ITU Artificial Intelligence & Data Engineering Department
YZV 211E Intro to Data Science & Engineering, Fall 2022
Homework #3
Due November 18, 2022 11:59pm

## Homework Rules

- You must submit your solution as a Jupyter (iPython) Notebook.

- Solution must include code and your analysis/comments. If the analysis is not self-explanatory and does not include adequate comments, a point penalty of up to 20 points will be applied.

- Do not share any code or text that can be submitted as a part of an assignment (discussing ideas is okay).

- Do not copy-paste code from internet sources. Do not post homework assignment problems to the internet to ask for help. If you need help, e-mail, come and talk to us.

- You may discuss the problems at an abstract level with your classmates, but you should not **share or copy code** from your classmates or the Internet. You should submit your **own and individual** homework.

- Only electronic submissions through Ninova will be accepted no later than the deadline.

- Academic dishonesty, including cheating, plagiarism, and direct copying, is unacceptable.

- If you have any questions about the homework, you can send an e-mail to Caner Özer (ozerc@itu.edu.tr).

- Note: The **submitted solutions WILL BE CHECKED WITH THE PLAGIARISM TOOLS!**

## Problem Definition

After collecting and clearing the data source, the next step is to analyse the dataset to understand the features and to infer some statements based on the data. As a result, we can find out the samples with anomaly or we can say that if one specific group is performing better than the other.

In this homework, we will be sorting out how to proceed exploratory and statistical data analysis. While we will provide the necessary guidance for this homework, we want you to satisfy a couple of requirements mentioned below.

## Requirements

- You should be complying with the instructions provided in the template Jupyter Notebook.

- Use Pandas to handle the CSV file, and **matplotlib** and **seaborn** libraries for plotting utilities. For statistical analysis, you may want to consult SciPy's Statistical Functions (stats).

- Make explanations on the plots you have generated to show what you have understand from the plot. **What does the graph tells you about the data?**

- Be aware that your plots should have **meaningful title, labels and legends** (if necessary).

- Graphs should be visually appealing such that it can show how much effort you put in to those graphs.

- Python 3.9.5 is recommended. Required libraries and preferred versions are listed below (Versions are not selected specifically. Therefore, you may work with other versions as long as they are close to the latest version. Old versions are not recommended.).

- matplotlib 3.5.2
- pandas 1.3.4
- numpy 1.22.3
- seaborn (optional)

## Project Deliveries

- Jupyter Notebook responsible for analysing the CSV file [100 pts]