

# Assignment 1

---

## *Course 55866*

Due date: 8<sup>th</sup> July, 2023. Assignment must be submitted electronically by 18:30 on the due date

### General Instructions

1. Assignments should be submitted electronically, as MS Word or PDF document containing text and embedded graphics. Label each plot, chart or table, add detailed title and reference it in your text (i.e. "Figure 1 shows ...").
2. Unless instructed otherwise, there is no need to submit your code.
3. Assignment should be submitted in pairs, except of one assignment of your choice (out of 5) that should be submitted alone.
4. The deadline is firm. Late submissions are not allowed.

Good luck



## Objective

In this assignment, you will tackle the critical business problem of predicting annual customer spending for a major bank, which seeks to enhance its competitive edge by understanding key factors influencing customer behavior. By applying and evaluating various predictive models on a rich dataset of customer demographics, behavior, and financial status, you will provide actionable insights that will help the bank tailor its products and services, boost customer engagement, and increase retention, ultimately driving customer satisfaction and loyalty.

## Dataset Description

You have been provided with a customers dataset, which includes the following features:

1. **RecordNumber**: Corresponds to the data record number.
2. **CustomerId**: Unique customer id in the bank.
3. **LastName**: The surname of a customer.
4. **RiskScore**: Customer credit score.
5. **Location**: Geographic location of the customer.
6. **Gender**: The gender of the customer.
7. **Age**: Customer's age.
8. **LoyaltyYears**: Number of years that the customer has been a client of the bank.
9. **AccountBalance**: Customer's account balance.
10. **ProductCount**: Number of products that a customer has purchased through the bank.
11. **HasCreditCard**: Whether or not a customer has a credit card in the bank.
12. **ActiveStatus**: Is the customer active or not.
13. **IncomeEstimate**: Client's estimated income.
14. **HasComplaint**: Indicates whether the customer has made a complaint or not.
15. **ComplaintSatisfaction**: Score provided by the customer for their complaint resolution.
16. **CardType**: Type of card held by the customer.
17. **CreditCardPoints**: The points earned by the customer for using the credit card.
18. **AnnualSpending (Target Variable)**: The annual spending of the customer.

## Tasks

1. **Data Exploration**: a. Load the dataset and perform an initial exploration to understand its structure, data types, and summary statistics. b. Identify any missing values, outliers, or inconsistencies in the data. Apply appropriate preprocessing techniques to handle these issues.
2. **Data Preprocessing**:
  - a. Split the dataset into training and testing sets, with a suitable ratio (e.g., 70:30).
  - b. Apply necessary preprocessing steps, such as feature scaling or encoding categorical variables, to prepare the data for model training.

3. **Regression Model Training and Evaluation:**
  - a. Train a linear regression model to predict the annual spending.
  - b. Evaluate the performance on the testing data using appropriate evaluation metrics.
4. **Convert the model to binary classification problem:**
  - a. Convert the target variable to binary variable indicating whether the annual spending is below average or not.
5. **Classification Model Training and Evaluation:**
  - a. Train a logistic regression model to predict if the annual spending is above average.
  - b. Choose the suitable evaluation criteria for this problem and explain why you selected this measurement.
  - c. Evaluate the performance on the testing data using appropriate evaluation metrics.
6. **Conclusion:** Write a conclusion summarizing your findings from the analysis, including the performance of the different models, their strengths, weaknesses, and any insights gained from the interpretation.

### Submission Guidelines

- Prepare a report documenting your analysis, including code snippets and explanations of the steps taken.
- Include the conclusion section summarizing your findings.
- Submit your report in PDF format along with the Python code as a separate file.