**CUSTOMER SEGMENTATION USING DATA SCIENCE**

|  |  |
| --- | --- |
| **Date** | **31-10-2023** |
| **Team ID** | **3884** |
| **Project Name** | **Age Based Customer Segmentation Using Data Science** |
| **Team Members** | **Threka M**  **Sowendarya P K**  **Swetha S M**  **Vipin R** |

**PROBLEM STATEMENT:**

A retail company wants to improve its marketing and product targeting strategies by understanding its one crucial aspect of customer segmentation is age-based categorization, as different age groups often have distinct preferences, purchasing behaviors, and communication preferences. The company aims to use data science techniques to segment its customers into meaningful age groups.

**INTRODUCTION:**

Age-based customer segmentation leverages data science methodologies and techniques to gain insights into customer behaviour, preferences, and purchasing patterns based on age groups. This approach recognizes that different age cohorts often exhibit distinct behaviours and respond differently to marketing strategies, making it a valuable tool for businesses seeking to tailor their efforts to specific customer segments.

In this process, businesses collect and analyse customer data, including age, in conjunction with other relevant information such as demographics, purchase history, and online interactions. Through data preprocessing, clustering algorithms, and data visualization, businesses can create meaningful customer segments, each representing a specific age group or generation.

**LITERATURE SURVEY:**

**1.”Multi-Behavior RFM Model Based on Improved SOM Neural Network Algorithm for Customer Segmentation”.**

Customer classification can help companies in understanding their customers better to make personalized recommendations and promote their applications. This could enable merchants to appropriately adopt various marketing strategies according to customer characteristics. Compared to traditional RFM models with one user-item interaction behaviour, the proposed MB-RFM model included multiple user-item interaction behaviours based on an improved SOM neural network, which is beneficial for effective customer segmentation. The information obtained in this study can help in developing marketing strategies (such as pricing policies, promotion strategies, and personalized service strategies) for applications to improve the utilization rate of customers and targeted item promotion.

**2.”An Extended Regularized K-Means Clustering Approach for High-Dimensional Customer Segmentation With Correlated Variables”.**

To the fast development of e-commerce and the customers’ acquaintance with multichannel shopping mode. Various business organizations have started to work on the omnichannel issue in order to satisfy the new trend of customer demand. The RFM model and the k-means clustering method are typical approaches to segmentation of customers. We apply the proposed method to a real example, but little interpretation of the clustering results is discussed which is also important for the sense of empirical study. This can be definitely one of our future research directions.

**3.”Integrated Churn Prediction and Customer Segmentation Framework for Telco Business”.**

In the telco industry, attracting new customers is no longer a good strategy since the cost of retaining existing customers is much lower. Churn management becomes instrumental in the telco industry. As there is limited study combining churn prediction and customer segmentation, aims to propose an integrated customer analytics framework for churn management. Churn customer segmentation is then carried out using K-means clustering. Customers are segmented into different groups, which allows marketers and decision makers to adopt retention strategies more precisely.

**4.”A Group Norm Regularized Factorization Model for Subspace Segmentation**”.

Subspace segmentation assumes that data comes from the union of different subspaces and the purpose of segmentation is to partition the data into the corresponding subspace. Low-rank representation (LRR) is a classic spectral-type method for solving subspace segmentation problems, that is, one first obtains an affinity matrix by solving a LRR model and then performs spectral clustering for segmentation. It proposes a group norm regularized factorization model (GNRFM) inspired by the LRR model for subspace segmentation and then designs an Accelerated Augmented Lagrangian Method (AALM) algorithm to solve this model.

**5.”Efficient Two-Stream Network for Online Video Action Segmentation”.**

Temporal action segmentation is a task of predicting frame-level classes of untrimmed long termed videos. It can be widely used in various applications including customer analysis, data collection, video indexing, and surveillance. Learning good representations from videos such as motion and spatial representation is a critical factor for model performance. We propose a two-stream action segmentation pipeline that can learn motion and spatial information efficiently and operate online. While the temporal stream combines frame-grouping and TSM for capturing short-term dynamics and long-term temporal information at the same time, the spatial stream captures information on colour and appearance complementary to representations from the temporal stream.

**DESIGN THINKING PROCESS:**

The design thinking process for age-based customer segmentation using data science involves a structured, human-centric approach to understanding customer needs and preferences across different age groups.

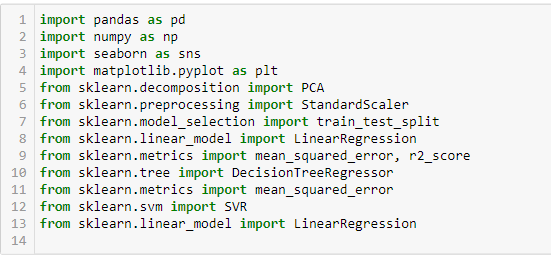
1. **Empathize:** Understand the diverse needs, behaviours, and challenges of customers across various age brackets. Gather qualitative and quantitative data to comprehend their preferences, buying patterns, and expectations.
2. **Define:** Define clear objectives for segmenting customers based on age, considering business goals and the specific insights needed. Formulate precise questions that data analysis should answer regarding different age groups.
3. **Ideate:** Brainstorm potential strategies and methods to segment customers without fixed age points. Consider various data features and attributes that might correlate with different age groups, allowing for dynamic segmentation.
4. **Prototype:** Develop data models or algorithms to categorize customers into age-based segments. Test these models against different datasets, using machine learning or statistical techniques to derive meaningful segmentations.
5. **Test:** Evaluate the effectiveness of the segmentation models. Analyse whether the segments created accurately represent distinct age groups and if they provide actionable insights for marketing, product development, or service improvements.
6. **Iterate:** Refine the segmentation models based on feedback and new data. Continuously improve the models to better align with the actual needs and behaviour of different age groups.

This design thinking process ensures a customer-centered approach to age-based segmentation, leveraging data science techniques to create dynamic, relevant, and actionable customer segments.

**PHASES OF DEVELOPMENT:**

**Phase 1: Importing Dependencies**

This phase involves necessary python libraries and modules . These libraries are required for data processing ,visualization and some tasks.

****

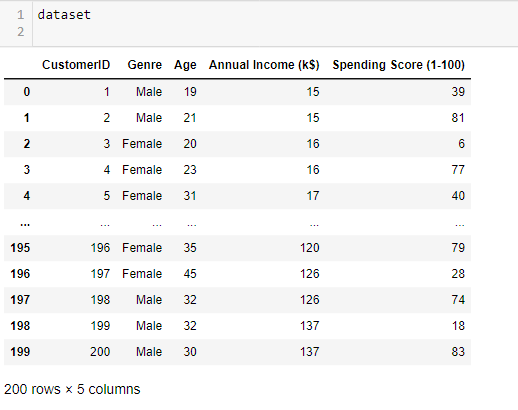
**Phase 2: Reading and Viewing Data**

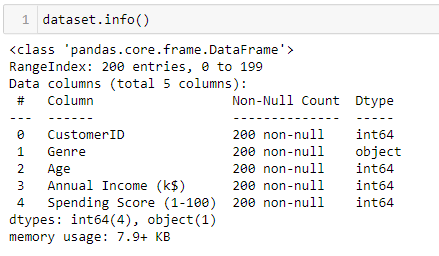
In this phase, the code reads a dataset from a CSV file named ‘Mall\_Customers.csv’ using Pandas.

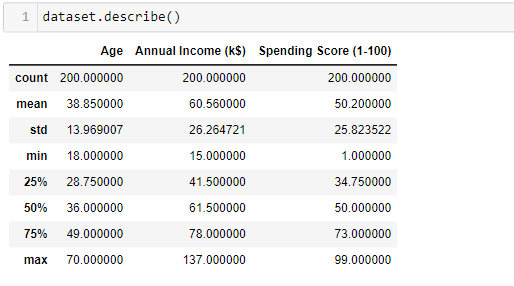
****

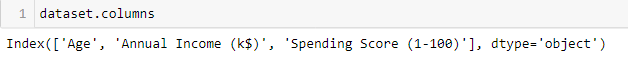
**Phase 3: Data Pre processing and Exploration**

Here, Data Pre processing and exploration is done



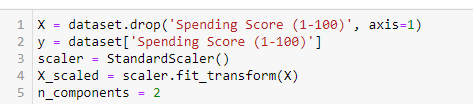


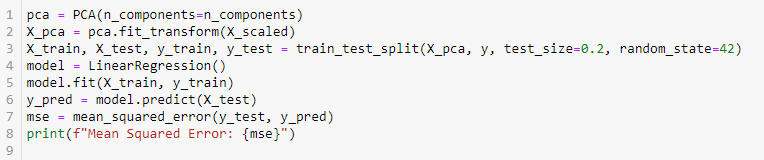


****

**Phase 4: Model Training**

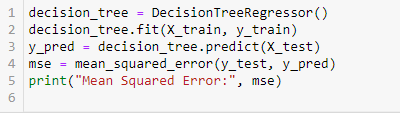
This phase includes data scaling, splitting the dataset into training and testing sets, and preparing the features and target values for customer segmentation models.

****

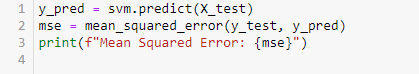
****

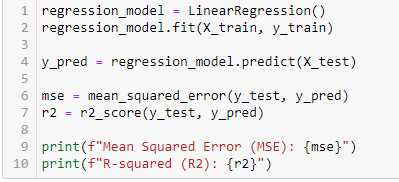
**Phase 5:Model Selection and Evaluation**

In this final phase, Age-based customer segmentation using data science, the model selection could involve algorithms such as K-means clustering or decision trees, which effectively group customers based on age-related patterns. Evaluation of these models might employ metrics like silhouette score for clustering or accuracy, precision, and recall for decision trees, assessing the model's ability to accurately segment customers into distinct age groups, aiding in targeted marketing strategies and personalized approaches.







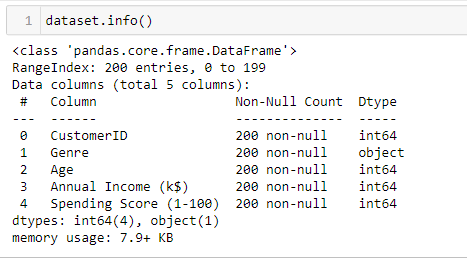


**DATASET USED:**

There are many datasets available online for customer segmentation using data science, in our analysis we used this dataset from Kaggle.comto train our model.

**Dataset Link:**[**https://www.kaggle.com/datasets/akram24/mall-customers**](https://www.kaggle.com/datasets/akram24/mall-customers).

Age-based customer segmentation using data science involves the process of categorizing customers into different groups or segments based on their age-related characteristics and behaviors. Data science techniques are employed to analyse customer data, such as age demographics, purchase history, online behaviour, and preferences, to create meaningful and distinct segments representative of various age groups. The goal is to gain insights into how different age cohorts behave, what products or services they prefer, and how their needs and behaviours vary, enabling businesses to tailor their marketing strategies, product development, and services to better suit the specific requirements of each age segment.



**DATA PREPROCESSING & FEATURE ENGINEERING:**

In the context of age-based customer segmentation using data science, the initial phase of data preprocessing involves several essential steps. This process includes cleaning the dataset to address missing or inaccurate values in the age-related information, ensuring data uniformity and accuracy for subsequent analysis. Additionally, data normalization or scaling may be applied to make age-related features comparable and consistent, especially when combined with other customer attributes.

Feature extraction focuses on identifying or engineering pertinent features related to age-based segmentation. This might involve transforming age data into categorical age groups or creating new features that encapsulate age-related behaviour patterns. For instance, one could derive features related to 'age range preference for products' or 'frequency of purchases based on age categories'. This phase aims to prepare the data by refining and extracting features that are crucial for robust age-based segmentation models, facilitating the subsequent application of data science techniques for meaningful customer group categorization.

**CHOICE OF ALGORITHM / Techniques:**

**1.PCA Algorithm**

Age-based customer segmentation using data science can benefit from the application of Principal Component Analysis (PCA). PCA is employed to reduce the dimensionality of age-related data while preserving its essential information. By transforming age-related variables into a smaller set of principal components, it enables a clearer understanding of age patterns and their influence on customer behaviour. This streamlined data can then be used for clustering or classification models to effectively segment customers by age, ultimately improving marketing strategies and product recommendations tailored to specific age groups.

**2.Decision Tree Regressor**

In the realm of age-based customer segmentation using data science, employing the Decision Tree Regressor algorithm proves beneficial. This algorithm partitions the customer dataset based on age-related features, forming a tree-like structure to predict or segment customers into distinct age groups. Decision trees split the data based on various features such as purchase history, browsing behaviour, or demographic information to predict the age range of customers. The evaluation of this model involves assessing its accuracy in predicting age, utilizing metrics like mean squared error or R-squared to gauge the model's effectiveness in accurately segmenting customers by age, enabling businesses to tailor personalized marketing strategies and services for different age groups.

Top of Form

**3.** **Support vector machine algorithm**

In age-based customer segmentation using data science, Support Vector Machine (SVM) algorithms can be employed to delineate distinct customer segments based on age-related patterns. SVM, a supervised learning model, can effectively classify customers into different age groups by identifying complex relationships within the data. By utilizing features related to customer behaviour, preferences, or purchase history, SVM can create a boundary that best separates various age clusters, allowing for accurate segmentation and targeted marketing strategies. Additionally, SVM's ability to handle non-linear data and its robustness in high-dimensional spaces makes it a valuable tool for this segmentation task, enabling businesses to tailor their services and marketing approaches to specific age demographics more precisely. Eval

**4.** **Linear Regression**

In age-based customer segmentation utilizing the linear regression algorithm, the focus is on predicting customer behaviours or preferences based on age-related patterns. Linear regression can help identify correlations between age and various customer traits, allowing businesses to forecast potential buying patterns or preferences across different age groups. By analys ing historical data, this method estimates how age influences certain characteristics, aiding in tailoring marketing strategies, product development, or service customization for specific age brackets. The model helps identify trends and relationships, enabling businesses to make informed decisions for more targeted and effective customer segmentation strategies

**The differences in the performance of the algorithm may be attributed to several factors, including the quality of data preprocessing and feature extraction.**

**PCA Algorithm:**

The Principal Component Analysis (PCA) algorithm works by transforming a dataset into a new coordinate system, aiming to identify the directions (principal components) where the data varies the most. It achieves this by finding the eigenvectors of the covariance matrix of the data and using them as the new axes. These eigenvectors represent the directions of maximum variance in the data, and the corresponding eigenvalues quantify the variance along each eigenvector. PCA then ranks these components by their explained variance and allows for dimensionality reduction by retaining the most significant components while eliminating redundant information

**Decision Tree:**

The efficacy of how the data is cleaned, normalized, and the relevant features extracted greatly impacts the decision tree's ability to accurately predict or segment customers by age. Improved preprocessing, such as handling missing values, scaling features, and selecting pertinent attributes, significantly influences the model's performance by enhancing its ability to discern age-related patterns within the data.

**Support vector machine:**

Support Vector Machine (SVM) algorithm, the accuracy and effectiveness in customer segmentation by age heavily rely on the appropriate handling of data normalization, feature scaling, and extraction techniques. A well-pre processed dataset with relevant features can significantly impact SVM's ability to distinguish age groups accurately. Optimal feature selection and transformation are pivotal for SVM to create an effective decision boundary, which is crucial for precise age-based customer segmentation in data science applications.

**Linear Regression:**

In the context of the linear regression algorithm, the quality of preprocessing steps such as handling missing values, scaling features, and selecting pertinent features significantly impacts its performance. Effective feature extraction techniques, like PCA or feature engineering, can enhance the algorithm's predictive capability, potentially resulting in more accurate age-based customer segmentation and improved insights into different age groups' behaviours and preferences.

**CONCLUSION:**

In the age-based customer segmentation using data science, this approach offers a valuable framework for businesses aiming to understand and cater to the diverse needs of customers across various age groups. Through the application of data science techniques, such as clustering or classification models, insights into the distinct behaviours, preferences, and purchasing patterns of different age segments can be uncovered. This allows companies to tailor their marketing strategies, product development, and service offerings more effectively, addressing the unique requirements of specific age cohorts.

By employing a customer-centric approach and leveraging data-driven insights, businesses can enhance customer satisfaction, foster personalized experiences, and optimize their operations to better serve the evolving demands of customers across different age brackets. Overall, age-based customer segmentation using data science proves instrumental in fostering more targeted and tailored approaches in a dynamic and ever-evolving market landscape.