LAB TEST
2.30 p.m. - 5.30 p.m.

1. About this data

AdvWorksCusts.csv

This file consist of customer demographic with the following fields:

- **CustomerID** (*integer*): A unique customer identifier.
- **Title** (*string*): The customer's formal title (Mr, Mrs, Ms, Miss Dr, etc.)
- **FirstName** (*string*): The customer's first name.
- **MiddleName** (*string*): The customer's middle name.
- **LastName** (*string*): The customer's last name.
- **Suffix** (*string*): A suffix for the customer name (Jr, Sr, etc.)
- **AddressLine1** (*string*): The first line of the customer's home address.
- **AddressLine2** (*string*): The second line of the customer's home address.
- **City** (*string*): The city where the customer lives.
- **StateProvince** (*string*): The state or province where the customer lives.
- **CountryRegion** (*string*): The country or region where the customer lives.
- **PostalCode** (*string*): The postal code for the customer's address.
- **PhoneNumber** (*string*): The customer's telephone number.
- **BirthDate** (*date*): The customer's date of birth in the format YYYY-MM-DD.
- **Education** (*string*): The maximum level of education achieved by the customer:
    - Partial High School
    - High School
    - Partial College
    - Bachelors
    - Graduate Degree
- **Occupation** (*string*): The type of job in which the customer is employed:
    - Manual
    - Skilled Manual

    - Clerical
    - Management
    - Professional
- **Gender** *(string):* The customer's gender (for example, M for male, F for female, etc.)
- **MaritalStatus** *(string):* Whether the customer is married (M) or single (S).
- **HomeOwnerFlag** *(integer):* A Boolean flag indicating whether the customer owns their own home (1) or not (0).
- **NumberCarsOwned** *(integer):* The number of cars owned by the customer.
- **NumberChildrenAtHome** *(integer):* The number of children the customer has who live at home.
- **TotalChildren** *(integer):* The total number of children the customer has.
- **YearlyIncome** *(decimal):* The annual income of the customer.

AW_AveMonthSpend.csv

Sales data for existing customers, consisting of the following fields:

- **CustomerID** *(integer):* The unique identifier for the customer.
- **AveMonthSpend** *(decimal):* The amount of money the customer spends with Adventure Works Cycles on average each month.

AW_BikeBuyer.csv

Sales data for existing customers, consisting of the following fields:

- **CustomerID** *(integer):* The unique identifier for the customer.
- **BikeBuyer** *(integer):* A Boolean flag indicating whether a customer has previously purchased a bike (1) or not (0).

2. Challenge 1: Data Exploration

To complete this challenge:

a. Download the Adventure Works data files
b. Clean the data by replacing any missing values and removing duplicate rows. In this dataset, each customer is identified by a unique customer ID. The most recent version of a duplicated record should be retained.
c. Explore the data by calculating summary and descriptive statistics for the features in the dataset, calculating correlations between features, and creating data visualizations to determine apparent relationships in the data.

d. Based on your analysis of the customer data **after** removing all duplicate customer records, solve challenge 2.

3. Challenge 2: Classification

a. You have explored and analyzed customer data collected by the Adventure Works Cycles company. Now you should be ready to apply what you have learned about the data to building, testing, and optimizing a predictive machine learning model. Specifically, you must use R to create a classification model that predicts whether or not a new customer will buy a bike.

b. Download the test data. This data includes customer features but does not include bike purchasing values. Use your model to predict the corresponding test dataset. Don't forget to try to do the data cleaning / data preprocessing to solve this problem

c. You need to upload your R code and your answer in .csv with 2 columns: Customer ID and BikeBuyer.