

APPILED DATA SCIENCE

Project Tittle: customer Segmentation using data science

Name:Thrisha.R

Reg no:712221104025

College: park college of engineering and technology



INTRODUCTION:

- ❖ Customer segmentation using data science is a strategic approach that businesses employ to divide their customer base into distinct groups based on specific characteristics, behaviors, or preferences. This technique leverages data analysis, statistical algorithms, machine learning, and artificial intelligence to identify patterns within large datasets, allowing businesses to understand their customers better and tailor their marketing strategies accordingly.
- ❖ By employing data science techniques, businesses can gain valuable insights into customer behavior, preferences, and purchasing patterns. These insights enable businesses to create targeted marketing campaigns, improve customer engagement, enhance customer satisfaction, and optimize their overall business operations.

- ❖ There are several methods used in customer segmentation, including demographic segmentation (age, gender, income), geographic segmentation (location, region), psychographic segmentation (lifestyle, interests, values), and behavioral segmentation (purchase history, brand loyalty, online interactions). Data scientists use advanced algorithms and tools to analyze these data points and cluster customers into meaningful segments.
- ❖ Overall, customer segmentation using data science empowers businesses to make data-driven decisions, personalize their offerings, and ultimately enhance customer experiences, leading to increased customer loyalty and higher revenue.

FEATURE ENGINEERING:

- ❖ Feature engineering plays a crucial role in customer segmentation using data science. It

involves selecting, transforming, or creating relevant features (variables) from raw data to improve the performance of machine learning models and enhance the accuracy of customer segmentation. Here are some key techniques and considerations for feature engineering in customer segmentation:

- ❖ **Domain Knowledge:** Understand the business domain and customer behavior to identify meaningful features. Domain experts can provide valuable insights into which variables might be relevant for segmentation.
- ❖ **Demographic Features:** Utilize demographic information such as age, gender, income, education, and marital status. These features can provide a foundation for understanding customer characteristics.

- ❖ Geographic Features: Include geographic data like location, city, region, or climate, especially if your business targets specific geographical areas.
- ❖ Behavioral Features: Incorporate customer behavior data such as purchase history, frequency of purchases, average transaction value, browsing patterns, and interactions with products or services.
- ❖ Recency, Frequency, Monetary (RFM) Analysis: RFM analysis quantifies customer behavior by evaluating how recently a customer made a purchase, how frequently they make purchases, and how much money they spend. These metrics can be powerful features for segmentation.

- ❖ Customer Engagement: Measure customer engagement through metrics like website visits, app usage, social media interactions, and email responses. Engaged customers might form a distinct segment.
- ❖ Loyalty and Churn Indicators: Include features related to customer loyalty, such as membership duration, loyalty program participation, and churn indicators. Identifying loyal customers and those at risk of churning is essential for targeted strategies.
- ❖ Social Media Data: If applicable, incorporate social media data, including likes, shares, comments, and followers. Social media activity can indicate customer interests and preferences.

- ❖ Sentiment Analysis: Analyze customer reviews and feedback using sentiment analysis techniques. Positive, negative, or neutral sentiments can serve as valuable features for understanding customer satisfaction.
- ❖ Feature Scaling and Transformation: Normalize or scale numerical features to ensure they have similar ranges, preventing certain features from dominating the segmentation process. Additionally, consider transforming skewed data using techniques like logarithmic transformations.
- ❖ Missing Value Handling: Address missing values in the dataset through techniques like imputation, ensuring that

missing data does not adversely affect the segmentation results.

- ❖ Interaction Features: Create interaction features by combining two or more existing features. For example, combining purchase frequency and average transaction value could provide insights into customer spending habits.

APPLYING CLUSTERING ALGORITHM:

- ❖ Applying clustering algorithms is a common technique in customer segmentation using data science. Clustering algorithms group similar data points together based on specific features, allowing businesses to identify distinct customer segments. Here are the steps to apply clustering algorithms for customer segmentation:
- ❖ Gather and preprocess relevant customer data, ensuring it is clean, consistent, and properly formatted.

- Normalize or scale the data to bring features to a similar scale, especially if you're using distance-based clustering algorithms like k-means.

❖ Selecting a Clustering Algorithm:

- Choose an appropriate clustering algorithm based on the nature of your data and business requirements. Common clustering algorithms include k-means, hierarchical clustering, DBSCAN, and Gaussian mixture models.

❖ Determining the Number of Clusters (k):

- For algorithms like k-means, you need to specify the number of clusters (k). Use techniques like the elbow method or silhouette score to find the optimal number of clusters that best represent the data's structure.

❖ Feature Selection:

- Select the relevant features for clustering. Consider using the engineered features from the

feature engineering step, focusing on those most likely to differentiate customer behavior.

❖ Applying the Clustering Algorithm:

- Apply the chosen clustering algorithm to the prepared data with the selected features and the determined number of clusters (k).

- For example, if you're using k-means, the algorithm will assign each customer to one of the k clusters based on their feature values.

❖ Interpreting the Clusters:

- Analyze the resulting clusters to understand the characteristics of each segment. Interpretation can involve looking at the mean values of features within clusters or visualizations to discern patterns.

Validate the clusters using domain knowledge or business experts. Ensure that the clusters make sense and align with the business context.

- Refine the clustering process if necessary. This might involve adjusting the features, reselecting the number of clusters, or trying a different clustering algorithm.

❖ Implementing Business Strategies:

- Once the clusters are validated and refined, businesses can implement targeted marketing strategies, personalized recommendations, and tailored customer experiences for each segment.

- Continuously monitor and update the clusters as new data becomes available to adapt to changing customer behavior patterns.

VISUALIZATION:

Visualizing customer segmentation results is essential for understanding the identified clusters and communicating the insights to stakeholders

effectively. Here are several visualization techniques commonly used in data science for visualizing customer segmentation:

❖ Scatter Plots:

- Create scatter plots for two selected features to visualize the clusters. Each point represents a customer, and the clusters can be distinguished by different colors or markers.

❖ Parallel Coordinates:

- Use parallel coordinates to visualize multivariate data in a comprehensible way. Each axis represents a feature, and lines connecting points across different axes can help identify patterns and differences between clusters.

❖ Histograms and Bar Charts:

- Create histograms or bar charts to display the distribution of a specific feature within each cluster. This visualization helps in understanding

the concentration of certain behaviors or characteristics within clusters.

❖ 3D Scatter Plots:

- If your data has more than two features, consider creating 3D scatter plots to visualize clusters in a three-dimensional space. This approach provides a more detailed view of the data distribution.

❖ Geospatial Visualizations:

- If your data includes geographic information, use maps to visualize clusters based on geographical locations. Geospatial visualizations provide insights into regional customer behavior patterns.

Remember, the choice of visualization method depends on the nature of your data and the specific insights you want to convey. Experiment with different visualization techniques to find the

cus most effective way to communicate the results of customer segmentation to your audience.

PROGRAM:

```
cluster = kmeans.fit_predict(X3)
df["label"] = cluster

from mpl_toolkits.mplot3d import Axes3D

fig = plt.figure(figsize=(20,10))
ax = fig.add_subplot(111,projection = '3d')

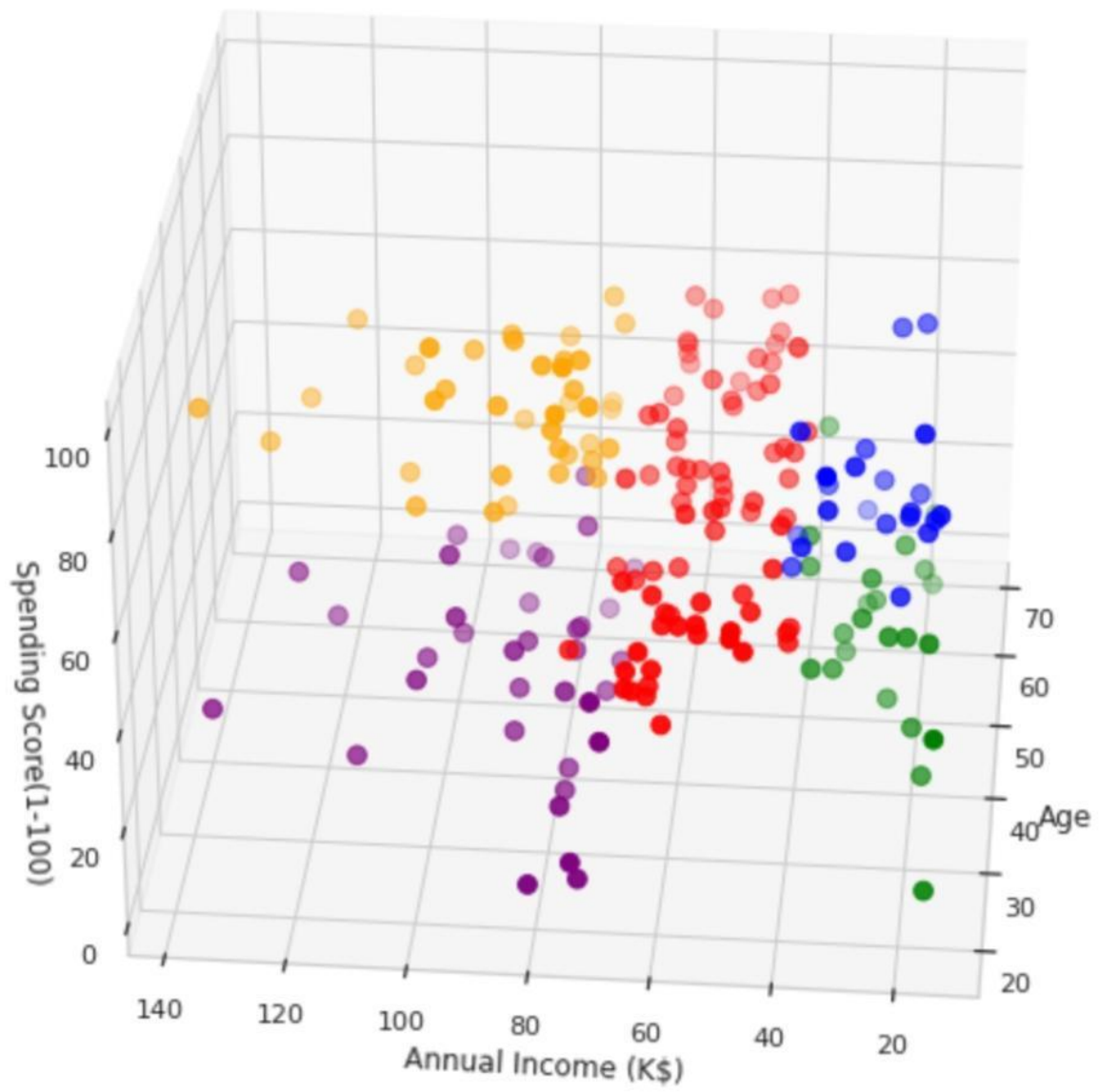
ax.scatter(df.Age[df.label == 0 ],df["Annual Income (k$)"][df.label == 0],df["Spending Score (1-100)"][df.label == 0], c = 'blue',s=60)
ax.scatter(df.Age[df.label == 1 ],df["Annual Income (k$)"][df.label == 1],df["Spending Score (1-100)"][df.label == 1], c = 'red',s=60)
ax.scatter(df.Age[df.label == 2 ],df["Annual Income (k$)"][df.label == 2],df["Spending Score (1-100)"][df.label == 2], c = 'green',s=60)
ax.scatter(df.Age[df.label == 3 ],df["Annual Income (k$)"][df.label == 3],df["Spending Score (1-100)"][df.label == 3], c = 'orange',s=60)
ax.scatter(df.Age[df.label == 4 ],df["Annual Income (k$)"][df.label == 4],df["Spending Score (1-100)"][df.label == 4], c = 'purple',s=60)

ax.view_init(30,185)

plt.xlabel("Age")
plt.ylabel("Annual Income (K$)")
ax.set_zlabel('Spending Score(1-100)')

plt.show()
```

OUTPUT:



INTERPRETATION:

Customer segmentation in data science refers to the process of dividing a company's customer base into distinct groups based on specific characteristics such as demographics, behavior, or purchasing patterns. This segmentation allows businesses to better understand their customers, target their marketing efforts, and enhance overall customer satisfaction. Data science techniques play a crucial role in this process by analyzing large datasets and extracting meaningful insights.

Here's how data science is typically applied to interpret customer segmentation:

- ❖ **Data Collection:** Gather relevant data from various sources, such as customer profiles,

transaction history, website interactions, and social media activity.

- ❖ Data Preprocessing: Cleanse and preprocess the data to handle missing values, outliers, and inconsistencies. This step ensures that the data used for segmentation is accurate and reliable.
- ❖ Feature Selection: Identify the most relevant features (variables) that will be used for segmentation. These features could include age, gender, purchase frequency, or any other relevant metrics.

- ❖ Exploratory Data Analysis (EDA): Conduct EDA to visualize the data and understand patterns, correlations, and trends within the dataset. This step helps in gaining initial insights into customer behavior.
- ❖ Segmentation Techniques: Apply various data science techniques such as clustering algorithms (like k-means clustering), decision trees, or neural networks to segment customers based on the selected features.

- ❖ These algorithms group customers with similar characteristics together.
- ❖ Interpretation: Interpret the results obtained from the segmentation models. Understand the characteristics of each segment, such as their preferences, behaviors, and needs. This interpretation helps in creating targeted marketing strategies for each segment.
- ❖ Validation: Validate the segmentation results to ensure their accuracy and reliability. This step might involve using different validation metrics depending on the technique used.

- ❖ Implementation: Implement the insights gained from customer segmentation into business strategies. This could involve personalized marketing campaigns, product
- ❖ recommendations, or tailored customer experiences for each segment.
- ❖ Monitoring and Iteration: Continuously monitor customer behavior and iteratively refine segmentation models as new data becomes available. Customer preferences and behaviors can change over time, so it's essential to adapt strategies accordingly.

