# Assignment II - Comprehensive Data Analysis System Using Python

By: Dr. Amir R

## Project Objectives

The primary objective of this project is to create a comprehensive data analysis system using Python that efficiently processes a large dataset containing information about individuals, including their age, gender, and occupation. The system will be designed to handle common data analysis tasks such as data cleaning, transformation, statistical analysis, and reporting. Additionally, the project will demonstrate the use of inheritance, dictionaries, loops, and file I/O operations to provide a robust and flexible solution.

### Detailed Objectives:

* Reading the Dataset:  
  - Develop a Python program that reads an input dataset from a CSV file. Dataset is provided in Moodle.  
  - Ensure the program can handle large datasets efficiently.  
  - The filename should be a parameter of the main function to allow for flexibility in input file selection.
* Cleaning the Data:  
  - Implement a method to iterate through the dataset and replace any missing values with "N/A."  
  - Ensure that the subsequent calculations are not affected by missing or empty data.  
  - Standardize gender values: if any values in gender are "F" or "f" or "M" or "m", then replace them with "Female" or "Male."
* Transforming the Data:  
  - Calculate the average age for males and females separately.  
  - Fill in missing age values in the dataset based on the gender-specific averages.  
  - Ensure that the transformations are applied correctly and consistently across the dataset.
* Analyzing the Data:  
  - Calculate the overall average age, and separate average ages for males and females.  
  - Determine the age range for males and females separately.  
  - Generate a gender distribution report.  
  - Ensure the statistical calculations are accurate and meaningful.
* Advanced Analysis:  
  - Identify the top N oldest and youngest individuals in the dataset.  
  - Count the number of individuals within specific age ranges.  
  - Group data by another attribute (e.g., occupation) to provide deeper insights.  
  - Implement methods to perform these advanced analyses efficiently.
* File I/O Operations:  
  - Implement functionality to save the cleaned and transformed dataset to a new CSV file.  
  - Ensure the program can load a dataset from a CSV file, enabling further analysis or modifications.  
  - Provide clear and user-friendly file I/O operations.
* Inheritance and Code Structure:  
  - Use inheritance to create a specialized analyzer class that extends the basic data processor with additional analysis capabilities.  
  - Ensure the code is well-structured, modular, and follows best practices for readability and maintainability.  
  - Demonstrate the use of dictionaries, loops, and other basic Python concepts throughout the project.
* Reporting and Output:  
  - Print comprehensive reports that include:  
   - Overall average age, average age for males, and average age for females.  
   - Age range for males and females.  
   - Gender distribution.  
   - Top N oldest and youngest individuals.  
   - Count of individuals within specific age ranges.  
   - Grouped data by another attribute.  
  - Ensure the reports are clear, concise, and provide valuable insights into the dataset.
* User Instructions and Documentation:  
  - Provide clear instructions on how to use the program, including how to specify the input file and interpret the output.  
  - Document the code thoroughly to explain the functionality of each component and method.  
  - Ensure the project is user-friendly and accessible to individuals with varying levels of programming experience.

### Expected Outcome:

By the end of this project, the system should be capable of processing large datasets, sample provided as csv file in Moodle that you can use that dataset, performing comprehensive data cleaning, transformation, and analysis tasks, and generating insightful reports. The use of inheritance, dictionaries, loops, and file I/O operations will be demonstrated, showcasing the flexibility and robustness of the solution.

Submission: code your project, run it and take screenshot from the source code and the outputs, then create a pdf file and submit only pdf file.