Customer Segmentation

- Kasyap Velampalli

Problem Statement

Effectively reaching diverse consumer groups through personalized marketing requires a focus on customer segmentation. This involves categorizing based on factors such as age, gender, interests, and spending habits to optimize marketing strategies. The goal is to tailor marketing materials accurately, recognizing that personalized messages resonate more effectively with specific customer preferences.

Customer Need Assessment

Customer segmentation enhances the customer experience by tailoring products, services, and communication to individual preferences. Through this approach, businesses can offer personalized solutions, communicate more effectively, and provide promotions that align with specific customer needs. This customer-centric strategy fosters satisfaction and loyalty, as it demonstrates a deep understanding of diverse customer preferences, ultimately creating a more meaningful and positive interaction between the business and its customers. Thereby, customers will be able to choose the businesses or firms that suit their interests.

Market Need Assessment

Customer segmentation empowers businesses by optimizing resource allocation and tailoring products and marketing strategies to specific customer segments. This targeted approach enhances efficiency, improves engagement, and identifies growth opportunities, ultimately boosting profitability through a more focused and effective business strategy, for example, by providing exclusive discounts, offers and promotions to specific set of customers.

Target Specifications and Characterisation

Customer segmentation will display the appropriate segments which a customer may be interested in, thereby a disguised recommendation system where customer gets to receive offers and suggestions of products and services from companies which will interest them. Our algorithm can be used to classify the customer into their particular segments, with their

respective probabilities which can give them insights on how to plan their shopping, etc..

External Sources

The dataset has been taken from the following link: https://datahack.analyticsvidhya.com/contest/janatahack-customer-segmentation

A preview of the dataset for a reference has been given below:

```
In [37]: ▶ df = pd.read_csv(r"C:\Users\kashy\Downloads\Train (3).csv")
            df

Out[37]:
```

	ID	Gender	Ever_Married	Age	Graduated	Profession	Work_Experience	Spending_Score	Family_Size	Var_1	Segmentation
0	462809	Male	No	22	No	Healthcare	1.0	Low	4.0	Cat_4	D
1	462643	Female	Yes	38	Yes	Engineer	NaN	Average	3.0	Cat_4	A
2	466315	Female	Yes	67	Yes	Engineer	1.0	Low	1.0	Cat_6	B
3	461735	Male	Yes	67	Yes	Lawyer	0.0	High	2.0	Cat_6	B
4	462669	Female	Yes	40	Yes	Entertainment	NaN	High	6.0	Cat_6	A
...	...	...	...	...	...	...	...	...	...	...	...
8063	464018	Male	No	22	No	NaN	0.0	Low	7.0	Cat_1	D
8064	464685	Male	No	35	No	Executive	3.0	Low	4.0	Cat_4	D
8065	465406	Female	No	33	Yes	Healthcare	1.0	Low	1.0	Cat_6	D
8066	467299	Female	No	27	Yes	Healthcare	1.0	Low	4.0	Cat_6	B
8067	461879	Male	Yes	37	Yes	Executive	0.0	Average	3.0	Cat_4	B

8068 rows × 11 columns

Bibliography:

Other references and readings have been taken from:

Customer Clustering: Cluster Segmentation Analysis | Optimove

Customer Segmentation: How to Segment Users & Clients Effectively (hubspot.com)

Benchmarking

There exists a significant difference between the performance of firms and companies, along with customer satisfaction with the employment of this Machine Learning model, which are summarised below:

- Targeted marketing strategies for specific customer segments, leading to higher relevance with improved customer engagement through targeted communication and offers for each segment unlike how generalized marketing strategies for the entire customer base were to be relied upon with limited ability to engage customers based on individual needs.
- Problems like limited customer retention efforts, affecting customer loyalty because of various factors, inherently as to how a mixed satisfaction levels because of not all customers receive tailored services are overcome by techniques which improve customer

retention through targeted retention strategies for different segments, therefore resulting in higher customer satisfaction as services and interactions align with specific segment expectations.

- A complex interplay of variety of such factors shall ultimately return higher profits which can be finetuned to be maximised to the firm with an efficient allocation of resources.

Applicable Regulations

Data Protection Laws:

- Compliance with GDPR or equivalent national data protection laws to safeguard customer data privacy.

Fair Lending Practices:

- Adherence to regulations preventing discriminatory practices in financial services, ensuring fairness in customizing financial products.

Transparency and Explainability Standards:

- Adherence to standards requiring transparency and explainability in algorithms, enabling customers to understand how their data is used for recommendations.

Applicable Constraints

Data privacy and ethical use:

- Adherence to data protection laws, requiring the responsible and secure handling of customer data
- Implementation of algorithms that avoid discriminatory practices, ensuring fair treatment of all customer segments.
- Requirement for explicit and informed consent from customers before utilizing their data for segmentation and personalized recommendations.

Expert verification necessities:

- The need for algorithms to be interpretable, allowing stakeholders to understand and validate the reasoning behind segmentation and recommendation outcomes.

Business Model

The emergence of machine learning applications for customer segmentation has created numerous business opportunities spanning diverse industries. These sophisticated tools empower businesses to attain a profound understanding of their customer base, enabling the implementation of targeted marketing strategies and the delivery of personalized customer experiences. E-commerce platforms stand to benefit by optimizing their product recommendations, thereby elevating the overall shopping experience and bolstering sales. In the financial sector, the application of customer segmentation allows banks to tailor financial products and services, ultimately enhancing customer satisfaction and fostering loyalty. Moreover, healthcare providers can elevate patient care through the customization of treatment plans based on individual needs and preferences. The introduction of a customer segmentation machine learning app has the potential to transform how businesses engage with their clientele, fostering heightened efficiency, competitiveness, and overall success in today's dynamic market.

In terms of monetization, businesses utilizing this app will be subject to a fee structure. They will be required to pay a standard 8% fee for accessing and utilizing the customer segmentation features. Additionally, for recommendations that result in successful matches between businesses and customers, a higher commission rate of 12% will be applied. The app shall be taken into higher heights by making it personalised for easier access.

Concept Generation

In customer segmentation, concept generalization involves identifying common traits and behaviors among specific customer groups to create overarching profiles that capture shared characteristics, enabling businesses to make generalized yet insightful decisions for targeted marketing and personalized strategies. This process extends to recommending similar businesses or enterprises to customers who share common traits, creating a more tailored experience and fostering a sense of relevance in the recommendations, ultimately enhancing customer satisfaction and engagement.

Concept Development

In the concept development of our machine learning-driven customer segmentation product, we integrate cutting-edge clustering algorithms to effectively group related customers based on shared characteristics. These clustering algorithms play a pivotal role in identifying patterns within customer data, enabling businesses to understand and target specific customer segments more precisely. Additionally, we employ classification algorithms to categorize new customers as they join, ensuring a seamless and dynamic segmentation process. This not only enhances the adaptability of the system but also allows businesses to respond promptly to evolving customer dynamics. The core of our concept lies in providing personalized recommendations and implementing targeted marketing strategies. Through the continuous analysis of customer behaviors and preferences, our product aims to deliver tailored content, promotions, and experiences, fostering a stronger connection between businesses and their customers. This personalized approach goes beyond conventional segmentation, ensuring that recommendations and strategies resonate with individual customers, thereby maximizing engagement and satisfaction.

Final Product Prototype

- The dataset will be dynamically updated and the model will keep updating its parameters from variety of sources having multiple features such as the customer's gender, age, profession, spending scores, product categories, etc.., which is further pre-processed with techniques such as imputing, feature-engineering, dimensionality reduction, etc.. which is used to train the model using various algorithms and ensemble techniques, primary ones being the clustering algorithms such as K-Means, DBSCAN, Gaussian Mixture Models, etc.. along with the classification algorithms such as K-Nearest Neighbours, Support Vector Machines, Decision trees, Naïve Bayes', etc.. with various ensemble techniques and hyperparameter tuning to continually update the elements of the model to yield high accuracies. The latter shall be made into an app, with personalized experience to both the company and its customers for those who register themselves.
- The companies may use this app to get insights on how the customers are spending and on what kind of goods they are

spending in order to plan the supply and stock; along with initiating policies and offers to attract more of them having targeted market strategies and enhanced customer retention with improved operational efficiencies.

- Customers benefit significantly from the customer segmentation app through personalized recommendations based on their interests. By categorizing and understanding customer segments, the app suggests businesses with similar interests, ensuring tailored and relevant suggestions. In e-commerce, this means discovering products aligned with preferences and the choices of like-minded individuals, simplifying decision-making, and fostering a sense of connection. Ultimately, the app empowers customers to make informed choices, contributing to a more satisfying and personalized consumer journey.

Schematic Diagram
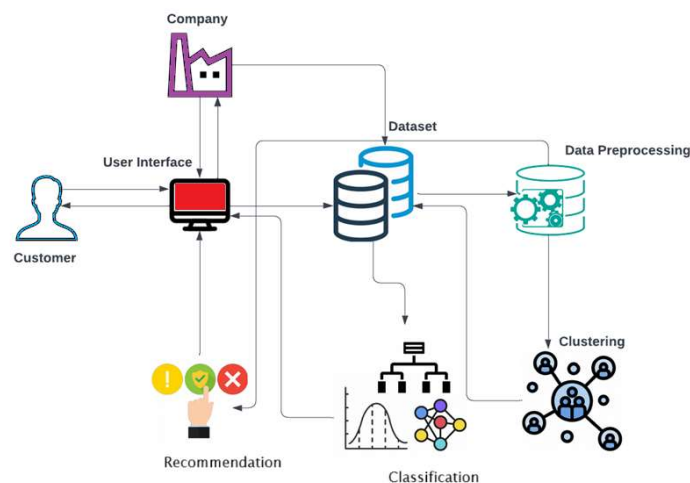
Code & Implementation

Exploratory Data Analysis:

Histogram of Age by Segmentation
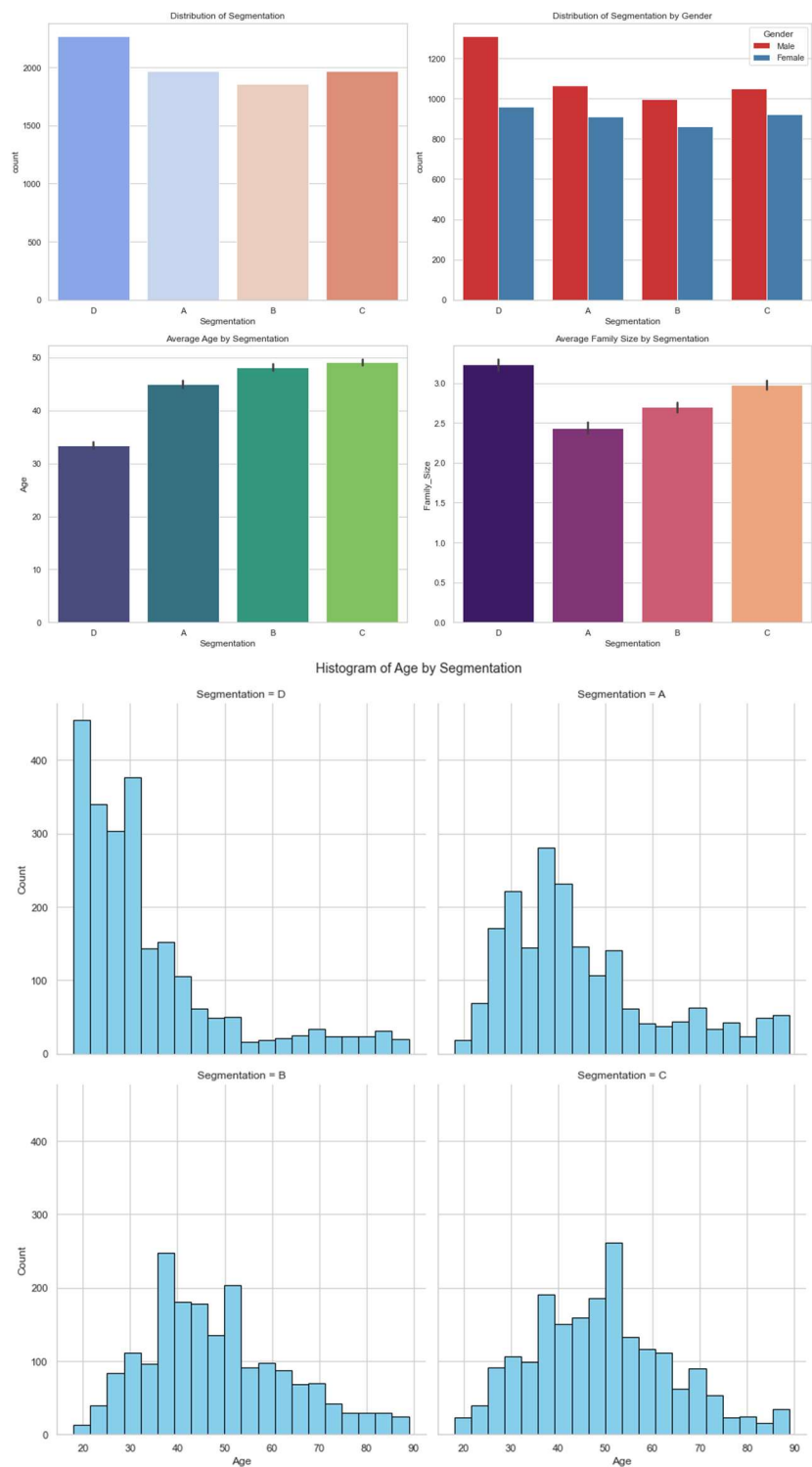
Preprocessing:

Dropping the ID column because it is of no relevance for training the model

```
In [3]:  df.drop('ID',axis=1,inplace=True)
```

OneHotEncoding and Label Encoding

```
In [4]:  df = pd.concat([df,pd.get_dummies(df[['Gender','Ever_Married','Graduated']],drop_first=True)],axis=1)
         df.drop(['Gender','Ever_Married','Graduated'],axis=1,inplace=True)
```

```
In [5]:  df['Spending_Score'].replace(df['Spending_Score'].unique(),[0,1,2],inplace=True)
```

```
In [6]:  df['Profession'].replace(df['Profession'].unique(),range(len(df['Profession'].unique())),inplace=True)
```

Extracting the number from string

```
In [9]:  df['Var_1'] = df['Var_1'].transform(lambda x:int(x[-1]))
```

Handling Null Values

```
In [8]:  from sklearn.impute import SimpleImputer
         si = SimpleImputer(strategy='most_frequent')
         df[df.columns] = si.fit_transform(df)
```

Encoding the dependent variable

```
In [11]:  df['Segmentation'] = df['Segmentation'].replace(df['Segmentation'].unique(),range(len(df['Segmentation'].unique())))
```

```
In [13]:  X = df.drop('Segmentation',axis=1)
          y = df['Segmentation']
```

Dimensionality reduction

```
In [14]:  from sklearn.decomposition import PCA
```

```
In [15]:  pca = PCA(n_components=2)
          X = pca.fit_transform(X)
```

```
In [16]:  from sklearn.model_selection import train_test_split
          X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.33, random_state=42)
```

```
In [17]:  from sklearn.linear_model import LogisticRegression
          from sklearn.tree import DecisionTreeClassifier
          from sklearn.ensemble import RandomForestClassifier
          from sklearn.svm import SVC
          from sklearn.neighbors import KNeighborsClassifier
          from sklearn.metrics import accuracy_score

          classifiers = [
              LogisticRegression(),
              DecisionTreeClassifier(),
              RandomForestClassifier(),
              SVC(),
              KNeighborsClassifier()
          ]
```

```
In [21]:  for clf in classifiers:
              clf.fit(X_train, y_train)
              y_pred = clf.predict(X_test)
              accuracy = accuracy_score(y_test, y_pred)
              print(f"{clf.__class__.__name__} Accuracy: {accuracy}")

          LogisticRegression Accuracy: 0.9276717987232446
          DecisionTreeClassifier Accuracy: 0.8901201652271874
          RandomForestClassifier Accuracy: 0.9340555764175742
          SVC Accuracy: 0.9761134059331581
          KNeighborsClassifier Accuracy: 0.9351821254224559
```

Best model: Support Vector Machines

Conclusion:

In conclusion, the ML customer segmentation project has effectively demonstrated the capability of machine learning in categorizing and grouping customers based on their behaviors, preferences, and characteristics. It's crucial to note that the dataset used in this project was chosen for demonstrative purposes, and in real-life applications, more robust ML pipelines with well-defined preprocessing steps and continuous hyperparameter tuning would be employed. This approach ensures the scalability and adaptability of the model to diverse and dynamic datasets.

The segmentation model has delivered actionable insights, enabling businesses to tailor marketing strategies for specific customer segments. This targeted approach enhances resource allocation, improves customer experiences, and ultimately contributes to business growth. The implementation of machine learning in customer segmentation underscores its role in making data-driven decisions and crafting customer-centric strategies. Ongoing monitoring and refinement of the segmentation model will be imperative to keep it aligned with evolving customer behaviors and market trends.

In essence, this project serves as a practical illustration of the potential impact of machine learning in customer segmentation, emphasizing the need for robust pipelines and continuous optimization in real-world applications.