



FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE DE
COIMBRA

INTELIGÊNCIA NO NEGÓCIO

FIFA ULTIMATE TEAM

Francisco Santiago
Renato Santos

2012132539
2015237457

fsantiago@student.dei.uc.pt
renatojms@student.dei.uc.pt

5 de junho de 2020

ÍNDICE

INTRODUÇÃO	3
CONTEXTO	3
O QUE É O FIFA ULTIMATE TEAM?	3
OBJETIVO	5
FONTE DE DADOS	6
DATA WAREHOUSE	7
MODELO ESTRELA	7
TABELA DE FACTOS	7
TABELA DIMENSÕES	8
ESCOLHA DE SOFTWARE	8
Base de Dados	9
Desenho do modelo da Base de Dados	9
ETL (Extract, Transform, Load)	10
OLAP (Online Analytical Processing)	10
DATA MINING	12
ETL	13
PLANO ETL	13
OLAP	14
DASHBOARDS	14
DATA MINING	18
MATRIZ DE CORRELAÇÃO	18
MODELO DE REGRESSÃO	20
MODELO DE CLASSIFICAÇÃO	22
CONCLUSÃO	24

1. INTRODUÇÃO

Estamos cada vez mais mergulhados numa vivência onde a importância e o tratamento de dados estão a ter um peso cada vez mais considerável no quotidiano. A Inteligência no Negócio (IN), está adjacente à evolução das tecnologias que presenciamos, em que a informação é cada vez mais preciosa. Por isso, deve ser tratada da melhor forma possível, com o objetivo de extrair o maior benefício possível para o negócio em causa.

Neste projeto, aplicámos um conjunto de técnicas e tecnologias, que fomos aprendemos ao longo deste semestre, com o propósito de construir uma solução *Business Intelligence (BI)* que pretende ajudar os jogadores mais aficionados de *FIFA Ultimate Team* espalhados pelos quatro cantos do mundo.

2. CONTEXTO

FIFA é um videojogo desenvolvido pela *EA Sports*, cujo primeiro título remonta a 1993. Retrata uma simulação do mundo futebolístico da maioria das ligas mundiais de futebol, com escalões principais e secundários. A grande maioria das equipas e ligas estão licenciadas, incluindo as competições europeus e sul-americanas, o que atrai milhões de jogadores a comprar todos os anos o novo título que é lançado sempre no final do mês de setembro.

2.1. O QUE É O FIFA ULTIMATE TEAM?

Dentro do jogo, existem inúmeros modos a ser explorados. Porém, em 2009, a *EA Sports* introduziu um novo modo que revolucionou a forma de jogar e divulgar o seu jogo junto da comunidade futebolística e dos adeptos: o modo *Ultimate Team*.



Figura 1: Equipa do FIFA ULTIMATE TEAM

O objetivo do modo é formar uma equipa com a maior pontuação geral possível e maximizando também sempre a equipa da equipa. A pontuação geral é calculada pela média de todos os jogadores presentes no plantel, e a química é construída com base nas ligações pessoais de cada jogador: ao colocar um jogo em campo, ele deve ter características em comum com os jogadores que jogam em posições vizinhas, tais como: a mesma nacionalidade, o mesmo clube ou a mesma liga. Uma maior química num jogador representa que os seus atributos vão ter um aumento, ou caso não tenha química com os outros jogadores, isso é refletido numa diminuição das pontuações nos seus atributos futebolísticos.

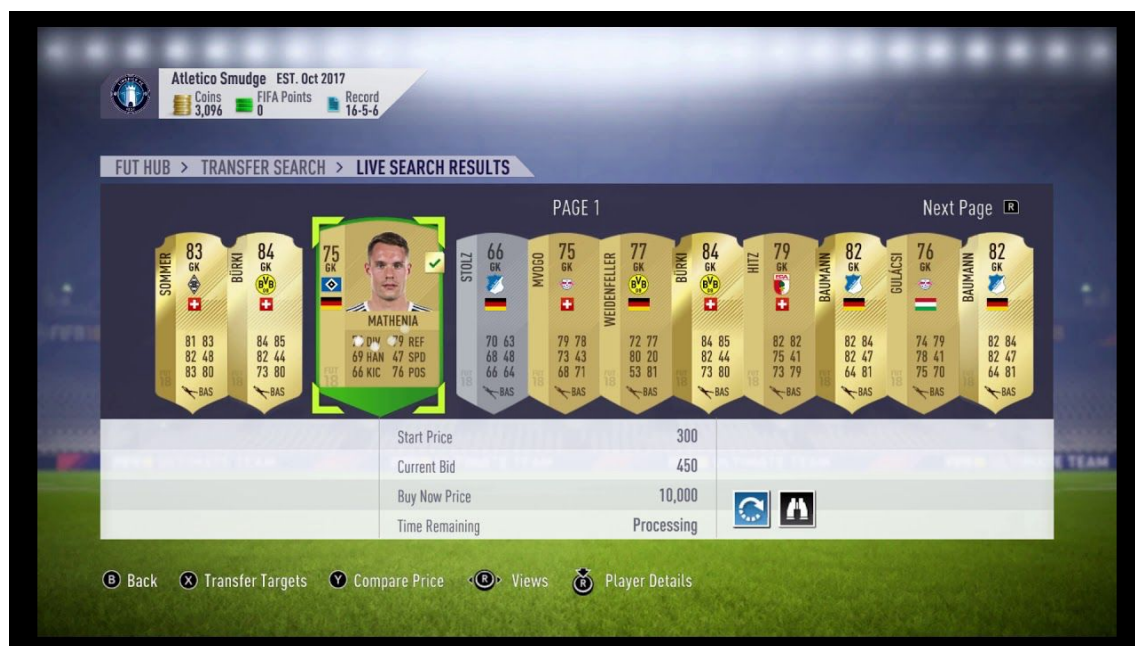


Figura 2: Mercado de transferências

Existe um mercado de transferências, onde todos os jogadores podem vender e comprar as suas cartas. Este mercado, baseia-se num sistema de leilão online, onde cada carta tem um preço mínimo e um preço máximo, e é vendida pela licitação máxima ou pelo preço máximo definido pelo vendedor.



Figura 3: Carta POTM do Bruno Fernandes

Um dos fatores deste modo de jogo ter tão sucesso prende-se à estratégia de marketing e comunicação implementada pela *EA Sports*, envolvendo e promovendo os principais protagonistas do mundo do futebol, havendo assim uma divulgação recíproca, uma vez que são elogiadas as melhores performances dos jogadores em campo na vida real, e os jogadores gostam de ter o seu trabalho reconhecido e partilham junto dos seus fãs as cartas especiais criadas para eles. Por exemplo, dentro do jogo a carta do Bruno Fernandes normal pode estar em milhares de equipa do Ultimate Team, podendo ser adquirida tanto no mercado de transferência ou através da compras de *packs* onde é a sorte que determina o que pode sair. Em fevereiro, o Bruno Fernandes foi eleito o melhor jogador da *Premier League*, e por isso saiu uma carta especial com uma subida da pontuação geral.

2.2. OBJETIVO

A nossa solução *BI* tem o objetivo de ajudar na tomada de decisão a construir uma equipa no *FIFA Ultimate Team*, para exista uma maximização de todas os recursos possíveis para que a equipa tenha um melhor desempenho em campo.

Com milhões de jogadores pelo mundo inteiro, seria inevitável não haver um torneio entre os melhores. Por isso, todos os anos é realizada a *FIFA eWorld Cup* onde se pretende apurar o campeão mundial. Com um prémio monetário de cerca de meio milhão de euros, cerca de 50 milhões de visualizações na final do torneio de 2019 e com mais de 20 milhões de jogadores a participar no torneio, a competitividade é muito elevada e todos os pormenores contam. Por isso, identificamos como público-alvo: os diversos clubes de *eSports* que estão cada vez mais a crescer em Portugal e pelo mundo, onde pertencem inúmeros profissionais de *eSports* que pretendem melhorar todos os dias, pretendemos ser também uma ferramenta de auxílio a jogadores que aspiram a ser profissionais e também aos inúmeros *Youtubers* que realizam conteúdo realizado com o *FIFA Ultimate Team*.

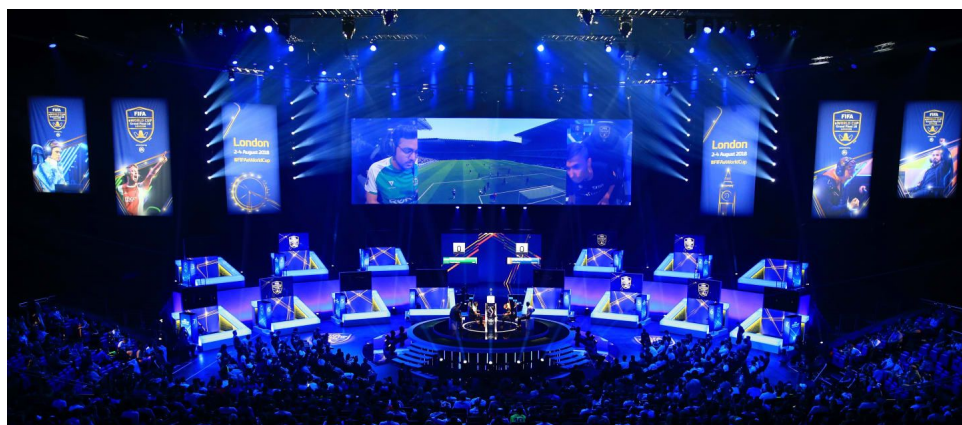


Figura 4: *FIFA eWorld Cup*

Com inúmeras possibilidades de formar uma equipa com 100% de química, é preciso ter em conta também se efetivamente é a melhor equipa possível ou se poderia poupar algum dinheiro e construir uma equipa ainda melhor, por isso pretendemos ajudar em:

- encontrar os melhores jogadores para cada posição, tendo em conta o preço, nacionalidade, liga, clube e qualidade da carta (bronze, prata, ouro)
- saber quais os jogadores que correspondem às características pretendidas (altura, peso, idade, qualidade de atributos específicos)
- Análise da variação do preço de uma carta
- Análise das características de cada liga e clube
- Como maximizar a performance de um jogador

2.3. FONTE DE DADOS

Para estudarmos os dados do videojogo, recorreremos a um dataset existente no *Kaggle*¹. Contém os dados de todas as cartas existentes no jogo, incluindo os preços, qualidade de todos os atributos, raridade das cartas e características individuais de cada jogador.

O *dataset* existente tem dados relativos aos últimos 4 anos, e uma análise de 44683 jogador, tendo no total 840 MB.

¹ [kaggle.com/stefanoleon992/fifa-20-ultimate-team-players-dataset](https://www.kaggle.com/stefanoleon992/fifa-20-ultimate-team-players-dataset)

3. DATA WAREHOUSE

Um data warehouse é um tipo de sistema de gestão de dados projetado para ativar e fornecer suporte às atividades de Inteligência no Negócio (*BI*), especialmente a análise avançada. Os data warehouses destinam-se exclusivamente a realizar consultas e análises avançadas e geralmente contêm grandes quantidades de dados, sendo estes geralmente derivados de uma ampla variedade de fontes.

3.1. MODELO ESTRELA

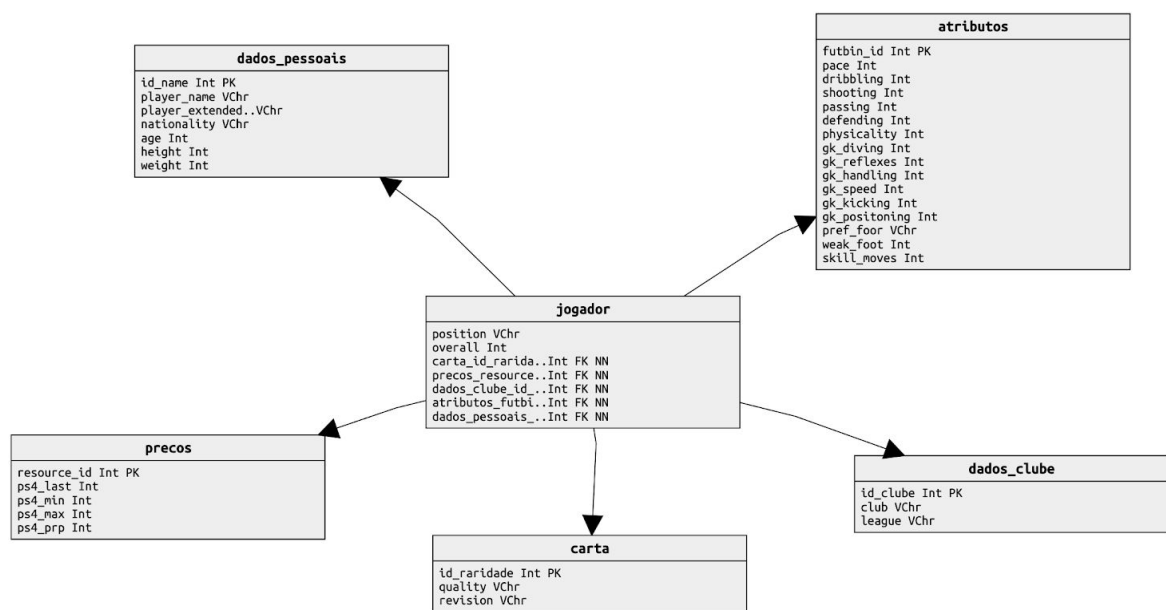


Figura 5: Modelo Estrela

No esquema estrela criado para o projeto temos uma tabela de factos e cinco tabelas de dimensões

3.1.1. TABELA DE FACTOS

A tabela **jogador** é a nossa tabela de factos onde se apresenta a posição e o overall (pontuação geral) da carta de cada jogador existente no jogo *FIFA*. De notar que os restantes atributos da tabela são as chaves estrangeiras, garantindo a integridade dos dados.

3.1.2. TABELA DIMENSÕES

De acordo com os dados, foram construídas cinco tabelas de dimensões, as quais se especificam abaixo.

- **dados_pessoais:**

Apresenta os dados biográficos que existem para cada jogador existente no *FIFA UT* como o nome que o jogador é conhecido, nome completo, nacionalidade, idade, peso e altura

- **preços:**

Tabela onde estão os preços de todos os jogadores tendo em conta que:

- *ps4_last*: último preço comprado
- *ps4_min*: preço mínimo atingido
- *ps4_max*: preço máximo atingido
- *ps4_prp*: percentagem de variação do preço

- **carta:**

Tabela que contém as diversas qualidades e respetiva raridade existentes das cartas. Podendo a qualidade (quality) variar entre ouro, prata e bronze e a raridade (revision) sendo cartas especiais criadas consoante uma alta performance do jogador no mundo real.

- **dados_clube:**

Apresenta informação sobre os clubes e ligas que existem no *FIFA*.

- **atributos:**

Apresenta todas as características de um jogador do ponto de vista futebolístico, guardando dados da aceleração (*pace*), dribble (*dribbling*), remate (*shooting*), passe (*passing*), defesa (*defending*), físico (*physicality*), “atirar” enquanto guarda-redes (*gk_diving*), reflexos enquanto guarda-redes (*gk_reflexes*), manuseio enquanto guarda-redes (*gk_handling*), velocidade enquanto guarda-redes (*gk_speed*), remate enquanto guarda-redes (*gk_kicking*), posicionamento enquanto guarda-redes (*gk_positioning*), pé dominante (*pref_foot*), qualidade do pior pé (*weak_foot*) e “skills” de jogador (*skill_moves*).

4. ESCOLHA DE SOFTWARE

Para o desenvolvimento da nossa solução, utilizámos técnicas de *ETL*, *OLAP* e *Data Mining*. Para isso, tivemos de definir primeiro o que pretendemos explorar e se as ferramentas se ajustavam à nossa realidade.

4.1. Base de Dados

Para a escolha da Base de Dados, definimos como requisito que deve permitir executar queries rapidamente. E entre as várias que cumpre o requisito, seleccionamos com base nos seguintes critérios:

- Facilidade de integração com ferramentas de *ETL*
- Rapidez de manipulação dos dados
- Experiência de utilização anterior

Entre as duas opções que mais agradavam aos elementos do grupo, fomos comparar com base na opinião de outros utilizadores:

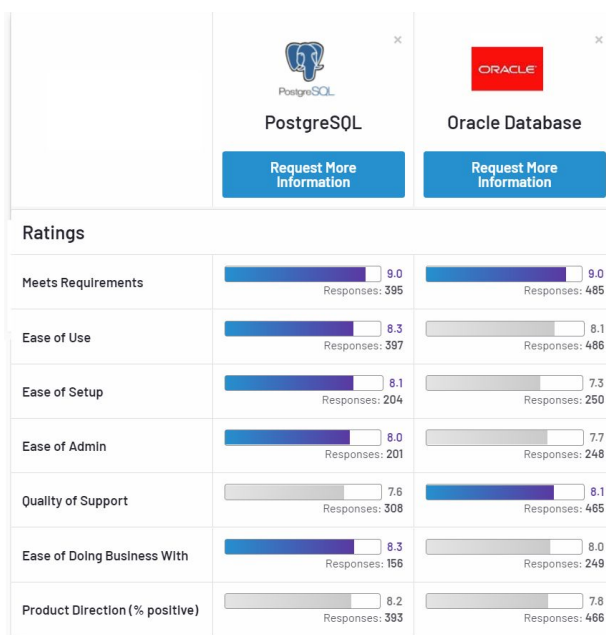


Figura 6: Comparação de PostgreSQL e Oracle Database²

Com base na informação recolhido, definimos o *PostgreSQL* como a ferramenta de Base de Dados a utilizar.

Desenho do modelo da Base de Dados

Para desenharmos o modelo estrela da nossa base de dados utilizámos o *Onda*³ a ferramenta disponibilizada pelo Departamento de Engenharia Informática da Universidade de Coimbra, uma vez que já foi utilizada em projetos de outras unidades curriculares, através do diagrama de Entidade-Relacionamento é gerado o diagrama físico e também permite obter logo um *script* com a criação das tabelas e as respetivas ligações.

² <https://www.g2.com/compare/postgresql-vs-oracle-database-vs-talend-data-integration>

³ <http://onda.dei.uc.pt/v3/>

4.2. ETL (Extract, Transform, Load)

O software escolhido para *ETL* deve:

- Permitir de forma rápida e intuitiva extrair dados e transformá-los
- Permitir carregar os nossos dados transformados para uma base de dados
- Facilitar a atualização de dados sempre que necessário

Com base nos requisitos definidos, tivemos de escolher um software com base em alguns critérios enquadrados com o projeto, tais como: a ferramenta ser gratuita, intuitiva, haver facilidade de integração com ferramentas *OLAP* e também as sugestões que o professor deu.




	 CloverDX	Pentaho	 Stitch <small>A Talend Company</small>
Focus	Data integration	Business intelligence, data integration, ETL	Data ingestion, ELT
Database replication	Full table; incremental replication via replication key fields	Full table; incremental via binary logs or SELECT/replication keys	Full table; incremental via change data capture or SELECT/replication keys
SaaS sources	Only Salesforce	Only Salesforce and Google Analytics	More than 100
Ability for customers to add new data sources	No	No	Yes
Connects to data warehouses? Data lakes?	No / No	Yes / Yes	Yes / Yes
Transparent pricing	No	No	Yes
G2 customer satisfaction 	4/5	4.2/5	4.8/5
Support SLAs	Yes	Yes	Available
Purchase process	Requires a conversation with sales	Requires a conversation with sales	Options for self-service or talking with sales. Also available from the AWS store .
Compliance, governance, and security certifications	GDPR	None	HIPAA, GDPR, SOC 2
Data sharing	No	Yes	Yes, through Talend Data Fabric

Figura 7: Comparação de ferramentas ETL⁴

Com base na pesquisa realizada, chegamos à conclusão que o *Pentaho Data Integration* seria uma boa opção para esta fase do projeto.

4.3. OLAP (Online Analytical Processing)

Um dos requisitos que o software para *OLAP* deve ter é a criação de forma rápida e fácil de diferentes tipos de visualização dos dados, de acordo com o objetivo em questão.

Dada a análise que pretendemos realizar, antes de começar a moldar esta fase do projeto, pesquisamos diversas alternativas para a utilização de uma ferramenta *OLAP*, tendo

⁴ <https://www.stitchdata.com/vs/cloverdx/pentaho>

sempre em conta que deve ser intuitiva, de fácil aprendizagem, gratuita, deve permitir visualizar os dados através de: gráficos, tabelas, bolhas, num mapa mundial; assim como a possibilidade de criar vários *dashboards* dinâmicos.

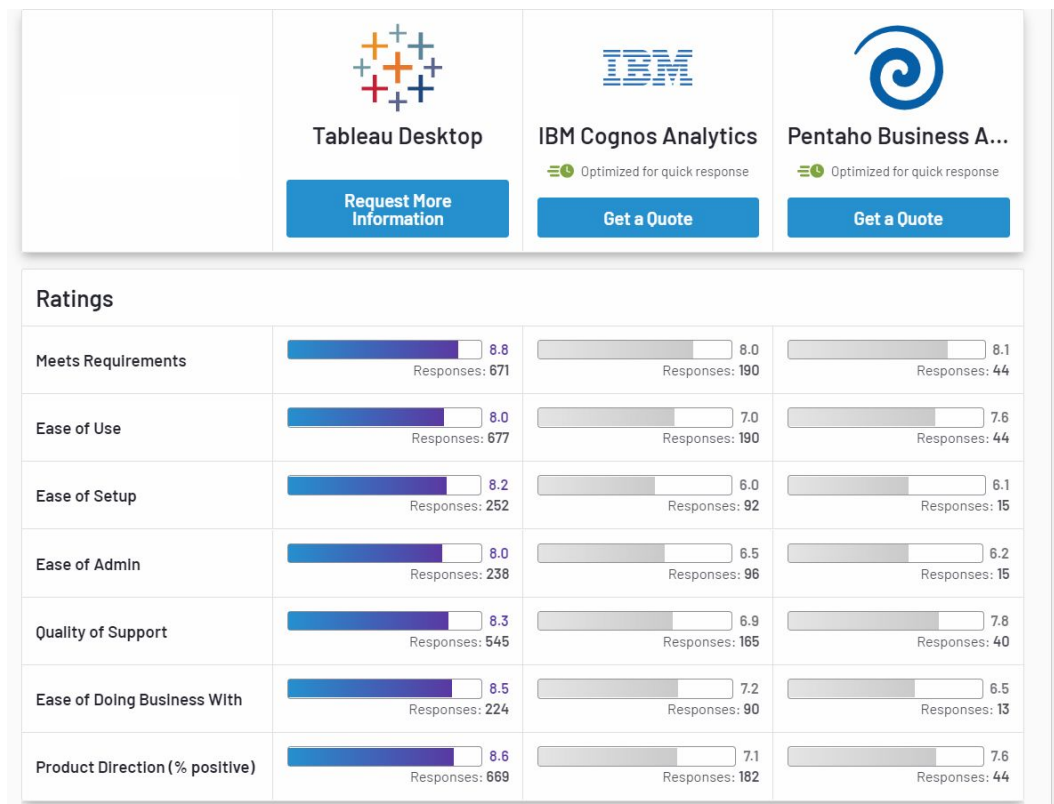


Figura 8: Comparação de ferramentas OLAP

Dada a simplicidade e a opinião de outros utilizadores, optámos por utilizar o *Tableau Desktop* que cumpriu com as nossas necessidades.

4.4. DATA MINING

O software destinado para *Data Mining* deve conseguir incorporar módulo que permitam visualizar os dados em análise e também módulos de machine learning.

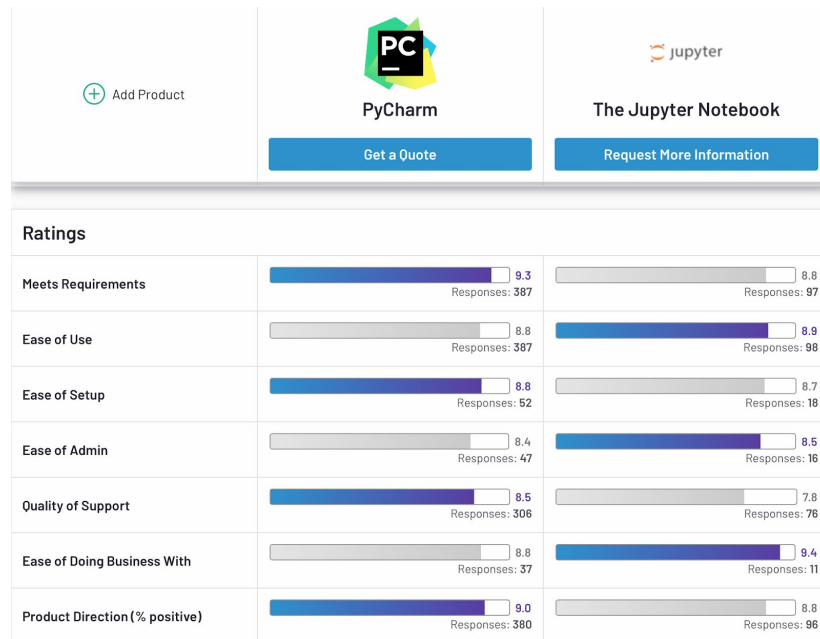


Figura 9: Comparação de software para *Data Mining*⁵

A nossa escolha para *Data Mining* foi o *PyCharm*, uma vez que já temos uma experiência anterior, e como estudantes temos acesso à versão profissional de forma gratuita.

⁵ <https://www.g2.com/compare/pycharm-vs-the-jupyter-notebook>

5. ETL

O *ETL* (*Extract, Transform, Load*) é um processo utilizado em soluções de Data Warehousing, que pressupõe a extração dos dados de uma fonte, a sua transformação e, por último, carregamento para uma *Data Warehouse*, através da qual poderão ser acedidos e utilizados.

5.1. PLANO ETL

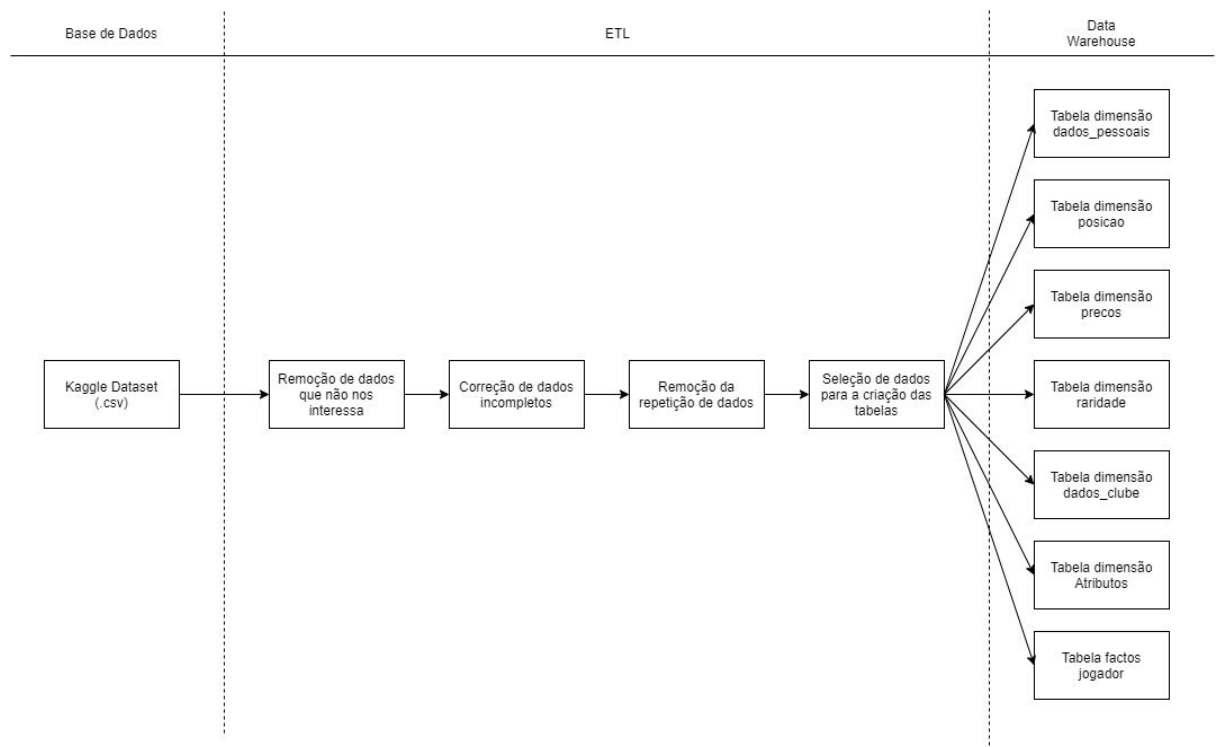


Figura 10: Plano *ETL*

O plano *ETL* é composto pelo *dataset* obtido no *Kaggle*, em ficheiro *.csv*, em que cada ficheiro contém 20981 jogadores (linhas) com 95 atributos (colunas), procedendo a execução das seguintes tarefas :

- Remoção de dados que não nos interessa - o *dataset* contém informação desnecessária para o nosso projeto e nesta etapa procedemos a remoção dessa informação, como por exemplo o ano de nascimento, reputação, etc... , tornando o dataset só com a informação essencial.
- Correção de dados incompletos - reparamos que este dataset continha campos onde a informação não existia, assim procedemos à substituição dos parâmetros a *NULL* por 0.

- Tratar repetições de dados - A informação de dados pessoais e dados do clube encontrasse repetida várias vezes portanto nós realizamos a remoção dessa informação.
- Seleção de dados para criar tabelas - nesta etapa, com o conteúdo do dataset previamente tratado, selecionamos os dados relacionados a cada tabela, previamente definida no modelo estrela

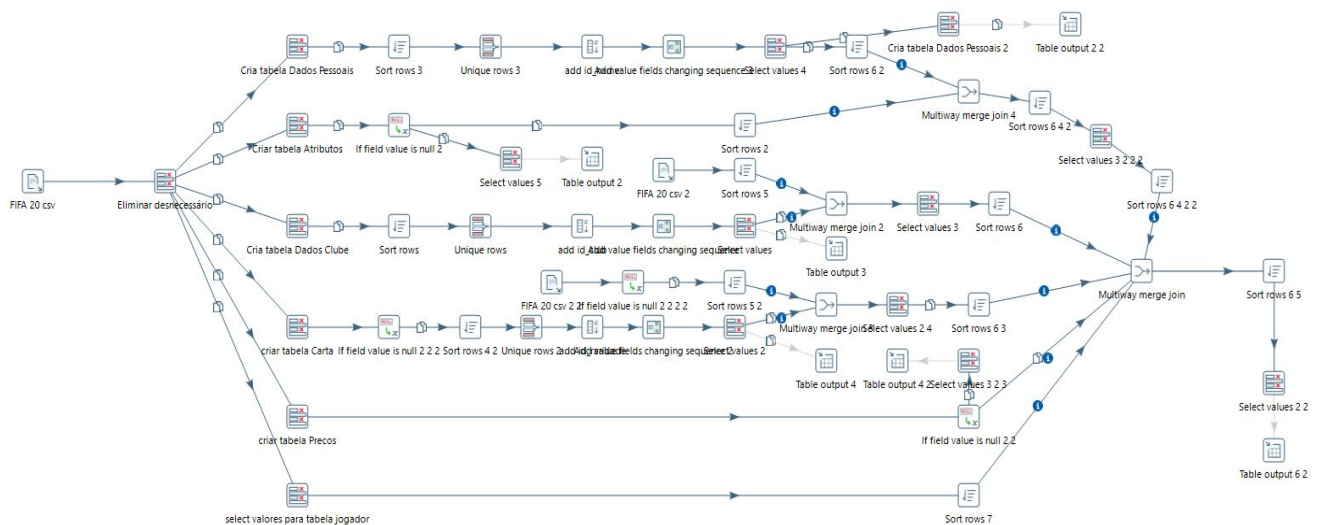


Figura 11: ETL para as tabelas definidas no modelo estrela

6. OLAP

6.1. DASHBOARDS

Com vista a analisar os dados pretendidos, adicionamos filtros personalizáveis para que seja possível pesquisar com o maior pormenor possível a matéria em questão, como por exemplo: preço, nacionalidade, atributos específicos, clube, liga, posição.

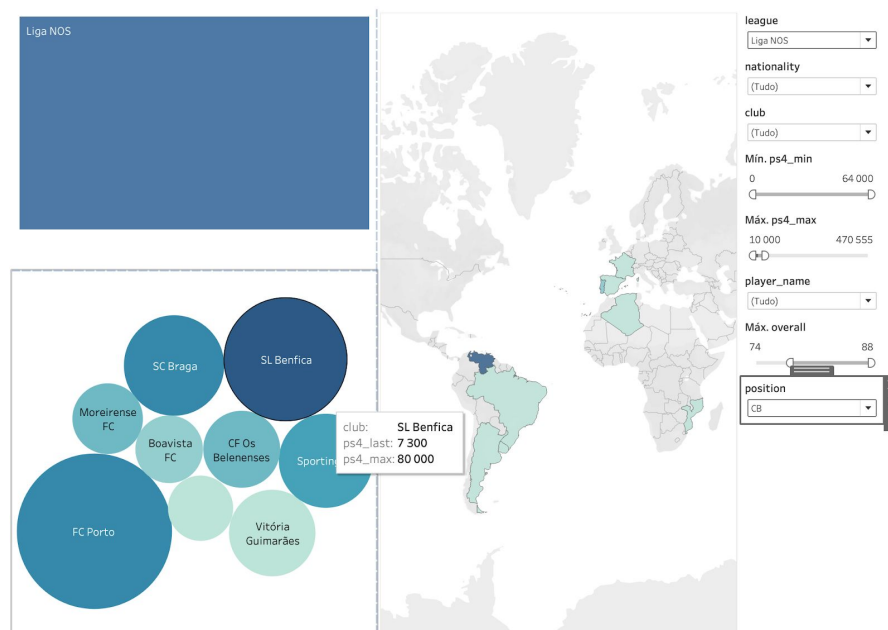


Figura 12: Análise da Liga Portuguesa

Na figura 12, ao aplicar os filtros apresentados, temos um resumo dos clubes com os jogadores mais caros e as suas respectivas nacionalidades.

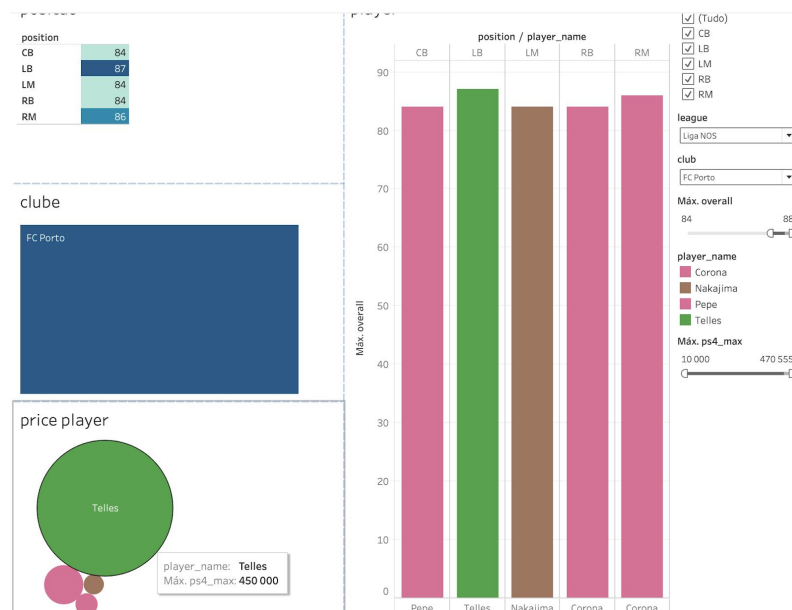


Figura 13: Análise do F.C. Porto

Podemos também ir mais pormenor dentro de um clube, tal como mostra a figura 13. Neste caso, é apresentado os melhores jogadores para as cinco posições escolhidas e os respetivos preços, sendo que o jogador Alex Telles é o melhor o jogador da equipa e também o que tem um preço de mercado mais elevado.

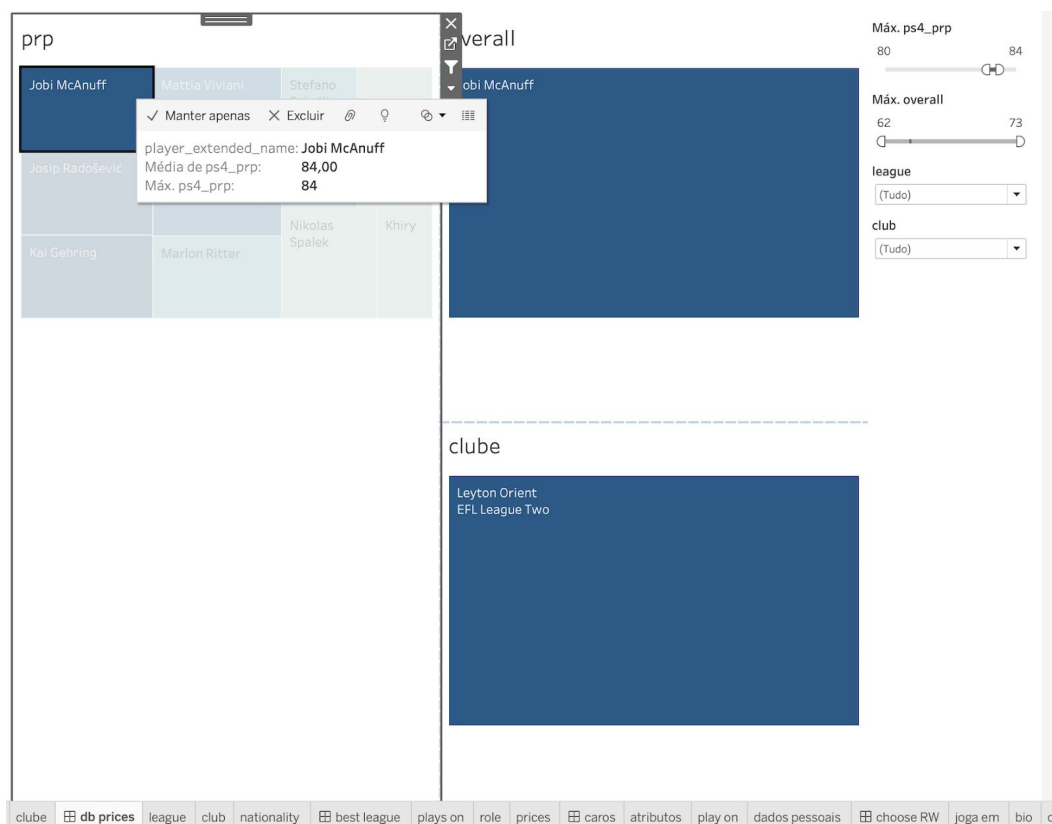


Figura 14: Análise da variação de preço

Com diversos desafios de construção de plantéis com tempo limitado e com uma recompensa no final, o preço de alguns jogadores estão longe de serem os melhores do jogo varia imenso, uma vez que existe uma procura muito elevada por estes, para concluir alguns *SBC (Squad Building Challenge)*. Porém, nem todos têm conhecimento do valor de mercado atual destes jogadores, acabando por achar que não valem muito e colocam o preço máximo de venda num valor muito baixo, acabando por vender o jogador rapidamente, o que depois vem a descobrir que foi um erro e que poderiam ter ganho muito mais moedas com aquela carta que possivelmente lhes saiu num pack de bronze ou de prata. Assim, ao analisar os jogadores com maior percentagem de variação de preço, sabe-se à partida que existe um desconhecimento geral do valor da carta, e um nicho específico do mercado que busca intensamente essa mesma carta, não para efetivamente jogar com ela, mas sim para concluir desafios. Posto isto, ao saberem quais os jogadores com maior variação, acabam por ter uma outra sensibilidade ao colocar as cartas à venda, colocando por exemplo um preço de compra muito mais elevado do que seria de esperado para o *overall* da carta em questão.

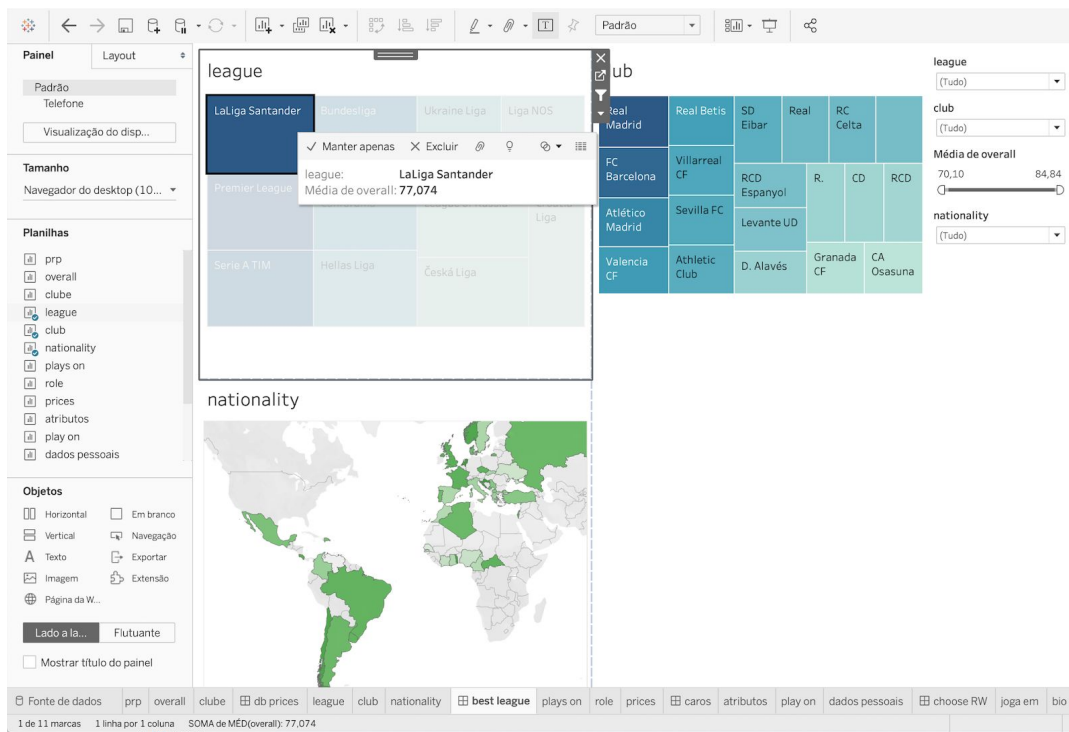


Figura 15: Análise das melhores ligas

Antes de começar a construir um plantel, é preciso definir um foco: pode ser só de uma liga, só de uma nacionalidade ou um misto de ambas, para ter 100% de química. Assim, é importante perceber onde estão os melhores jogadores para conseguirmos ter a equipa o mais cedo possível.

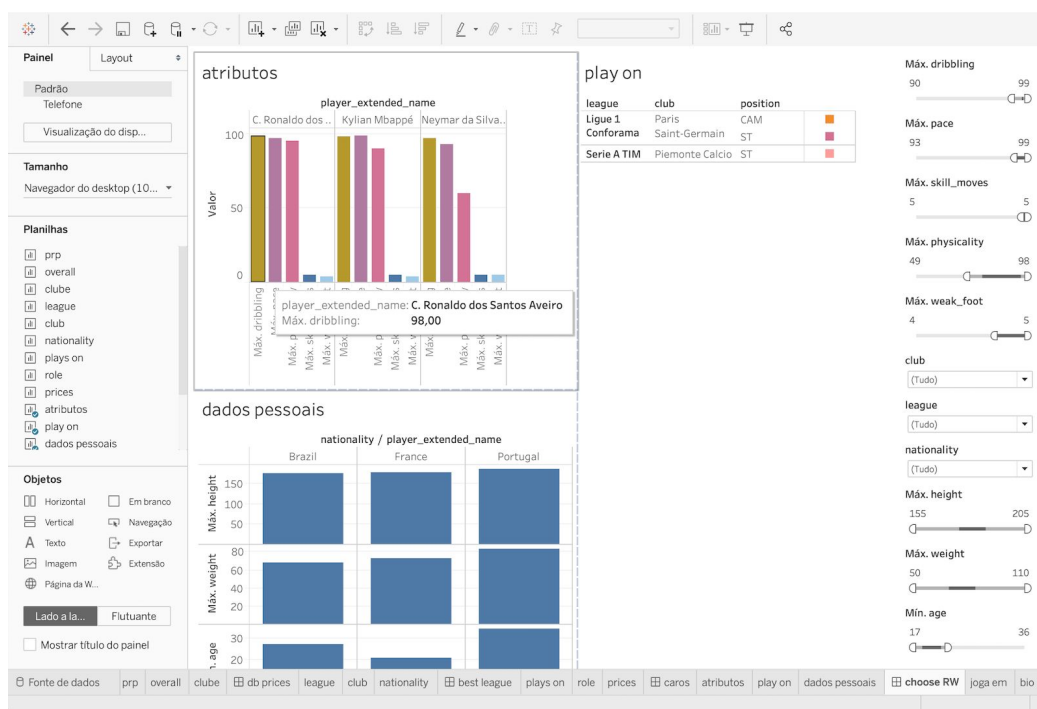


Figura 16: Pesquisa por atributos específicos

Dado o estilo de jogo de cada um, muitas vezes é preciso um certo jogador com aquelas habilidades que gostamos. Na figura 16, isso mesmo está representado. Procuramos um jogador com cinco estrelas nos movimentos técnicos e que seja rápido e com bom drible. Como resultado final, temos os três jogadores no *FIFA 20* que cumprem com os critérios da pesquisa.

7. DATA MINING

7.1. MATRIZ DE CORRELAÇÃO

A matriz de Correlação permite uma rápida visualização da relação entre todas as features dos dados.

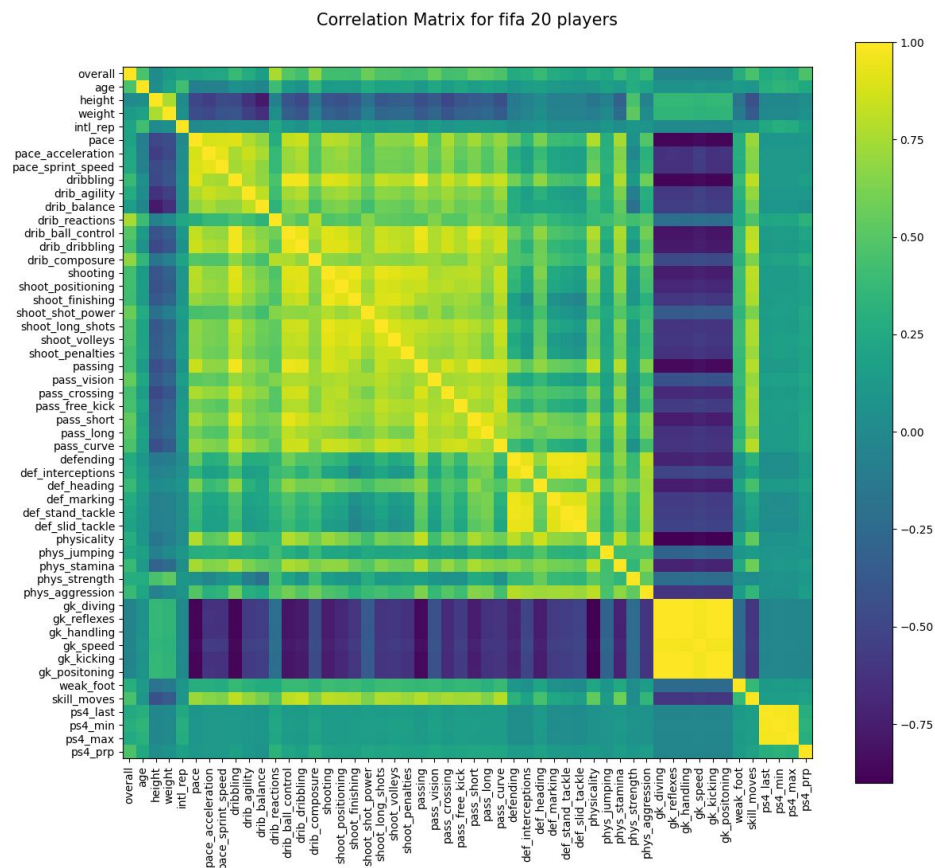


Figura 17: Matriz correlação de todos os atributos

Através do gráfico, observa-se que os atributos específicos dos jogadores de campo estão relacionados entre si. Os atributos de guarda-redes (*GK*) estão também estão fortemente relacionados entre si, o que era de esperar visto que os guarda-redes ficam sempre na baliza fazendo com que os restantes atributos tenham uma correlação negativa para esta posição.

Com base nisto, focamos o nosso estudo nos atributos do jogador.



Figura 18: Detalhes dos atributos

Dentro do FIFA 20, temos acesso aos detalhes dos atributos que permite ver com maior pormenor as características futebolísticas de cada carta, e também quanto é que um estilo de química faz aumentar ou não essas características.

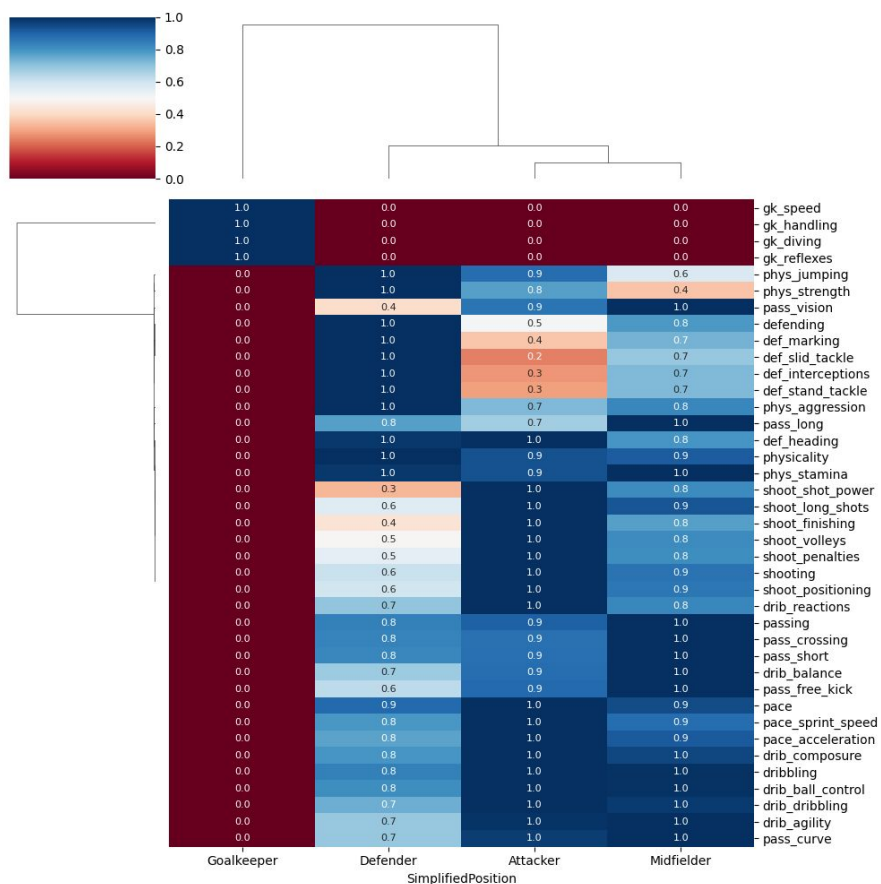


Figura 19: Correlação por setor de campo

Através desta figura, gerada através do módulo *Seaborn*, podemos verificar através da correlação dos atributos por posição que há diversos atributos que têm mais correlação pela posição do jogador, o que é de esperar visto que cada posição tem função diferente dentro de campo.

Dividimos assim todas as posições existentes no jogo entre quatro classes: *Goalkeeper*, *Defender*, *Midfielder*, *Attacker*.

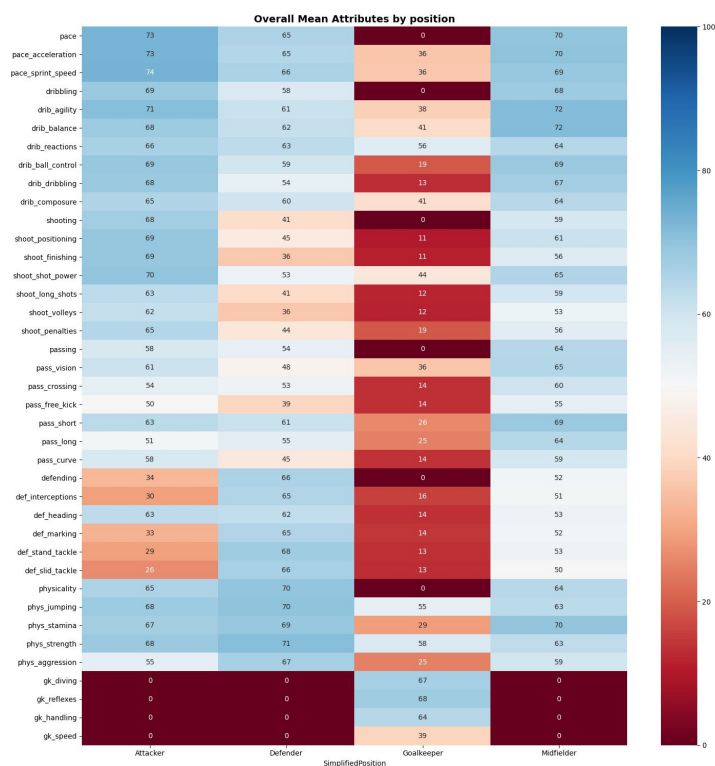


Figura 20: Média por setor de campo

Analizamos também as média por setor de campo para cada atributo específico, para que tenhamos uma melhor noção das funções a desempenhar em campo por cada classe.

7.2. MODELO DE REGRESSÃO

Para o decidirmos qual seria o nosso modelo de regressão, aplicamos o *Train_test_split*, que é função do Sklearn para seleção de modelo, em que dividimos em 20% para teste e 80% para treino. Com isto, comparamos o comportamento de diversos algoritmos, através da técnica de *Cross-validation*, em que avaliamos a capacidade de generalização de um modelo a partir da nossa fonte de dados. E consoante os valores gerados, escolhemos o melhor.

Testamos os seguintes algoritmos:

- Linear Regression
- LASSO
- Elastic Net
- K Neighbors Regressor
- Classification and Regression Trees
- Random Forest Regressor

A base para uma comparação justa dos algoritmos de *Data Mining* é assegurar que cada algoritmo é avaliado da mesma forma sobre os mesmos dados. Na figura seguinte está representada graficamente a “*performance*” dos algoritmos apresentados anteriormente.

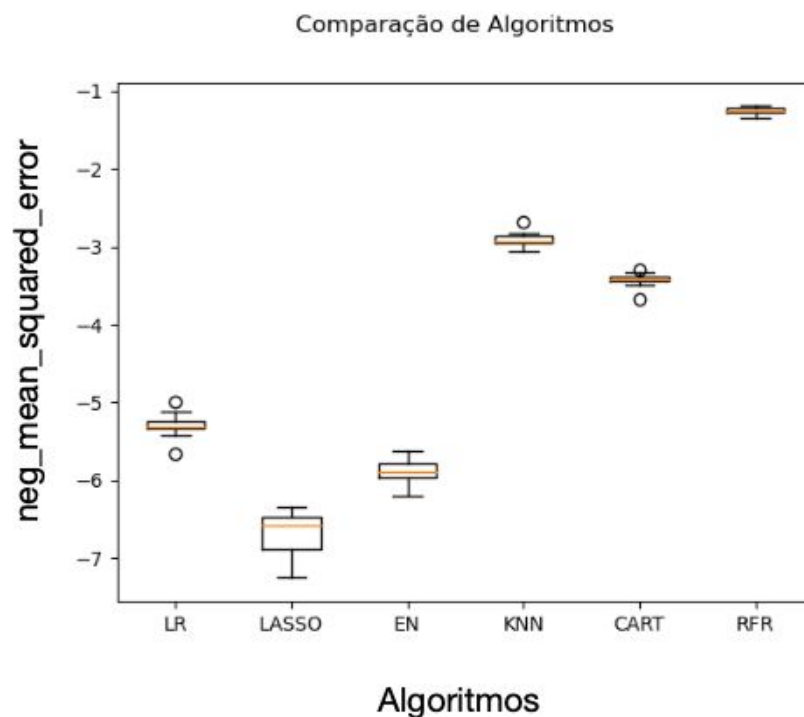


Figura 21: Comparação de algoritmos de regressão

Na figura X, podemos constatar que o *Random Forest Regressor* é o algoritmo que apresenta menor scoring no *neg_mean_squared_error*, sendo por isso considerado o melhor algoritmo para o nosso caso. Utiliza *Multiple Decision Trees* e também uma técnica de *bootstrap aggregation*, conhecida como bagging, na qual envolve treinar cada decision tree em diferentes amostras de dados onde depois cada amostragem é feita com substituição. A ideia é combinar múltiplas árvores de decisão na determinação do resultado final, em vez de depender de árvores de decisão individuais.

Ao aplicar este algoritmo sobre os nossos dados, obtivemos os seguintes resultados:

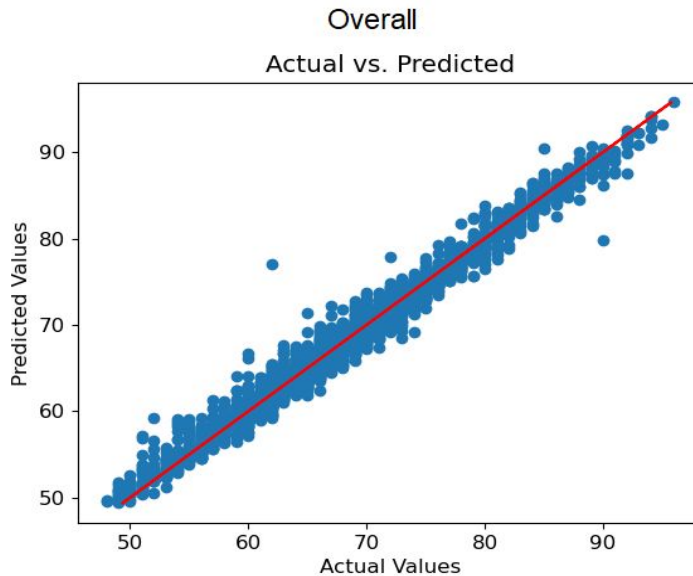


Figura 22: Overall actual vs previsto

Na figura 22, podemos ver que o nosso modelo se ajusta na generalidade, conseguindo prever com uma grande eficácia a maioria dos valores, tendo como target o valor contínuo do *overall*.

De modo a verificar se existiu um bom desempenho do algoritmo utilizado recorreremos às métrica de avaliação R^2 e *Mean Square Error (MSE)* para observar o erro das previsões que o modelo poderia ter e a respectiva magnitude.

```
MAE: 0.8353262316910787  
R2 score: 0.9795922147648647
```

Figura 23: R^2 e MSE

7.3. MODELO DE CLASSIFICAÇÃO

Para escolhermos um modelo de classificação, procedemos de modo análogo à etapa inicial que referimos anteriormente no modelo de regressão. Sendo que desta vez, testamos os seguintes algoritmos:

- Logistic Regression
- Linear Discriminant Analysis
- K Neighbors Classifier
- Classification and Regression Trees
- Naive Bayes

Na figura seguinte está retratado o desempenho dos algoritmos de classificação.

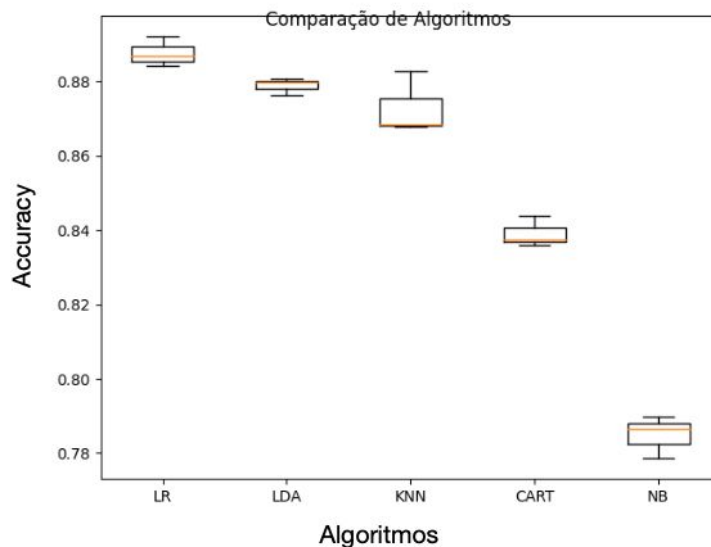


Figura 24: Comparação de algoritmos de classificação

Analisando o gráfico, verificamos que o algoritmo *Logistic Regression* é o que tem a maior exatidão. Este algoritmo assume uma distribuição gaussiana para os inputs numéricos das variáveis e consegue modelar problemas de classificação binária. Prevê a possibilidade de um resultado que só pode ter dois valores (0 ou 1), usa least square regression ao método para estimar o coeficiente e faz uso de *Maximum Likelihood Estimation (MLE)* para obter o modelo de coeficientes que relacionam as previsões com o target.

Usando este algoritmo e extraindo o relatório de classificação e a matriz de confusão alcançamos o seguinte:

Classification Report				
	precision	recall	f1-score	support
Attacker	0.86	0.84	0.85	770
Defender	0.93	0.93	0.93	1206
Goalkeeper	1.00	1.00	1.00	380
Midfielder	0.85	0.86	0.86	1399
accuracy			0.89	3755
macro avg	0.91	0.91	0.91	3755
weighted avg	0.89	0.89	0.89	3755

Matriz confusão				
[[648	0	0	122]
[0	1123	0	83]
[0	0	380	0]
[105	89	0	1205]]

Figura 25: Relatório de classificação e matriz de confusão

Para avaliar os resultados usamos a métrica *accuracy* (Figura 26) na qual verificamos o número de previsões corretas feitas como o ratio de todas as previsões, ou seja, quanto maior for o número de *True Positives* e *True Negatives*, maior será o valor da *accuracy*.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$



Figura 26: Accuracy

Através deste modelo, conseguimos garantir, uma vez que apresenta uma boa *accuracy*, que os jogadores testados pertencem à sua classe, ou seja, ao seu setor de campo. Concluimos que, sabendo apenas os atributos dos jogadores, conseguimos prever a posição em que joga com uma *accuracy* de 89%, o que é verificado na matriz de confusão (Figura 25), onde reparamos que na primeira linha, sobre os 770 jogadores atacantes, identificamos 648 com características *Attacker* e 122 como sendo *Midfielder*.

8. CONCLUSÃO

A nossa solução de BI pretende auxiliar o jogador nas diversas tomada de decisão que tem de fazer ao construir e gerir uma equipa no modo de jogo do FIFA Ultimate Team.

Na fase de data mining, através das *feature importances* do modelo de regressão, geramos o seguinte gráfico:

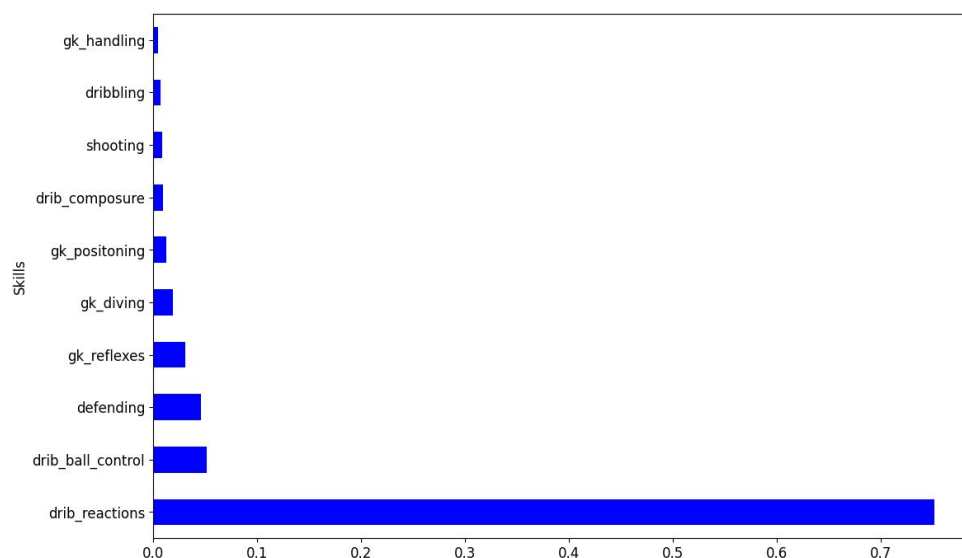


Figura 27: Top 10 features retiradas do algoritmos de regressão escolhido

Através da análise das dez características mais importantes no jogo, podemos concluir que as reacções de drible são de longe a mais importante. E uma vez que o nosso target é o valor contínuo do overall, estas são então as dez características que mais contribuem para que um jogador tenha um melhor comportamento em campo, e consequentemente um melhor overall. Por exemplo: no gráfico da figura 27, reparamos que o *drib_reaction* tem o valor de cerca 0.75, ou seja, esse atributo subir um valor, o overall do jogador sobe 0.75. Em suma, é quanto é que estas características contribuem para um aumento geral do overall da carta.



Figura 28: Estilos de química no mercado de transferências

Com inúmeras combinações de estilos, importa reconhecer então qual será o mais eficaz. E como vimos anteriormente, é o drible o mais importante. Por isso, ao pesquisar no mercado de transferência por um estilo, o jogador deve optar por comprar estilos que favorecem o drible, uma vez que vai influenciar positivamente o desempenho em campo. E também por exemplo, se pesquisar no mercado de transferência por um jogador específico que pretende comprar, e esse jogador já tiver estilos de química aplicados, no caso do preço da carta ser idêntico, o jogador opta por a carta que já tem o estilo de química aplicado que favorece o drible.

Para finalizar, queremos destacar a enorme versatilidade que a nossa solução *BI* apresenta, uma vez que cada jogador tem o seu estilo de jogo, quer seja de transições rápida, posse de bola ou ser mais compacto a defender, por isso existe uma panóplia de combinações possíveis e o segredo para o sucesso vem sempre com um foco bem definido e não perdendo tempo em ações que no futuro podem ser desnecessárias e que se revelem como uma perda de dinheiro. Investindo assim no mercado nas melhores opções existentes e tendo sempre consciência do valor de cada carta, para que seja cada vez mais fácil ganhar jogos à medida que se vai adquirindo melhores jogadores para a equipa.