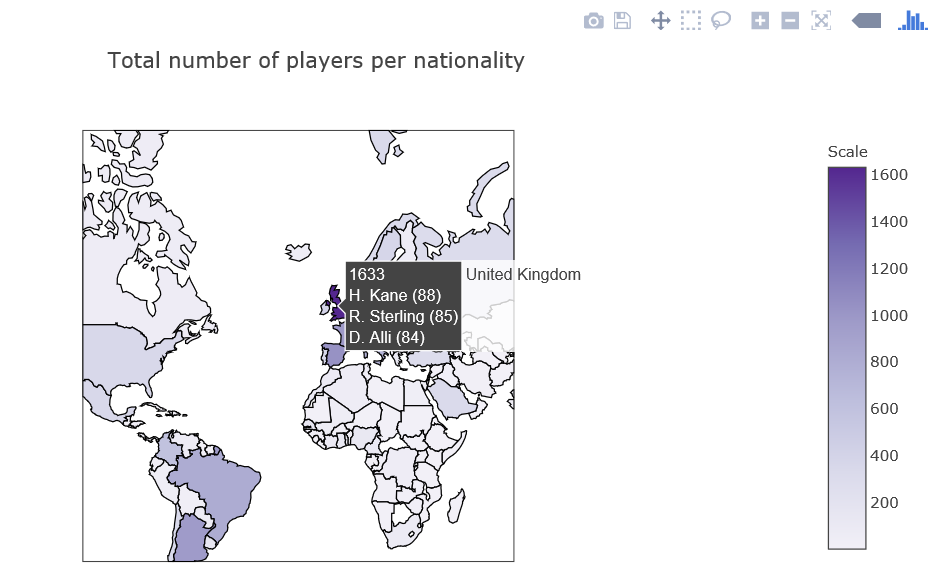
**Assignment 1 – Machine Learning (INFO 7390) Review**

**Gaurang Davda |** [**davda.g@husky.neu.edu**](mailto:davda.g@husky.neu.edu)

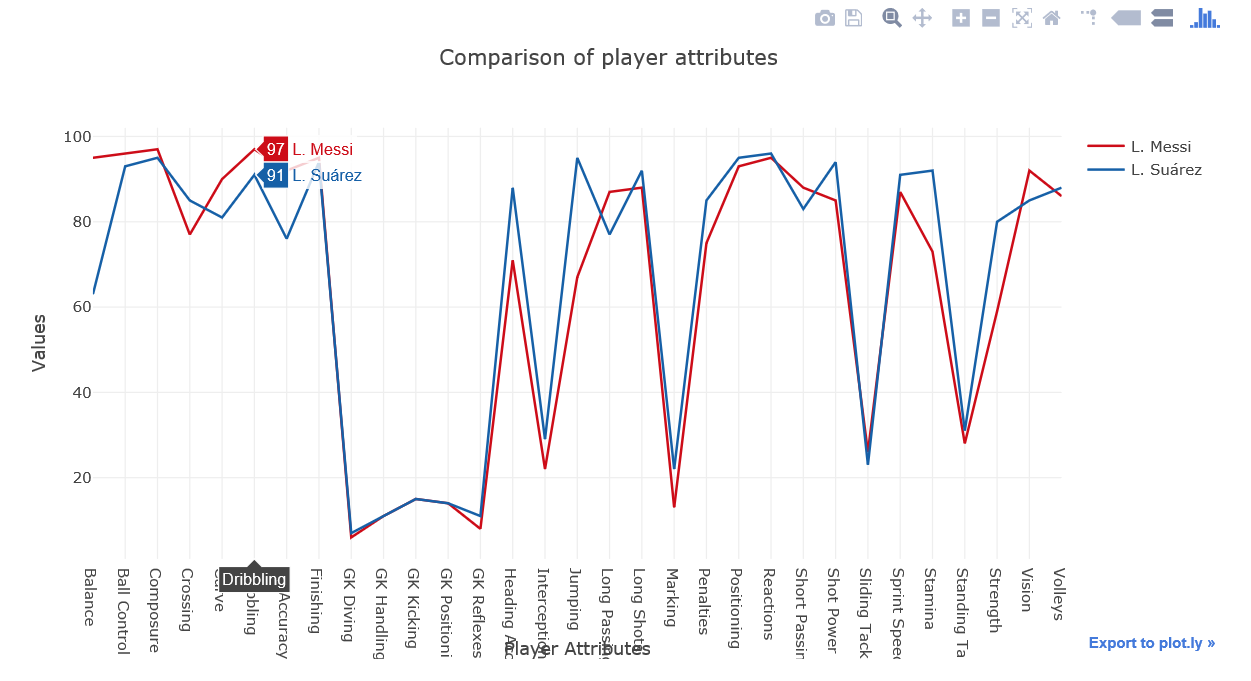
**Introduction**

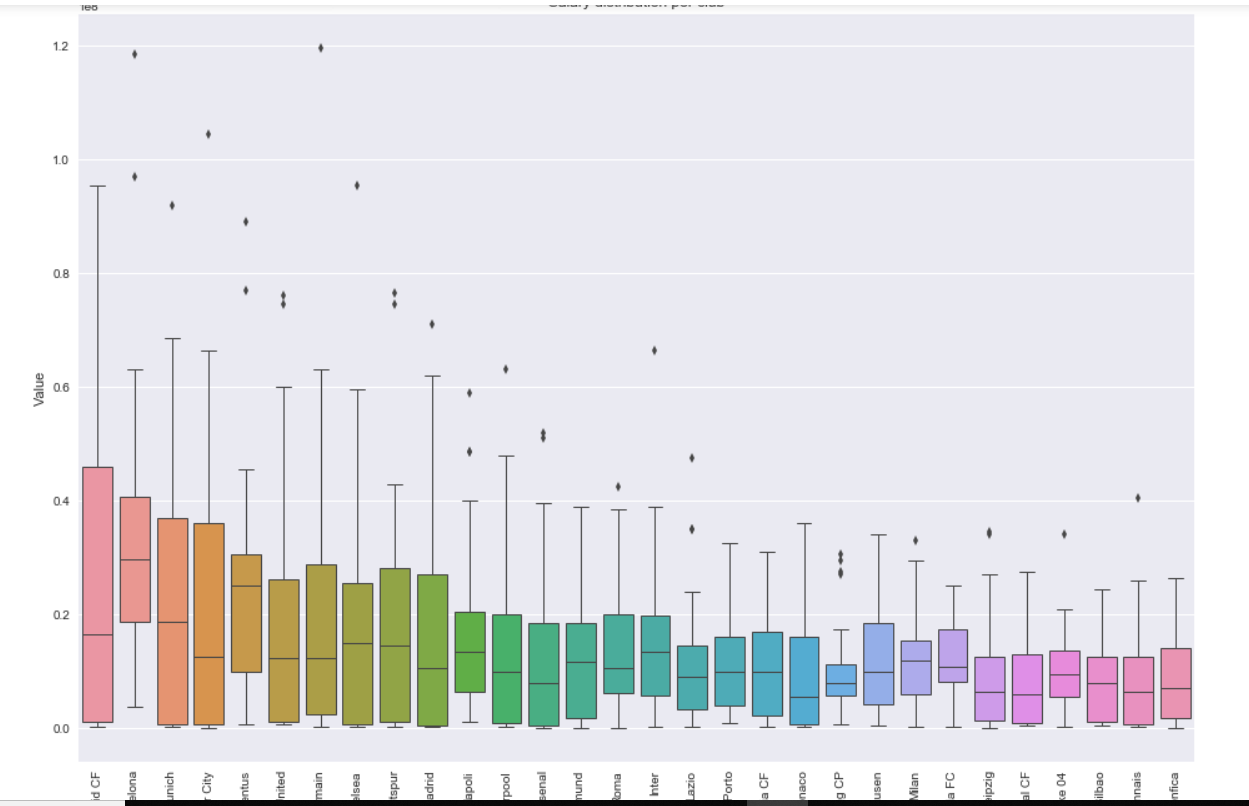
This assignment involves performing Exploratory Data Analysis on the FIFA-18 dataset and then predicting player valuation by performing regression using multiple Algorithms. The dataset includes 18000 players from across the world, hence I have plotted total number of players representing the country, top 3 players along with their overall on the world map using plotly. Players are also compared based on their attributes like Aggression, Acceleration and many more which are used by the clubs for finding which player is the best in a certain skill.

**Plotting Data on world map using plotly:**



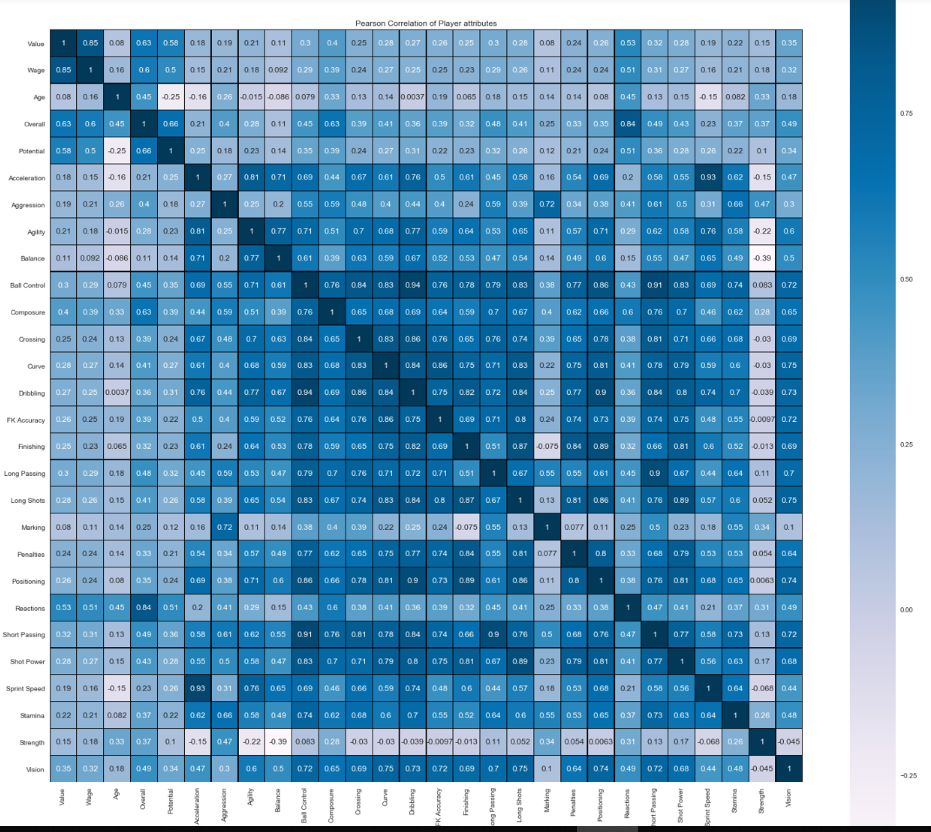
**Comparing Players Based on Attributes:**



**Anomaly Detection**

We can see outliers in terms of wages in few clubs. These players are exceptional and have a very high overall as compared to other players in FIFA-18.

**Correlation among columns**



This was totally unexpected result, the heatmap shows that no particular ability is strongly related with wages

and value of a player. Only overall rating is somewhat strongly related with both of them which was kind of expected.

Dribbling and ball control shows a very strong relationship. Same goes with Acceleration and sprint speed. Also short passing and ball control have a strong relationship

**Regression Results**

By applying regression, I am predicting value of a player during transfer market with wages and overall as the predictors. On top of it, I also used Hyperparameter tuning using Grid Search to check if there is an improvement in accuracy with different parameter combinations.

Accuracy for the 3 Machine Learning Algorithms before and after applying Grid Search is listed below.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Random Forest** | **Multilayer Perceptron** | **Gradient Boosting** |
| **Before Grid Search** | 90.16 % | 72.54 % | 89.01 % |
| **After Grid Search** | 90.07 % | 71.97 % | 86.31 % |

**Hence I conclude that Random Forest Algorithm is the best for FIFA-18 dataset as per the analysis done above and the hyperparameters selected.**

**References**

1. <https://www.kaggle.com/piyushgandhi811/fifa-18-updated-dataset>
2. <https://plot.ly/python/choropleth-maps/>
3. <https://plot.ly/python/line-charts/>