Course Title:	Data Science	**Course Code:**	CSAI 351
Semester:	Spring 2025	**Assignment Due Date & Time:**	30/03/2025 11:57 PM
Student ID		**Mark**	
Name			

Assignment – 10% Individual

Instructions:

Submit the completed assignment through Blackboard with well documented Python code. Any similarity with other student's work will lead to a zero for both students.

Individual Assignment Description [10 Marks]

In this assignment, you will work with a real-world classification dataset using Python to build a classification model for it then evaluate the model using classification metrics such as accuracy, precision, recall, and F1 score using Python.

I. Find your **own dataset** (it should be related to a binary classification problem where the task is to predict a target class **and different from those datasets used by any other student** in the class). You can also select from the datasets provided by the popular websites for datasets used in data science competitions, here is a short list of some of these sites ordered by popularity:
   1. Kaggle - https://www.kaggle.com/datasets
   2. UCI Machine Learning Repository - https://archive.ics.uci.edu/ml/index.php
   3. Google Dataset Search - https://datasetsearch.research.google.com/
   4. AWS Open Data Registry - https://registry.opendata.aws/
   5. Data.gov - https://www.data.gov/

   Or select from those datasets uploaded with the assignment on Blackboard

II. Perform the following tasks **on your chosen dataset**:
- Load your dataset into a suitable data structure
- Perform the necessary **EDA**: use visualization techniques to preprocess your dataset for cleaning, imputation and feature reduction if needed.

III. Write a report summarizing your findings and providing insights into the data analysis and preprocessing (visualize your results)

Submission Details:

- This is an **individual** assignment. You may **not** work in groups. Your completed assignment is Due: **Sunday March 30, 2025 by 11:57pm**, avoid late submissions!
- To submit your assignment, prepare a PDF containing your work under your name and ID as follows:

ADS-FirstnameLastnameID.PDF

- Administrative overhead penalty of up to 20% will be imposed if additional processing work is required to handle your assignment, such as: did not follow answer template, unreadable file, wrongly named file, using different format, … etc.

The assignment will be marked based on the following Rubric

Criteria	Level 5 3 points	Level 4 2 points	Level 3 1 point	Level 1 0 point	Mark
Finding proper dataset	**Finding real world dataset with >=900 records**	**Finding academic dataset with >=900 records**	**Finding any data with <=300 records** **Or** **Synthesizing a random dataset**	**No dataset**	**/3**
	4 points	**3 points**	**2 points**	**1 point**	
(EDA)	Successfully using **4 EDA techniques**	Successfully using 3 EDA techniques	Successfully using 2 EDA techniques	one EDA technique is deployed	**/4**
	3 points	**2 point**	**1 point**	**0 points**	
Evaluation and Result Analysis Quality	**High quality report with detailed EDA evaluation and analysis**	**Correctly using analysis techniques with high quality report**	**Basic Quality**	**Poor Quality**	**/3**