

INSTITUTO POLITÉCNICO DO CÁVADO E DO AVE
ESCOLA SUPERIOR DE TECNOLOGIA
ENGENHARIA EM SISTEMAS INFORMÁTICOS

Trabalho Prático 1

Trabalho realizado por: Guilherme Azeredo – Nº 23510

Docente: Óscar Ribeiro

Unidade Curricular: Integração de Sistemas de Informação

2024/2025

Índice

1.Introdução	3
2. Estratégia Utilizada.....	4
3. Transformações	5
3.1. Workflow Principal	5
3.2. Metanode Data Processing	6
3.3. Metanode Data Cleaning.....	7
3.4. Metanode Analysing Data	8
3.4.1. Rule Engine	8
3.4.2. Analysing Data nodes	9
4. Table View	10
5. Conclusão	11
4. Referências	11

1. Introdução

O futebol é um dos desportos mais populares do mundo, com uma vasta base de fãs e inúmeras competições a nível global. A análise de dados no futebol tem vindo a ganhar relevância, principalmente na otimização de equipas, avaliação de desempenho de jogadores e análise de dados físicos, e outras características relevantes para o desempenho em campo. Este projeto visa analisar a base de dados de jogadores de futebol, utilizando a plataforma KNIME.

O objetivo deste trabalho é aplicar técnicas de manipulação de dados, criação de workflows que permitam extrair informações úteis para análise dos jogadores. Ferramentas como JSON, Excel, XML e gráficos visuais serão usadas para interpretar os dados.

2. Estratégia Utilizada

Como estratégia para a obtenção dos dados dos jogadores no Spotify em 2023 foi necessária a utilização de um ficheiro CSV com os dados de 32405 jogadores, retirado da base de dados do site Transfermarkt, referenciado no relatório. Com isto é necessário realizar a extração dos dados, limpeza, processamento e análise dos mesmos. Isto permite que sejam obtidos os insights mais importantes. Para a realização deste trabalho foi utilizada a aplicação Knime.

3. Transformações

3.1 Workflow Principal

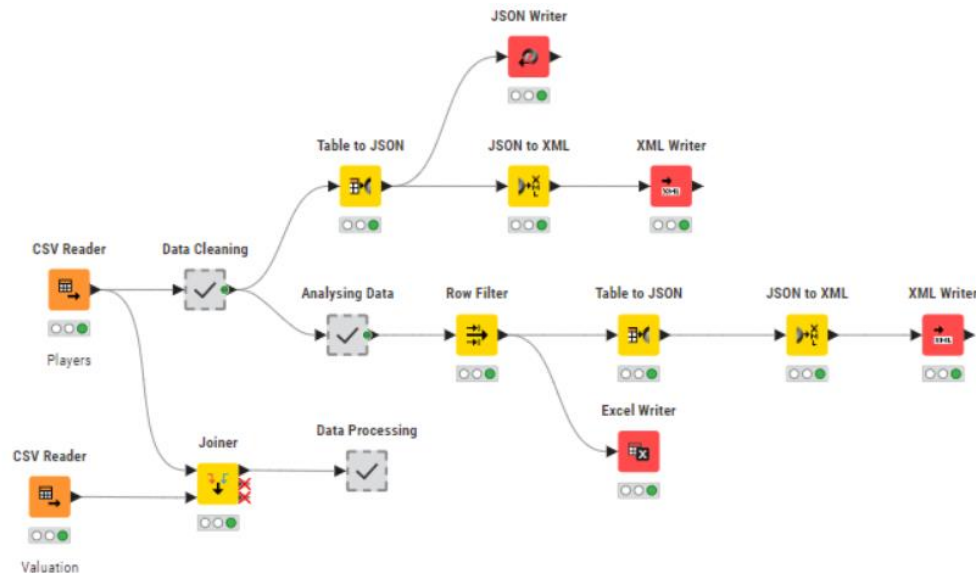


Figura 1 – Workflow Principal

Na Figura 1, podemos observar o workflow principal utilizado para o processamento dos dados dos jogadores de futebol. Este workflow faz a leitura dos dados de um ficheiro CSV e exporta-os para vários formatos, como Excel, JSON e XML. Os principais nodes envolvidos no processo são:

- **CSV Reader:** Lê o ficheiro CSV com os dados dos jogadores.
- **Joiner:** Une as duas bases de dados numa só.
- **Row Filter:** Filtra jogadores com base em critérios específicos. Neste caso, guarda apenas jogadores do Futebol Clube do Porto.
- **Table to JSON:** Converte os dados em formato JSON para maior flexibilidade.
- **JSON to XML:** Transforma os dados JSON para XML.
- **Excel Writer:** Guarda os dados em formato Excel.
- **JSON Writer:** Guarda os dados em formato JSON.

3.2 Metanode Data Processing

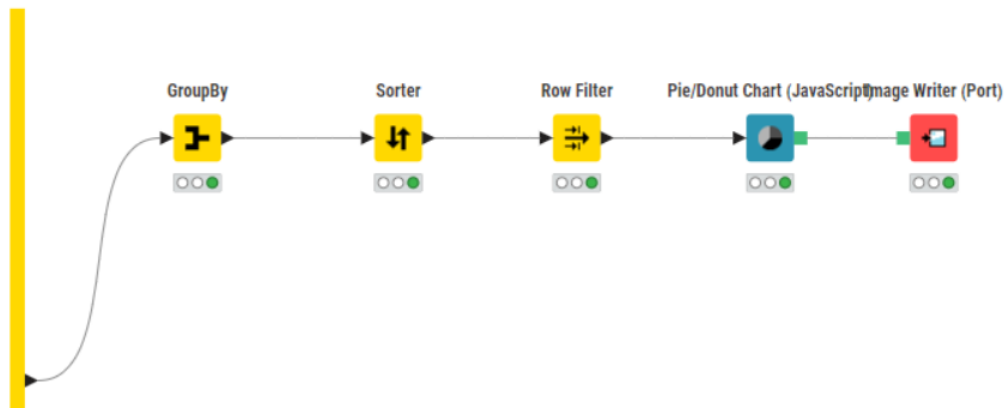


Figura 2 - Metanode Data Processing

No metanode "Data Processing" (Figura 2), realizam-se várias operações de processamento dos dados, tais como:

- “GroupBy” exclui todas as colunas exceto a coluna “player_id” e “name” e “highest_market_value_in_eur”;
- “Sorter” ordena os dados pela coluna “highest_market_value_in_eur” pela ordem decrescente de valor de mercado;
- “Row Filter” é usado para a remoção de várias linhas de jogadores e deixa apenas os jogadores com valor de mercado superior a 115 milhões de euros;
- “Pie Chart” Gera um pie chart que mostra os jogadores com valor de mercado superior a 115 milhões de euros, como demonstrado na Figura 3;
- “Image Writer (Port)” guarda uma imagem do pie Chart em uma pasta específica.

Pie Chart



Figura 3 - Pie Chart Valor Mercado

3.3 Metanode Data Cleaning

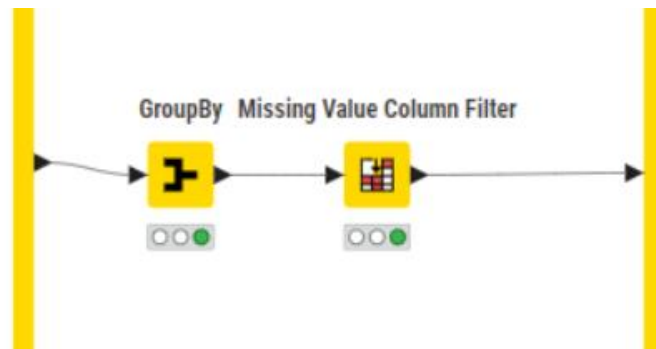


Figura 4 - Metanode Data Cleaning

Na Figura 4, é possível visualizar o metanode “Data Cleaning” na qual existe 6 nodes, eles são:

- GroupBy: Foram removidas as colunas com informações redundantes, como “city_of_birth”, “sub_position” e “agente_name”.
- Missing Value Column Filter: Retira as colunas que não estejam preenchidas em mais de 5% dos jogadores. No caso foi removida a coluna “foot”.

3.4 Metanode Analysing Data

3.4.1 Rule Engine

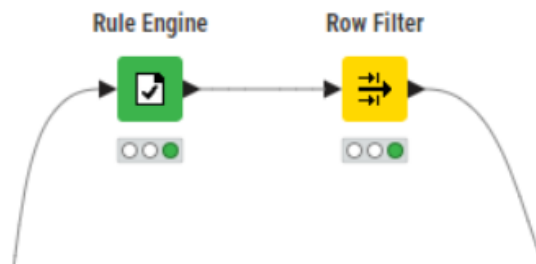


Figura 5 - Rule Engine and Column Resorter

Expression	
S	1 MISSING \$height_in_cm\$ => "Excluir"
S	2 \$height_in_cm\$ <= 175 => "Estatura Baixa"
S	3 \$height_in_cm\$ >= 176 AND \$height_in_cm\$ <= 187 => "Estatura Média"
S	4 \$height_in_cm\$ > 187 => "Estatura Alta"

Figura 6 - Expressão do Rule Engine

Foi utilizado um **Rule Engine** para classificar os jogadores em três categorias de altura:

- **Estatura Baixa:** Jogadores com altura menor ou igual a 175 cm.
- **Estatura Média:** Jogadores com altura entre 176 cm e 187 cm.
- **Estatura Alta:** Jogadores com altura superior a 187 cm.

Foi utilizado o Row Filter para excluir as linhas dos jogadores que não têm registo da altura.

Foi gerado um novo Pie Chart para a média da estatura dos jogadores.



Figura 7 - Pie Chart Estatura Média

3.4.2 Analysing Data nodes

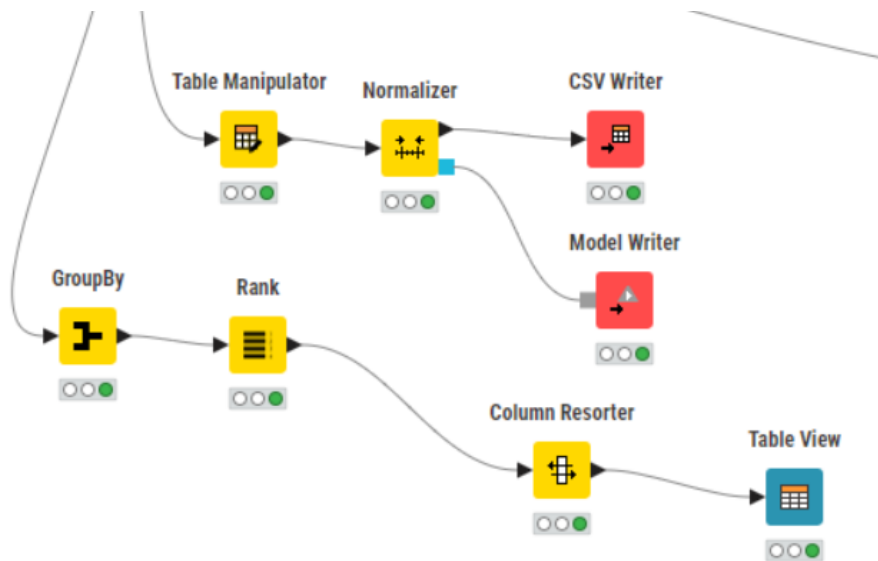


Figura 8 - Analysing Data Nodes

Na Figura 8, é possível visualizar outra parte do metanode “Analysing Data”, onde é possível identificar:

- “Table Manipulator” realiza a manipulação de dados das tabelas;
- “Normalizer” normaliza os dados da coluna “player_id” com a escala mínima 0 e máxima 1;
- “Model Writer” guarda um modelo para um ficheiro.
- “Rank” classifica os dados pelo campo “highest_market_value_in_eur”, na qual ordena os jogadores por valor de mercado;
- “Column Resorter” ordena as colunas na tabela.
- “Table View” possibilita a visualização da tabela com o ranking.

4. Table View

rank <small>Number (integer)</small>	name <small>String</small>	highest_... <small>Number (integer)</small>	player_id <small>Number (integer)</small>	position <small>String</small>	current_c... <small>String</small>
1	Kylian Mbapp...	200000000	342229	Attack	Real Madrid ...
1	Vinicius Junior	200000000	371998	Attack	Real Madrid ...
1	Erling Haaland	200000000	418560	Attack	Manchester ...
4	Lionel Messi	180000000	28003	Attack	Paris Saint-G...
4	Neymar	180000000	68290	Attack	Paris Saint-G...
4	Jude Bellingh...	180000000	581678	Midfield	Real Madrid ...
7	Raheem Sterl...	160000000	134425	Attack	Arsenal Foot...
8	Eden Hazard	150000000	50202	Attack	Real Madrid ...
8	Philippe Cout...	150000000	80444	Midfield	Aston Villa Fo...
8	Kevin De Bruy...	150000000	88755	Midfield	Manchester ...
8	Antoine Griez...	150000000	125781	Attack	Club AtlÃ©tic...
8	Harry Kane	150000000	132098	Attack	FC Bayern M...
8	Mohamed Sal...	150000000	148455	Attack	Liverpool Foo...
8	Sadio ManÃ©	150000000	200512	Attack	FC Bayern M...
8	Phil Foden	150000000	406635	Attack	Manchester ...
8	Lamine Yamal	150000000	937958	Attack	Futbol Club B...
17	Bukayo Saka	140000000	433177	Attack	Arsenal Foot...
18	Rodri	130000000	357565	Midfield	Manchester ...
18	Federico Valv...	130000000	369081	Midfield	Real Madrid ...
18	Jadon Sancho	130000000	401173	Attack	Chelsea Foot...
18	Jamal Musiala	130000000	580195	Midfield	FC Bayern M...
18	Florian Wirtz	130000000	598577	Midfield	Bayer 04 Lev...
23	Cristiano Ron...	120000000	8198	Attack	Manchester ...
23	Ousmane De...	120000000	288230	Attack	Paris Saint-G...
23	Declan Rice	120000000	357662	Midfield	Arsenal Foot...
23	Victor Osimh...	120000000	401923	Attack	Galatasaray S...
27	Paulo Dybala	110000000	206050	Attack	Associazione...
27	Trent Alexand...	110000000	314353	Defender	Liverpool Foo...
27	Martin Ã³deg...	110000000	316264	Midfield	Arsenal Foot...

5. Conclusão

Conclui-se que o KNIME Analytics Platform é uma ferramenta eficaz para a análise e visualização de dados. O projeto permitiu obter insights relevantes sobre os dados dos jogadores, o que pode ser útil para análises futuras. Como trabalho futuro, sugere-se a análise de outras características físicas e estatísticas dos jogadores e clubes de futebol, além da implementação de modelos para avaliação de desempenho.

6. Referências

- <https://docs.knime.com/>
- <https://www.knime.com/>
- <https://www.transfermarkt.com/>