Data Wrangling Academy Project Brief

Challenge!

The Muskets Football team is preparing for the next games season and have asked you to help clean and preprocess their data and hopefully help predict number of hits by each player

TASK:

1. Conduct a full range cleaning of the data. Provide explanations and justifications for any actions you take.
2. Preprocess the cleaned data from task 1 above and transform it into a well behaved data.
3. Select input features for an outcome feature of HITS.

Sub Tasks

1. Extract the player names from the **PlayerUrl** column and create a new column name **Player Name** from the extracts
2. Create a new column titled **Player Status** from the **CONTRACT** column with 3 labels ;
   a. **'Active'**   If the player has an active contract
   b. **'Free'**, if the player is free
   c. **'On Loan'** if the player is on loan
3. Unpack the **POSITIONS** column into as many columns as there are positions and assign Boolean values in the columns for each player as appropriate. Name the columns the play position
4. **Weight** and **Height**, **W/F, SM** and **IR** Columns: convert to integers
5. **Value,   Wage and Release Clause** columns: convert to Float
6. Inspect the **HITS** column and ensure its float
7. Create 5 new categorical columns for the **Height, Weight, Release Clause, Value and Wage** into which you convert the respective values into clusters/labels as follows
   a. Height: Bucket intervals of 10 years
   b. Weight: Bucket intervals of 10 kg
   c. Wage: bucket intervals of 50K
   d. Value: bucket intervals of 50M
   e. Release Clause: bucket intervals of 50M

A. BONUS: Build a prediction model to predict the number of **Hits** of each payer

SUBMISSION DEADLINE:

July 20, 2023

Submit on https://globalcampus.onecampus.co