

# Customer Segmentation Project

Data source:

<https://www.kaggle.com/datasets/abisheksudarshan/customer-segmentation>

- It is a data that study main features for customers and classify them into 4 classes
- Train data is labeled , but I removed labels to cluster it

Data preprocessing and cleaning:

- Nulls:  
Manipulation with nulls by study other features and get the median or mode that can replace these nulls
- Outlier:  
Check outliers by histplot and boxplot visualization  
Numerical features are normally distributed
- EDA:  
Explore data by figures like countplot – scatterplot -vilion plot
- Feature Engineering:  
Select most important features and study correlation
- Data preparation:  
Prepare data to model by label encoding and normalization
- Unsupervised learning:  
Use Elbow method to determine best clusters to segmentation  
Use k-means algorithm to cluster data

- Supervised learning

Use decision tree algorithm to learn our model

Exclude logistic regression cause it for binary classification

And SVM ,KNN cause they are very strong to our data

Achieve 92% accuracy