**REPORT**

***FIFA 21 ANALYSIS***

**Introduction**

International Federation of Association Football (FIFA) is a global body that governs the entire football on this globe. From organizing events such as world cups and other competitions it also Rank the Teams as well as Individual players every year based on their skills and achievements.

In this Capstone Project we will be analysing individual players abilities, skills and based on that will be looking for some insights.

****

**About the Dataset**

The FIFA 2021 dataset consists of more than 17,000 active players around the world and more than 100 attributes for every player.

The dataset has been obtained online from Kaggle and is uncleaned and needs a lot of cleaning and preprocessing.

Among 107 attributes in the data, most of them are the skillsets of the players such as pace, jumping, attack, defence, speed, passing, shooting, dribbling etc.

A few other attributes consist of Players information such as Name, country, club, contracts, photo, rating, potential etc.

There are a lot of attributes which are of no use during our analysis so during pre-processing we will be dropping them.

**OBJECTIVES FOR ANALYSIS**

We will be working on a bunch of objectives for this project which will include data visualization and machine leaning models.

These are our objectives for this project:

* Top 5 highest rated players.
* Which player has the highest potential in game?
* Top 5 players for every playing position.
* Team of the year according to every highest rated position.
* Most valuable player?
* Which players are highly valuable but still underpaid.
* Most preferred foot by players?
* Longest stay in the Club

Machine Leaning models objectives:

* A regression model to predict rating based on skills of a player.

**Tools to be used for analysis:**

* Excel (Data understanding)
* Python (Data preprocessing, ML models/Data Visualization)
* Tableau (Data Visualization)

**Data Preprocessing and Cleaning:**

After understanding the data through Excel sheet, out of 107 attributes we will be dropping 42 columns directly which won’t be helpful in our analysis.

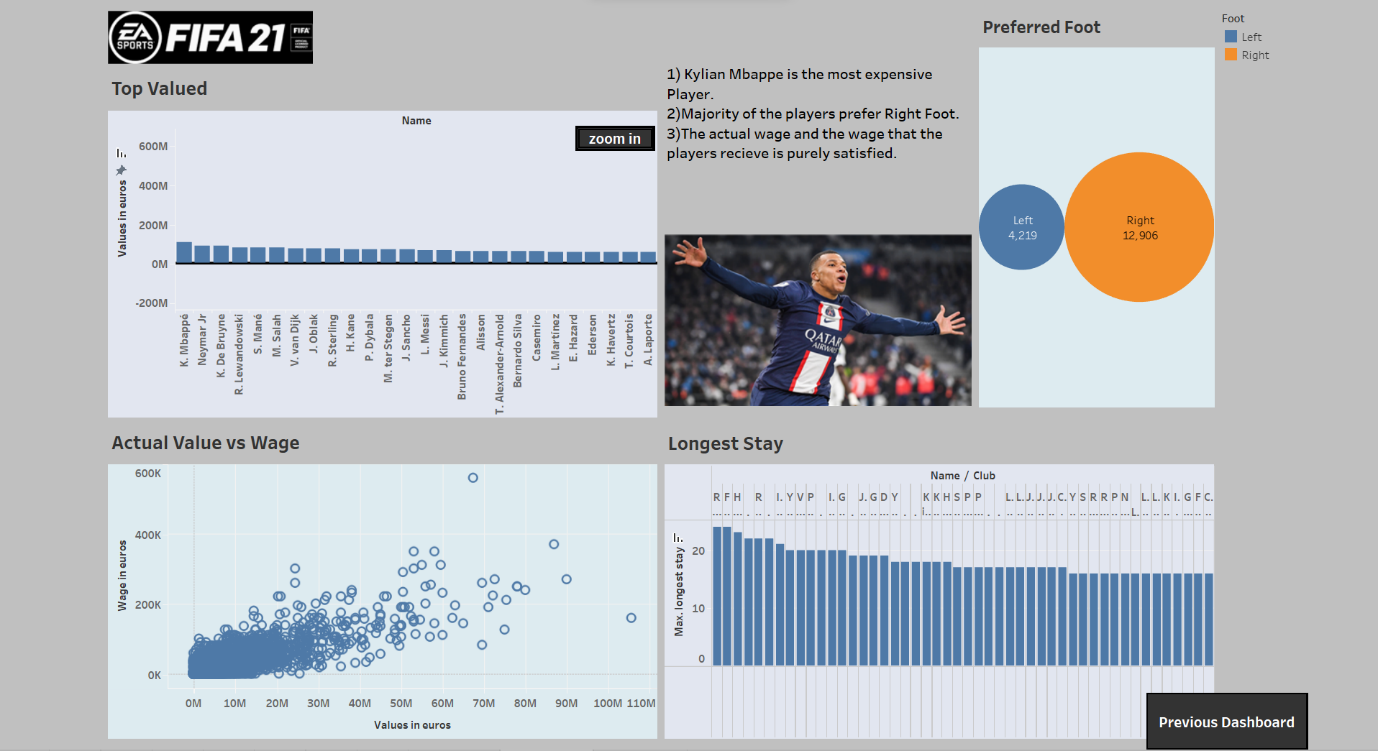
Then we have some missing values in and noise in our dataset such as:

* Height and Weight attributes need to be converted into numeric.
* Value, Wage, Release Clause have unexpected signs which make it object datatype so we will handle those columns too.
* Splitting the Contract column into Two columns.
* Exporting the Cleaned Dataset.

**VISUALIZTIONS**

* Importing Cleaned dataset in Tableau and getting some insights on the players and their stats.
* Making interactive dashboards.

****



**MACHINE LEARNING MODEL**

For a Machine Learning model we need to further preprocess the dataset.

The dataset is divided into input and output variables x and y.

We have 50 features in our input variable which contains missing values. These missing values needs to be handled properly.

* The first technique is to replace missing values with Mean value.

After handling the missing values we import Lasso Regression class from scikit learn library and do all the necessary steps to train and fit our model.

Lasso Regression is helpful when we have a lot of input features and make a generalised model.

The trained model have an Accuracy score of 88%.

After Cross Validating for 5 folds we still get some decent accuracy and can confirm that our model is generalised with some decent accuracy score.

**Conclusion:**

From Visualization in Tableau, we got most of our answers such as Lionel Messi, Cristiano Ronaldo are top 2 players of FIFA 2021.

Kylian Mbappe is the most Expensive Player as well as with the Highest Potential in the Game.

Most Players prefer Right foot while playing.

From the Machine Learning Model, we concluded that our model is capturing most of the features well and is giving as good prediction on unseen dataset.

**Dataset from: Kaggle**