# PyCaret Project to Build and Deploy an ML App using Streamlit

**Business Objective**

Customers are the lifeblood of any firm. Customer retention should be implemented with the firm's long-term goals in any organization that serves a large number of customers. In general, businesses spend a lot of money to attract new consumers, but often don't know that their existing clients account for the majority of their sales and earnings.

Customer analytics, which serves as the foundation for all marketing efforts, includes approaches such as predictive modeling, data visualization, information management, and segmentation. Customer segmentation and a product recommendation engine are employed effectively to maximize sales and conversions. We shall concentrate on customer segmentation in this project. Marketers can use segmentation to better customize their marketing efforts to different audience segments.

Customer segmentation is the process of breaking customers into multiple groups based on shared criteria in order for businesses to promote each group efficiently and appropriately. We'll be working on an end-to-end project in which we'll create a model with the help of the Pycaret library, Streamlit, Github, and the Streamlit cloud.

PyCaret is a Python-based low-code machine learning library that allows you to go from data preparation to model deployment in minutes in your preferred notebook environment. In this particular project, K means clustering will be used to segment the data. We shall then build a Streamlit application and deploy it on the Streamlit cloud by linking it to the Github repository containing our project.

**Data description**

The dataset that we will use for this project is a part of PyCaret datasets and is called ‘jewellery’. It comprises customer profiles with columns such as age, income, spending score, and savings, and we're aiming to classify customers based on these so that our marketing team can target potential product buyers.

**Aim**

To build an unsupervised learning model with the help of PyCaret library in python that can categorize the customers into different segments and deploy the model using streamlit.

**Tech Stack**

Language - Python
Libraries - pycaret, pandas, streamlit

**Approach**

1. Importing the required libraries and packages
2. Open the config file. (This is a configuration file that can be edited according to your dataset)
3. Get the dataset
4. Setup PyCaret environment
5. Model Creation
6. Model Assigning
7. Plotting model
8. Making predictions
9. Saving Model
10. Creating streamlit application
11. Creating a github repository for the project
12. Connecting streamlit cloud to github
13. Deploying the project

**Modular code overview**

```
input
|_ config.yaml
|_ jewel_data.csv

src
|_ engine.py
|_ ml_pipeline
        |_utils.py
        |_processing.py
        |_model.py
|_ streamlit_app
        |_app.py
|_ requirements.txt

lib
|_Customers Segmentation with Pycaret.ipynb

output
|_ Final_kmeans_model.pkl
|_ test.csv
```

Once you unzip the modular_code.zip file you can find the following folders within it.

1. input
2. src
3. output
4. lib

1) input folder - It contains all the data that we will need for analysis.
    - A config file, with some basic configuration parameters which can be edited according to your dataset.
    - A jewel_data.csv file, that has 505 rows of customer data with 4 columns as- Age, Income, Spending score and Savings.
2) src folder - This is the most important folder of the project. This folder contains all the modularized code for all the above steps in a modularized manner. This folder consists of:

- engine.py
- ml_pipeline

The ml_pipeline is a folder that contains all the functions put into different python files which are appropriately named. These python functions are then called inside the engine.py file.

- streamlit_app

This folder contains the python file for the streamlit application that will be used for deployment of this project.

- requirements.txt

  This file will help you install all the packages that are required to run the project successfully. You can install theses packages by using the command → **pip install -r requirements.txt**

3) output folder - The output folder contains model that we trained for this data. This model can be easily loaded and used for future use and the user need not have to train all the models from the beginning.

4) lib folder - This is a reference folder. It contains the original ipython notebook that we saw in the videos.

**Project Takeaways**

1. Understanding customer segmentation and its usage
2. Understanding the working of PyCaret
3. How to perform data preprocessing using PyCaret?
4. How to build a model using PyCaret?
5. Understanding streamlit and its usage
6. Understanding the importance of deploying a model
7. Learning how to productionize a model
8. How to build a streamlit app?
9. Understanding what is github
10. Understanding github repositories and branches
11. How to create a pull request in github?
12. How to create a feature branch in github?
13. Learn to create a github repository for your project

14. Understanding the steps to connect streamlit cloud to github
15. Understanding the project deployment process using streamlit