

1. Introduction

1.1 Problem Statement

To analyse FIFA data using python.

1.2 Purpose of the project

The main objective of this sport analysis is to get insights and correlations between players values , wage , age , special attributes and performance. This integrated data can be further converted into information. By analysing it, I have derived some statistics for teams , clubs and player through extensive football experience.

1.3 Features

- 1) The insights provide in the results along with the understanding and contextualize information , enables user act smartly when playing Free-Fall picking a better team for fantasy premier league or increasing their betting odds.
- 2) It tells us insights about your player before forming a dream team or betting.

1.4 Detailed Description

This Sports Analysis enables detailed analysis of the game. Such granularity can offer predictors more information on how the players can perform in said conditions and their Expected performance .It helps user to discover counter-intuitive facts about football , upon which one can build a winning team.

Instead of looking at single variable and metrics we can start aggregating them , Creating specific performance indications. By the analysis one can clearly identify the best player in one specific , performance level during season or even throughout their entire career. Using performance analysis it is easier to sport outliers. Data analysis becomes a tool to help predict and cultivate players' potential. At the end of the ,showing objective facts is a more efficient and persuasive way to state a point.

2.Minimum Requirements

2.1 Hardware Requirements

- 1 GB RAM
- CPU - INTEL PENTIUM, 1.8GHz
- HDD - 100 GB

2.2 Software Requirements

- WINDOWS XP/7
- JUPYTER NOTEBOOK
- LAIBRARIES – PANDAS , NUMPY,MATPLOTLIB,SEABOARN

3. Proposed Work

3.1 DATASET

The data is crap from sofifa website using python crawler, the data that comprises every player featured in FIFA18 .

It has 9 attributes from player name , flag images , playing position data , attributes based on actual data of the latest FIFA18.

Other attributes include players' style statistics also players' personal data.

EDA is used to conclude following points.

1. World Statistics
2. Clustering players by nationality
3. Value and wages of players with age
4. Value of players with position
5. Overall & Potential with wage
6. Variation of overall & potential by country for top countries
7. Variation of wages for top clubs
8. Make your dream team
9. Predicting player position using player statistics.

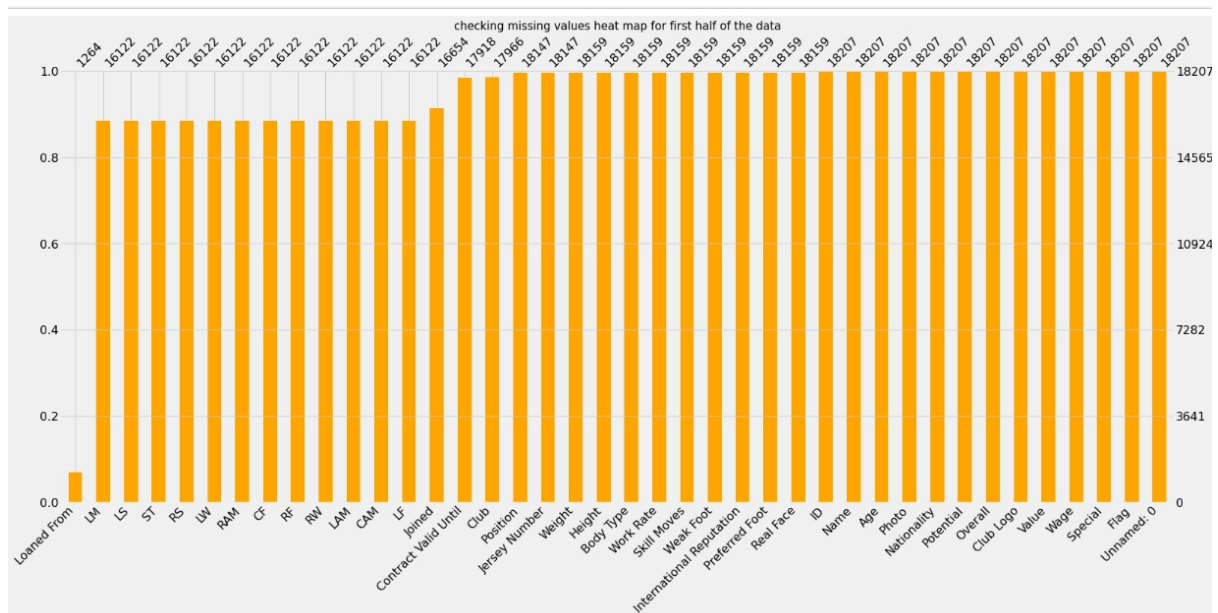
DATA CLEANING

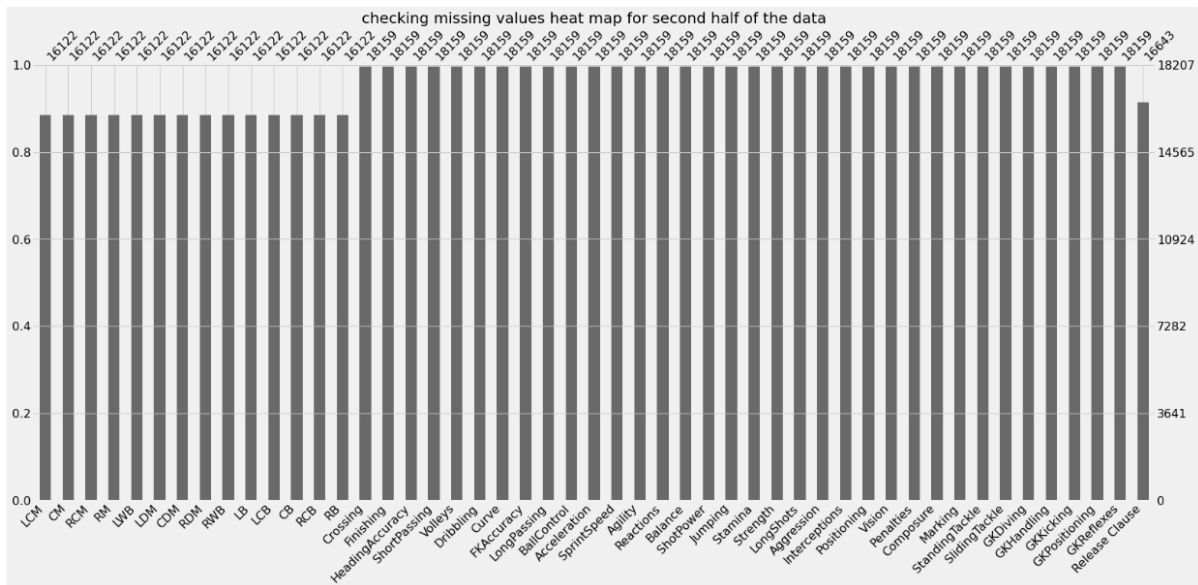
The section includes detecting and correcting or removing the corrupt data or inaccurate records from data set. And then replacing , modifying or deleting the dirty or coarse data. All unnecessary data is cleaned at the end of data cleaning process.

Data cleaning is crucial and emphasized because wrong data can drive to wrong decisions, conclusions.

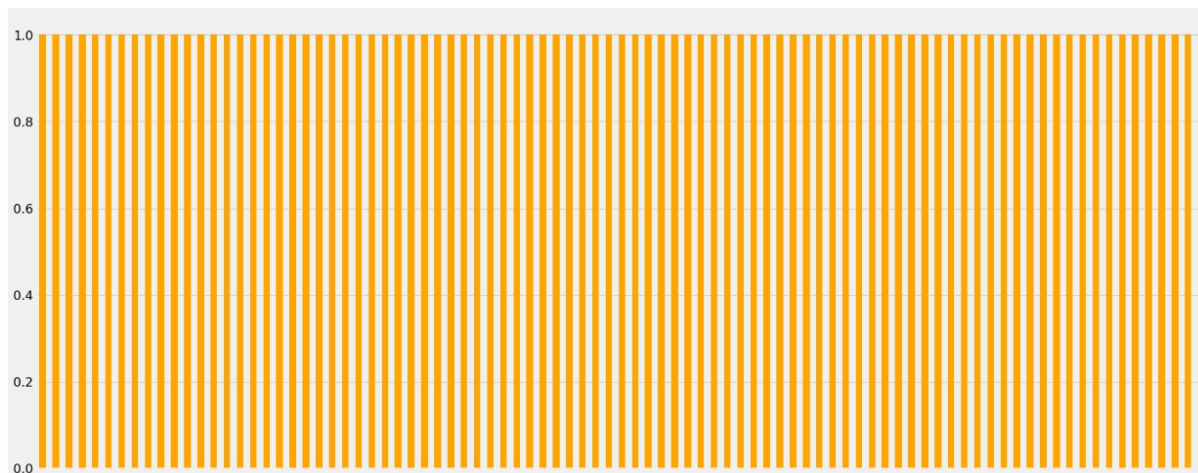
Values before cleaning:

For first half and second half:





Values after cleaning:



FEACHER ENGINEERING

The section includes the process of using domain knowledge to extract features from raw data via data mining techniques.

```
data.columns #after feacher engineering
```

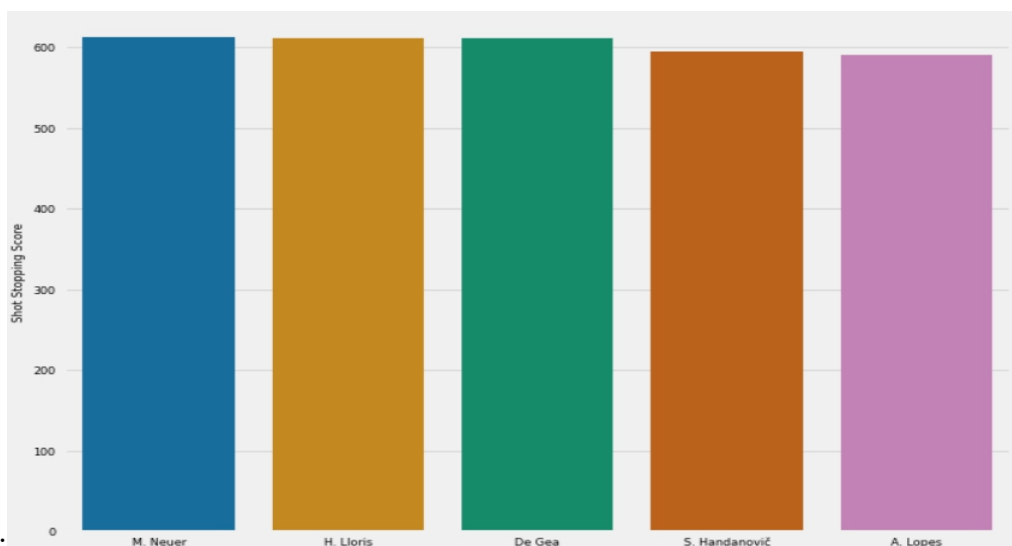
```
Index(['Unnamed: 0', 'ID', 'Name', 'Age', 'Photo', 'Nationality', 'Flag',
      'Overall', 'Potential', 'Club', 'Club Logo', 'Value', 'Wage', 'Special',
      'Preferred Foot', 'International Reputation', 'Weak Foot',
      'Skill Moves', 'Work Rate', 'Body Type', 'Real Face', 'Position',
      'Jersey Number', 'Joined', 'Loaned From', 'Contract Valid Until',
      'Height', 'Weight', 'LS', 'ST', 'RS', 'LW', 'LF', 'CF', 'RF', 'RW',
      'LAM', 'CAM', 'RAM', 'LM', 'LCM', 'CM', 'RCM', 'RM', 'LWB', 'LDM',
      'CDM', 'RDM', 'RWB', 'LB', 'LCB', 'CB', 'RCB', 'RB', 'Crossing',
      'Finishing', 'HeadingAccuracy', 'ShortPassing', 'Volleys', 'Dribbling',
      'Curve', 'FKAccuracy', 'LongPassing', 'BallControl', 'Acceleration',
      'SprintSpeed', 'Agility', 'Reactions', 'Balance', 'ShotPower',
      'Jumping', 'Stamina', 'Strength', 'LongShots', 'Aggression',
      'Interceptions', 'Positioning', 'Vision', 'Penalties', 'Composure',
      'Marking', 'StandingTackle', 'SlidingTackle', 'GKDividing', 'GKHandling',
      'GKKicking', 'GKPositioning', 'GKReflexes', 'Release Clause',
      'Defending', 'General', 'Mental', 'Passing', 'Mobility', 'Power',
      'Rating', 'Shooting'],
      dtype='object')
```

DATA VISUALIZATION

The section includes visualizing data in order to take insights from it.

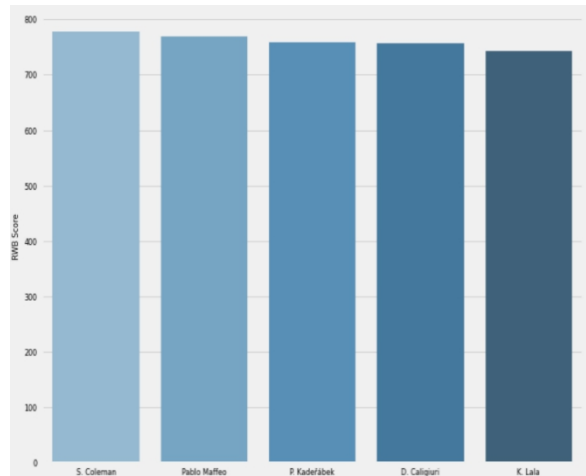
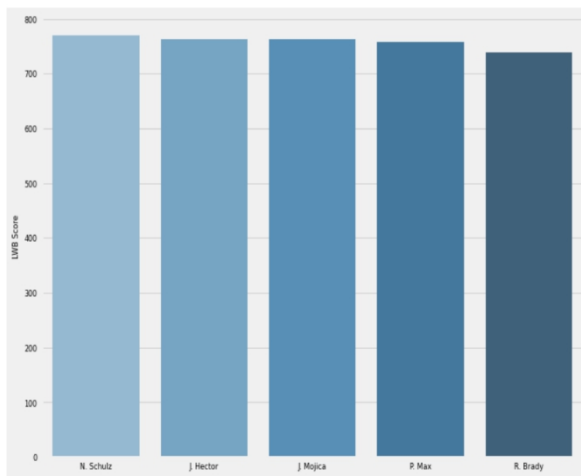
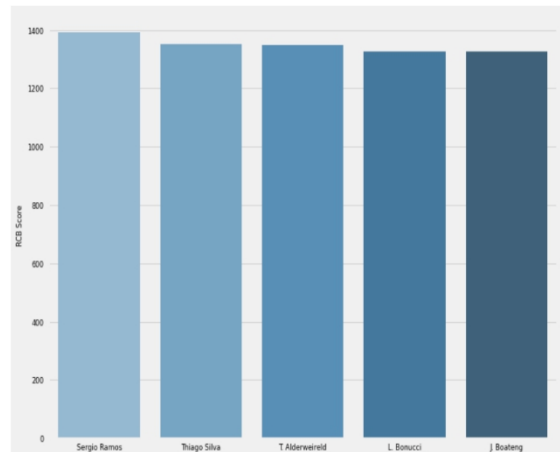
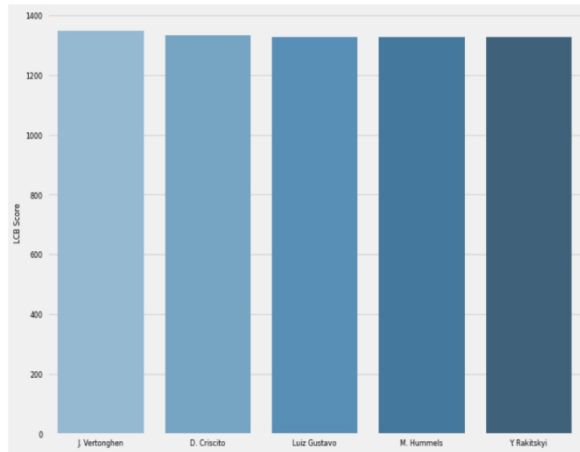
By the data visualization technique it is concluded that :

Best

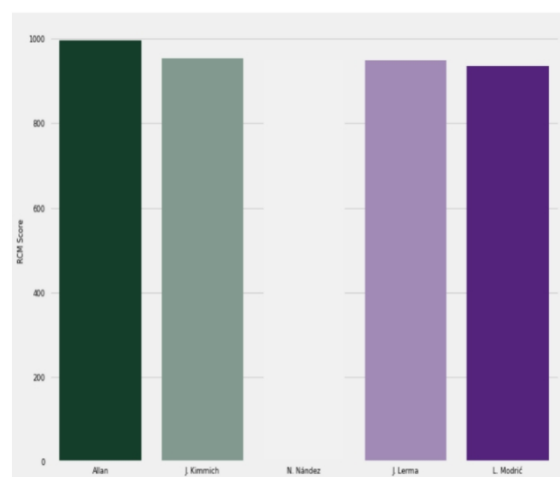
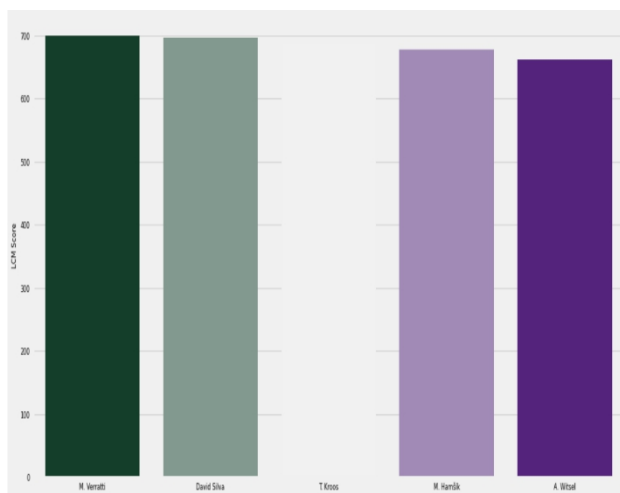


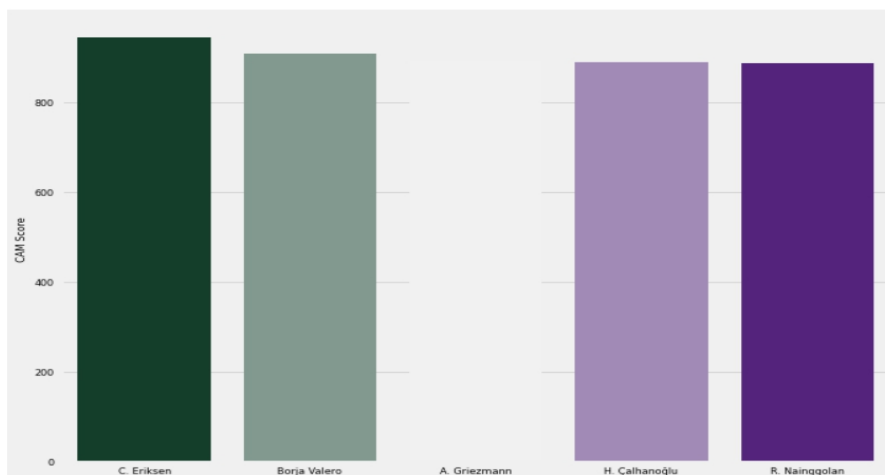
Goalkeeper:

Best Defenders (4):

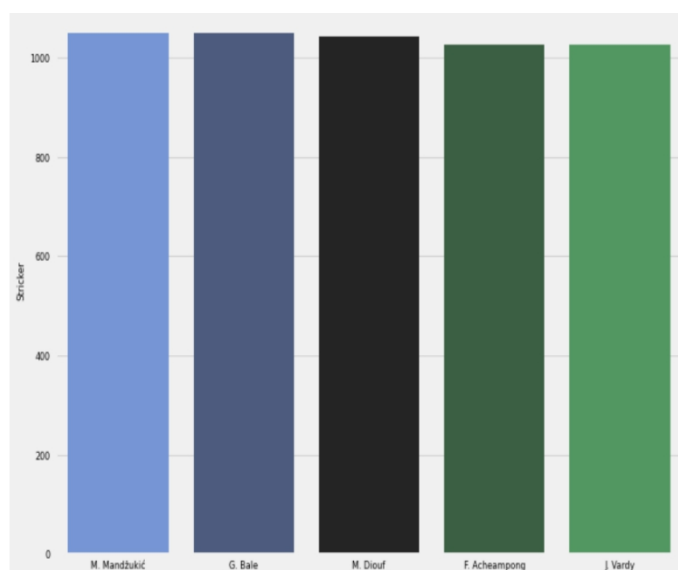
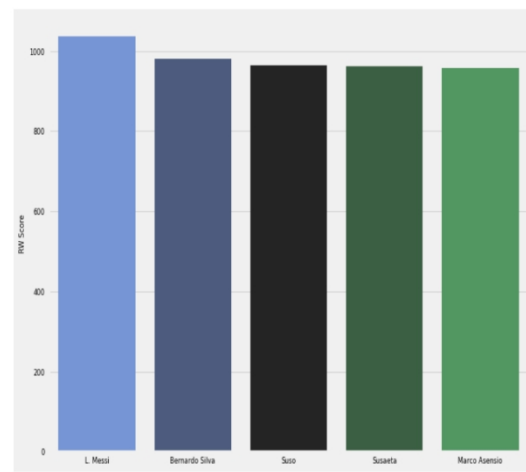
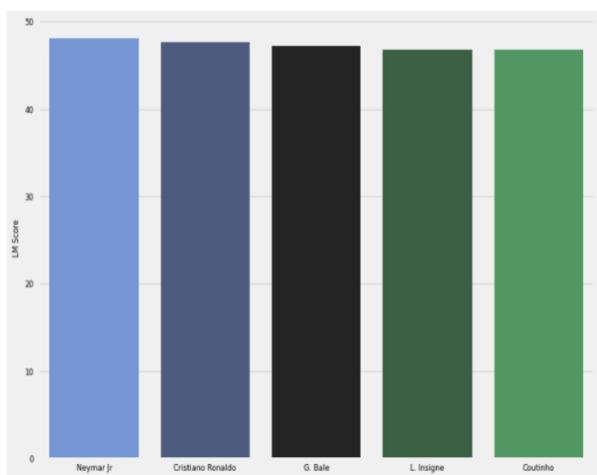


Best Mid-Fielders (3):





Best Attackers (3):



The dream team is :



6.Observation and conclusion

1. EDA can give us a good insight into a dataset. It is a crucial part of any data science pipeline before we learning machine learning models.
2. Data visualizations are most significant part to provide quick & clear understanding of the information .
3. Graphic representations we can visualize large volumes of data in a understandable way Helps us to comprehend the information and draw conclusion .

6. References

Websites/ Web Links

<https://www.udemy.com/>

<https://www.kaggle.com/>

Date: 27/05/2021