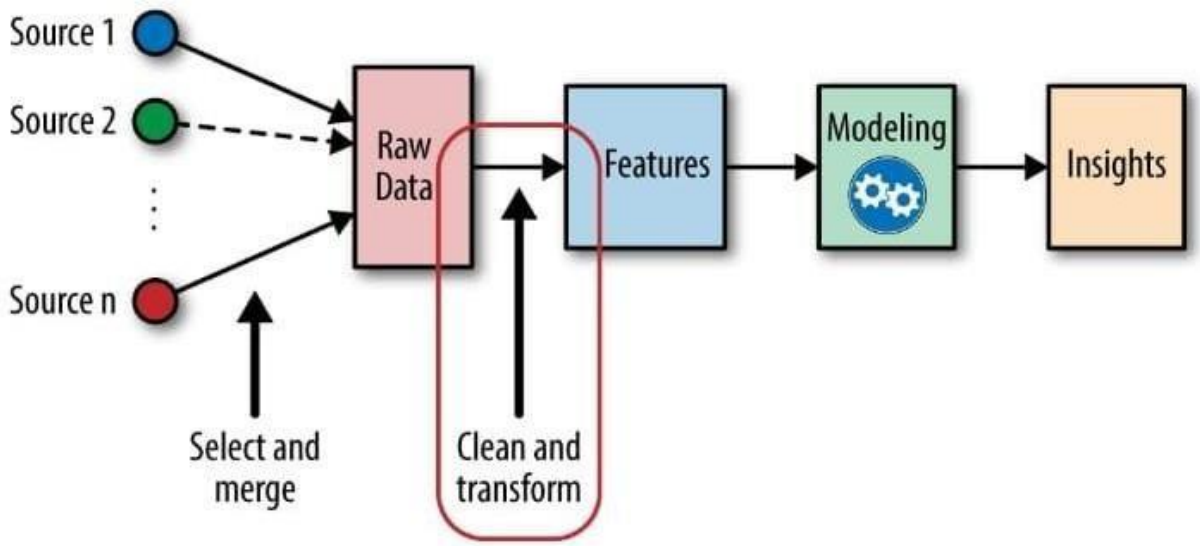
- Example: Considering Mall dataset we have to find out if any outliers or any missing values. The perfect data will output a perfect result.



Phase 3: Feature Engineering

- The main objective of the feature engineering stage is to create the relevant features that capture customer behavior and preferences.

- It also generate new features based on customer interactions and demographics.
- It reduce the dimensionality in case of the necessary situation.



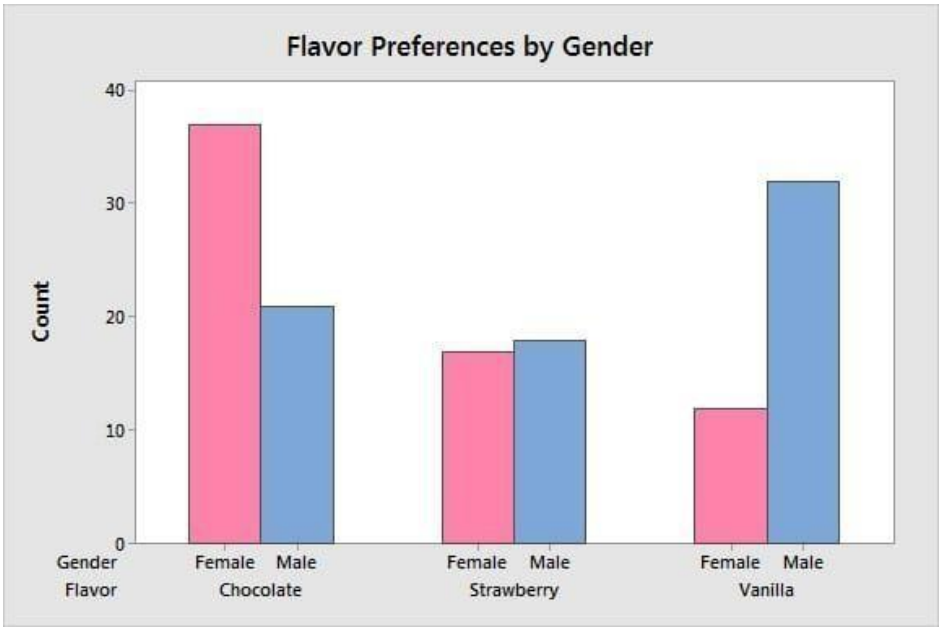
Phase 4: Visualization

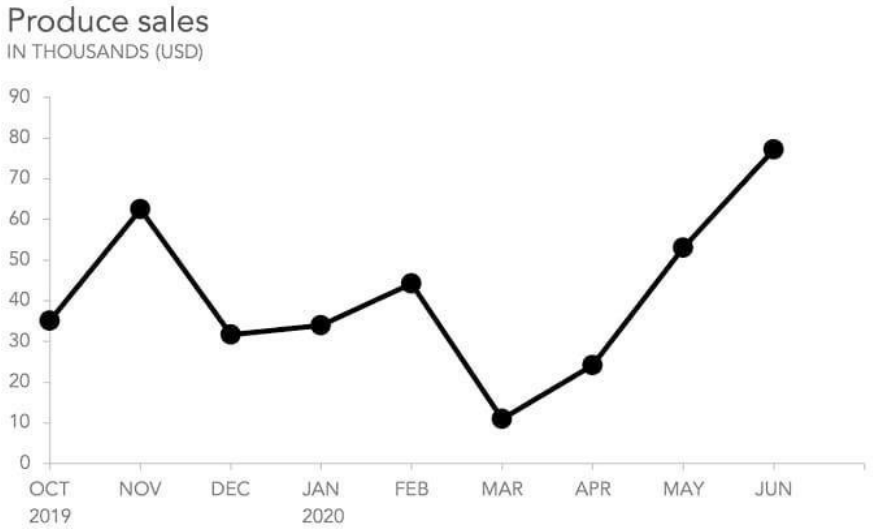
- Visualization is one of the key concept in data science which can be used for give the pictorial or virtual representation about the data.
- Several plots are used for visualize the customer segments.
- Plots example:

1. Bar Chart
2. Scatter Plot
3. Pie Plot
4. Line Plot
5. Histogram

- In python, Matplotlib library used for the visualization.
- Using these charts we can clearly virtualize our Mall Dataset especially Bar chart is used in popularly for virtualize the dataset.
- Syntax,

" Import matplotlib.pyplot as plt "





Phase 5: Interpretation of Results

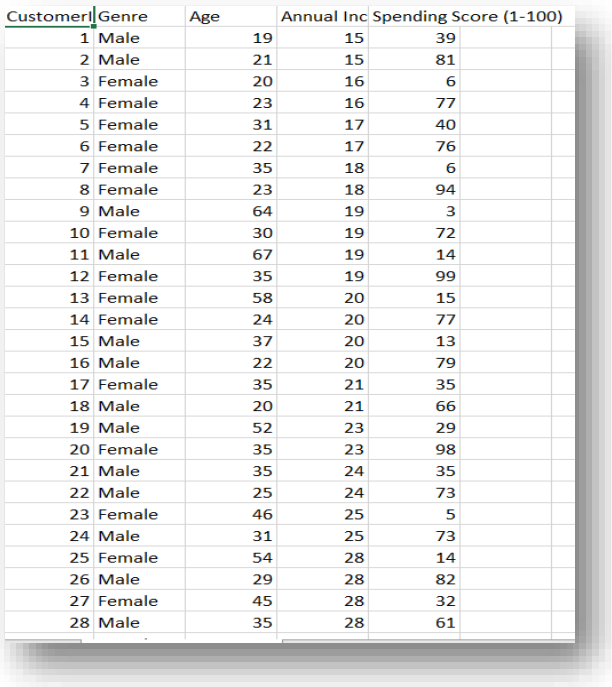
- The goal of this phase is to interpret customer segments and derive actionable insights.
- It identifies the distinguishing characteristics for each segment.
- Profile each customer segment regarding behavior, preferences, and demographics.
- It also formulates the personalized marketing strategies for each segment.

Dataset:

Mall Customers Dataset

Considering the mall Customers dataset we can analysis the customer's spending time.

Dataset link :(https://www.kaggle.com/datasets/akram24/mall-customers)

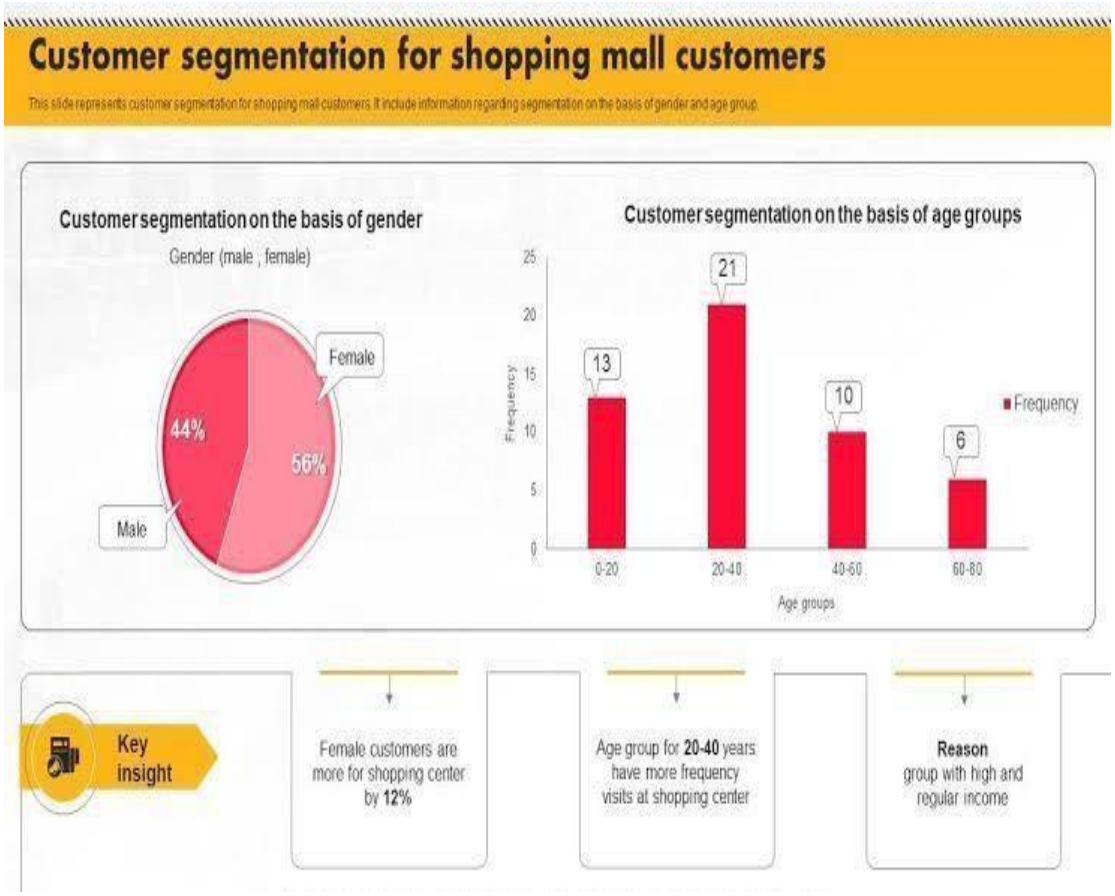


CustomerI	Genre	Age	Annual Inc	Spending Score (1-100)
1	Male	19	15	39
2	Male	21	15	81
3	Female	20	16	6
4	Female	23	16	77
5	Female	31	17	40
6	Female	22	17	76
7	Female	35	18	6
8	Female	23	18	94
9	Male	64	19	3
10	Female	30	19	72
11	Male	67	19	14
12	Female	35	19	99
13	Female	58	20	15
14	Female	24	20	77
15	Male	37	20	13
16	Male	22	20	79
17	Female	35	21	35
18	Male	20	21	66
19	Male	52	23	29
20	Female	35	23	98
21	Male	35	24	35
22	Male	25	24	73
23	Female	46	25	5
24	Male	31	25	73
25	Female	54	28	14
26	Male	29	28	82
27	Female	45	28	32
28	Male	35	28	61

Given data,

1. Customer_Id
2. Age
3. Annual Income
4. Spending Score

- If the customer’s annual income will increases then they are spending more time comparing to before.
- Considering our dataset, female having a majority for spending more time in a mall.



Data Preprocessing :

Data preprocessing is a important step in data analysis and machine learning. It involves cleaning, integrating, and transforming raw data into a format suitable for analysis or model training. Data preprocessing is essential because real-world data is often incomplete and inconsistent. By preparing the data properly, we can improve the quality and reliability of our analysis or machine learning models.

Techniques :

1) Data Cleaning
2) Data Transformation
3) Data Reduction

Data Cleaning :

- Data cleaning is also known as data cleansing or data scrubbing. It is the process of identifying and correcting errors, inconsistencies, and inaccuracies in a dataset. It is a crucial step in data preprocessing, necessary to ensure that the data used for analysis or machine learning is accurate, reliable, and free from noise.
- **Handling Missing Data:** Identifying and dealing with missing values in the dataset. This can involve filling in missing values, removing rows or columns with too many missing values
- **Removing Duplicates:** Identifying and removing duplicate records or observations from the dataset.

Data Transformation :

Data transformation is the context of data analysis, refers to the process of converting data from one format, structure, or representation into another. This transformation is performed to make the data suitable for analysis, reporting, visualization, or modeling.

Data Reduction :

Data reduction is a process used in data analysis and data mining to decrease the volume but produce the same or similar analytical results.It is also called as “Dimensionality Reduction”.

Exploratory Data Analysis:

Exploratory Data Analysis (EDA) for customer segmentation in a mall dataset involves first importing and cleaning the data. Then, it requires summarizing key variables and employing data visualization techniques to understand customer characteristics. EDA helps identify patterns, correlations, and outliers, which are essential for informed

decision-making. Subsequently, relevant variables are chosen for segmentation, and clustering techniques are applied to create customer segments. These segments are then interpreted to understand the distinct characteristics of each group. EDA facilitates insights into shopping behavior and assists in tailoring marketing strategies and store layouts to better meet the diverse needs of mall customers, enhancing business outcomes.

Key findings :

Customer segmentation yields valuable insights for businesses, including the recognition of diverse customer profiles and distinct behavior patterns. By studying demographic data, companies can uncover surprising trends and preferences within specific groups. This segmentation also provides insights into customer loyalty and churn rates, enabling the implementation of targeted retention strategies. Identifying cross-selling opportunities, optimal marketing channels, and the need for customized content enhances customer engagement and maximizes the potential for increased sales. Furthermore, product development can be guided by segment-specific preferences, potentially leading to market expansion and a competitive advantage. Customer segmentation ultimately empowers businesses to allocate resources more efficiently and to better understand and serve their customers' unique needs and desires.

Insight :

Customer segmentation through data science is a process that involves collecting and analyzing customer data to divide them into distinct groups or segments. It begins with the collection of customer information from various sources and then cleaning and selecting the most relevant details. Utilizing clustering algorithms, customers with similar characteristics are grouped together. These groups are then evaluated to ensure their quality, and profiles are created to describe each segment. The real power of customer segmentation lies in its ability to enable targeted marketing efforts, personalized experiences, and proactive customer

retention strategies. By continuously monitoring and adjusting these segments, businesses can enjoy benefits such as a deeper understanding of their customers, increased sales, cost efficiency, improved customer loyalty, reduced churn, and a competitive edge through more tailored offerings. Customer segmentation empowers businesses to make data-driven decisions and create more meaningful relationships with their customers.

Recommendation :

Customer segmentation is a valuable strategy that involves categorizing your customer base into distinct groups based on shared characteristics or behaviors. By implementing customer segmentation, businesses can offer personalized marketing, products, and services tailored to the unique preferences of each group. This not only leads to more effective and efficient marketing efforts but also boosts customer satisfaction and loyalty. By continuously analyzing and adapting these segments, companies can stay ahead in a competitive market and build stronger, long-lasting relationships with their customers.

conclusion :

Customer segmentation using data science on a mall dataset helps businesses understand their customers better by grouping them into distinct categories based on their preferences, behaviors, and demographics. This enables targeted marketing strategies, personalized product recommendations, and improved customer experiences, ultimately leading to increased customer satisfaction and higher profitability for the business.