

CSE 523 Machine Learning

Progress Report - 2

Section 1

Date of Submission: 16th February 2022

Group Name: **Bug Smashers**

Group Details

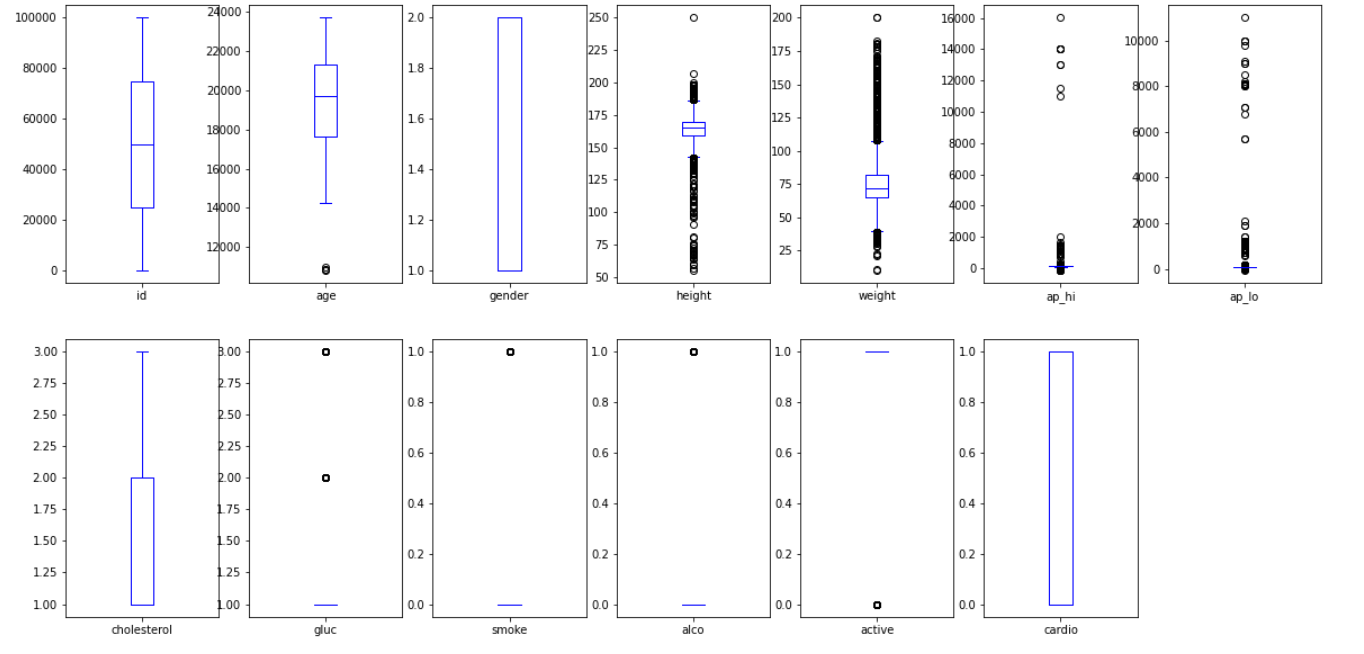
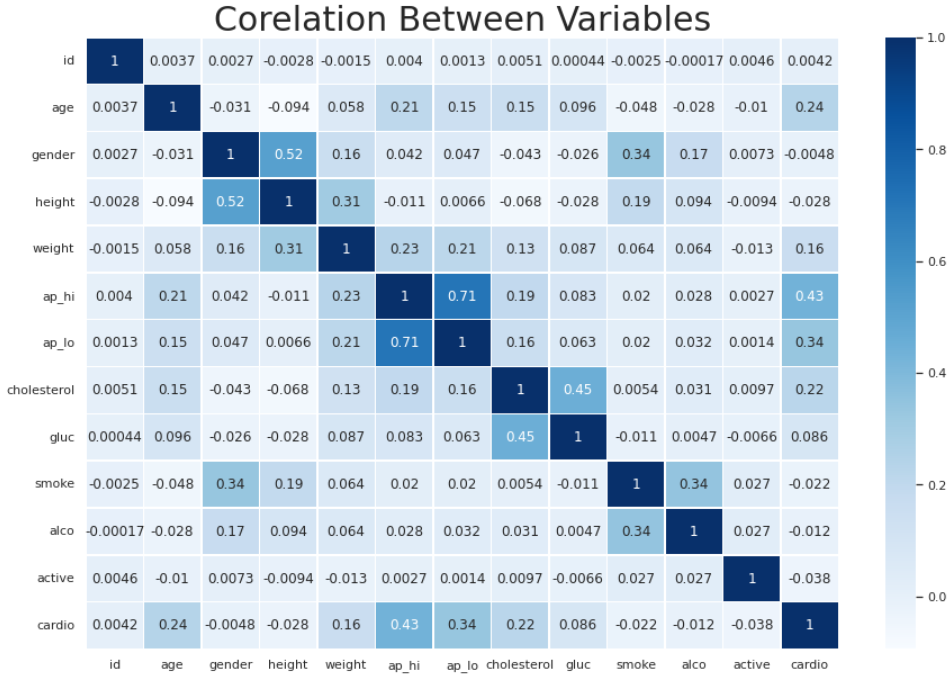
| Roll No | Name |
| --- | --- |
| AU1940028 | Moksh Doshi |
| AU1940120 | Jaimik Patel |
| AU1940130 | Nandish Patel |
| AU1940164 | Jenil Bagadiya |

2022 Winter Semester

# Tasks Performed in the week

* + This week we focused on pre-processing of data.
  + First we tried to find various sources from which we could gather relevant data and then after finding similarities between datasets we selected a few of them which could be useful for our project and eliminated the rest. The option of creating our own data was risky as medical science demands high precision and inserting incorrect values could hamper future results.
  + Since data was from multiple sources we had to change formats for unification and even eliminate features, using only those which were present in all of them.
  + Later we had to remove the noise by cleaning the data. We did so by removing values which were not useful like duplicate and empty ones. In cases where very little data was missing we filled the mode value.
  + We did all the basic tasks manually using excel features and made a draft csv file. Then we set up a public google colab and shared among group members and imported the mentioned file. Further steps which could easily be carried out using python libraries were done on the platform.

# Outcomes of the tasks performed

* + Obtained a clean dataset with no undesired values (duplicates, null, etc.)
  + Looked into the distribution patterns to gain insights about the appropriate model selection
  + 
  + 
  + [Google Colab](https://colab.research.google.com/drive/1sFlAk80RdzKh5-wykBfXMNb8BCkfNQup?usp=sharing)

# Tasks to be performed in the upcoming week

* + Rigorous Statistical Analysis to gain even deeper insights
  + Finalize the features to be included in the model training
  + Research different classification models