

Text  Description automatically generated with low confidence

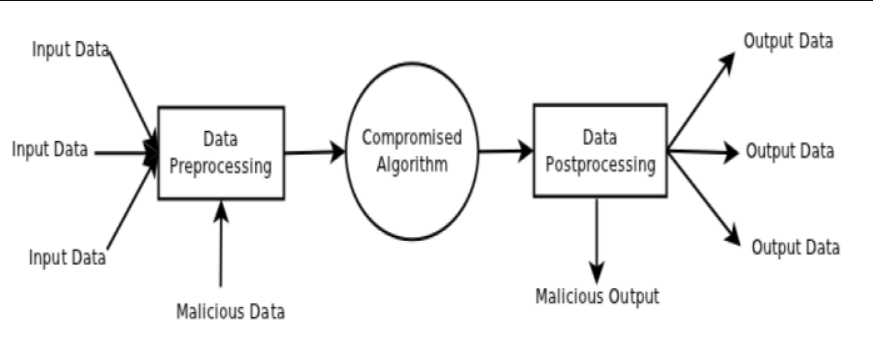
**Credit Card Customer Segmentation Using Clustering Algorithm**

Customer Segmentation

# Project Description:

The "Credit Card Customer Segmentation Using Clustering Algorithm" project aims to analyze and group credit card customers based on their spending behaviour, credit usage patterns, and demographic information. By applying clustering algorithms such as K-means or hierarchical clustering, the project seeks to identify distinct customer segments within the credit card user base. This segmentation can provide valuable insights for financial institutions and credit card companies, enabling them to tailor their marketing strategies, design targeted products, and enhance customer satisfaction. Ultimately, the project aims to improve customer retention and acquisition strategies by understanding the unique needs and preferences of different customer groups in the credit card market.

# Technical Architecture:



# Pre requisites:

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

* **Anaconda navigator and VSCode:**
  + 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>
  + KMeans: <https://www.javatpoint.com/k-means-clustering-algorithm-in-machine-learning>
  + PCA: <https://www.javatpoint.com/principal-component-analysis>
  + Evaluation metrics: <https://www.analyticsvidhya.com/blog/2019/08/11-important-model-evaluation-error-metrics/>
* **Flask Basics** : <https://www.youtube.com/watch?v=lj4I_CvBnt0>

# Project Objectives:

By the end of this project you will:

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

# Project Flow:

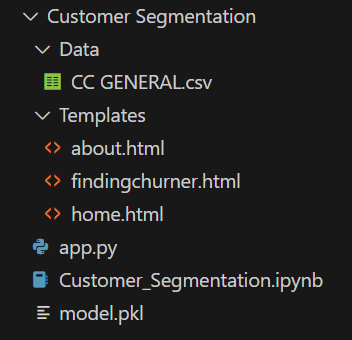
* User interacts with the UI to enter the input.
* Entered input is analyzed by the model which is integrated.
* Once model analyses the input the prediction is showcased on the UI

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
* Model building
  + Import the model building libraries
  + Initializing the model
  + Training 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.
* Data folder contains dataset and Templates folder contains html files.

# Milestone 1: Define Problem / Problem Understanding

**Activity 1: Specify the business problem**

The business problem addressed by the "Credit Card Customer Segmentation Using Clustering Algorithm" project is the need for a more precise understanding of the diverse customer base within a credit card company's portfolio. Traditional, one-size-fits-all marketing and product offerings often fall short in meeting the specific needs and preferences of individual customers. This lack of personalization can result in reduced customer satisfaction, lower retention rates, and missed revenue opportunities. By utilizing clustering algorithms to segment credit card customers into distinct groups based on their spending behavior and demographic characteristics, businesses can tailor their marketing campaigns and product offerings to cater to the unique requirements of each segment. This data-driven approach can lead to more effective customer engagement, increased customer loyalty, and ultimately improved profitability for the credit card company.

## Activity 2: Business requirements

Here are some potential business requirements for Customer Segmentation.

**Data Collection and Integration:** Gather comprehensive data from various sources, including transaction histories, demographic information, and credit usage data, and integrate it into a unified dataset for analysis.

**Data Preprocessing:** Clean, preprocess, and transform the data to ensure accuracy, consistency, and compatibility for clustering algorithms. Handle missing values, outliers, and data quality issues.

**Clustering Algorithm Selection:** Choose an appropriate clustering algorithm, such as K-means, hierarchical clustering, or DBSCAN, based on the project's objectives and data characteristics..

**Segmentation:** Apply the chosen clustering algorithm to group credit card customers into distinct segments or clusters based on spending behavior, credit usage, and demographic attributes.

**Segment Profiling:** Develop detailed profiles for each customer segment, including their spending habits, credit utilization patterns, age groups, income levels, and any other relevant attributes.

## Activity 3: Literature Survey (Student Will Write)

A literature survey would involve researching and reviewing existing studies, articles, and other publications on the topic of project. The survey would aim to gather information on current systems, their strengths and weaknesses, and any gaps in knowledge that the project could address. The literature survey would also look at the methods and techniques used in previous projects, and any relevant data or findings that could inform the design and implementation of the current project.

**Activity 4: Social and Business Impact.**

The Customer Segmentation project can have both social and business impacts.

**Social Impact:**

The "Credit Card Customer Segmentation Using Clustering Algorithm" project carries significant social impact by positively influencing both consumers and financial institutions. On the consumer side, it fosters greater financial inclusivity and personalized financial experiences. By understanding individual and diverse customer needs, financial institutions can offer tailored services and products that align with customers' unique financial situations. This can lead to improved financial well-being for customers, reduced debt burdens, and a greater sense of control over their finances. Additionally, the project can have a positive impact on responsible lending practices by helping identify high-risk customers who may need financial education or assistance.

Moreover, from a societal perspective, such segmentation can also contribute to reducing predatory lending practices and promoting ethical financial behavior. Financial institutions can use customer insights to discourage reckless spending and encourage responsible credit card usage. This can result in lower default rates, reducing the overall financial stress on consumers and contributing to a more stable financial ecosystem.

**Business Impact:**

The "Credit Card Customer Segmentation Using Clustering Algorithm" project is poised to deliver substantial business impact for financial institutions and credit card companies. By effectively segmenting their customer base, these organizations can unlock a multitude of benefits. First and foremost, they can enhance their marketing strategies by tailoring campaigns to target specific customer segments with precision. This can lead to higher conversion rates, increased customer engagement, and ultimately, improved revenue streams.

Risk management is another critical area of impact. By identifying high-risk customer segments early on, organizations can implement proactive measures to mitigate potential credit defaults, thereby safeguarding their financial health. Additionally, the project can help in resource allocation, allowing businesses to allocate their marketing, customer service, and sales resources more efficiently, reducing operational costs, and enhancing profitability.

Ultimately, the "Credit Card Customer Segmentation Using Clustering Algorithm" project has the potential to revolutionize how financial institutions engage with their customers, drive revenue growth, mitigate risks, and optimize resource allocation, all of which contribute significantly to their long-term success and competitiveness in the market.

# Milestone 2: Data Collection and Visualizing and analyzing the data

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 anime.csv & rating.csv data. This data is downloaded from kaggle.com. Please refer the link given below to download the dataset.

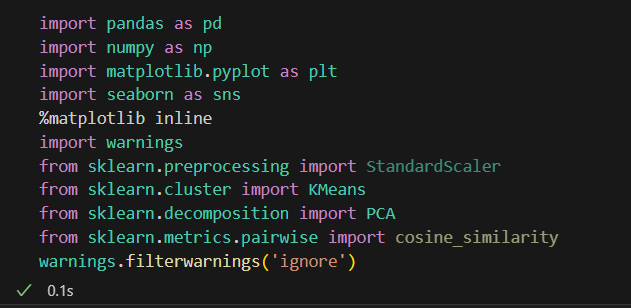
Link: <https://www.kaggle.com/datasets/arjunbhasin2013/ccdata>

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 2: Importing the libraries**

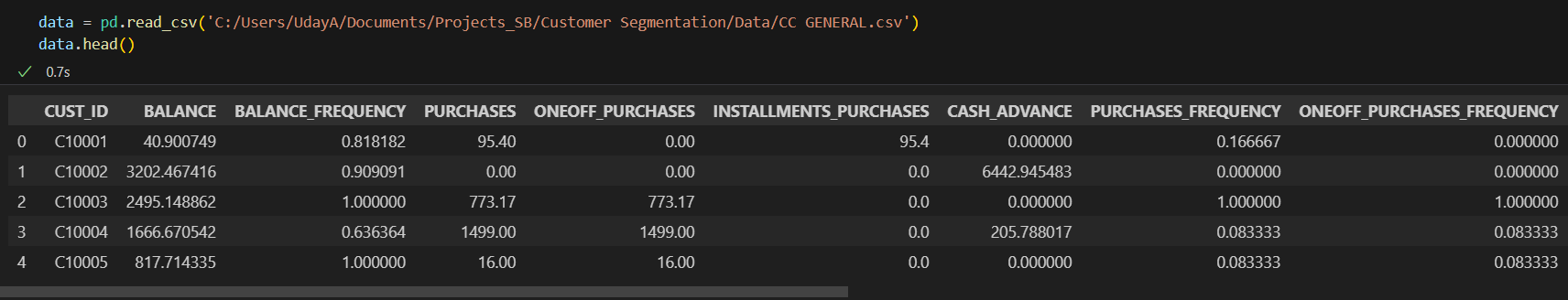
Import the necessary libraries as shown in the image. Here we have used visualization style as five thirty eight.



**Activity 3: 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.

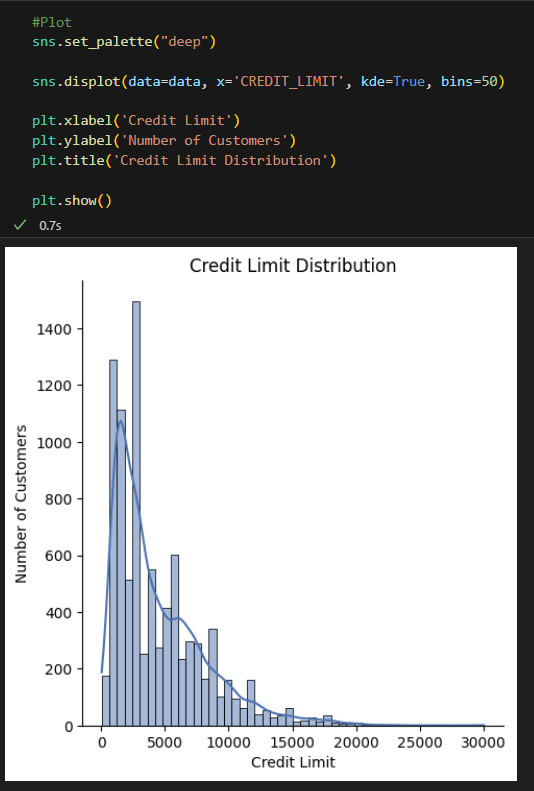
In pandas we have a function called read\_csv() to read the dataset. As a parameter we have to give the directory of csv file.



**Activity 4: Univariate analysis**

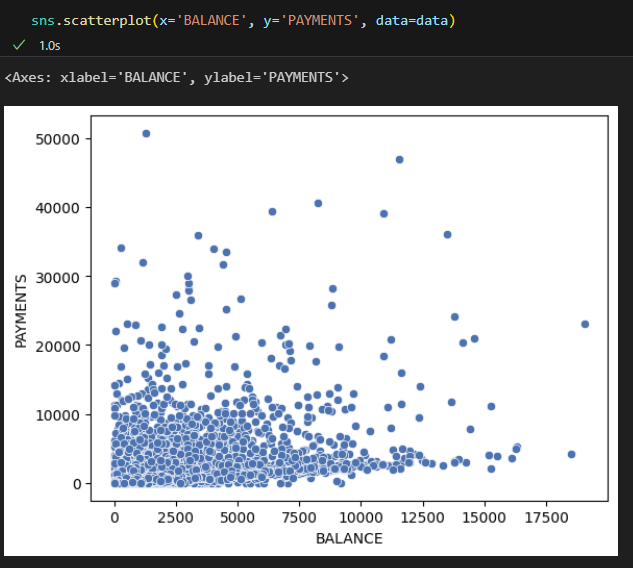
In simple words, univariate analysis is understanding the data with single feature. Here we have displayed two different graphs such as distplot and countplot.

* 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.



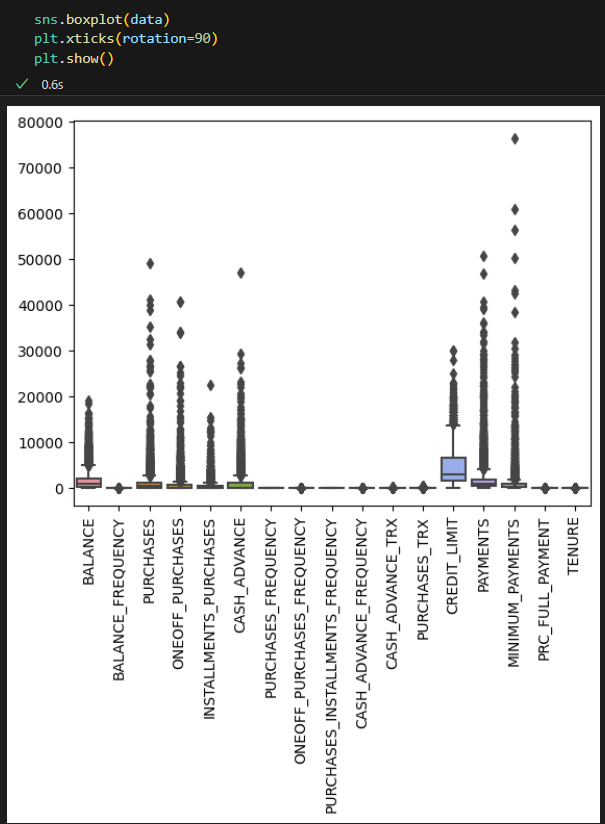
**Activity 5: Bivariate analysis**

To find the relation between two features we use bivariate analysis. Here we are visualizing the relationship between ‘charges’ and ‘bmi’ variables using scatterplot.



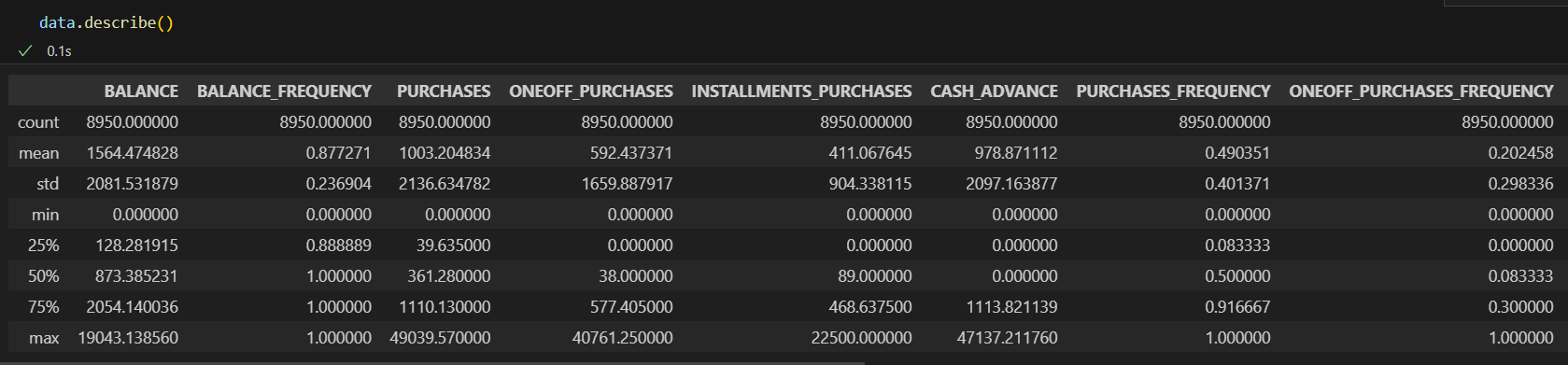
**Activity 6: Multivariate analysis**

In simple words, multivariate analysis is to find the relation between multiple features. Here we have used boxplot from seaborn package.



**Activity 7: Descriptive analysis**

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 3: 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.

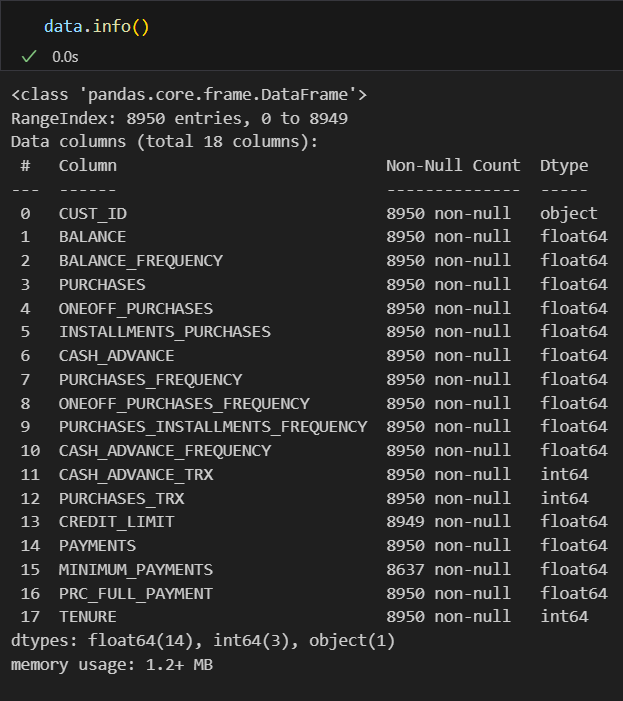
* Handling missing values
* Handling categorical data
* Relationship between Charges and Weight Status
* BMI for Smokers Status by Age Category
* Charges based on Age

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.

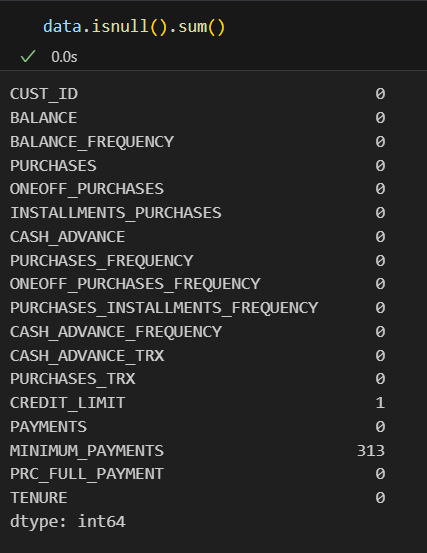
**Activity 1: Checking for null values**

* 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.

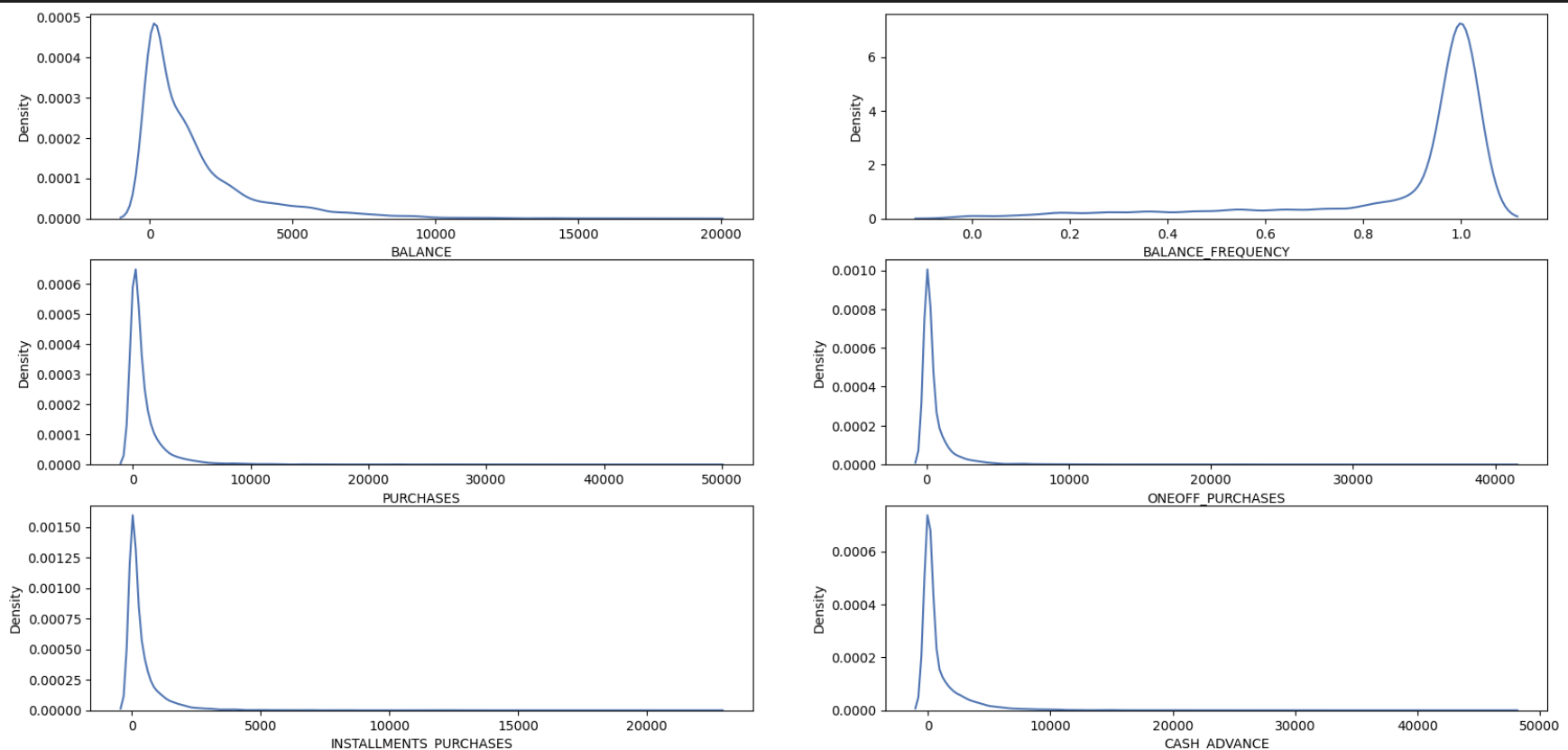




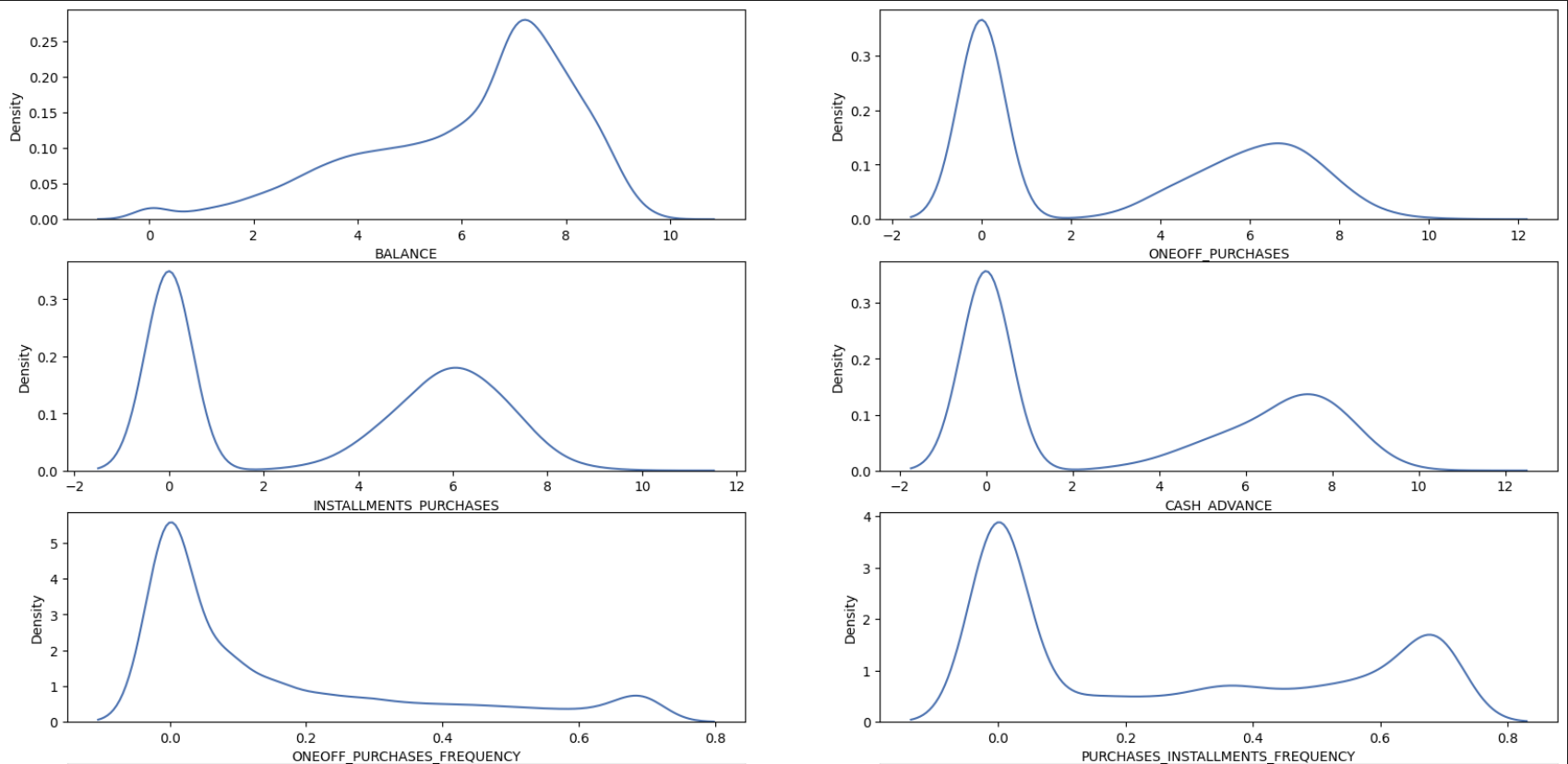
* For checking the null values, df.isnull() function is used. To sum those null values we use .sum() function to it. From the below image we found that there are no null values present in our dataset. So we can skip handling of missing values step.



**Activity 2: KDE plot for different features**



**Activity 3: BKDE plot with logarithmic transformation**

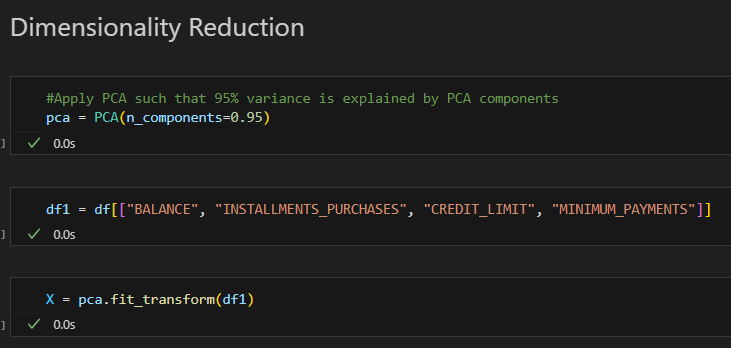


# Milestone 4: Model Building

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 four classification algorithms. The best model is saved based on its performance.

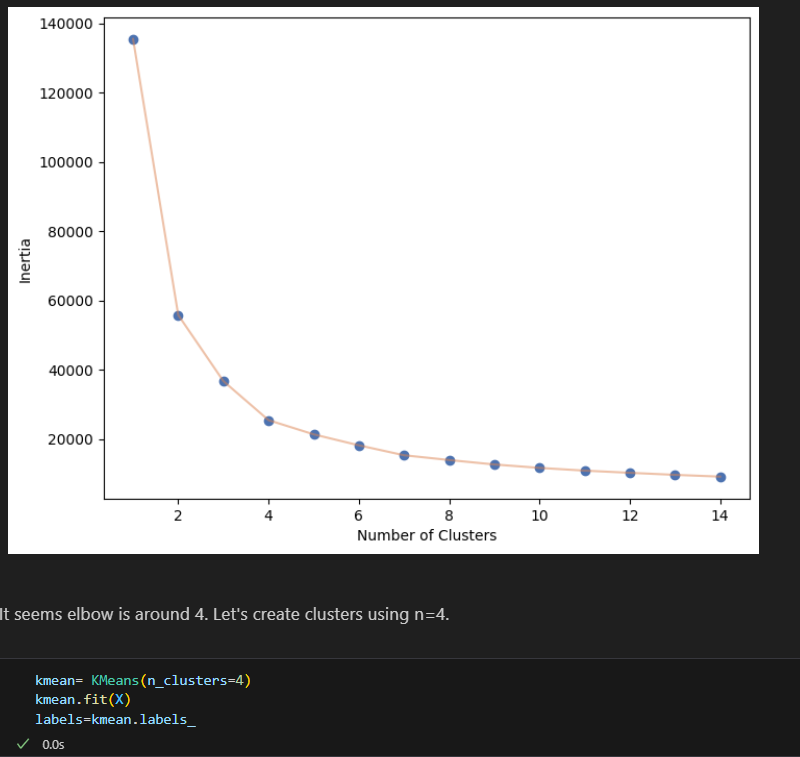
**Activity 1: PCA**

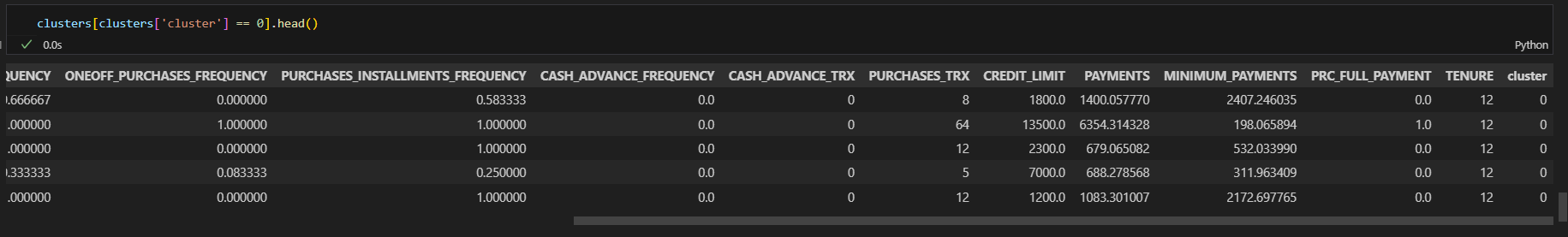
Principal Component Analysis (PCA) is a dimensionality reduction technique widely used in data analysis and machine learning. Its primary purpose is to transform a high-dimensional dataset into a lower-dimensional one while preserving as much of the original data's variance as possible. PCA achieves this by identifying and extracting a set of orthogonal axes, called principal components, that capture the most significant variations in the data. The first principal component explains the largest amount of variance, followed by the second, third, and so on. By selecting a subset of these components, you can reduce the dimensionality of the data while retaining its essential information. PCA is valuable for visualizing data, reducing noise, and improving the efficiency and interpretability of machine learning models, especially when dealing with high-dimensional datasets.



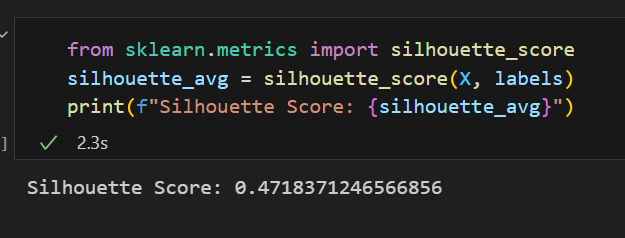
**Activity 2: Kmeans Clustering**

K-means clustering is a fundamental unsupervised machine learning technique employed to group data points into clusters based on their similarity. This algorithm starts by selecting an initial set of cluster centroids and then iteratively assigns each data point to the nearest centroid, recalculating the centroids as the means of the data points within each cluster. This process continues until convergence, resulting in K clusters where data points within each cluster are more similar to each other than to those in other clusters. K-means clustering has a wide range of applications, from customer segmentation in marketing to image compression in computer vision. However, it has limitations, such as sensitivity to initial centroid selection and the assumption of spherical, equally sized clusters, making it important to choose the right value of K and preprocess data appropriately for effective clustering.



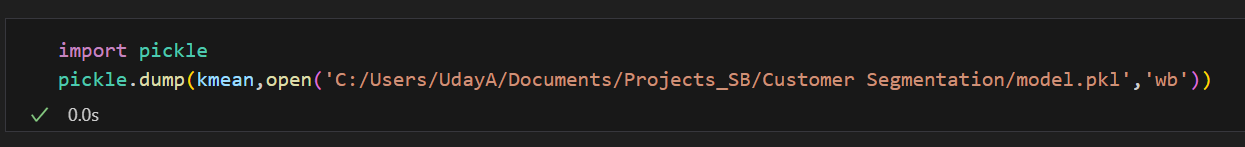


**Activity 3: Evaluating performance of the model and saving the model**



**Activity 4: Saving the Model**

Our model is performing well. So, we are saving the model by pickle.dump().



# Milestone 5: 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

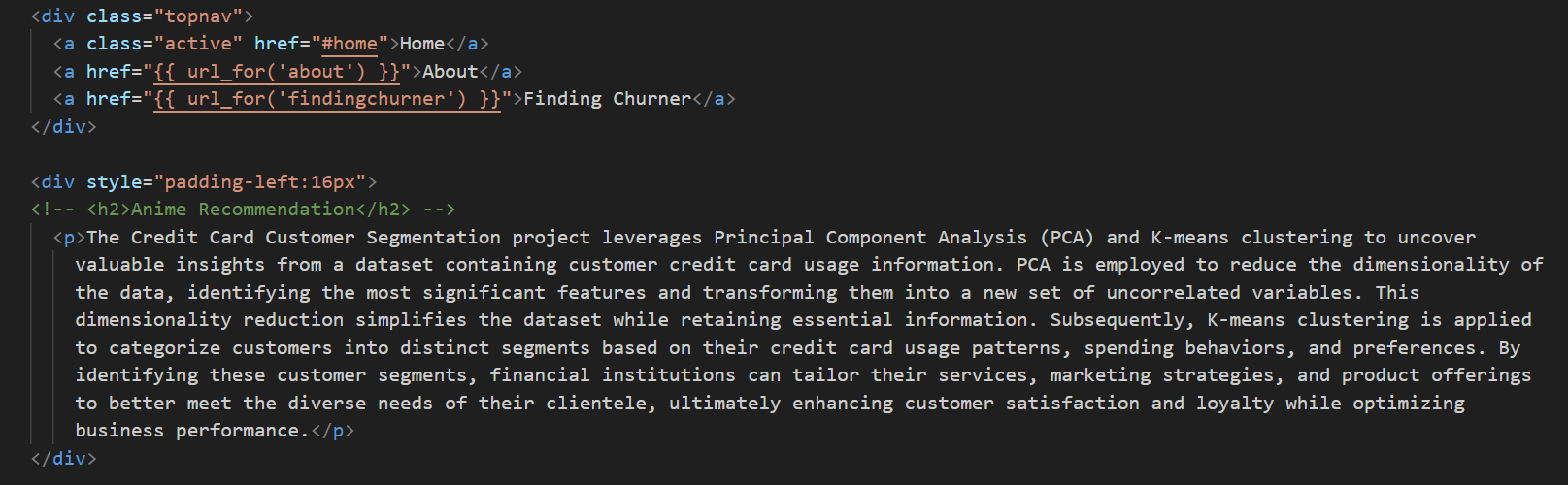
* Building HTML Pages
* Building serverside script

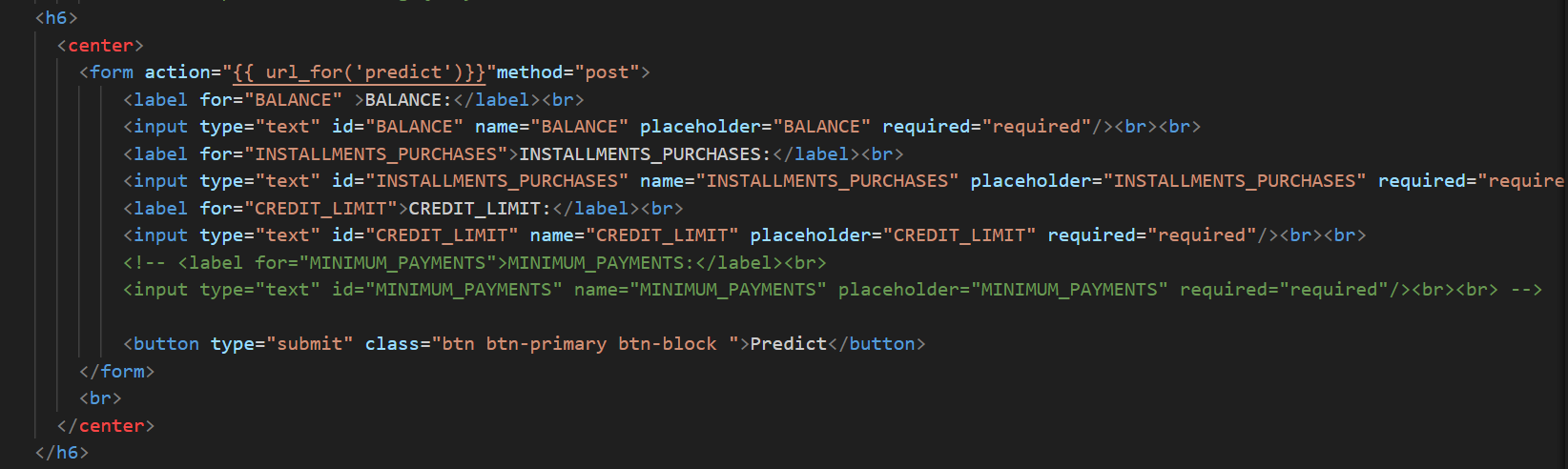
**Activity1: Building Html Pages:**

For this project create three HTML files namely

* home.html
* about.html
* findingcluster.html

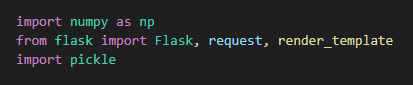
and save them in Templates folder.



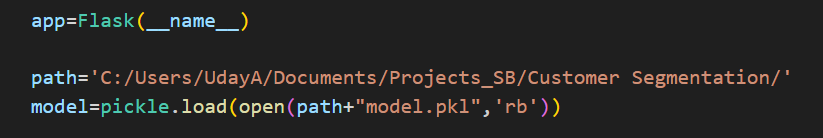


**Activity 2: Build Python code:**

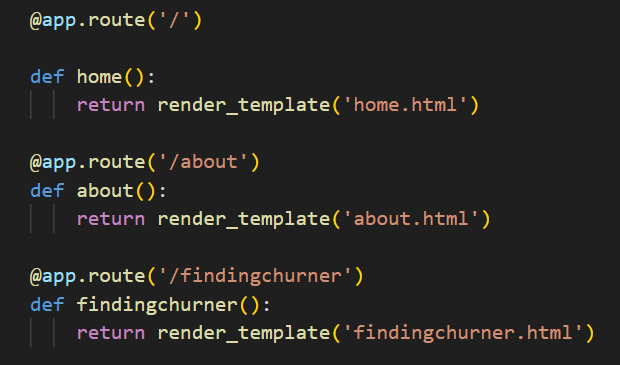
Import the libraries



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.



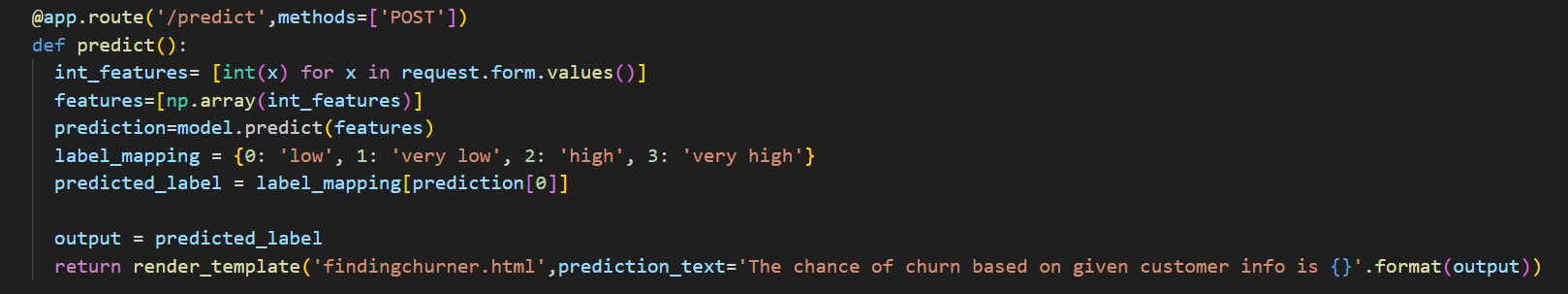
Render HTML page:



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:



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 rendered to the text that we have mentioned in the submit.html page earlier.

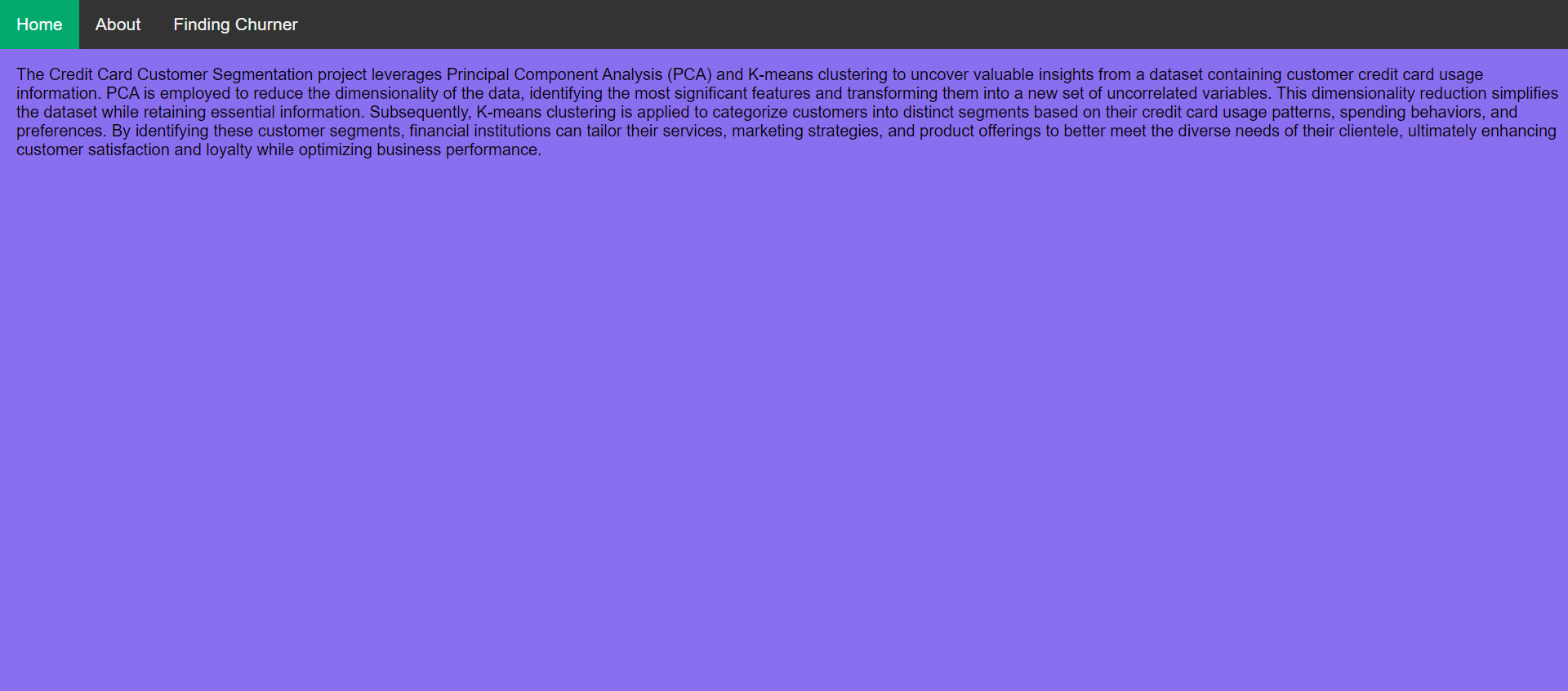
Main Function:



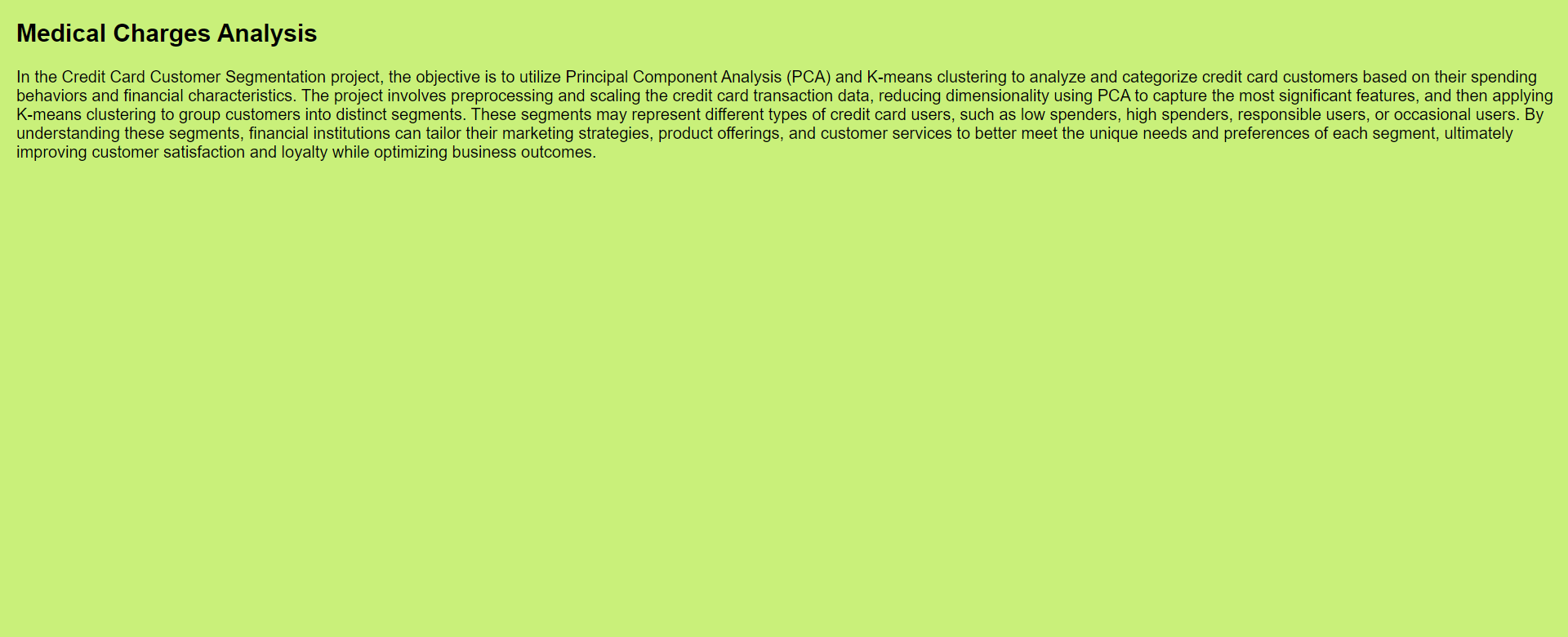
**Activity 3: Run the application**

Open Visual studio code and Import all the project folders.

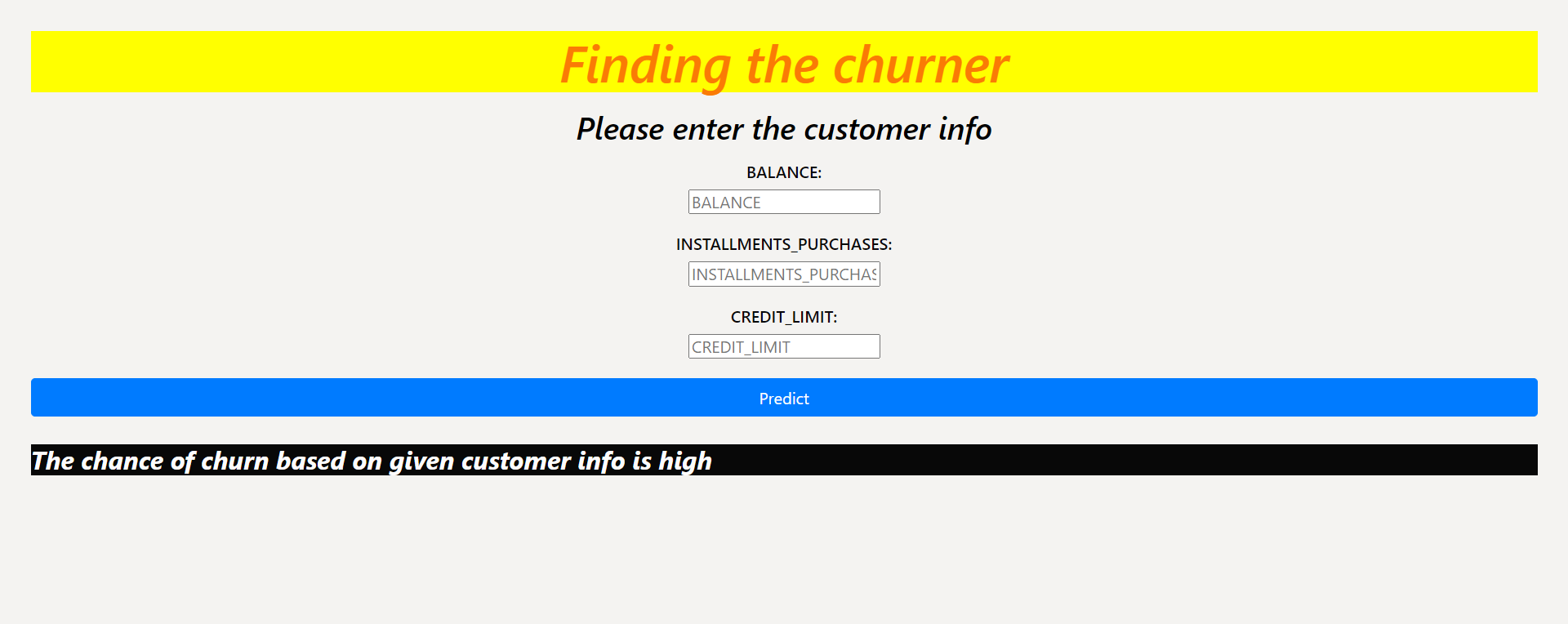
When you run the app.py file and click on the server url in terminal, you will redirected to home page. The home page will looks like:



If you click on About option, you will redirected to about.html page and it looks like:



Then click on ‘Finding Churner’ option from top left corner, you will redirected to findingchurner.html page. Enter the factors and click on predict button, the output looks like:



**Conclusion:**

In conclusion, the "Credit Card Customer Segmentation Using Clustering Algorithm" project represents a strategic opportunity for financial institutions and credit card companies to enhance their operations and customer engagement significantly. By leveraging advanced clustering algorithms and data-driven insights, businesses can gain a deeper understanding of their customer base, leading to more effective marketing strategies, personalized product offerings, and improved risk management. This not only fosters customer satisfaction and loyalty but also has the potential to drive revenue growth, reduce operational costs, and strengthen the overall financial health of these organizations. In an era where personalization and data-driven decision-making are paramount, this project stands as a powerful tool for reshaping the way financial services are delivered, benefitting both businesses and consumers alike. As the project advances, it is crucial to maintain a commitment to ethical data handling and privacy practices to ensure that these advancements are made responsibly and with the best interests of all stakeholders in mind.