



## Introduction to Data Science

## Project – Phase 1

Instructors: **Dr. Bahrak, Dr. Yaghoobzadeh**

TA(s): **Hamid Salemi,**  
**Mohammad Araghi**

Deadline: Friday, Khordad 4th,  
11:59 PM

### Introduction

In this phase, we aim to analyze our dataset further and get a deeper understanding, so that it can help us in the next phase. Analyzing the variables and their relationship would probably help you a lot in achieving the insight you'll need.

### Task

To get insights from the dataset, you'll need to extract meaningful statistics. Also, as you've learned, EDA and visualization can help you understand the underlying mystery of your data! You are free to use whatever tools you've learned so far for this phase, but pay attention! Your task will be considered successful if and only if you can explain your dataset scientifically. Note that modeling is not the goal of this phase.

### Hint

The procedure you should take for this phase is similar to what you have done in CA1, CA2 and CA3. Make sure you have logical reasons for each step you take and each investigation you do during this phase. For example, useless plots wouldn't probably help you gain insights from the data. Don't forget to do preprocessing steps whenever needed, for better analysis of data. Moreover, it is recommended to use statistical tools that you've learned, such as statistical tests, to find out complex relationships in your dataset. Follow a scientific pattern, and enjoy your journey!

### Notes

- Upload your work in this format on the website: DS\_Project\_P1\_[Std number].zip. If the project is done in a group, include all of the group members' student numbers in the name.
- If the project is done in a group, only one member must upload the work.
- This phase will not be accepted after its deadline i.e. there will be no late and grace policy!