FIREBLAZE AI SCHOOL

ENLIGHTEN YOUR CAREER

# Mini Project - Machine Learning

**Domain :** Logistic

**About Dataset**

- A company that sells some of the product, and you want to know how well the selling performance of the product. You have the data that we can analyze, but what kind of analysis can we do? Well, we can segment customers based on their buying behavior on the market.
- Keep in mind that the data is really huge, and we can not analyze it using our bare eyes. We will use machine learning algorithms and the power of computing for it.
- This project will show you how to cluster customers on segments based on their behavior using the clustering algorithm in Python.
- I hope that this project will help you on how to do customer segmentation step-by-step from preparing the data to cluster it.

**Features Information :**

| Variable | Definition |
|---|---|
| **InvoiceNo :** | InvoiceNo |
| **StockCode :** | StockCode |
| **Description :** | Product Description |
| **Quantity:** | Quantity of product |

- **InvoiceDate :** Invoice Date
- **UnitPrice :** Per unit product price
- **CustomerID:** Customer unique ID
- **Country :** Customer belong to which country

**Deliverables :**

**Data Preprocessing :** This step performs all pre-processing steps such as data manipulation, data filling,converting categorical into numeric, and all processes.

**Exploratory Data Analysis :**

- The data preprocessing stage extracts useful information statistically.
- Such as check outliers, skewness, compare the features by
- graph and many more. Do all the required steps as well.
- The EDA process involves performing

1. **Univariate Analysis :** In this part, first check target features and start univariate analysis.
2. **Bivariate analysis :** This analysis involves studying two variables and their relationship, recalling some of the hypotheses that we generated earlier.

3. **Removing Missing values if any / Outlier treatment :** After exploring all the variables in our data, we can now impute the missing values and treat the outliers because missing data and outliers can have an adverse effect on the model performance and accuracy.
4. **Machine Learning :**
    - Understand the Dataset & cleanup (if required).
    - Build a clustering model to segment the customer-based similarity.
    - Also fine-tune the hyperparameters & compare the evaluation metrics of various classification algorithms.

**Deadline to submit the project :** One Week.

**ALL THE BEST**