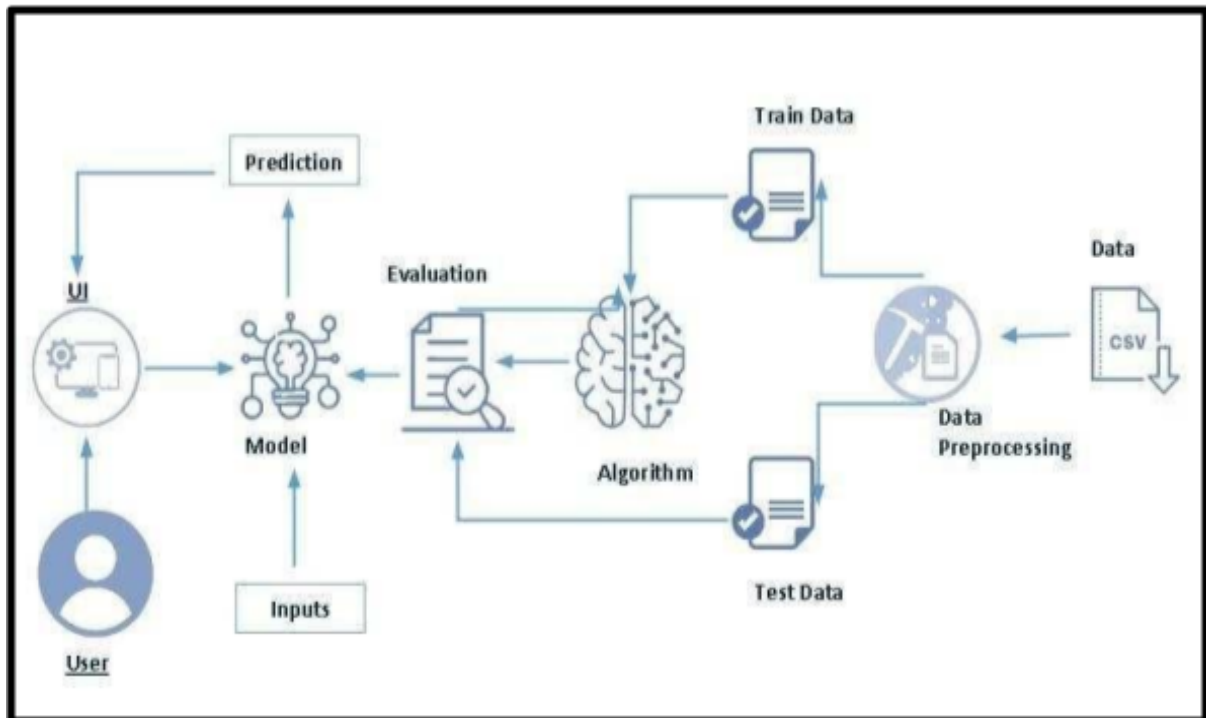


# Market Segmentation Analysis Using ML

## Project Description:

- In this report, we are going through the Steps of Market segmentation and covering the basic idea of market segmentation. We will be covering all the instruction or procedure you have to keep in mind during the market segmentation.
- The purpose of marketing is to match the genuine needs and desires of consumers with the offers of suppliers particularly suited to satisfy those needs and desires. This matching process benefits consumers and suppliers, and drives an organization's marketing planning process.

## Technical Architecture:



## Pre requisites:

To complete this project, you must required following software's, concepts and packages

- **Anaconda navigator and pycharm:**
  - Refer the link below to download anaconda navigator
    - Link : <https://youtu.be/1ra4zH2G4o0>
- **Python packages:**

- Open anaconda prompt as administrator ○ Type “pip install numpy” and click enter. ○ Type “pip install pandas” and click enter. ○ Type “pip install scikit-learn” and click enter. ○ Type “pip install matplotlib” and click enter. ○ Type “pip install scipy” and click enter. ○ Type “pip install pickle-mixin” and click enter. ○ Type “pip install seaborn” and click enter. ○ Type “pip install Flask” and click enter.

## Prior Knowledge:

You must have prior knowledge of following topics to complete this project.

- **ML Concepts**
  - Supervised learning: <https://www.javatpoint.com/supervised-machine-learning> ○ Unsupervised learning: <https://www.javatpoint.com/unsupervised-machine-learning>
  - KNN: <https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machinelearning>
  - Logistic Regression: <https://www.analyticsvidhya.com/blog/2018/09/an-end-toend-guide-to-understand-the-math-behind-logistic-regression/>
  - Evaluation metrics: <https://www.analyticsvidhya.com/blog/2019/08/11-important-model-evaluation-error-metrics/>
- **Flask Basics** : [https://www.youtube.com/watch?v=Ij4l\\_CvBnt0](https://www.youtube.com/watch?v=Ij4l_CvBnt0)

## Project Objectives:

By the end of this project you will:

- Know fundamental concepts and techniques used for machine learning.
- Gain a broad understanding about data.
- Have knowledge on pre-processing the data/transformation techniques on outlier and some visualization concepts.

## Project Flow:

- User interacts with the UI to enter the input.
- Entered input is analyzed by the model which is integrated.
- Once model analyses the input the prediction is showcased on the UI

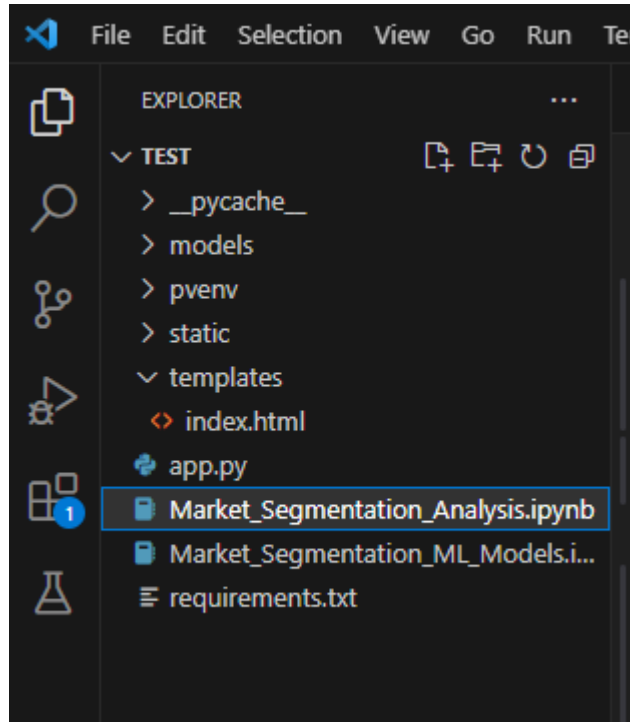
To accomplish this, we have to complete all the activities listed below,

- Data collection ○ Collect the dataset or create the dataset
- Visualizing and analyzing data ○ Univariate analysis ○ Bivariate analysis ○ Multivariate analysis ○ Descriptive analysis
- Data pre-processing ○ Checking for null values ○ Handling outlier ○ Handling categorical data ○ Splitting data into train and test

- Model building
  - Import the model building libraries
  - Initializing the model
  - Training and testing the model
  - Evaluating performance of model
  - Save the model
- Application Building
  - Create an HTML file
  - Build python code

## Project Structure:

Create the Project folder which contains files as shown below:



- We are building a flask application which needs HTML pages stored in the templates folder and a python script app.py for scripting.
- model.pkl is our saved model. Further we will use this model for flask integration.

## Milestone 1 : Define Problem/Problem Understanding

### Activity 1: Specify the business problem

- The Market Segmentation Analysis Using ML aims to analyze the spending behaviour of customers and identify opportunities for growth. The data set consists of spending ('yummy','convenient','spicy','fattening','greasy','fast','cheap','tasty','expensive', 'healthy', 'disgusting', 'Like','Age','VisitFrequency','Gender',) of the customer.
- Using unsupervised machine learning techniques, specifically clustering algorithms, the project seeks to group customers with similar spending patterns together. By identifying customer segments with distinct spending behaviours, the project aims to provide insights on how w businesses can tailor their marketing strategies and product offerings to better serve each customer segment. The project also aims to identify opportunities for growth,

such as which products or product categories are underrepresented among customers, and which segments may be receptive to new product offerings.

- Overall, the project seeks to provide valuable insights for wholesale businesses on how to optimize their operations and increase customer satisfaction and retention.

## **Activity 2: Business requirements**

Here are some potential business requirements for Market Segmentation Analysis Using ML.

**Accurate forecasting:** The predictor must be able to accurately forecast the spending behaviour of customers.

**User-friendly interface:** The predictor must have a user-friendly interface that is easy to navigate and understand. The interface should present the results of the predictor in a clear and concise manner to provide valuable insights for businesses on how to optimize their operations and increase customer satisfaction and retention.

## **Activity 3: Literature Survey**

Market Segmentation is the process of dividing customers into groups based on their shared characteristics, such as spending habits, location, or industry. This can be a valuable tool for wholesale businesses to better understand their customers and tailor their marketing and sales strategies accordingly.

There is a growing body of literature on wholesale customer segmentation. A 2019 study by the Aberdeen Group found that businesses that use customer segmentation are more likely to achieve their revenue and profit goals than those that do not. The study also found that businesses that use customer segmentation are better able to:

Target their marketing campaigns more effectively  
Develop products and services that meet the needs of their customers  
Increase customer satisfaction and retention

There are a number of different ways to segment customers. Some common methods include:

- **Geographic segmentation:** This involves dividing customers into groups based on their location. This can be a useful way to target customers with local marketing campaigns or to tailor product offerings to meet the needs of customers in different regions.
- **Demographic segmentation:** This involves dividing customers into groups based on their age, gender, income, or other demographic characteristics. This can be a useful way to target customers with specific products or services.

- **Behavioral segmentation:** This involves dividing customers into groups based on their buying habits, such as the products they purchase, the frequency of their purchases, or the amount they spend. This can be a useful way to identify customers who are most likely to respond to a particular marketing campaign or to develop new products or services that meet the needs of these customers.

Wholesale customer segmentation can be a valuable tool for businesses of all sizes. By understanding their customers and their needs, businesses can better tailor their marketing and sales strategies to achieve their goals.

#### **Activity 4: Social or Business Impact.**

The social and business impact of the Wholesale Customer Segmentation project are as follows:

- **Increased customer satisfaction and retention:** By understanding the spending behavior of their customers, wholesale businesses can tailor their marketing strategies and product offerings to better meet the needs of each customer segment. This can lead to increased customer satisfaction and retention, as customers are more likely to do business with companies that understand their needs and preferences.
- **Improved operational efficiency:** By identifying opportunities for growth, such as which products or product categories are underrepresented among customers, and which segments may be receptive to new product offerings, wholesale businesses can improve their operational efficiency. This can be done by streamlining their supply chain, optimizing their inventory management, and allocating resources more effectively.
- **Increased profitability:** By improving customer satisfaction and retention, and by improving operational efficiency, wholesale businesses can increase their profitability. This can be done by generating more revenue from existing customers, by acquiring new customers, and by reducing costs.

### **Milestone 2: Data Collection**

ML depends heavily on data, It is most crucial aspect that makes algorithm training possible. So this section allows you to download the required dataset.

#### **Activity 1: Download the dataset**

There are many popular open sources for collecting the data. Eg: kaggle.com, UCI repository, etc.

In this project we have used The Market Segmentation Analysis Using ML McDonald's data. This data is downloaded from kaggle.com. Please refer the link given below to download the dataset.

Link: <https://homepage.boku.ac.at/leisch/MSA/datasets/mcdonalds.csv>

## Milestone 3: Visualizing and analysing the data

As the dataset is downloaded. Let us read and understand the data properly with the help of some visualization techniques and some analysing techniques.

**Note: There is n number of techniques for understanding the data. But here we have used some of it. In an additional way, you can use multiple techniques.**

### Activity 1: Importing the libraries

Import the necessary libraries as shown in the image.

```
import pandas as pd
import numpy as np
import os
import matplotlib.pyplot as plt
from sklearn.decomposition import PCA
```

[107]

### Activity 2: Read the Dataset

Our dataset format might be in .csv, excel files, .txt, .json, etc. We can read the dataset with the help of pandas.

In pandas we have a function called `read_csv()` to read the dataset. As a parameter we have to give the directory of csv file.

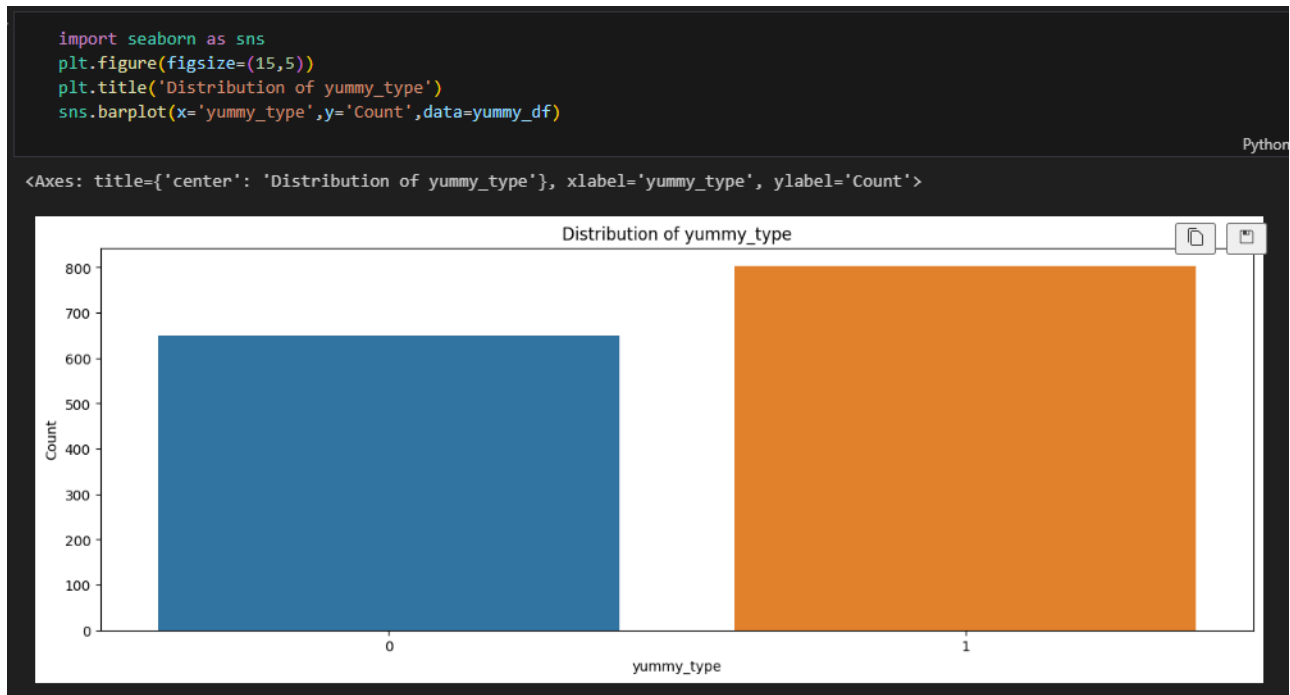
```
#print first 5 rows of the dataset
df.head()
```

	yummy	convenient	spicy	fattening	greasy	fast	cheap	tasty	expensive	healthy	disgusting	Like	Age	VisitFrequency	Gender
0	No	Yes	No	Yes	No	Yes	Yes	No	Yes	No	No	-3	61	Every three months	Fem
1	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No	No	+2	51	Every three months	Fem
2	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	No	+1	62	Every three months	Fem
3	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	No	No	Yes	+4	69	Once a week	Fem
4	No	Yes	No	Yes	Yes	Yes	Yes	No	No	Yes	No	+2	49	Once a month	M

### Activity 3: Univariate analysis

In simple words, univariate analysis is understanding the data with single feature. Here we have displayed two different graphs such as distplot and countplot.

- Seaborn package provides a wonderful function distplot. With the help of distplot, we can find the distribution of the feature. To make multiple graphs in a single plot, we use subplot.



### Activity 4: Bivariate analysis

To find the relation between two features we use bivariate analysis.

### Activity 5: Multivariate analysis

In simple words, multivariate analysis is to find the relation between multiple features.

## Activity 6: Descriptive analysis

```
df.describe(include= "all")
```

	yummy	convenient	spicy	fattening	greasy	fast	cheap	tasty	expensive	healthy	disgusting	Like	Age	VisitFrequ
count	1453	1453	1453	1453	1453	1453	1453	1453	1453	1453	1453	1453	1453.000000	
unique	2	2	2	2	2	2	2	2	2	2	2	11	NaN	
top	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	No	No	No	+3	NaN	Once a m
freq	803	1319	1317	1260	765	1308	870	936	933	1164	1100	229	NaN	
mean	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	44.604955	
std	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	14.221178	
min	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	18.000000	
25%	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	33.000000	
50%	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	45.000000	
75%	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	57.000000	
max	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	71.000000	

Descriptive analysis is to study the basic features of data with the statistical process. Here pandas has a worthy function called describe. With this describe function we can understand the unique, top and frequent values of categorical features. And we can find mean, std, min, max and percentile values of continuous features.

## Milestone 4: Data Pre-processing

As we have understood how the data is lets pre-process the collected data.

The download data set is not suitable for training the machine learning model as it might have so much of randomness so we need to clean the dataset properly in order to fetch good results. This activity includes the following steps.

- Handling missing values
- Handling categorical data
- Handling outliers
- Scaling Techniques
- Splitting dataset into training and test set

Note: These are the general steps of pre-processing the data before using it for machine learning.

Depending on the condition of your dataset, you may or may not have to go through all these steps.

### Activity 1: Checking for null values

- Let's find the shape of our dataset first, To find the shape of our data, df.shape method is used. To find the data type, df.info() function is used.
- For checking the null values, df.isnull() function is used. To sum those null values we use .sum() function to it. From the below image we found that there are no null values present in our dataset. So we can skip handling of missing values step.



Let's look for any outliers in the dataset

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1453 entries, 0 to 1452
Data columns (total 15 columns):
#   Column              Non-Null Count  Dtype  
---  -
0   yummy               1453 non-null   object 
1   convenient          1453 non-null   object 
2   spicy               1453 non-null   object 
3   fattening           1453 non-null   object 
4   greasy              1453 non-null   object 
5   fast                1453 non-null   object 
6   cheap               1453 non-null   object 
7   tasty               1453 non-null   object 
8   expensive            1453 non-null   object 
9   healthy             1453 non-null   object 
10  disgusting           1453 non-null   object 
11  Like                 1453 non-null   object 
12  Age                  1453 non-null   int64  
13  VisitFrequency       1453 non-null   object 
14  Gender               1453 non-null   object 
dtypes: int64(1), object(14)
memory usage: 170.4+ KB
```

```
#check NaN Values
df.isna().sum()

yummy      0
convenient  0
spicy       0
fattening   0
greasy      0
fast        0
cheap       0
tasty       0
expensive   0
healthy     0
disgusting  0
Like        0
Age         0
VisitFrequency  0
Gender      0
dtype: int64
```

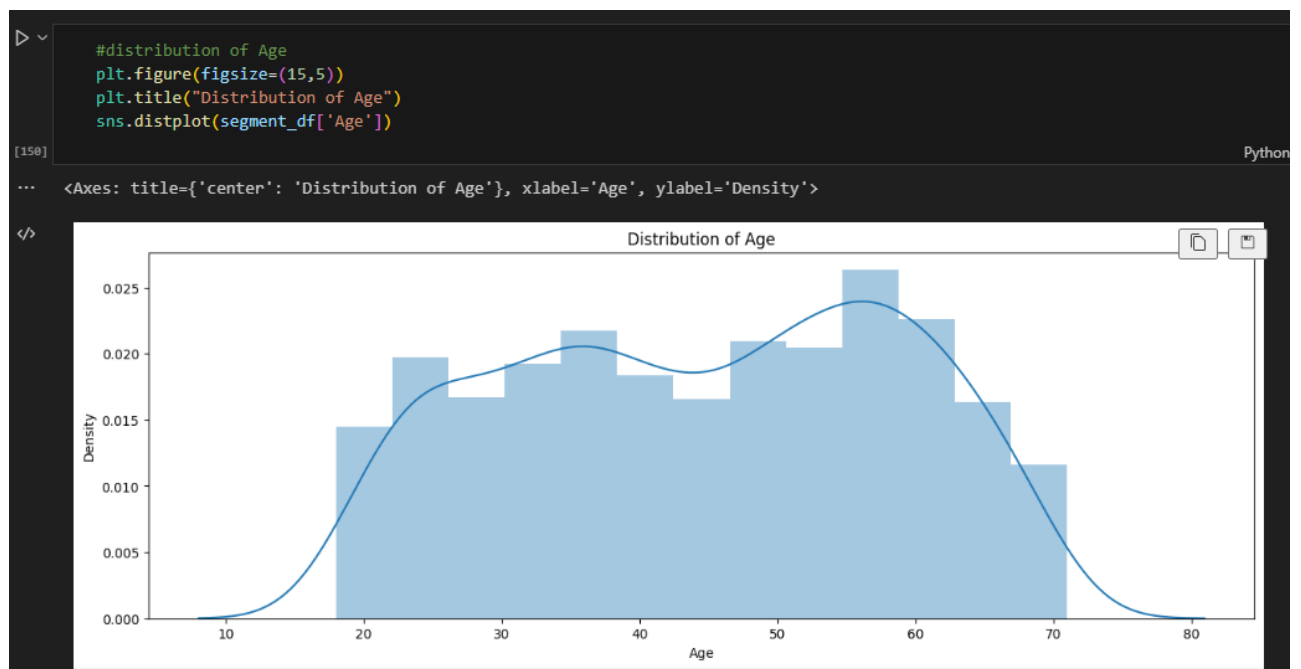
## Activity 2: Handling outliers

With the help distribution, outliers are visualized. And here we are going to find upper bound and lower bound of all features with some mathematical formula.

- From the below diagram, we could visualize that Distribution of feature has outliers or Not.

```
segment_df.Age.describe([.75,.90,.95,.99])
```

count	1453.000000
mean	44.604955
std	14.221178
min	18.000000
50%	45.000000
75%	57.000000
90%	63.000000
95%	66.000000
99%	70.000000
max	71.000000
Name: Age, dtype: float64	



### Activity 3: Splitting data into train and test

Now let's split the Dataset into train and test sets. First split the dataset into x and y and then split the data set

Here x and y variables are created. On x variable, df is passed with dropping the target variable. And on y target variable is passed. For splitting training and testing data we are using `train_test_split()` function from sklearn. As parameters, we are passing x, y, `test_size`, `random_state`.

## Train ,Test and Cross-Validation Dataset Construction

```
# split the data into test and train by maintaining same distribution of output variable 'y_true' [stratify=y_true]
X_train, test_df, y_train, y_test = train_test_split(X, y, stratify=y, test_size=0.2)
# split the train data into train and cross validation by maintaining same distribution of output variable 'y_train' [stratify=y_train]
train_df, cv_df, y_train, y_cv = train_test_split(X_train, y_train, stratify=y_train, test_size=0.2)
```

1]

Python

### Milestone 5: Model Building

Now our data is cleaned and it's time to build the model. We can train our data on different algorithms. For this project we are applying Three classification algorithms KNN, Logistic Regression and Random Forest Classifier. The best model is saved based on its performance.

### Milestone 6: Application Building

In this section, we will be building a web application that is integrated to the model we built. A UI is provided for the uses where he has to enter the values for predictions. The enter values are given to the saved model and prediction is showcased on the UI.

This section has the following tasks

- Building HTML Pages
- Building serverside script

#### Activity1: Building Html Pages:

For this project create three HTML files namely

- Index.html and save them in templates folder.

```

index.html X app.py Market_Segmentation_Analysis.ipynb Market_Segmentation_ML_Models.ipynb
templates > index.html > html > body > main#main > section#features.features > div.container > div.row > div.image.col-lg-6
76 </div>
77 </section><!-- End Hero -->
78
79 <main id="main">
80
81 <!-- ===== Features Section ===== -->
82 <section id="features" class="features">
83 <div class="container" data-aos="fade-up">
84 <div class="row">
85 <div class="image col-lg-6" style='background-image: url("https://www.start.io/wp-content/uploads/2022
86 <div class="col-lg-6" data-aos="fade-left" data-aos-delay="100">
87 <div class="icon-box mt-5 mt-lg-0" data-aos="zoom-in" data-aos-delay="150">
88 <i class="bx bx-receipt"></i>
89 <h4>Data Collection and Preprocessing</h4>
90 <p>The first step involves collecting customers data and preprocessing it to handle missing values
91 </div>
92 <div class="icon-box mt-5
93 " data-aos="zoom-in" data-aos-delay="150">
94 <i class="bx bx-cube-alt"></i>
95 <h4>Feature Engineering and Model Selection</h4>
96 <p>The second step involves selecting relevant features and transforming them into a format suitab
97 </div>
98 <div class="icon-box mt-5" data-aos="zoom-in" data-aos-delay="150">
99 <i class="bx bx-images"></i>
100 <h4>Model Training and Evaluation</h4>
101 <p>The third step involves training the selected model using the preprocessed data and evaluating
102 </div>
103 <div class="icon-box mt-5" data-aos="zoom-in" data-aos-delay="150">
104 <i class="bx bx-shield"></i>
105 <h4>Model Deployment</h4>
106 <p>The final step involves deploying the model in a real-world scenario To predict the customer be
107 </div>
108 </div>

```

## Activity 2: Build Python code:

Import the libraries

```

from flask import Flask, render_template, url_for, request
import pickle as p
import pickle
from flask import Flask, request, jsonify, render_template
import numpy as np
import pandas as pd
from sklearn.preprocessing import StandardScaler

```

Load the saved model. Importing flask module in the project is mandatory. An object of Flask class is our WSGI application. Flask constructor takes the name of the current module (`__name__`) as argument.

```

modelfile = 'models/final_prediction.pickle'
model = p.load(open(modelfile, 'rb'))
scaler= pickle.load(open('models/scaler.pickle', 'rb'))
app = Flask(__name__)

```

Render HTML page:

```
@app.route('/')
def welcome():
    return render_template('index.html')
```

Here we will be using declared constructor to route to the HTML page which we have created earlier.

In the above example, '/' URL is bound with home.html function. Hence, when the home page of the web server is opened in browser, the html page will be rendered. Whenever you enter the values from the html page the values can be retrieved using POST Method.

Retrieves the value from UI:

```
@app.route('/predict',methods=['GET','POST'])
def predict():
    yummy = float(request.form["yummy"])
    convenient =float(request.form['convenient'])
    spicy = float(request.form['spicy'])
    fattening=float(request.form['fattening'])
    greasy = float(request.form['greasy'])
    fast = float(request.form['fast'])
    cheap= float(request.form['cheap'])
    tasty =float(request.form['tasty'])
    expensive = float(request.form['expensive'])
    healthy=float(request.form['healthy'])
    disgusting = float(request.form['disgusting'])
    Age = float(request.form['Age'])
    Gender= float(request.form['Gender'])

    total = [[yummy, convenient, spicy, fattening, greasy, fast, cheap,
    |   tasty, expensive, healthy, disgusting, Age, Gender]]
    prediction = model.predict(scaler.transform(total))
    prediction = int(prediction[0])

    if prediction==0:
        return render_template('index.html',predict="Predicts Customer belong to cluster 0")

    if prediction==1:
        return render_template('index.html',predict="Predicts Customer belong to cluster 1")
    if prediction==2:
        return render_template('index.html',predict="Predicts Customer belong to cluster 2")

    else:
        return render_template('index.html',predict="Predicts Customer belong to cluster 3")
```

Here we are routing our app to predict() function. This function retrieves all the values from the HTML page using Post request. That is stored in an array. This array is passed to the model.predict() function. This function returns the prediction. And this prediction value will be rendered to the text that we have mentioned in the submit.html page earlier.

Main Function:

```
67
68 if __name__ == "__main__":
69     app.run(debug = True)
```

Ln 69, Col 26 UTF-8 CRLF Python 3.10.6 64-bit Go Live

### Activity 3: Run the application

- Open anaconda prompt from the start menu
- Navigate to the folder where your python script is.
- Now type “python app.py” command
- Navigate to the localhost where you can view your web page.
- Click on the predict, enter the inputs, click on the submit button, and see the result/prediction on the web.

### Final Output :

MARKET SEGMENTATION ANALYSIS USING ML.

Market Segmentation Analysis Using ML

Home About Contact Predict

yummy

convenient

spicy

fattening

greasy

fast

cheap

tasty

expensive

Type here to search

33°C Haze 12:57 26-08-2023