**PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS  
NÚCLEO DE EDUCAÇÃO A DISTÂNCIA**

**Pós-graduação *Lato Sensu* em Ciência de Dados e Big Data**

**Alan Felix**

**Modelo de analise exploratória de predição do resultado de campeonatos do Premier League**

Belo Horizonte

2020

**Alan Felix**

**MODELO DE analise exploratória de predição do resultado de campeonatos do Premier League**

Trabalho de Conclusão de Curso apresentado ao Curso de Especialização em Ciência de Dados e Big Data como requisito parcial à obtenção do título de especialista.

Belo Horizonte

2020**SUMÁRIO**

[1. Introdução 4](#_Toc37607636)

[1.1.Contextualização 4](#_Toc37607637)

[1.2. O problema proposto 4](#_Toc37607638)

[2. Coleta de Dados 6](#_Toc37607639)

[3. Processamento/Tratamento de Dados 11](#_Toc37607640)

[3.1. Dados Brutos 11](#_Toc37607641)

[3.1. Dados Preparados 14](#_Toc37607642)

[4. Análise e Exploração dos Dados 16](#_Toc37607643)

[4.1 Média geral do clube por estatura dos jogadores 16](#_Toc37607644)

[4.2 Média geral do clube por idade dos jogadores 18](#_Toc37607645)

[4.3 Média geral do clube por avaliação geral dos jogadores 20](#_Toc37607646)

[4.4 Média geral do clube por valor de mercado dos jogadores 22](#_Toc37607647)

[5. Apresentação dos resultados 25](#_Toc37607648)

[6. Links 26](#_Toc37607649)

[REFERÊNCIAS 27](#_Toc37607650)

# 1. Introdução

## Contextualização

Este material é o resultado de uma analise exploratória do jogo FIFA 2019 Ultimate Team da EA Sports com os resultados do Premier League 2018/2019 estudo que tenta identificar os principais requisitos para prever um possível campeão da próxima edição do Premier League 2019/2020.

## 1.2. O problema proposto

**Por quê?**

Na área apostas esportiva é sempre importante descobrir as principais caraterísticas do elenco de um clube para possibilitar a previsão dos resultados do campeonato, o jogo FIFA Ultimate Team da EA Sports tem dados extremamente próximo à realidade das qualificações das equipes, que pode ser muito útil para predição do resultado do Premier League.

**Para quem?**

Este estudo visa apresentar um modelo de analise exploratória de predição de resultado de campeonatos do Premier League, para apostadores das casas de apostas nos países onde existe a regulamentação para este tipo de aposta esportiva.

**O que?**

Este estudo trata da apresentação de um modelo estatístico que utilizou como base o DataSet do jogo FIFA 2019 Ultimate Team da EA Sports e DataSet do Premier League 2018/2019 de domínio público, que podem ser encontrados no Site kaggle, que relaciona as caraterísticas dos jogadores com os resultados das equipes no Premier League.

**Onde?**

Este modelo estatístico atende apenas a resultados de apostas esportivas na liga europeia de futebol, porém poderá ser utilizada por apostadores das casas de apostas em qualquer país que tenha a regulamentação para apostas esportiva.

**Quando?**

Foram analisados dados do jogo FIFA 2019 e do Premier League 2018/2019, visando estabelecer um modelo de analise exploratória do Premier League 2019/2020, com base nos dados jogo FIFA 2020.

# 2. Coleta de Dados

Os dados do jogo FIFA Ultimate Team da EA Sports utilizados foram coletados no Site kaggle: <https://www.kaggle.com/karangadiya/fifa19>

|  |  |  |  |
| --- | --- | --- | --- |
| **Field Name** | **Nome do Campo** | **Descrição** | **Tipo** |
| ID | ID | ID único para cada jogador | Número |
| Name | Nome | nome | Texto |
| Age | Idade | idade | Número |
| Photo | Foto | URL para a foto do jogador | Texto |
| Nationality | Nacionalidade | nacionalidade | Texto |
| Flag | Bandeira | URL para a bandeira do país dos jogadores | Texto |
| Overall | Macacão | avaliação geral | Número |
| Potential | Potencial | classificação potencial | Número |
| Club | Clube | clube atual | Texto |
| Club Logo | Logotipo do clube | URL para o logotipo do clube | Texto |
| Value | Valor | valor atual de mercado | Moeda |
| Wage | Salário | salário atual | Moeda |
| Special | Especial | especial | Número |
| Preferred Foot | Pé preferido | esquerda direita | Texto |
| International Reputation | Reputação internacional | classificação na escala de 5 | Número |
| Weak Foot | Pé fraco | classificação na escala de 5 | Número |
| Skill Moves | Movimentos de habilidades | classificação na escala de 5 | Número |
| Work Rate | Ritmo de trabalho | taxa de trabalho de ataque / taxa de trabalho de defesa | Texto |
| Body Type | Tipo de corpo | tipo de corpo do jogador | Texto |
| Real Face | Face real | verdadeiro ou falso | Boleano |
| Position | Posição | posição em campo | Texto |
| Jersey Number | Número da camisa | número da camisa | Número |
| Joined | Cadastrado | data de ingresso | Data |
| Loaned From | Emprestado | nome do clube, se aplicável | Texto |
| Contract Valid Until | Contrato válido até | data de término do contrato | Data |
| Height | Altura | altura do jogador | Texto |
| Weight | Peso | peso do jogador | Texto |
| LS | LS | classificação na escala de 100 | Número |
| ST | ST | classificação na escala de 100 | Número |
| RS | RS | classificação na escala de 100 | Número |
| LW | LW | classificação na escala de 100 | Número |
| LF | LF | classificação na escala de 100 | Número |
| CF | CF | classificação na escala de 100 | Número |
| RF | RF | classificação na escala de 100 | Número |
| RW | RAM | classificação na escala de 100 | Número |
| LAM | LAM | classificação na escala de 100 | Número |
| CAM | CAM | classificação na escala de 100 | Número |
| RAM | RAM | classificação na escala de 100 | Número |
| LM | LM | classificação na escala de 100 | Número |
| LCM | LCM | classificação na escala de 100 | Número |
| CM | CM | classificação na escala de 100 | Número |
| RCM | RCM | classificação na escala de 100 | Número |
| RM | RM | classificação na escala de 100 | Número |
| LWB | LWB | classificação na escala de 100 | Número |
| LDM | LDM | classificação na escala de 100 | Número |
| CDM | CDM | classificação na escala de 100 | Número |
| RDM | RDM | classificação na escala de 100 | Número |
| RWB | RWB | classificação na escala de 100 | Número |
| LB | LB | classificação na escala de 100 | Número |
| LCB | LCB | classificação na escala de 100 | Número |
| CB | CB | classificação na escala de 100 | Número |
| RCB | RCB | classificação na escala de 100 | Número |
| RB | RB | classificação na escala de 100 | Número |
| Crossing | Cruzamento | classificação na escala de 100 | Número |
| Finishing | Acabamento | classificação na escala de 100 | Número |
| HeadingAccuracy | Precisão no rumo | classificação na escala de 100 | Número |
| ShortPassing | Passe curto | classificação na escala de 100 | Número |
| Volleys | Voleios | classificação na escala de 100 | Número |
| Dribbling | Drible | classificação na escala de 100 | Número |
| Curve | Curva | classificação na escala de 100 | Número |
| FKAccuracy | Precisão FK | classificação na escala de 100 | Número |
| LongPassing | Passe longo | classificação na escala de 100 | Número |
| BallControl | Controle de bola | classificação na escala de 100 | Número |
| Acceleration | Aceleração | classificação na escala de 100 | Número |
| SprintSpeed | Velocidade de sprint | classificação na escala de 100 | Número |
| Agility | Agilidade | classificação na escala de 100 | Número |
| Reactions | Reações | classificação na escala de 100 | Número |
| Balance | Equilíbrio | classificação na escala de 100 | Número |
| ShotPower | Força de tiro | classificação na escala de 100 | Número |
| Jumping | Salto | classificação na escala de 100 | Número |
| Stamina | Vigor | classificação na escala de 100 | Número |
| Strength | Força | classificação na escala de 100 | Número |
| LongShots | LongShots | classificação na escala de 100 | Número |
| Aggression | Agressão | classificação na escala de 100 | Número |
| Interceptions | Intercepções | classificação na escala de 100 | Número |
| Positioning | Posicionamento | classificação na escala de 100 | Número |
| Vision | Visão | classificação na escala de 100 | Número |
| Penalties | Pênaltis | classificação na escala de 100 | Número |
| Composure | Composição | classificação na escala de 100 | Número |
| Marking | Marcação | classificação na escala de 100 | Número |
| StandingTackle | Equipamento de pé | classificação na escala de 100 | Número |
| SlidingTackle | Equipamento de deslizamento | classificação na escala de 100 | Número |
| GKDiving | Mergulho GK | classificação na escala de 100 | Número |
| GKHandling | Manuseio GK | classificação na escala de 100 | Número |
| GKKicking | Chute GK | classificação na escala de 100 | Número |
| GKPositioning | Posicionamento GK | classificação na escala de 100 | Número |
| GKReflexes | Reflexos GK | classificação na escala de 100 | Número |
| Release Clause | Cláusula de Liberação | valor da cláusula release | Moeda |

Os dados do Premier League 2018/2019 utilizados foram coletados no Site kaggle: <https://www.kaggle.com/thesiff/premierleague1819>

|  |  |  |  |
| --- | --- | --- | --- |
| **Field Name** | **Nome do Campo** | **Descrição** | **Tipo** |
| Team | Equipe | Nome da Equipe | Texto |
| category | categoria | Lugares na Liga dos Campeões, Lugares e Qualificações na Liga Europa, bem como equipes rebaixadas. | Texto |
| general league position | posição geral da liga | Posição geral na Liga | Número |
| finance live games televised | jogos ao vivo televisionados | Número de jogos ao vivo televisionados | Número |
| finance tv revenue | receitas de tv | Receita de contratos de TV | Número |
| general matches played | partidas gerais jogadas | Número partidas jogadas | Número |
| general won | geral ganho | Número jogos vencidos | Número |
| general draw | geral empates | Número jogos empatados | Número |
| general lost | geral perdido | Número jogos perdidos | Número |
| attack scored | ataque marcado | Número gols marcados | Número |
| defence goals conceded | gols de defesa sofridos | Número gols sofridos | Número |
| general goal difference | diferença geral de gols | Golos marcados - Golos sofridos | Número |
| general points | pontos gerais | Total de pontos - 3 pontos por vitória; 1 pontos por empate; 0 pontos por derrota | Número |
| general squad size | tamanho geral do esquadrão | Número de jogadores registrados para a liga premier | Número |
| general squad average age | idade média do esquadrão geral | Média de idade dos jogadores registrados para a liga premier | Número |
| general squad foreigners | estrangeiros do esquadrão geral | Número de jogadores estrangeiros | Número |
| finance team market | mercado de equipe financeira | Valor de mercado do time | Número |
| finance market average | média do mercado financeiro | Média do valor de mercado por jogador | Número |
| attack passes | passes de ataque | Número de passes de ataque | Número |
| attack passes through | passes por ataque | Número de passes por ataque | Número |
| attack passes long | passes longo de ataque | Número de passes longo de ataque | Número |
| attack passes back | passes de volta de ataque | Número de volta de ataque | Número |
| attack crosses | passes de cruzamento de ataque | Número de passes de cruzamento de ataque | Número |
| attack corners taken | escanteio | Número de escanteios | Número |
| attack shots | chutes a gol | Número de chutes a gol | Número |
| attack shots on target | chutes a gol no alvo | Número de chutes a gol no alvo | Número |
| attack goals headed | gols de cabeça | Número de gols de cabeça | Número |
| attack goals penalty | gols de pênalti | Número de gols de pênalti | Número |
| attack goals box | gols dentro da grande aréa | Número de gols dentro da grande área | Número |
| attack goals outsidebox | gols no ângulo | Número de gols no ângulo | Número |
| general card yellow | cartão geral amarelo | Número de cartão geral amarelo | Número |
| general card red | cartão geral vermelho | Número de cartão geral vermelho | Número |
| attack goals counter | gols contra | Número de gols contra | Número |
| attack goals freekick | gols de falta | Número de gols de falta | Número |
| defence saves | defesa em cima da linha | Número de defesa em cima da linha | Número |
| defence blocks | bloqueio de defesa | Número de bloqueio de defesa | Número |
| defence interceptions | interceptações de defesa | Número de interceptações de defesa | Número |
| defence tackles | contra ataque | Número de contra ataque | Número |
| defence tackles last man | defesa do último homem | Número de defesa do último homem | Número |
| defence clearances | defesas | Número de defesa | Número |
| defence clearances headed | defesas de cabeça | Número de defesas de cabeça | Número |
| defence penalty conceeded | penalidade de defesa | Número de penalidade de defesa | Número |
| attack posession | posse de ataque | Percentual de posse de bola no ataque | Número |
| attack pass accuracy | precisão do passe de ataque | Percentual de passes precisos de ataque | Número |

# 3. Processamento/Tratamento de Dados

Na análise foi seguido o seguinte processo:

# 3.1. Dados Brutos

**FIFA Ultimate Team 2019**

Conjunto de dados completo de jogadores de FIFA 19, mais de 18 mil jogadores do jogo FIFA 19 e 90 atributos extraídos do banco de dados mais recente da FIFA.

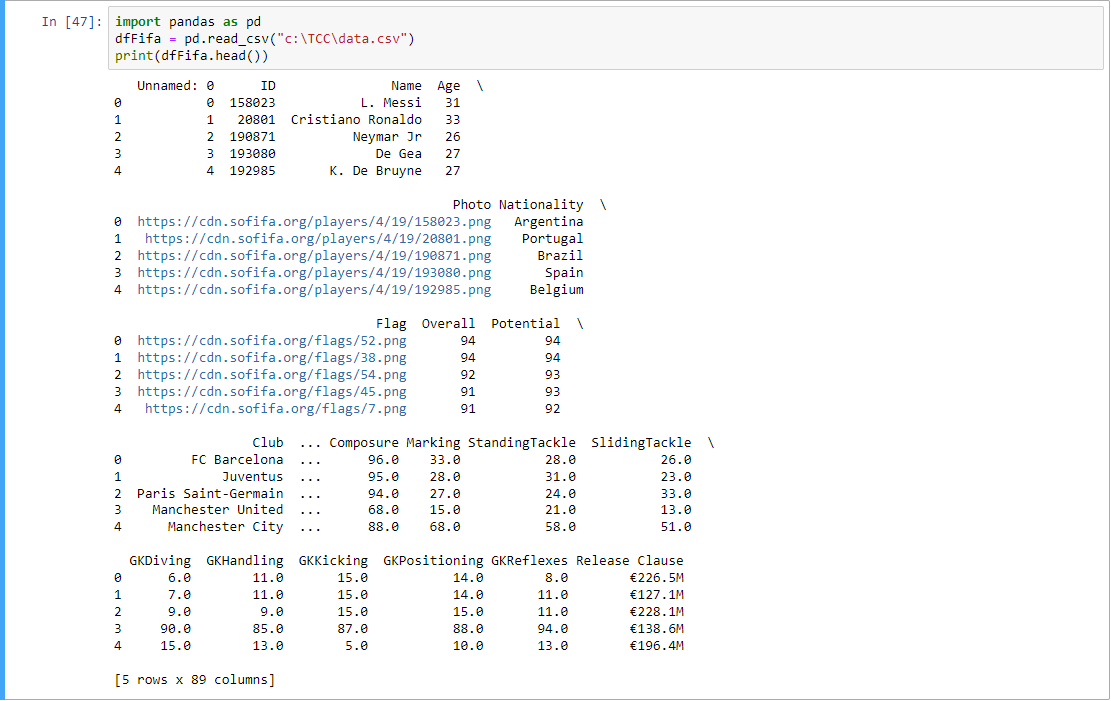
Contexto

Análise de futebol

Conteúdo

Atributos detalhados para todos os jogadores registrados na última edição do banco de dados do FIFA 19.

Carregando dados de Origem



**Premier League 2018/2019**

Dados do Premier League 2018-2019, dados de jogos do Premier League, bem como dados financeiros de todos os clubes da 2018-2019.

Contexto

Olhando para a temporada 2018-2019 e procurando aprofundar-se em percepções mais consistentes. Usando os dados para ver como os clubes são semelhantes estatisticamente, na maneira de passes de bola, ataque, marcação gols e mais.

Conteúdo

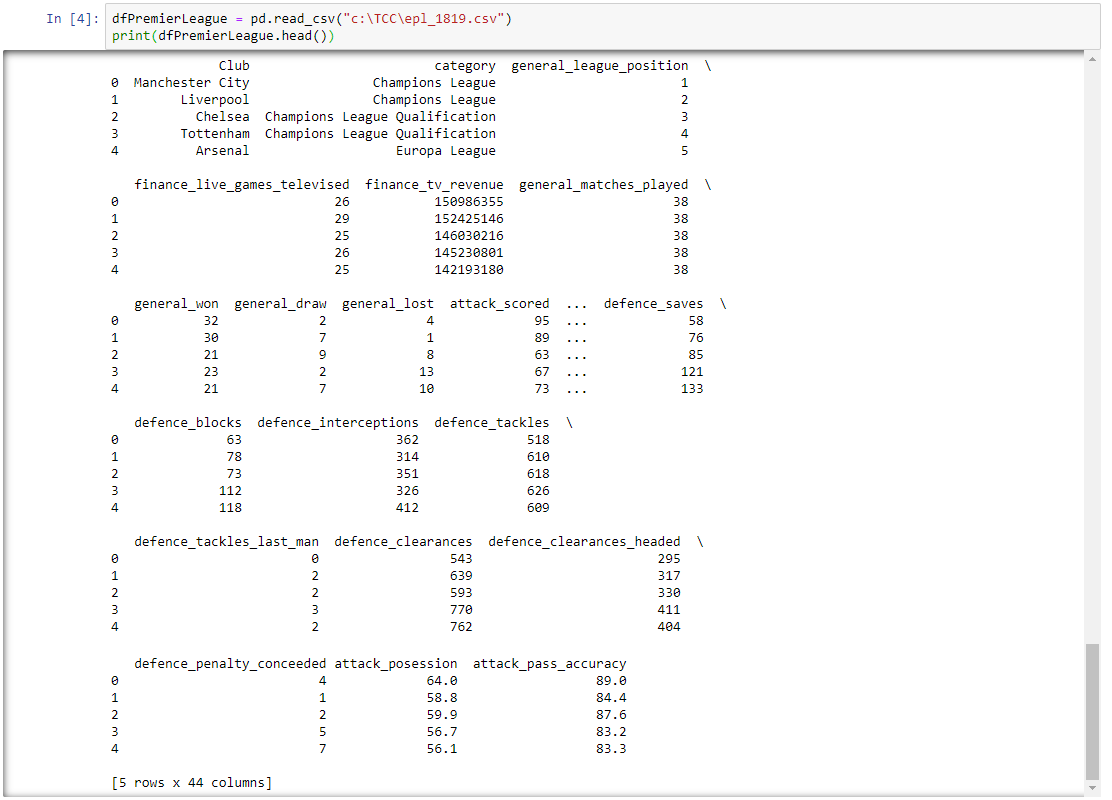
Esse conjunto de dados é amplo, no sentido em que abrange estatísticas vistas em uma tabela regular da liga, mas vai além de ver como as equipes passam e mantêm a posse de bola, como elas defendem, atacam e também analisam os valores de mercado de uma equipe e quanto dinheiro cada equipe foi alocada no acordo de direitos de TV.

Esses dados foram coletados de

1. BBC Sports Football,
2. Premierleague.com
3. Transfermarkt.co.uk

Esses dados não foram descartados no sentido convencional e aparecem de maneira aleatória. Para combater isso, incluí descritores de categoria no início de cada nome de variável, isso deve ajudar a fornecer uma compreensão mais coesa do conjunto de dados, bem como auxiliar na configuração.

Carregando dados de Origem

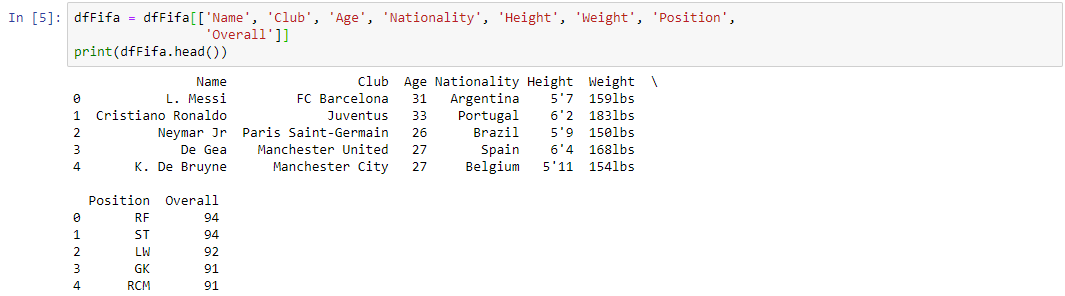


# 3.1. Dados Preparados

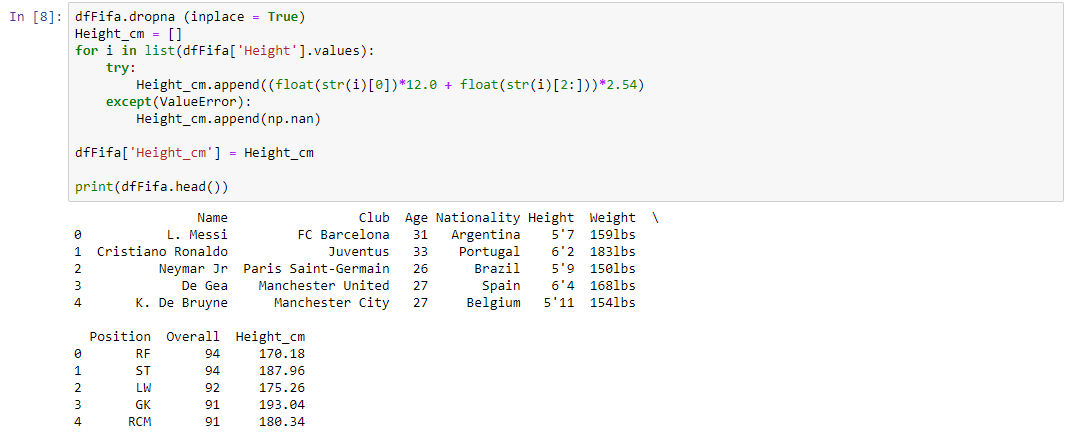
A partir do DataSet brutos foram feitos os tratamentos dados aos atributos utilizados na análise

**FIFA Ultimate Team 2019**

O DataFrame origem foi filtrado os dados a serem utilizados em nossa analise exploratória.

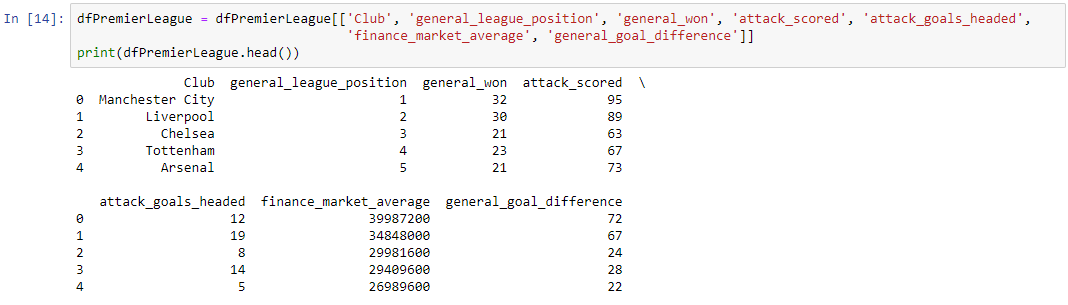


O atributo “Height” tem a altura do jogador em polegadas, será criado um novo atributo “Height\_cm” com a altura do jogador em centímetros e numérico para que possamos realizar os cálculos utilizados na análise.



**Premier League 2018/2019**

O DataFrame origem foi filtrado os dados a serem utilizados em nossa analise exploratória.



A partir dos DataFrame dfFifa e dfPremierLeague foi gerado um novo DataFrame dfFifaPremierLeague, através do atributo de ligação Club, filtrado apenas os jogadores do clubes participante da edição do Premier League 2018/2019.

# 4. Análise e Exploração dos Dados

Após o tratamento dos dados nulos, vazios e inválidos de origem foi gerado uma DataStage com dados do jogo FIFA Ultimate Team e outro DataStage com os dados do Premier League, onde foi feito um cruzamento obrigatório (INNER JOIN) entre a o DataFrame dfFifa com o DataFrame dfPremierLeague por Club. Nesse cruzamento apenas ficarão disponíveis para a análise os registros dos jogadores dos 20 clubes participantes da edição do Premier League 2019. O resultado do merge destes dois DataFrame é o DataFrame dfFifaPremierLeague, que a partir deste DataFrame foi gerado a analise exploratória das seguintes características dos clubes:

* Média geral do clube por estatura dos jogadores
* Média geral do clube por idade dos jogadores
* Média geral do clube por avaliação geral dos jogadores
* Média geral do clube por valor de mercado dos jogadores

# 4.1 Média geral do clube por estatura dos jogadores

Abaixo segue a analise da média de estatura dos jogadores pelo total de gols marcados de cabeça agrupados por clube e ordenados pela classificação geral do clube na edição do Premier League dos dados analisados.

**Comando de Consulta:**



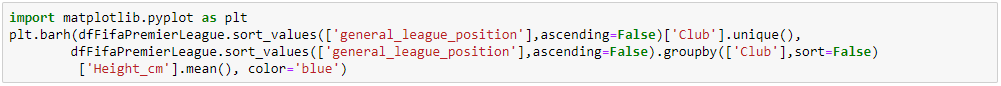
**Resultado:**



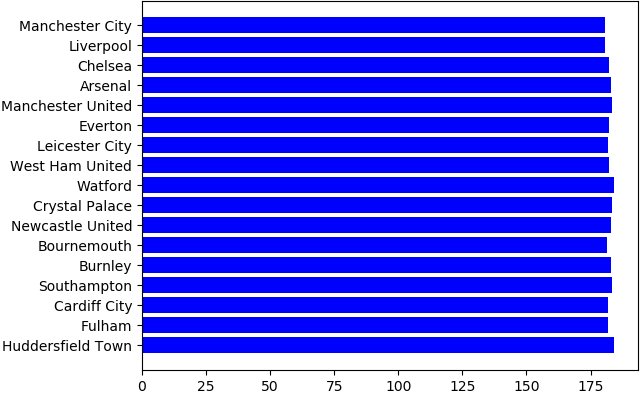
Observamos que a estatura média dos jogares do clube não reflete com exatidão ao número de gols marcados de cabeça, mostrado que os clubes a maior média de estatura dos jogadores não tem resultados expressivos na quantidade de gols marcados de cabeça.

Agora vamos realizar a analise através de um gráfico de barras, com objetivo de identificar o nível de influência do atributo estatura dos jogadores no resultado do Premier League.

**Comando de Geração do Gráfico:**



**Resultado:**



Observamos que os dois clubes com a menor média de estatura dos jogadores atingiram os melhores resultados na edição do Premier League dos dados analisados, assim concluímos que estatura média dos jogadores tem influencia no resultado final do campeonato, tornando assim um parâmetro de qualificação a ser considerado em nossa analise exploratória.

# 4.2 Média geral do clube por idade dos jogadores

Segue a analise da média de idade dos jogadores pelo total de vitorias agrupado por clube e ordenados pela classificação geral do clube na edição do Premier League dos dados analisados.

**Comando de Consulta:**



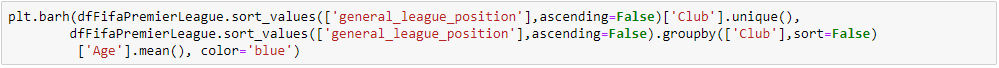
**Resultado:**



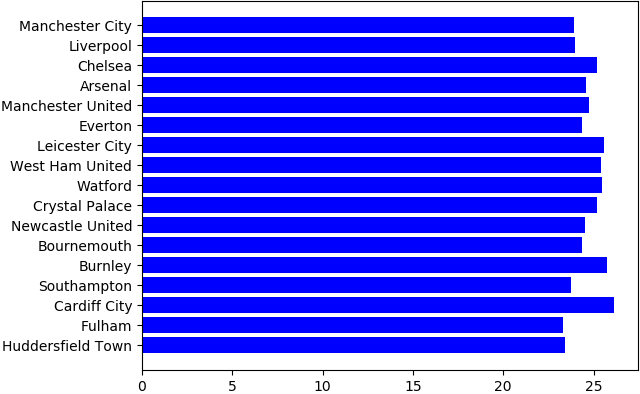
Nesta analise podemos observar que a idade média dos jogares do clube não tem o mínimo de interferência ao total de vitorias do clube nesta edição do Premier League, mostrado que media de idades dos jogadores é aleatória em comparação com a quantidade total de vitorias do clube, ou seja, o nível de experiência dos jogadores não afeta diretamente no total de vitorias do clube no decorrer do campeonato.

Vamos a analise através de um gráfico de barras, a comparação entre a média de idade dos jogadores por clube e a classificação final do campeonato, com objetivo de identificar o nível de influência do atributo idade dos jogadores no resultado do Premier League.

**Comando de Geração do Gráfico:**



**Resultado:**



Observamos que a média de idade jogadores de cada grupo não tem a mínima relação com o resultado da edição do Premier League dos dados analisados, concluímos que idade média dos jogadores afeta diretamente o resultado final do campeonato, assim desconsideramos este parâmetro em nossa analise exploratória.

# 4.3 Média geral do clube por avaliação geral dos jogadores

Abaixo segue a analise da média da avalição geral dos jogadores pelo total de gols marcados no campeonato agrupados por clube e ordenados pela classificação geral do clube na edição do Premier League dos dados analisados.

**Comando de Consulta:**



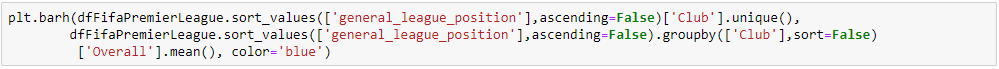
**Resultado:**



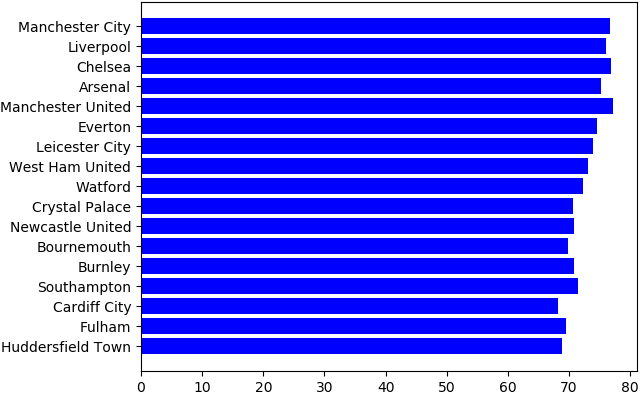
Observamos que a avalição geral dos jogadores do clube reflete com exatidão ao número de gols marcados, mostrado que os clubes a maior média de avalição geral dos jogadores tem resultados expressivos em relação na quantidade de gols marcados.

Agora vamos realizar a analise através de um gráfico de barras, com objetivo de identificar o nível de influência do atributo avaliação geral dos jogadores no resultado do Premier League.

**Comando de Geração do Gráfico:**



**Resultado:**



Observamos que a avaliação geral dos jogadores tem uma relação aleatória com classificação dos clubes nesta edição do Premier League dos dados analisados, assim concluímos que avaliação geral dos jogadores não tem influencia no resultado final do campeonato, tornando assim um parâmetro de qualificação a não ser considerado em nossa analise exploratória.

# 4.4 Média geral do clube por valor de mercado dos jogadores

Segue a analise da média de valor de mercado dos jogadores pelo saldo de gols agrupados por clube e ordenados pela classificação geral do clube na edição do Premier League dos dados analisados.

**Comando de Consulta:**



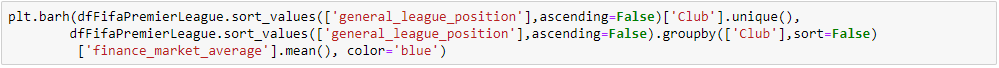
**Resultado:**



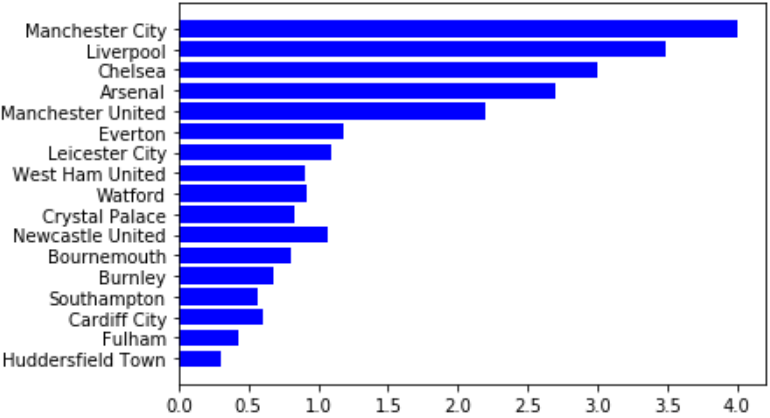
Observamos que a média de valor de mercado dos jogares do clube reflete com exatidão ao saldo de gols do clube, mostrado que os clubes a maior média de valor de mercado dos jogadores tem relação direta no saldo de gols com alto nível de precisão.

Agora vamos realizar a analise através de um gráfico de barras, com objetivo de identificar o nível de influência do atributo média de valor de mercado dos jogadores no resultado do Premier League.

**Comando de Geração do Gráfico:**



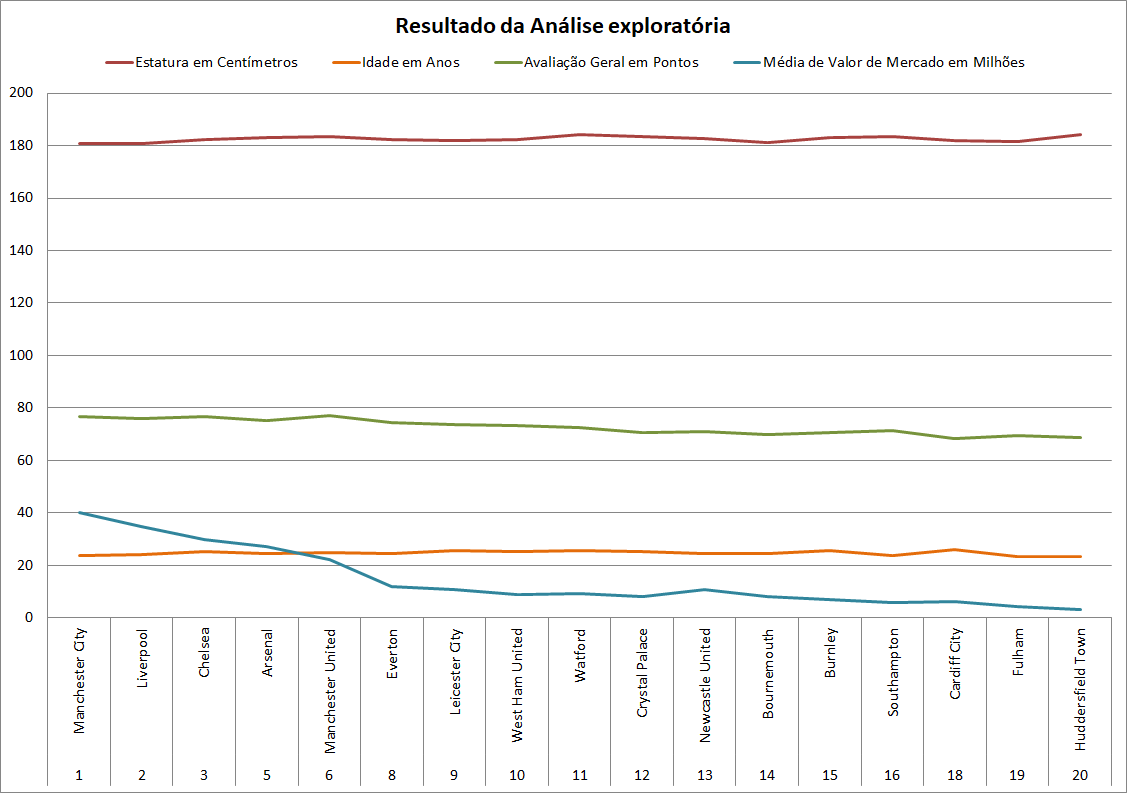
**Resultado:**



Observamos que os clubes com maior a média de valor de mercado dos jogares atingiram os melhores resultados na edição do Premier League dos dados analisados, assim concluímos que média de valor de mercado dos jogares tem a maior precisão no resultado final do campeonato, tornando assim o parâmetro mais assertivo de qualificação a ser considerado em nossa analise exploratória.

# 5. Apresentação dos resultados

Com base na analise exploratória atingimos os seguintes resultados no gráfico a abaixo:



O gráfico apresenta o resultado da análise comparativa das seguintes características dos jogadores: média geral do clube por estatura, idade, avaliação geral e valor de mercado dos jogadores, onde constatamos que entre as características analisadas o valor de mercado é a que mais se destaca, pois podemos observar que a linha azul é a que mais se diferencia conforme o resultado final do clube na edição da premier league analisada.

# 6. Links

* Projeto no Github:

<https://github.com/alanfelikis/CDBDPUC>

* Apresentação do Projeto no Google Drive: <https://drive.google.com/open?id=1ZToHMm2e11e_mjPfxfs16FTjxwIneX3I>

# REFERÊNCIAS

Pierre, Sadrach, **Exploratory Data Analysis of the FIFA 19 DataSet in Python**. Internet: Site da Towards Data Science, 2019. Link: [https://towardsdatascience.com/exploratory-data-analysis-of-the-fifa-19-DataSet-in-python-24eb27de9e59](https://towardsdatascience.com/exploratory-data-analysis-of-the-fifa-19-dataset-in-python-24eb27de9e59). Acessado em 10 de Abril de 2020.

Doan Nguyen, Tuan, **The Beautiful Game: Predicting the Premier League with a random model**. Internet: Site da Towards Data Science, 2018. Link: <https://towardsdatascience.com/o-jogo-bonito-predicting-the-premier-league-with-a-random-model-1b02fa3a7e5a>. Acessado em 10 de Abril de 2020.

Fordham, Stephen, **Pandas in the Premier League**. Internet: Site da Towards Data Science, 2019. Link: <https://towardsdatascience.com/pandas-in-the-premier-league-78f1d0c21b4f>. Acessado em 10 de Abril de 2020.