Business Intelligence and Data Analytics Mini Project.

Sarvesh Gulhane	**(E1) 321018**	**21810873**
Shreya Jadhav	**(E1) 321019**	**21810101**
Kartik Mawa	**(E1) 321025**	**21810747**

Aim:

To analyze the data of a player's performance statistics in order to predict their stakes, rates and positioning done by the football club management during the game seasons, in order to help football clubs during auction.

Objectives:

1) To get overall analysis of every player by creating a dashboard.
2) To analyze physical fitness and abilities of the player and compare them.
3) To predict the minimum base value of the player in order to help club management during the player auctioning.

Theory:

What id Data Analytics?

Data analytics is the science of analysing raw data in order to make conclusions about that information. Many of the techniques and processes of data analytics have been automated into mechanical processes and algorithms that work over raw data for human consumption.

Data analytics techniques can reveal trends and metrics that would otherwise be lost in the mass of information. This information can then be used to optimize processes to increase the overall efficiency of a business or system.

What tool has been used?

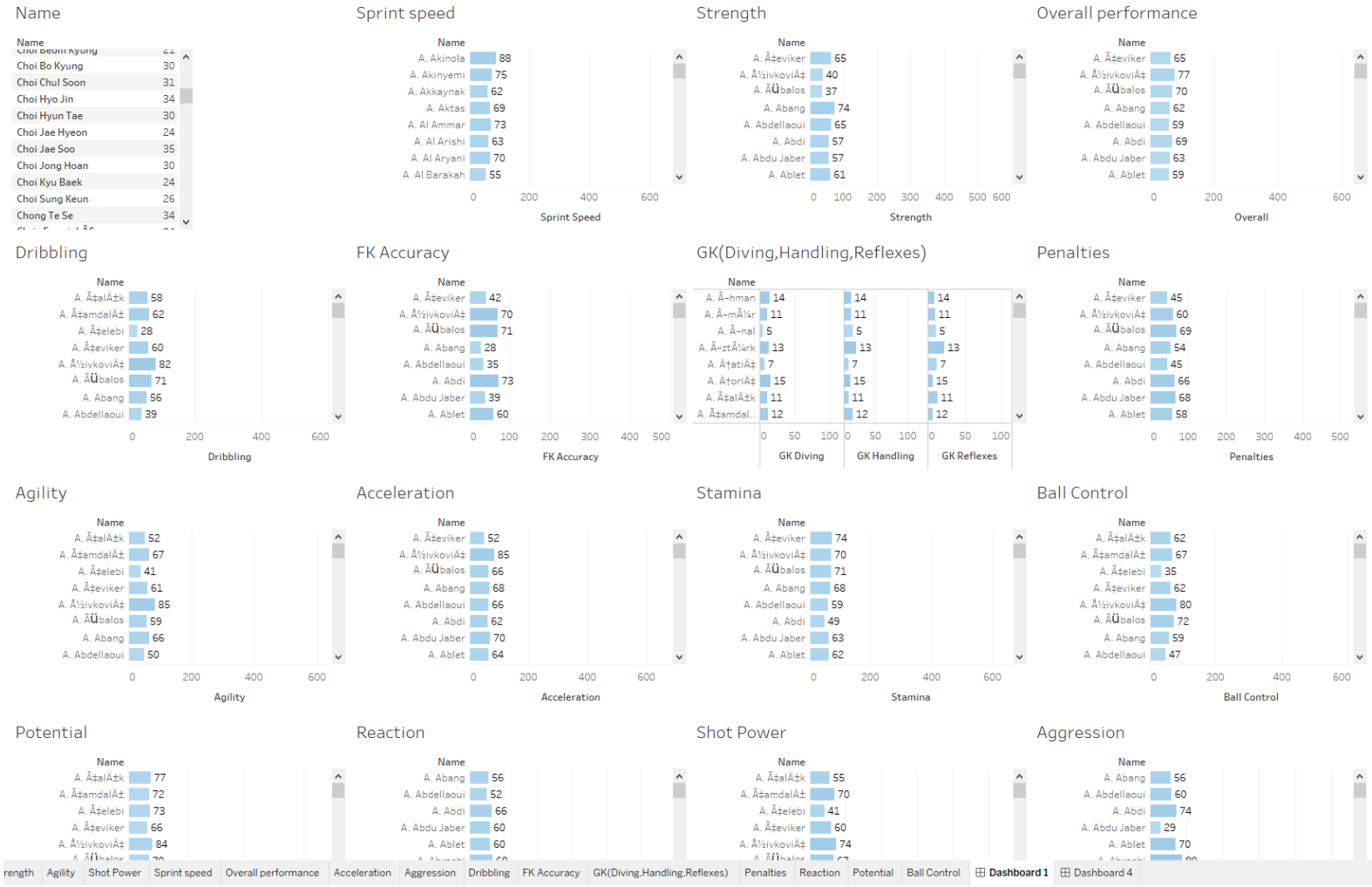
- **RapidMiner** is a data science software platform developed by the company of the same name that provides an integrated environment for data preparation, machine learning, deep learning, text mining, and predictive analytics. It is used for business and commercial applications as well as for research, education, training, rapid prototyping, and application development and supports all steps of the machine learning process including data preparation, results visualization, model validation and optimization. RapidMiner is developed on an open core model.

- **Tableau** is a data visualization and data analytics tool that aims to help people see and understand data. In order to transform the way people, use data to solve problems, tableau software ensures to meet strict requirements. In other words, it simply converts raw data into a very easily understandable format.
  Data analysis is great, as it is a powerful visualization tool in the business intelligence industry. Data that is created using this software becomes so easy that it allows even a non-technical user to create a customized dashboard. It provides top class interactive data visualization with the purpose to help organizations solve their data problems.

Input Data:

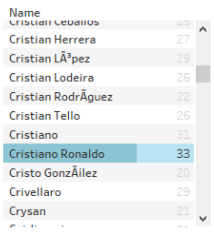
The dataset provides us with the information about all the football players on the basis of their basic information (like name, age, etc.), their physic information like weight and height, their skills and their wages through their career.

Output:
Dashboard:

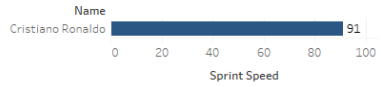


Stats of Ronaldo

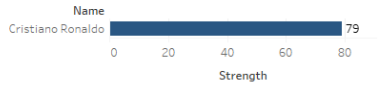
Name



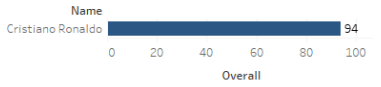
Sprint speed



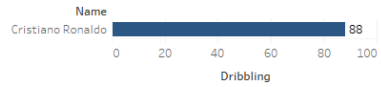
Strength



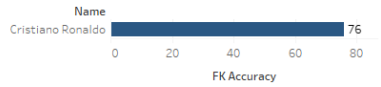
Overall performance



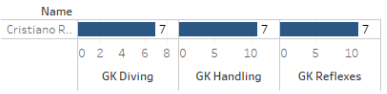
Dribbling



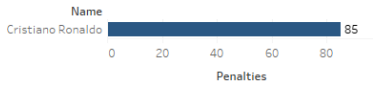
FK Accuracy



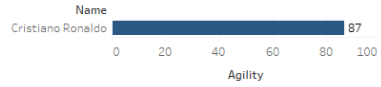
GK(Diving,Handling,Reflexes)



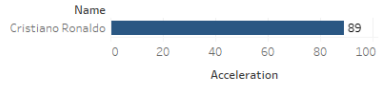
Penalties



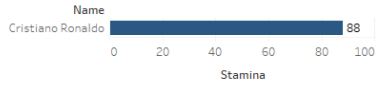
Agility



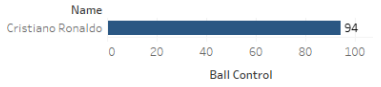
Acceleration



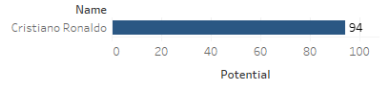
Stamina



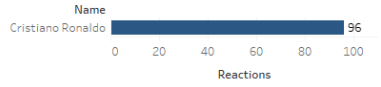
Ball Control



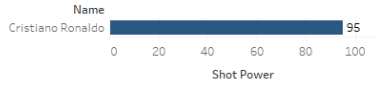
Potential



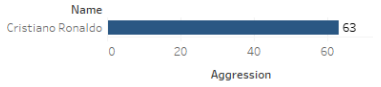
Reaction



Shot Power

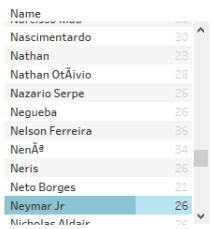


Aggression

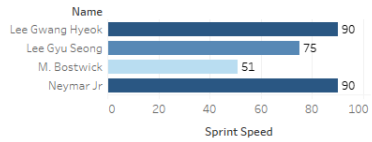


Comparing stats of players:

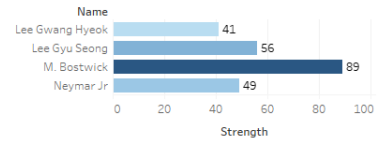
Name



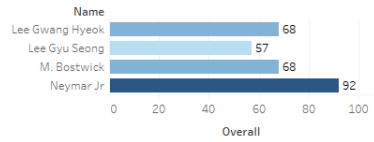
Sprint speed



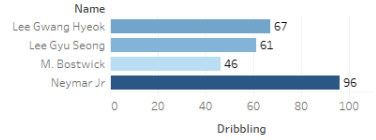
Strength



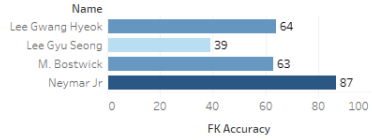
Overall performance



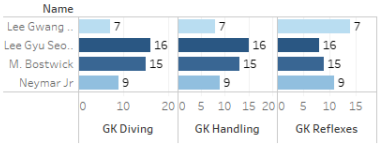
Dribbling



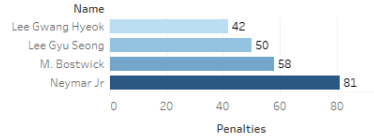
FK Accuracy



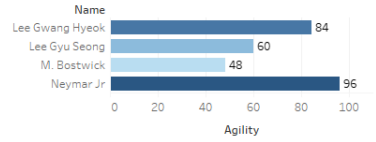
GK(Diving,Handling,Reflexes)



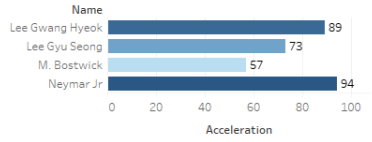
Penalties



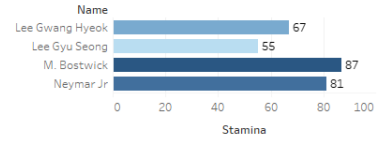
Agility



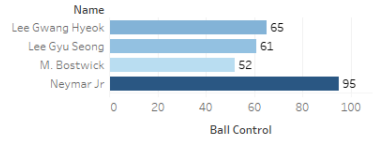
Acceleration



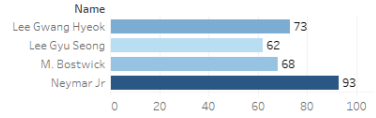
Stamina



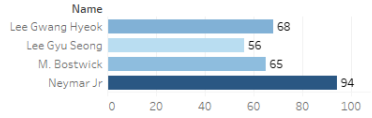
Ball Control



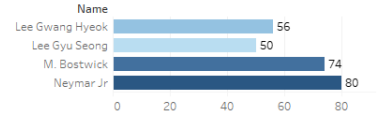
Potential



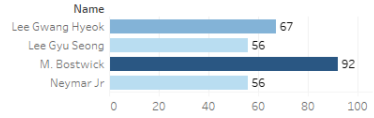
Reaction



Shot Power

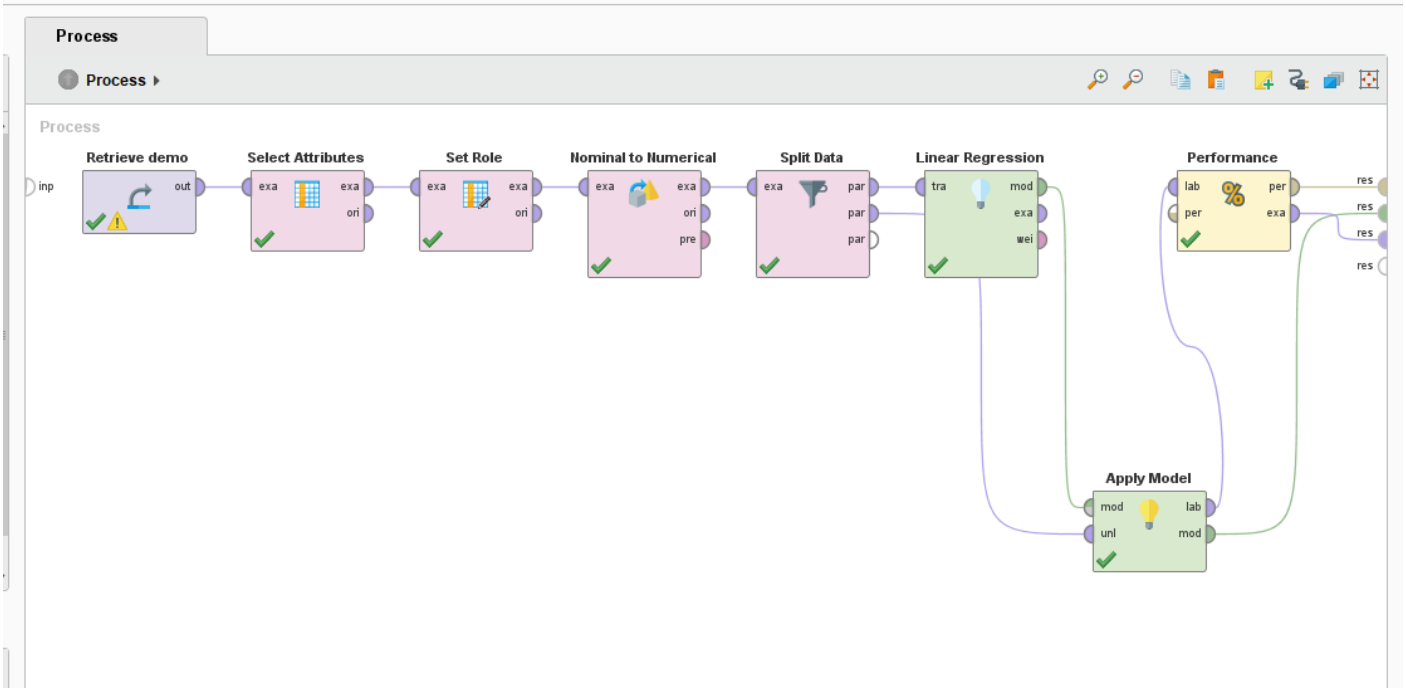


Aggression



Rapidminer Process:

Applying Regression using rapidminer:



- First we inserted dataset
- Select Attribute: Important attribute were selected from data.
- Set Role: We need to label target variable in this case release clause and Wage are labelled.
- Nominal to Numerical: Convert nominal data into numeric data.
- Split Data: Splitting data for training and testing purpose
- Linear Regression: Data given to Linear Regression model for training.
- Apply Model: Model applied on testing data.
- Performance: Checked performance score.

Predicted values (Release Clause):

Row No.	Release Cla...	prediction(R...	Position = RF	Position = ST	Position = LW	Position = R...	Position = LF	Position = RS	Position = R...	Position = L...	Position = CB	Position = L...
1	226.500	207.228	1	0	0	0	0	0	0	0	0	0
2	164	143.483	0	0	0	0	0	1	0	0	0	0
3	104.600	94.825	0	0	0	0	0	0	1	0	0	0
4	160.700	152.326	0	1	0	0	0	0	0	0	0	0
5	165.800	143.066	0	0	0	0	0	0	0	0	0	0
6	111	125.754	0	0	0	0	0	0	0	0	0	0
7	119.300	122.616	0	1	0	0	0	0	0	0	0	0
8	147.700	145.951	0	0	1	0	0	0	0	0	0	0
9	88.200	79.480	0	0	0	0	0	0	0	0	0	0
10	123	121.547	0	1	0	0	0	0	0	0	0	0
11	44.400	40.194	0	0	0	0	0	0	1	0	0	0
12	105.900	95.189	0	0	0	0	0	0	0	1	0	0
13	60.200	61.397	0	0	0	0	0	0	0	0	0	0
14	74.300	90.704	0	0	0	0	0	0	0	0	0	0
15	49.500	57.653	0	0	0	0	0	0	1	0	0	0
16	75.100	79.726	0	0	0	0	0	0	1	0	0	0
17	72.600	83.573	0	0	0	0	0	0	0	0	0	0
18	33.300	38.950	0	0	0	0	0	0	0	0	0	0
19	89.600	103.959	0	0	0	0	0	0	0	0	0	0
20	69.900	83.544	0	0	0	1	0	0	0	0	0	0
21	107.300	104.621	0	0	0	1	0	0	0	0	0	0
22	88.900	88.382	0	1	0	0	0	0	0	0	0	0
23	66.600	53.823	0	0	0	0	0	0	0	0	0	0
24	52.700	64.312	0	0	0	0	0	0	0	0	1	0
25	78	67.377	0	0	0	0	0	0	0	0	0	0

PerformanceVector

```
PerformanceVector:
root_mean_squared_error: 10.374 +/- 0.000
squared_correlation: 0.930
```

Predicted values (Wage):

Row No.	Wage	prediction(...	Position = RF	Position = ST	Position = LW	Position = R...	Position = LF	Position = RS	Position = R...	Position = L...	Position = CB	Position = L...
1	565	431.457	1	0	0	0	0	0	0	0	0	0
2	455	515.303	0	0	0	0	0	1	0	0	0	0
3	380	311.950	0	0	0	0	0	0	1	0	0	0
4	205	275.992	0	1	0	0	0	0	0	0	0	0
5	145	260.292	0	0	0	0	0	0	0	0	0	0
6	200	254.189	0	0	0	0	0	0	0	0	0	0
7	300	293.065	0	1	0	0	0	0	0	0	0	0
8	340	275.996	0	0	1	0	0	0	0	0	0	0
9	285	261.624	0	0	0	0	0	0	0	0	0	0
10	355	259.296	0	1	0	0	0	0	0	0	0	0
11	165	173.934	0	0	0	0	0	0	1	0	0	0
12	135	41.202	0	0	0	0	0	0	0	1	0	0
13	160	89.114	0	0	0	0	0	0	0	0	0	0
14	135	183.520	0	0	0	0	0	0	0	0	0	0
15	160	117.525	0	0	0	0	0	0	1	0	0	0
16	150	144.671	0	0	0	0	0	0	1	0	0	0
17	180	140.873	0	0	0	0	0	0	0	0	0	0
18	185	63.159	0	0	0	0	0	0	0	0	0	0
19	73	82.054	0	0	0	0	0	0	0	0	0	0
20	92	119.162	0	0	0	1	0	0	0	0	0	0
21	77	178.234	0	0	0	1	0	0	0	0	0	0
22	180	144.438	0	1	0	0	0	0	0	0	0	0
23	96	155.015	0	0	0	0	0	0	0	0	0	0
24	170	97.824	0	0	0	0	0	0	0	0	1	0
25	28	92.744	0	0	0	0	0	0	0	0	0	0

PerformanceVector

```
PerformanceVector:
root_mean_squared_error: 72.337 +/- 0.000
squared_correlation: 0.579
```