Project - Phase 3-

<u>Description:</u> The project is split into three phases that match the learning outcomes throughout the course. Each phase accounts for 10% of your total grade.

<u>Guidelines:</u> The aim of this project is to demonstrate your ability to apply and discuss the outcomes of various data mining techniques on a problem and a dataset of your interest.

- The dataset must include quantitative and qualitative attributes.
- Your work should not be limited to what you learn in the practical sessions of the course.
- You must submit an R markdown, knitted as a pdf file, for every phase.
- You can work in a group of two same group in all phases.
- Your grade will be subject to a 5% penalty for every day of submission delay.
- <u>Phase III:</u> (10%) due Wednesday, Dec. 7, 11:59pm.
 - Use the dataset that you picked in Phase 2 or choose a new dataset discuss your choice with me in that case. (1%)
 - **N.B.** Your dataset should not be associated with any existing work related to the required tasks e.g., on kaggle, Github, ...
 - Apply tree-based approaches including decision trees, random forest, bagging, and boosting.
 (4%)
 - Apply unsupervised techniques including k-means and hierarchical clustering, as well as principal component analysis. Analyze and comment on your results. (6%)

For each phase, make sure to highlight the following in your R markdown pdf file:

- Dataset description including context and features
- Data mining tasks
- Model performance
- Results
- Comparison of results
- Comments and interpretation

Name of your R markdown pdf file following this template: NameOfTeamMember1-NameOfTeamMember2 Phase PhaseNumber.