

**COLLEGE OF COMPUTING AND INFORMATION SCIENCES**

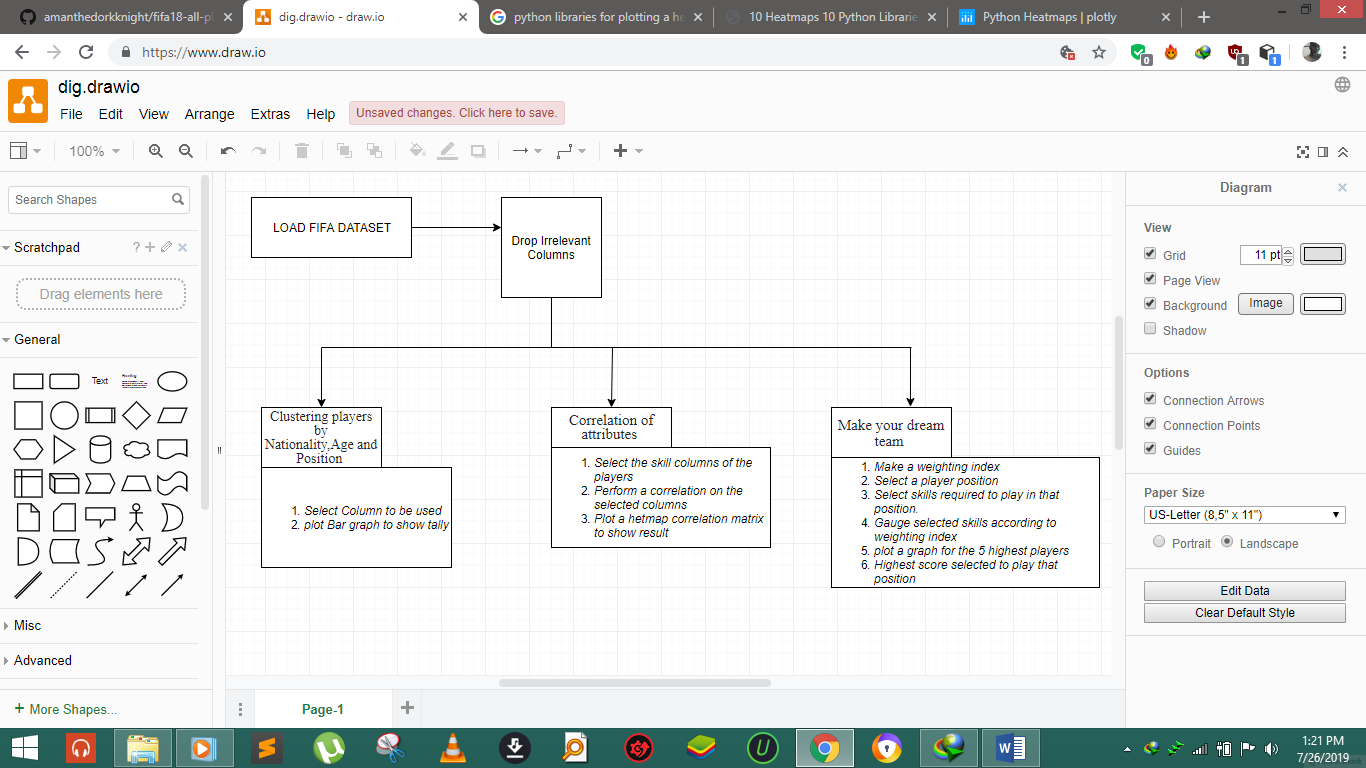
**SCHOOL OF COMPUTING AND INFORMATION SCIENCES**

**BACHELOR OF SCIENCE IN SOFTWARE ENGINEERING**

**RECESS GROUP DATA SCIENCE MINI PROJECT *DESIGN DOCUMENT***

|  |  |  |
| --- | --- | --- |
| **NAME** | STUDENT NUMBER | **REGISTRATION NUMBER** |
| MUSINGUZI JOSEPH | **217015788** | 17/U/6671/PS |
| MUKWATSE COLLIN | **217012303** | 17/U/6509/PS |
| KALANZI TONNY | **217015739** | 17/U/4463/EVE |
| BETARIZA ALVIN | **217012330** | 17/U/3574/PS |

**Pipeline Overview**

****

**Pipeline Components.**

* ***Clustering players by Nationality, Age and Position*** – After preparing the dataset for analysis, explorations are to be made on a macro level like determining number of players of a given nationality, age and the various positions they play.

*Libraries used*- ***Matplotlib*** (This is a python library used to create 2D graphs and plots by using python scripts. It has a module named pyplot which makes things easy for plotting by providing feature to control line styles, font properties, formatting axes etc.)

* ***Correlation of attributes*** - Speaking of player attributes, there are 35 skills across which all the players have been rated. These ratings contribute to the player’s overall rating and potential. The idea is to understand which skills are highly correlated with the overall and potential ratings of players. To help perform and visualize the analysis, a correlation matrix will be plotted and displayed as a heatmap as it can be useful when you want to see which intersections of the categorical values have higher concentration of the data compared to the others. The heatmap, whose color-coded scale of correlation coefficient gives a quick overview of which skills are highly correlated and which are negatively correlated. The darker the color of the square, the more is the correlation.

*Libraries used***- *Seaborn*** *(*Seaborn is a library for making statistical graphics in Python. It is built on top of Matplotlib and closely integrated with pandas data structures. Seaborn aims to make visualization a central part of exploring and understanding data.*)*

* ***Make your dream team*** – The aim in this section is to use statistical analysis on our data to find out the best squad for a dream team given a user selected formation. For every given playing formation the program computes the 5 best players for each position in that formation. This will be done using the skills given in the dataset and each will be given a weight basing on the nature of its importance when playing a given position. The 5 highest scores will be plotted as bar plots as bar graphs can be used for comparison amongst different things. The highest score will be selected as the best player.

We opted for a 4-3-3 formation ie. Goal Keeper, 4 defenders, 3 mid fielders and 3 attackers using the following criteria:

*Goal keeper* - In order to get the best goalkeeper, we are to analyze the data for the below mentioned parameters. **Shot Stopper (**A goalkeeper who is strong in stopping shots taken by opponents) and **Sweeper** (A goalkeeper who is strong in playing with his feet and making passes).

*Defenders –* In order to find the best defenders, we shall opt for the following. **Centre Backs (**We need two center-backs. One who plays LCB and the other who plays RCB).**Wing Backs (**We again need two wing backs. One who plays on the Left and the other who plays on the right)

*Midfielders -* As per our game formation 4-3-3, we have to choose 3 midfielders basing on the following: ***Playmaker (***A playmaker is someone who will move the ball to the attacking 3rd from defence or midfield.) ***Beast*** (A beast is a typical box-to-box player with loads of energy and who can boss the midfield.) ***Controller*** (A controller is the person who is orchestrating your midfield engine by either sitting back or going forward based on dynamic needs.

*Attackers -* In order to find the best attackers, we considered the following: **Attacking Left Wing (**He is a player, attacking from the left flank.) **Attacking Right Wing (**He is a player, attacking from the right flank) **Striker (**He is a player attacking from the center)

*Attribute Selection*

Shot stopper – GKReflexes, GKDiving, Jumping, GKPositioning, GkHandling, Composure, Reactions, SprintSpeed, Strength

Sweeper – GKKicking, GKReflexes, Vision, GKHandling, GKDiving, GKPositioning, Jumping, Composure, Reactions, SprintSpeed, Short Passing, Long Passing

Centre Backs – Reactions, Sliding tackle, Standing tackle, Stamina, Jumping, Heading Accuracy, Marking, Aggression, Acceleration, Interceptions, Vision, Composure, Crossing, LongPassing, SprintSpeed, Longshots, Short Passing.

Wing Backs (Left & Right) – Acceleration, SlidingTackle, StandingTackle, Vision, Crosssing, Acceleration, SprintSpeed, LongPassing, Stamina, BallControl, ShortPassing, Dribbling, Marking, Positioning, Finishing.

Playmaker – BallControl, Dribbling, Reactions, Vision, ShortPassing, Positioning, Crossing, LongPassing, Curve, FKAccuracy, LongShots, Marking.

Beast – Agility, Stamina, ShortPassing, Reactions, Balance, Strength, Acceleration, Aggression, Jumping, Marking, StandingTackle, SlidingTackle, Interceptions, SprintSpeed

Controller – BallControl, ShortPassing, LongPassing, Vision, Composure, Dribbling, Marking, Reactions.

Attacking Wingers (Left & Right) – Acceleration, Finishing, BallControl, Dribbling, SprintSpeed, Crossing, LongShots, Vision, ShortPassing, LongPassing, Aggression, Agility, FKAccuracy, Curve.

Striker – Balance, Finishing, HeadingAccuracy, Jumping, Dribbling, BallControl, Aggression, Agility, Vision, Curve, LongShots

*Libraries used***- *NumPy*** (This is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.) ***Matplotlib*** (This is a python library used to create 2D graphs and plots by using python scripts. It has a module named pyplot which makes things easy for plotting by providing feature to control line styles, font properties, formatting axes etc.)