CUSTOMER SEGMENTATION USING DATASCIENCE

TEAM MEMBERS

T.TAMIL SELVAN

V.MUTHUMARIYAPPAN

K.RAHUL

R.PANDIYARAJAN

A.ARUNKUMAR

**INTRODUCTION**

Today many of the businesses are going online and, in this case, online marketing is becoming essential to hold customers, but during this, considering all customers as same and targeting all of them with similar marketing strategy is not very efficient way rather it's also annoys the customers by neglecting his or her individuality, so customer segmentation is becoming very popular and also became the efficient solution for this existing problem. Customer segmentation is defined as dividing company's customers on the basis of demographic (age, gender, marital status) and behavioral (types of products ordered for customer segmentation as its focus on individuality and we can do proper segmentation with the help of it., annual income) aspects. Since demographic characteristics does not emphasize on individuality of customer because same age groups may have different interests so behavioral aspects is a better approach

**DESIGN THINKING** :

Customer segmentation using data science is the process of dividing a customer base into distinct groups or segments based on their shared characteristics, behaviors, or preferences. This approach helps businesses tailor their marketing, product development, and customer engagement strategies to better meet the needs of each segment, leading to more personalized and effective interactions. Here's how data science can be applied to customer segmentation:

1. **Data Collection**: The first step is to gather relevant data about your customers. This can include demographic information, purchase history, website behavior, social media interactions, and more. The more data you collect, the better your segmentation can be.
2. **Data Preprocessing**: Raw data often needs to be cleaned and prepared for analysis. This involves handling missing values, dealing with outliers, and converting data into a suitable format for analysis.
3. **Feature Selection:** Not all collected data may be relevant for segmentation. Data scientists typically choose the most important features or attributes to include in the analysis. Feature selection techniques can help identify the most relevant variables.
4. **Data Analysis and Modeling:** This is where the core of data science comes into play. Various algorithms and techniques can be used for customer segmentation, including:

a. **Clustering:** Clustering algorithms like K-Means, Hierarchical Clustering, or DBSCAN can group customers with similar characteristics together. These clusters represent customer segments.

b. **Classification:** Classification algorithms, like decision trees or support vector machines, can be used to predict customer segments based on specific criteria.

c. **Dimensionality Reduction:** Techniques like Principal Component Analysis (PCA) can reduce the dimensionality of the data, making it easier to visualize and work with.

d. **Neural Networks:** Deep learning methods, such as neural networks, can be used for more complex segmentation tasks when the data is high-dimensional or the relationships are nonlinear.

1. **Model Evaluation:** It's important to assess the quality of the segmentation model. Common evaluation metrics include silhouette score, Davies-Bouldin index, and others depending on the algorithm used.
2. **Interpretation:** Once the segments are identified, it's crucial to interpret what each segment represents. What are the common characteristics of customers in each segment? This information is vital for crafting marketing strategies and personalized campaigns.
3. **Implementation:** Use the identified segments to tailor marketing messages, product recommendations, and customer experiences to the specific needs and preferences of each segment. This can involve creating targeted ad campaigns, product offerings, and communication channels.
4. **Continuous Improvement:** Customer segments are not static; they can change over time. Regularly update and refine your segmentation model as new data becomes available to ensure it remains relevant.
5. **Privacy and Data Ethics:** It's essential to handle customer data ethically and ensure compliance with data privacy regulations (e.g., GDPR). Customer data should be anonymized and secured to protect individual privacy.

**PHASE OF DEVELOPMENT:**

**Phase 1: Problem Definition and Design Thinking**

In this part We will understand the problem statement and created a document on what we have understood and we proceed ahead with solving the problem. We think on a design and present in form of the document.

**Phase 2: Innovation**

In this section we put our design into innovation to solve the problem and Created a document around it.

**Phase 3: Development Part 1**

In this section begin building our project by loading and preprocessing the dataset.

**Phase 4: Development Part 2**

In this section continue building the project by performing different activities like feature engineering, model training, evaluation etc .

**Phase 5: Project Documentation & Submission**

In this section we document the completed project and prepare it for submission.

**DATA COLLECTION:**

**DATASET:**

The data set for customer segmentation using data science typically includes a wide range of information about your customers, which can be both demographic and behavioral. The specific data variables and the volume of data you collect depend on the goals of your segmentation and the sources of your data. Here are some common types of data that might be included in a customer segmentation data set:

1. **Demographic Data:**
   * AGE: The age of the customer.
   * GENDER: The customer's gender.
   * LOCATION: Geographic information such as city, state, or country.
   * INCOME: Customer income or income bracket.
   * EDUCATION: Educational level or qualifications.
2. **Behavioral Data:**
   * PURCHASE HISTORY: Information about past purchases, including product categories, frequency, and monetary value.
   * WEBSITE BEHAVIOR: Data related to online interactions, such as pages viewed, time spent on the website, and click-through rates.
   * EMAIL INTERACTION: Data on email open rates, click-through rates, and responses to marketing emails.
   * SOCIAL MEDIA ACTIVITY: Engagement on social media platforms, including likes, shares, comments, and posts.
   * CUSTOMER SUPPORT INTERACTIONS: Data on customer service inquiries, complaints, and resolutions.
   * LOYALTY PROGRAM PARTICIPATION: Information about customers enrolled in loyalty programs and their activity within these programs.
3. **Psychographic Data:**
   * LIFESTYLE: Hobbies, interests, and activities.
   * VALUES: Beliefs, attitudes, and principles.
   * PERSONALITY TRAITS: Customer personality characteristics.
4. **Purchase Intent Data:**
   * SHOPPING CART ABANDONMENT: Data on customers who add items to their cart but do not complete the purchase.
   * WISHLISTS: Products saved for future purchase.
   * PRODUCT PREFERENCES: Preferred product categories, brands, or attributes.
5. **Technographic Data:**
   * DEVICE USAGE: Information on the devices (e.g., desktop, mobile, tablet) customers use to interact with your platform.
   * BROWSING HISTORY: Websites visited, online services used, and apps downloaded.
6. **Customer Feedback:**
   * CUSTOMER SURVEYS: Data from customer satisfaction surveys, Net Promoter Score (NPS) surveys, or other feedback mechanisms.
   * REVIEWS AND RATINGS: Feedback and ratings left by customers on your products or services.
7. **Time-Stamped Data:**
   * DATE AND TIME: Timestamps for customer interactions, such as purchase dates, website visits, or email clicks. These can be used to track behavior over time.
8. **Third-Party Data:**
   * DATA FROM EXTERNAL SOURCES: Data from external providers, such as credit bureaus or data enrichment services, which can provide additional information about customers.
9. **Customer Identifiers:**
   * CUSTOMER ID OR UNIQUE IDENTIFIER: A unique identifier for each customer to associate their various interactions and data points.
10. **Geospatial Data:**
    * LATITUDE AND LONGITUDE: Geographic coordinates for location-based analysis, such as understanding the physical locations of customers.
11. **Interaction Channels:**
    * Data on how customers interact with your business, such as in-store, online, mobile app, phone, or email
    * When working with customer segmentation data, it's important to ensure data privacy and security, especially if you're collecting sensitive customer information. Additionally, data quality and accuracy are critical, as the effectiveness of your customer segmentation model depends on the quality of the input data. Data preprocessing and cleaning are often necessary to address missing values, outliers, and inconsistencies before performing segmentation analysis.

**DATA SOURCE :**

**-** Data for this dataset can be collected from various sources, including Customer segmentation using data science official datasets, publicly available movie databases, or by web scraping Customer segmentation using data science

DATA SOURCE LINK : <https://www.kaggle.com/datasets/akram24/mall-customers>

**DATA PREPROCESSING:**

Customer segmentation is a crucial task in marketing and business strategy, and data preprocessing is a fundamental step in the process of using data science for this purpose. Data preprocessing involves cleaning and preparing the data to make it suitable for analysis. Here are the key steps for customer segmentation using data science and data preprocessing:

**1. Data Collection:**

- Gather data from various sources, such as transaction records, customer profiles, surveys, website interactions, and social media. This data may include demographic information, purchase history, behavior, and more.

**2. Data Cleaning:**

- Remove or handle missing values, as missing data can affect the quality of your segmentation analysis.

- Handle outliers that can distort the segmentation results.

**3. Data Transformation:**

- Normalize or standardize numerical features to ensure that they are on the same scale.

- Encode categorical variables by converting them into numerical values using techniques like one-hot encoding or label encoding.

**4. Feature Selection:**

- Identify and select relevant features (attributes) for segmentation analysis. Feature selection can help reduce dimensionality and improve model performance.

**5. Data Exploration and Visualization:**

- Visualize data using plots and charts to gain insights into the distribution and relationships between variables. This can help you understand your customer data better.

**6. Data Preprocessing for Text and Image Data (if applicable):**

- If you have text or image data, you may need to preprocess it separately. For text data, this can include tokenization, stemming, and sentiment analysis. For image data, techniques like resizing, normalization, and augmentation may be necessary.

**7. Feature Engineering:**

- Create new features that might be more informative for segmentation. For example, you can calculate customer recency, frequency, and monetary (RFM) values for e-commerce data.

**8. Dimensionality Reduction (if necessary):**

- If you have high-dimensional data, consider using techniques like Principal Component Analysis (PCA) or t-SNE to reduce the number of features while preserving meaningful information.

**9. Data Splitting:**

- Split the dataset into training and testing sets to evaluate the performance of your segmentation models.

**10. Model Selection:**

- Choose appropriate clustering algorithms for customer segmentation, such as K-means, hierarchical clustering, or DBSCAN.

- Alternatively, you can use supervised learning algorithms for segmentation, such as decision trees, random forests, or neural networks, with the target variable being the customer segment.

**11. Model Training:**

- Train the selected clustering or classification model on the training data.

**12. Model Evaluation:**

- Assess the quality of the segmentation using appropriate evaluation metrics, such as silhouette score, Davies-Bouldin index, or confusion matrix (for classification-based segmentation).

**13. Interpretation:**

- Interpret the results of the segmentation to gain insights into different customer segments and their characteristics.

**14. Implementation:**

- Use the customer segments to tailor marketing strategies, product recommendations, or other business decisions to better meet the needs of each group.

**15. Continuous Monitoring and Refinement:**

- Continuously update and refine the segmentation model as more data becomes available and as customer behaviour changes.

Effective customer segmentation can help businesses improve customer targeting, product recommendations, and overall customer satisfaction. Data preprocessing is a critical step in this process to ensure that the data used for segmentation is accurate, relevant, and suitable for analysis.

Customer segmentation using data science often involves feature extraction as a crucial step to create meaningful and informative attributes from your data. Feature extraction helps in reducing dimensionality, enhancing the performance of segmentation models, and uncovering valuable insights about your customers. Here's a step-by-step guide on customer segmentation with a focus on feature extraction:

**Feature extraction:**

**1. Data Collection:**

Gather and compile data from various sources, including customer profiles, transaction records, online behavior, and any other relevant information.

**2. Data Cleaning:**

Clean the data by handling missing values, dealing with outliers, and ensuring data quality.

**3. Data Transformation:**

Normalize or standardize numerical features to bring them to a consistent scale. Apply one-hot encoding or label encoding to categorical variables.

**4. Data Exploration and Visualization:**

Visualize and explore the data to understand its distribution and relationships between variables. This can help you identify potential features for extraction.

**5. Feature Selection:**

Identify relevant features for segmentation and remove irrelevant or redundant ones. Feature selection can help reduce dimensionality.

**6. Feature Engineering:**

Create new features that capture meaningful information about your customers. Common techniques for feature extraction include:

a. RECENCY, FREQUENCY, AND MONETARY (RFM) ANALYSIS:

Calculate RFM values for each customer, representing how recently they made a purchase, how often they buy, and how much they spend. These can be used as features.

b. CUSTOMER BEHAVIOR FEATURES:

Create features that capture specific customer behaviors, such as the number of product views, cart additions, or social media interactions.

c. TEXT FEATURES (IF APPLICABLE):

If you have textual data, use techniques like TF-IDF (Term Frequency-Inverse Document Frequency) or word embeddings (e.g., Word2Vec or GloVe) to extract relevant information from text.

d. IMAGE FEATURES (IF APPLICABLE):

For image data, extract features using deep learning models like convolutional neural networks (CNNs) or use pre-trained models for feature extraction.

e. GEOGRAPHIC FEATURES (IF APPLICABLE):

Extract geographical features like location-based clusters, distance to physical stores, or regional preferences.

f. TIME-BASED FEATURES:

Create features related to temporal patterns, such as the time of day, day of the week, or seasonal trends in customer behavior.

**7. Dimensionality Reduction (if necessary):**

Apply dimensionality reduction techniques like Principal Component Analysis (PCA) or t-Distributed Stochastic Neighbor Embedding (t-SNE) to reduce the number of features while preserving important information.

**8. Model Selection:**

Choose an appropriate segmentation algorithm, such as K-means, hierarchical clustering, DBSCAN, or a machine learning classification algorithm, depending on your specific business goals and data.

**9. Model Training:**

Train your chosen segmentation model on the feature-engineered dataset.

**10. Model Evaluation:**

Evaluate the quality of customer segments using relevant metrics, such as silhouette score or Davies-Bouldin index for clustering algorithms, or accuracy, F1-score, or AUC for classification-based segmentation.

**11. Interpretation:**

Interpret the customer segments to gain insights into their characteristics and preferences, using the newly created features for understanding.

**12. Implementation:**

Use the customer segments and extracted features to tailor marketing strategies, product recommendations, or other business decisions to meet the needs of each segment.

Feature extraction is a critical part of customer segmentation, as it helps you uncover meaningful patterns in your data and create more effective customer segments. The choice of features and extraction techniques should align with your business goals and the nature of your data.

**MACHINE LEARNING ALGORITHM :**

1. **Introduction:**

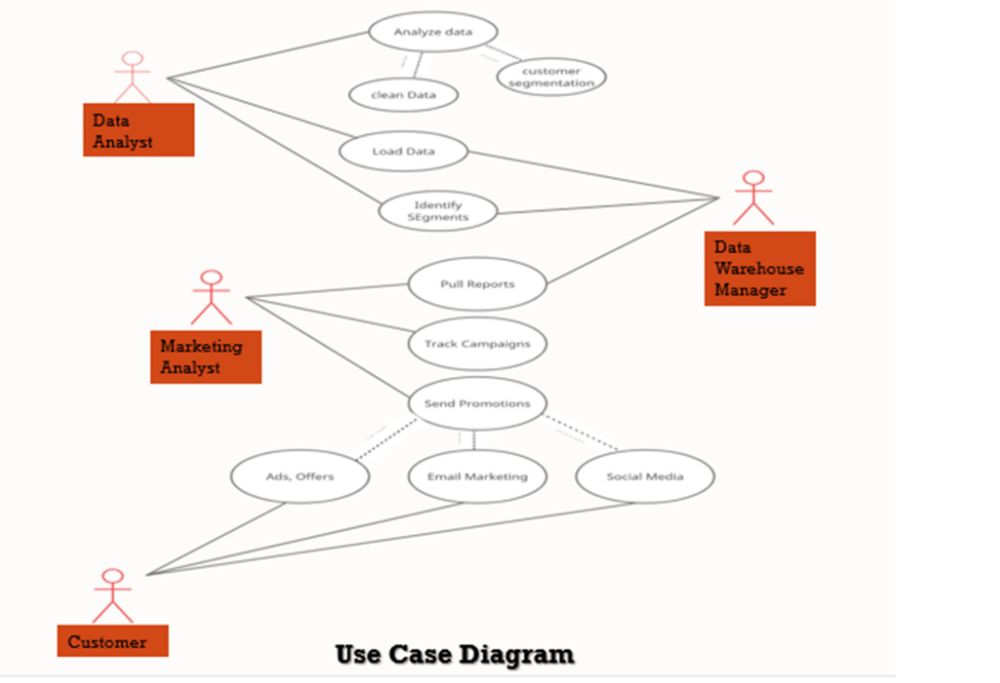
Today many of the businesses are going online and, in this case, online marketing is becoming essential to hold customers, but during this, considering all customers as same and targeting all of them with similar marketing strategy is not very efficient way rather it's also annoys the customers by neglecting his or her individuality, so customer segmentation is becoming very popular and also became the efficient solution for this existing problem. Customer segmentation is defined as dividing company's customers on the basis of demographic (age, gender, marital status) and behavioral (types of products ordered, annual income) aspects. Since demographic characteristics does not emphasize on individuality of customer because same age groups may have different interests so behavioral aspects is a better approach for customer segmentation as its focus on individuality and we can do proper segmentation with the help of it

1. **Literature Survey** :

A solution is proposed as distinguish the customers group into two groups named as premium and standard with the help of machine learning methods named as NEM, LiRM and LoRM [2].

Tushar Kansal, Suraj Bahuguna, Vishal Singh, Tanupriya Choudhury. “Customer Segmentation using K-means Clustering”, International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS).2018, In this paper customer segmentation on Telecom customers is achieved by using information such as age, interest, etc. with the help of cluster analysis method.

1. **Use Case Diagram**

****

Use case diagram of proposed system consist of 4 users 1. Data Analyst 2. Marketing Analyst 3. Data Warehouse Manager 4.

1**. Analyze Data:** analyst has the access to loaded data and analyst clean the data and perform analysis to form clusters.

**2. Load Data:** analyst log into database & view data & load into memory to work on it.

**3. Identify Segments:** analyst form report for segmented customer data and send to data warehouse and marketing analyst can access that data to form marketing strategies.

1. **Pull Reports:** marketing team can view & make edits on the reports, data for report is pulled from DW system.

**5. Track Campaigns:** The customer’s interaction tracked by marketing team for success report.

6**. Send Promotions:** Marketing team send promotions through mail, social media ads, paid ads, coupons.

1. **K-means Clustering Algorithm**

K-means Clustering is a clustering Algorithm in which we are given with data points with its data set and features and the mechanism is to categories those data points into clusters as per their similarities.

The algorithm forms K clusters based on its similarity. To calculate the similarity K-means uses Euclidean distance measurement method.

**Steps**

1. In first step, we randomly initialize k points.
2. K-means classifier categorizes each data point to its nearest mean and rewrite the mean’s coordinates.
3. Iteration is continuing up till all data points are classified.
4. **Proposed System**

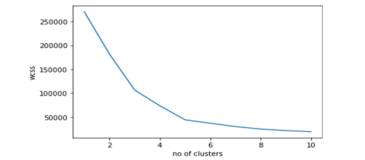
In our system we including annual income and total spending as a feature for classification in

**1. Data Gathering:** first, Data analyst fetch data required for analysis from database, format data i.e., remove all NA values from data & make data ready for processing.

**2. Feature Extraction:** Selects features which makes model more accurate, in our case features are annual income and spending score for efficient analysis.

**3. K-means Classifier:** After that, K means classifier performs clustering with respect to features provided to it,

**4.** **Hyper Parameter Tuning:** during forming groups to select optimal no of clusters we applied hyper parameter tuning which is achieved by Elbow method to choose optimal no of clusters. below graph is for elbow method which shows curve is getting flatter after 5 which indicates that 5 is optimal no of clusters we can form for better classification.

**FIG: ELB0W METHOD :** 

**5. Data Visualization:** With the formed clusters marketing team can make different strategies for better targeting customers

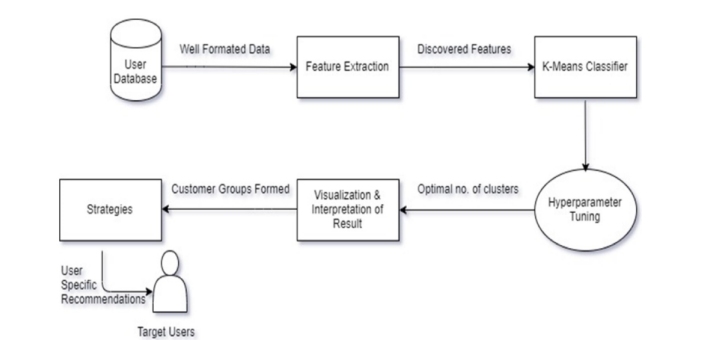


FIG: **FLOW OF OPERATION**

1. **Results**

After analysis of data and classifying customers with features annual income and spending score, we got clusters of customers & with formed clusters marketing team form strategies for customers specific recommendation to make value out of them

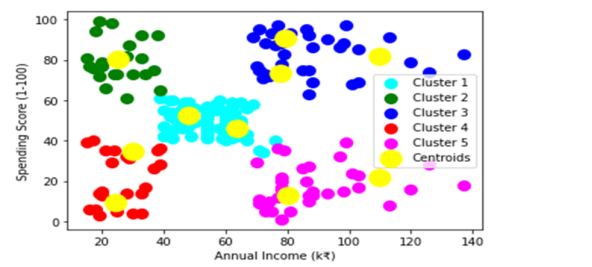
****

Fig: **FINAL CLUSTER FORMED**

**Program:**

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

**---------------------------------------------------------------------------**

**ModuleNotFoundError** Traceback (most recent call last)

[**c:\Users\dhari\AppData\Local\Microsoft\Windows\INetCache\IE\8QGR3A8L\Notebook**](file:///C:\Users\dhari\AppData\Local\Microsoft\Windows\INetCache\IE\8QGR3A8L\Notebook) **- Customer Segmentation using Data Science[1].ipynb Cell 3** line 1

**---->** [**1**](vscode-notebook-cell://c%3A/Users/dhari/AppData/Local/Microsoft/Windows/INetCache/IE/8QGR3A8L/Notebook%20-%20Customer%20Segmentation%20using%20Data%20Science%5B1%5D.ipynb#W2sZmlsZQ%3D%3D?line=0) import numpy as np

[2](vscode-notebook-cell://c%3A/Users/dhari/AppData/Local/Microsoft/Windows/INetCache/IE/8QGR3A8L/Notebook%20-%20Customer%20Segmentation%20using%20Data%20Science%5B1%5D.ipynb#W2sZmlsZQ%3D%3D?line=1) import pandas as pd

[3](vscode-notebook-cell://c%3A/Users/dhari/AppData/Local/Microsoft/Windows/INetCache/IE/8QGR3A8L/Notebook%20-%20Customer%20Segmentation%20using%20Data%20Science%5B1%5D.ipynb#W2sZmlsZQ%3D%3D?line=2) import matplotlib.pyplot as plt

**ModuleNotFoundError**: No module named 'numpy'

df=pd.read\_csv(r'C:\Users\Admin\Desktop\sivaash\Naan Mudhalvan\CustomerSegmentation-AppliedDataScience\Mall\_Customers.csv')

df.head()

| **CustomerID** | **Gender** | **Age** | **Annual Income (k$)** | **Spending Score (1-100)** |
| --- | --- | --- | --- | --- |
| 0 | 1 | Male | 19 | 15 | 39 |
| 1 | 2 | Male | 21 | 15 | 81 |
| 2 | 3 | Female | 20 | 16 | 6 |
| 3 | 4 | Female | 23 | 16 | 77 |
| 4 | 5 | Female | 31 | 17 | 40 |

#### This dataset is used in retail and marketing analytics to understand customer behavior and preferences. It includes the following types of information:

##### Customer ID

##### Gender

##### Age

##### Annual Income

##### Spending Score

A "customer ID" (Customer Identification) is a unique identifier assigned to each customer in a database or system. It is used to distinguish one customer from another and track their activities, purchases, interactions, and other relevant information.

Gender is one of the key factors in segmenting customers into distinct groups. For example, stores may tailor their product offerings and marketing strategies differently for male and female customers.

Age is a fundamental factor for segmenting customers into groups. Different age groups may have distinct preferences, shopping behaviors, and income levels. For example, retailers often distinguish between teenagers, young adults, middle-aged individuals, and seniors.

The annual income of mall customers is a crucial demographic variable that helps businesses and mall operators understand the spending capacity and shopping preferences of their customer base.

Spending score is a metric used to assess and quantify a customer's purchasing behavior within a mall.

Index(['CustomerID', 'Gender', 'Age', 'Annual Income (k$)',

'Spending Score (1-100)'],

dtype='object')

The dataset contains 5 columns:

-> CustomerID

-> Gender

-> Age

-> Annual Income

-> Spending Score

df.describe()

| **CustomerID** | **Age** | **Annual Income (k$)** | **Spending Score (1-100)** |
| --- | --- | --- | --- |
| count | 200.000000 | 200.000000 | 200.000000 | 200.000000 |
| mean | 100.500000 | 38.850000 | 60.560000 | 50.200000 |
| std | 57.879185 | 13.969007 | 26.264721 | 25.823522 |
| min | 1.000000 | 18.000000 | 15.000000 | 1.000000 |
| 25% | 50.750000 | 28.750000 | 41.500000 | 34.750000 |
| 50% | 100.500000 | 36.000000 | 61.500000 | 50.000000 |
| 75% | 150.250000 | 49.000000 | 78.000000 | 73.000000 |
| max | 200.000000 | 70.000000 | 137.000000 | 99.000000 |

df.isnull().sum()

CustomerID 0

Gender 0

Age 0

Annual Income (k$) 0

Spending Score (1-100) 0

dtype: int64

df[df.duplicated()]

|  | **CustomerID** | **Gender** | **Age** | **Annual Income (k$)** | **Spending Score (1-100)** |
| --- | --- | --- | --- | --- | --- |

df.nunique()

CustomerID 200

Gender 2

Age 51

Annual Income (k$) 64

Spending Score (1-100) 84

dtype: int64

df.columns

Index(['CustomerID', 'Gender', 'Age', 'Annual\_Income', 'Score'], dtype='object')

df.Gender.value\_counts()

Gender

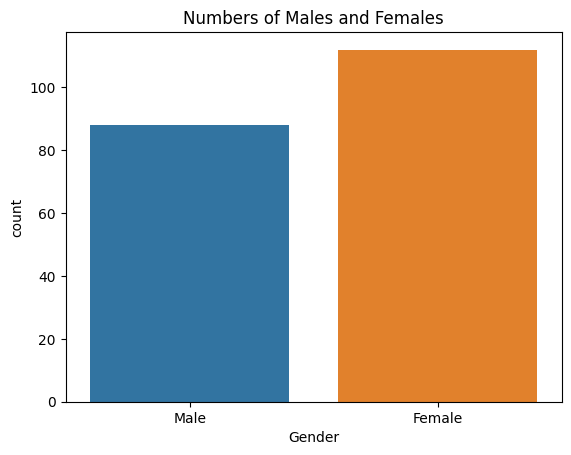
Female 112

Male 88

Name: count, dtype: int64

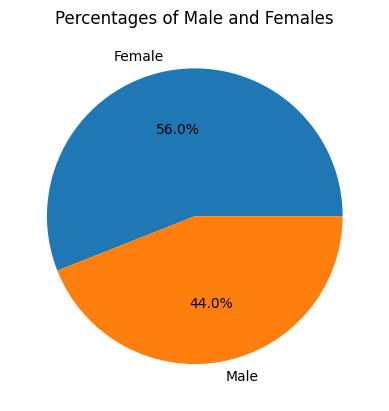
sns.countplot(x = 'Gender', data = df, hue = 'Gender')

plt.title("Numbers of Males and Females");

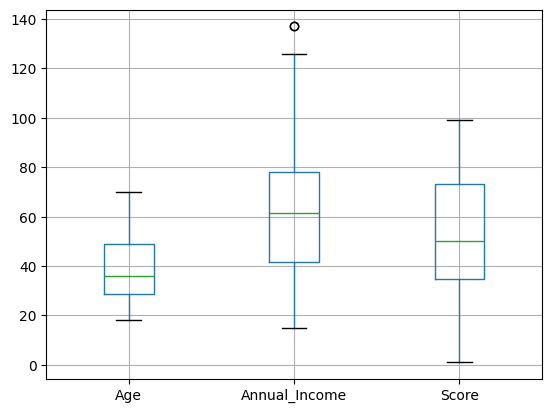


plt.pie(df.Gender.value\_counts(), labels = ['Female', 'Male'], autopct ="%.01f%%")

plt.title('Percentages of Male and Females' );



df.iloc[:,1:].boxplot();



plt.figure(figsize = (15,3))

plt.subplot(1,3,1)

sns.histplot(df.Age, kde = True)

plt.title("Age")

plt.xlabel("Age");

plt.subplot(1,3,2)

sns.histplot(df.Annual\_Income, kde = True)

plt.title("Annual Income")

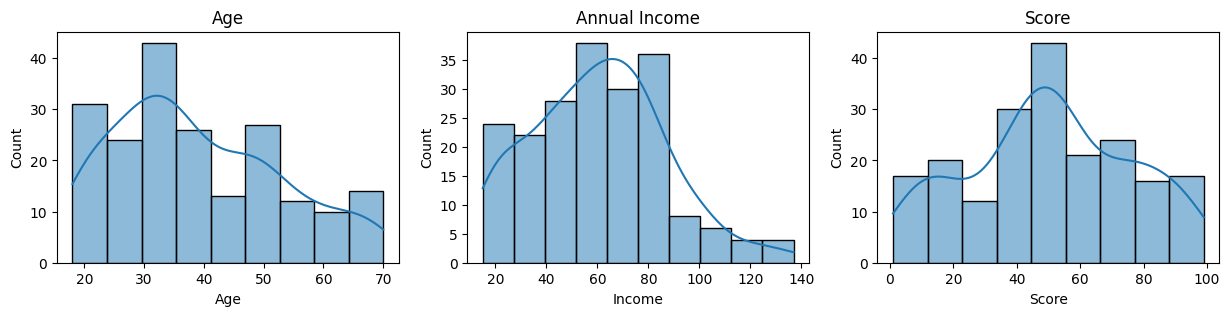
plt.xlabel("Income");

plt.subplot(1, 3, 3)

sns.histplot(df.Score, kde = True)

plt.title("Score")

plt.xlabel("Score");



male = df[df.Gender == "Male"]["Age"]

female = df[df.Gender != "Male"]['Age']

plt.figure(figsize = (10,4))

plt.subplot(1,2,1)

sns.histplot(male, color='#0066ff', bins = range(15,75,5), kde = True)

plt.title("Male age distribution ")

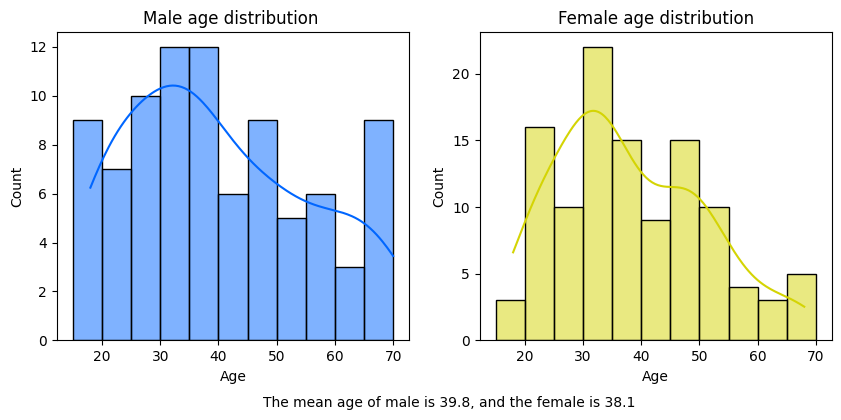
plt.subplot(1,2,2)

sns.histplot(female, color = '#D4D404', bins = range(15,75,5), kde = True)

plt.title("Female age distribution");

plt.text(-25,-5,f"The mean age of male is {round(male.mean(),1)}, and the female is {round(female.mean(),1)}")

plt.show()



income\_male = df[df.Gender == "Male"]["Annual\_Income"]

income\_female = df[df.Gender != "Male"]["Annual\_Income"]

plt.figure(figsize = (15, 4))

plt.subplot(1,3,1)

sns.histplot(income\_male, color = '#A8D10E')

plt.title("Annual Income of Male")

plt.xlabel("Annual income (k$)")

plt.subplot(1,3,2)

sns.histplot(income\_female, color = '#1DE1B8')

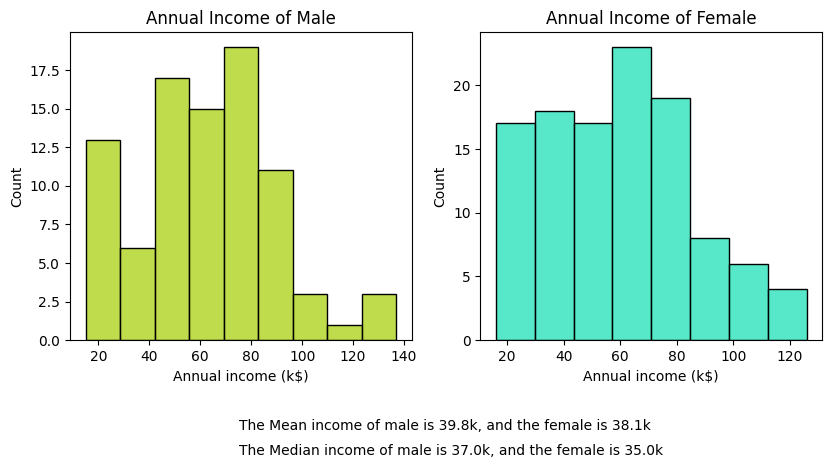
plt.title("Annual Income of Female")

plt.xlabel("Annual income (k$)")

plt.text(-75,-7,f"The Mean income of male is {round(male.mean(),1)}k, and the female is {round(female.mean(),1)}k")

plt.text(-75,-9,f"The Median income of male is {round(male.median(),1)}k, and the female is {round(female.median(),1)}k")

plt.show()



score\_male = df[df.Gender == "Male"]["Score"]

score\_female = df[df.Gender != "Male"]["Score"]

plt.figure(figsize = (10,4))

plt.subplot(1,2,1)

sns.histplot(score\_male, color = "#7CA3B1")

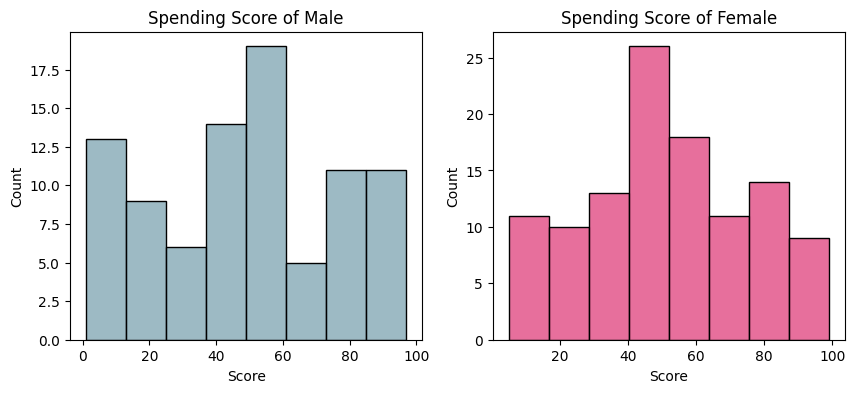
plt.title("Spending Score of Male")

plt.subplot(1,2,2)

sns.histplot(score\_female, color = '#DF3F7B')

plt.title("Spending Score of Female")

plt.show()



df['Age\_Group'] = pd.cut(df.Age, bins = [18, 25, 35, 45, 55, 65, 70], labels = ['18-24', '25-34', '35-44', '45-54', '55-64', '65-70'])

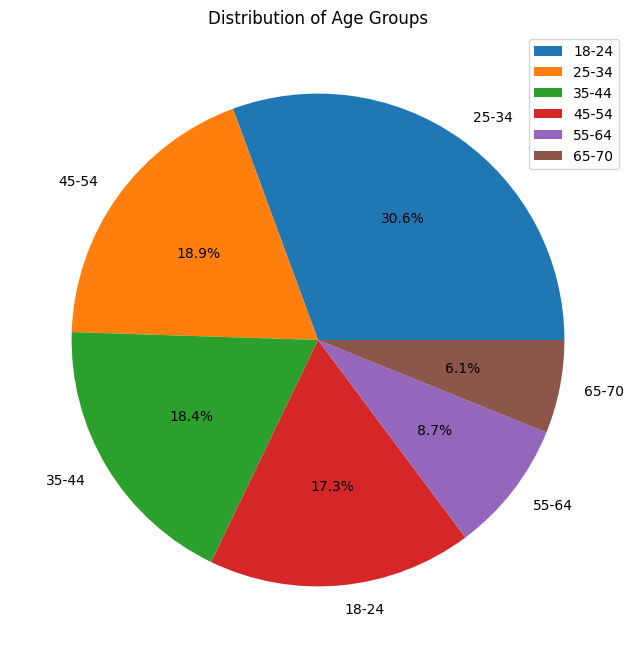
plt.figure(figsize = (8, 8))

plt.pie(df.Age\_Group.value\_counts(), labels = df.Age\_Group.value\_counts().index, autopct='%1.1f%%')

plt.title('Distribution of Age Groups')

plt.legend(['18-24', '25-34', '35-44', '45-54', '55-64', '65-70'])

plt.show()



mean\_score = df.groupby('Age\_Group')['Score'].mean()

sns.barplot(x = mean\_score.index, y = mean\_score.values, palette = 'vlag')

plt.title("Average annual spending score by Age Group")

plt.ylabel("Average Spending Score")

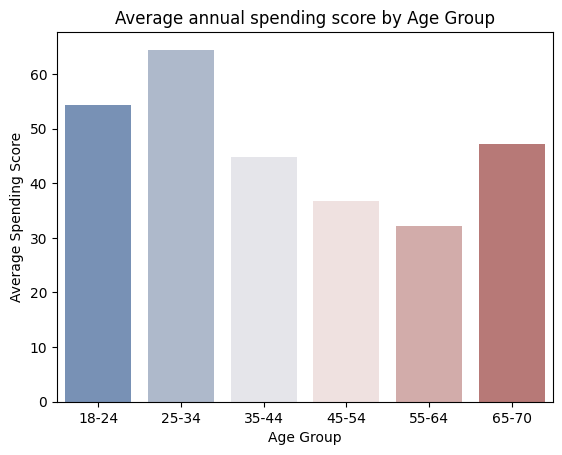
plt.xlabel("Age Group")

plt.show()

[C:\Users\Admin\AppData\Local\Temp\ipykernel\_10296\885886067.py:3](file:///C:\Users\Admin\AppData\Local\Temp\ipykernel_10296\885886067.py:3): FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

sns.barplot(x = mean\_score.index, y = mean\_score.values, palette = 'vlag')



**CONCLUSION**

Customer segmentation is performed on the company's customers data and with the help of K-means clustering machine learning algorithm customers are divided using features like total spending and annual income, this study also proves that the dividing customers on the basis of behavioral characteristics is a better solution for existing customer segmentation problem and K-means clustering algorithm is identified as a good choice for this approach.