**AI LAB 2 2023**

**Assignment DUE by 8 Feb 23:45 H**

**Instructor : Moinul Zaber**

Now it's time to look into Unsupervised Learning techniques, and this dataset is a great start:

https://kaggle.com/vjchoudhary7/customer-segmentation-tutorial-in-python

## 

## 

Image from: <https://businessweekly.co.bw/columns/banking-metamorphosis-2/the-science-and-sense-in-segmenting-customers>

## Content

You own a supermarket mall and through membership cards, you have some basic data about your customers like Customer ID, age, gender, annual income and spending score.

A spending Score is something you assign to the customer based on your defined parameters like customer behavior and purchasing data.

## Problem Statement

You want to understand the customers who can easily converge [Target Customers] so that the sense can be given to the marketing team and plan the strategy accordingly.

Your goal here is to determine which customers are a good target for your marketing department.

Objective:

* Learning to code ML/AI with Python
* Learning to search API documentation and online resources

Evaluation: (total score: 100)

* Coding basic ML pipeline
  + Data preparation: input, output, and cleaning (30)
  + Feature engineering: preparing raw data for ML (20)
  + Implement the [KMeans Clustering Algorithm](https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html) and achieve customer segmentation and target customers whom you can easily converge (20)
* Documenting your code (20)
  + Code description in simple English
  + The reasoning behind code logic
  + Inference from model parameters
* Bonus marks for achieving a score >= 80th percentile in the leaderboard (10)

Steps:

1. Prepare a [new kaggle notebook](https://www.kaggle.com/competitions/titanic/code) with the pipeline and documentation
   1. Use python [pandas](https://pandas.pydata.org/docs/getting_started/index.html) and sci-kit learn ([sklearn](https://scikit-learn.org/stable/getting_started.html)). These are likely to be installed in your kaggle notebook.
2. Submit predictions to the competition and check score (iteratively improve pipeline if necessary)
3. Submit the following information via [form](https://forms.gle/p3ERriLs2iUsxn3S7)
   1. ID
   2. Name
   3. URL to your notebook (please remember to share with code output)
   4. Kaggle username

Tip: You can find many online resources to learn about the tools and Kaggle. Here are a few videos from the top of my google search:

* [Corey Schafer, Pandas Tutorial](https://www.youtube.com/watch?v=ZyhVh-qRZPA),
* [Kunal Naik, sklearn logistic regression tutorial](https://www.youtube.com/watch?v=tI_Pco7snZw)
* [DS Dojo, submitting predictions in Kaggle](https://www.youtube.com/watch?v=68l47Zu4Yxg).

Feel free to look up different tutorials, resources, and titanic competition kernels (starter notebooks) to learn more. Good luck!