

Entrepôt de données : Analyse des Données de Football

SOMMAIRE

I. Contexte	2
II. Le modèle de données	2
III. Le datamart	3
A. Dimensions	3
B. Le fait	4
C. Modèle en étoile	4
IV. La qualité des données	5
A. Données manquantes	5
B. Données non uniformes	6
V. Résultats	6
A. Composition des équipes au fil des saisons	6
B. Caractéristiques des joueurs au fil des saisons	9
C. Matches gagnés par équipes	13
1. Par saison	13
2. Par ligue	15
3. Par équipes	17
D. Buts marqués par équipes	20
1. Par saison	20
2. Par ligue	22
3. Par équipes	25
E. Radar des joueurs par saison	28
1. Par rapport à tous les autres joueurs	28
2. Par rapport à leur équipe	31
V. Les difficultés rencontrées	38

I. Contexte

Pour ce projet, nous avons accès à une base de données sur les statistiques de football. Cette base contient des informations sur les matchs, les ligues, les pays, les joueurs et les équipes. A partir de cette base, le but est de construire un datamart à l'aide de Talend afin d'extraire des informations pertinentes et réaliser des grâce d'évolution et comparatifs.

II. Le modèle de données

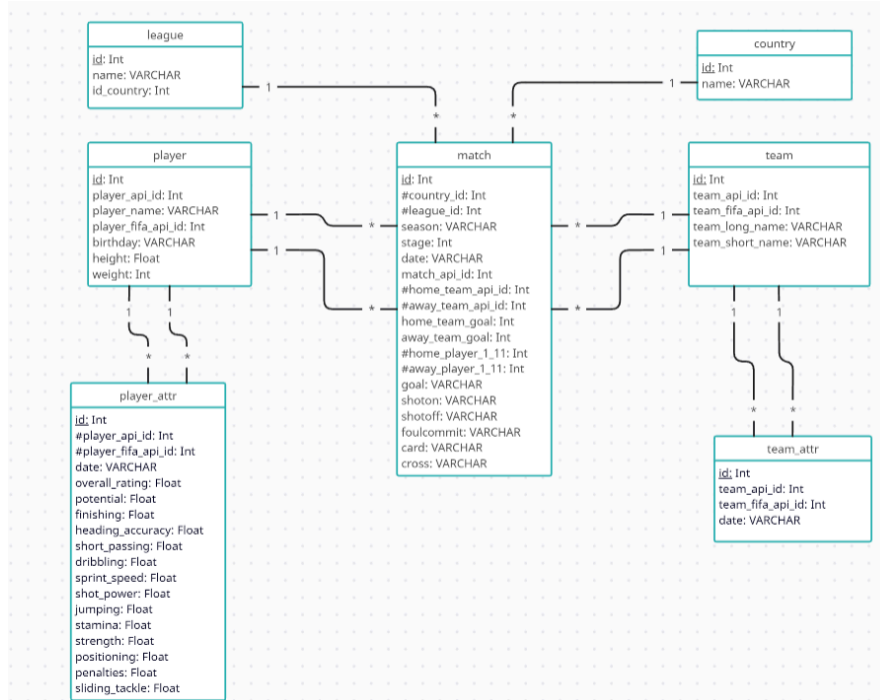


Figure : Modèle de données

Dans un 1er temps, on a construit le modèle de données pour l'exploiter dans Talend. Pour cela, nous nous sommes basées sur le modèle déjà existant. Nous avons simplement ajouté des attributs dans le table "player_attr" qui nous sont nécessaires pour réaliser les radars de compétences des joueurs.

III. Le datamart

Ensuite, nous avons déterminé le schéma du datamart à partir des restitutions qui nous ont été demandées ainsi que du modèle de données.

A. Dimensions

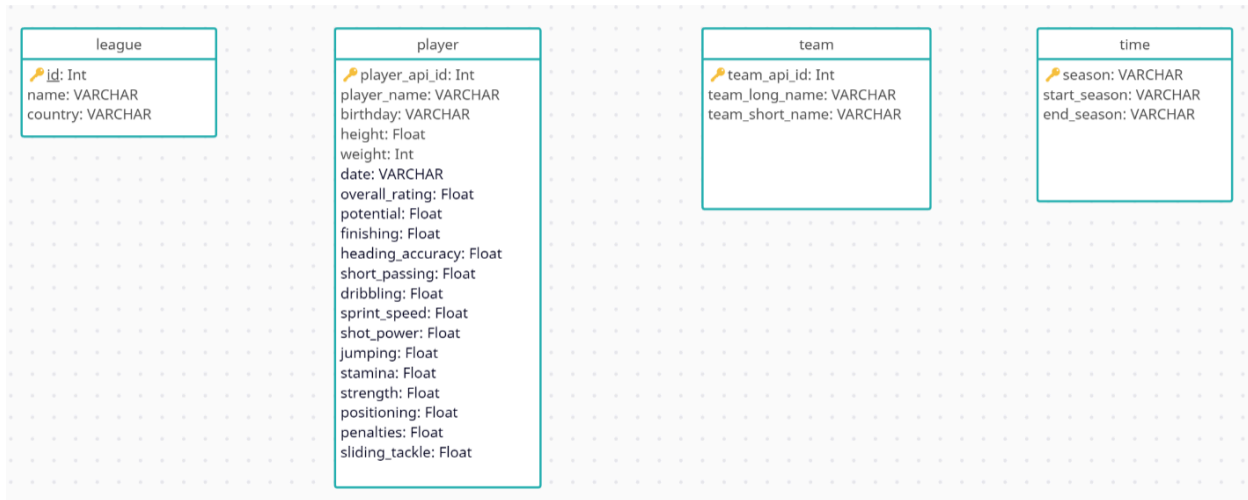


Figure : Tables de dimensions

Dans un premier temps, on a déterminé 4 dimensions:

- league: qui fusionne les tables league et country
- Player qui fusionne les tables player et player_attr
- Time qui contient la saison ainsi que les dates de début et de fin de cette saison
- Team: qui contient le nom des équipes

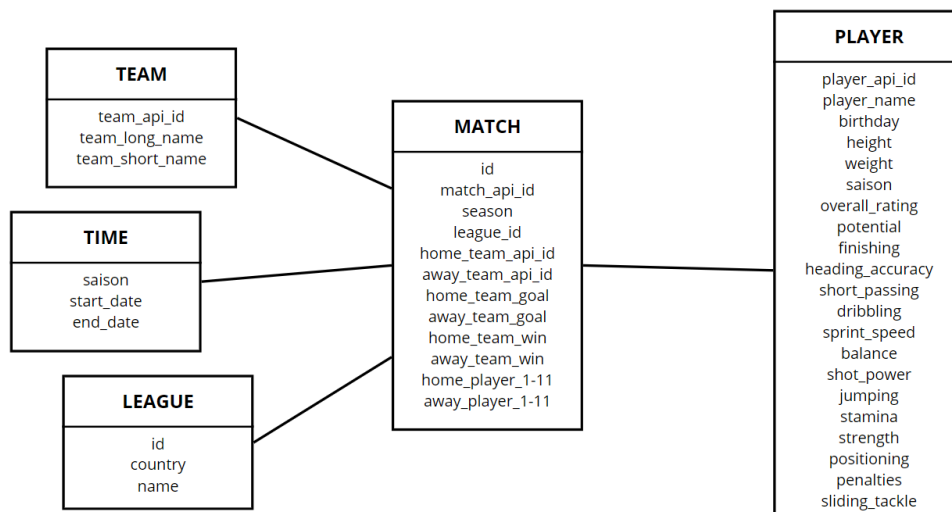
B. Le fait

match
id: Int
#league_id: Int
#season: VARCHAR
match_api_id: Int
#home_team_api_id: Int
#away_team_api_id: Int
#home_player_1: Int
#home_player_2: Int
#home_player_3: Int
#home_player_4: Int
#home_player_5: Int
#home_player_6: Int
#home_player_7: Int
#home_player_8: Int
#home_player_9: Int
#home_player_10: Int
#home_player_11: Int
#away_player_1: Int
#away_player_2: Int
#away_player_3: Int
#away_player_4: Int
#away_player_5: Int
#away_player_6: Int
#away_player_7: Int
#away_player_8: Int
#away_player_9: Int
#away_player_10: Int
#away_player_11: Int
home_team_goal: Int
away_team_goal: Int
home_team_win: Int
away_team_win: Int

Figure : Table de fait

Pour le fait, on a sélectionné la table match et on a supprimé les attributs qui ne nous ont pas servis pour ce projet. En plus des clés étrangères, on a 4 mesures. Pour chaque équipe, on a le nombre de goals et si l'équipe a gagné. Si l'équipe a gagné, la valeur de la mesure est de 1, sinon c'est de 0. Pour déterminer cette valeur, on compare le nombre de goals.

C. Modèle en étoile



IV. La qualité des données

A. Données manquantes

Dans la table MATCH, certaines lignes ne contiennent pas d'informations sur les joueurs qui ont participé :

```
id;season;league_id;home_team_api_id;away_team_api_id;home_team_goal;away_team_goal;home_team_win;away_team_win;home_player_1;home_player_2;home_r
3/2009;1;9987;9993;1;1;0;0;,,,,,,,,,,,,,,,,,,,,,
3/2009;1;10000;9994;0;0;0;0;,,,,,,,,,,,,,,,,,,,,,
3/2009;1;9984;8635;0;3;0;1;,,,,,,,,,,,,,,,,,,,,,
3/2009;1;9991;9998;5;0;1;0;,,,,,,,,,,,,,,,,,,,,,
3/2009;1;7947;9985;1;3;0;1;,,,,,,,,,,,,,,,,,,,,,
3/2009;1;8203;8342;1;1;0;0;,,,,,,,,,,,,,,,,,,,,,
3/2009;1;9999;8571;2;2;0;0;,,,,,,,,,,,,,,,,,,,,,
3/2009;1;4049;9996;1;2;0;1;,,,,,,,,,,,,,,,,,,,,,
3/2009;1;10001;9986;1;0;1;0;,,,,,,,,,,,,,,,,,,,,,
3/2009;1;8342;8571;4;1;1;0;,,,,,,,,,,,,,,,,,,,,,
3/2009;1;9985;9986;1;2;0;1;,,,,,,,,,,,,,,,,,,,,,
3/2009;1;10000;9991;0;2;0;1;,,,,,,,,,,,,,,,,,,,,,
3/2009;1;9994;9998;0;0;0;0;,,,,,,,,,,,,,,,,,,,,,
3/2009;1;7947;10001;2;2;0;0;,,,,,,,,,,,,,,,,,,,,,
3/2009;1;8203;9999;1;2;0;1;,,,,,,,,,,,,,,,,,,,,,
3/2009;1;9996;9984;0;1;0;1;,,,,,,,,,,,,,,,,,,,,,
3/2009;1;4049;9987;1;3;0;1;,,,,,,,,,,,,,,,,,,,,,
3/2009;1;9993;8635;1;3;0;1;,,,,,,,,,,,,,,,,,,,,,
3/2009;1;8635;9994;2;3;0;1;,,,,,,,,,,,,,,,,,,,,,
3/2009;1;9998;9996;0;0;0;0;,,,,,,,,,,,,,,,,,,,,,
3/2009;1;9986;8342;2;2;0;0;,,,,,,,,,,,,,,,,,,,,,
3/2009;1;9984;10000;2;0;1;0;,,,,,,,,,,,,,,,,,,,,,
3/2009;1;9991;7947;1;1;0;0;,,,,,,,,,,,,,,,,,,,,,
```

Figure : données manquantes dans la tables MATCH (partie)

Mais le home_team_api_id et le away_team_id, ainsi que le nombre de buts marqués par chaque équipe pour chaque match sont tous enregistrés :

The screenshot displays the Talend Studio interface for configuring a filter on the 'MATCH' table. The filter is named 'tFilterRow_1' and is set to 'Ou' (OR) logic. The conditions are defined as follows:

Colonne d'entrée	Fonction	Opérateur	Valeur
away_team_api_id	Vider	Égal à	null
away_team_goal	Vider	Égal à	null
home_team_goal	Vider	Égal à	null
away_team_goal	Vider	Égal à	null

The interface also shows the number of rows affected: 25979 rows in the input and 0 rows in the output, indicating that all rows are filtered out due to missing data.

Figure : Vérification des manques de données avec Talend

Par conséquent, les données dans la table MATCH ne présentent pas de problème pour les analyses liées aux équipes. Pour les analyses des joueurs, ce dont nous avons surtout besoin, ce sont les données des tables PLAYER et PLAYER_ATTRIBUTE. Le rôle de la table MATCH dans l'analyse des joueurs est de trouver l'équipe à laquelle le joueur appartient et ses collègues.

De plus, nous constatons qu'il y a un total de huit enregistrements de données dans la table match, tandis que la table PLAYER_ATTRIBUTE contient en tout dix saisons de données.

B. Données non uniformes

Sur la table PLAYER, nous constatons que le format des noms des joueurs (player_name) n'est pas uniforme. Certains joueurs ont leur prénom et nom, certains n'ont que leur prénom ou nom et d'autres ont un chiffre à la fin qui est peut-être leur numéro de maillot.

A screenshot of a database record showing a player's name as '37450;Duda;'. The text is displayed in a monospaced font with a blue background and white text.

Figure : player_name n'ayant qu'un prénom

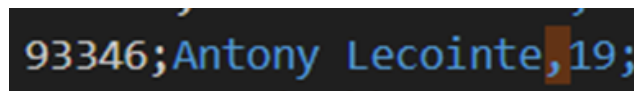
A screenshot of a database record showing a player's name as '93346;Antony Lecoïnte,19;'. The text is displayed in a monospaced font with a blue background and white text.

Figure : player_name ayant un chiffre à la fin

Pour les enregistrements qui n'ont que le prénom, nous devons utiliser player_api_id pour déterminer s'ils appartiennent à la même personne que les autres enregistrements. Pour les enregistrements suivis de chiffres, nous devons enlever ces chiffres, cela sera utile pour notre travail à venir. Pour cela, on a utilisé Excel qui a une fonctionnalité permettant de scinder une colonne en 2 grâce à un caractère de séparation, ici la virgule. Ensuite, on a supprimé la colonne contenant les nombres.

V. Résultats

A. Composition des équipes au fil des saisons

Le 1er résultat qu'on a cherché à obtenir est la composition des équipes au fil des saisons. Pour faire cela, on a d'abord créé un Excel où chaque ligne contient une saison et les identifiants d'un joueur et de son équipe qui a joué pendant cette saison. Il n'y a pas de doublons (ce fichier se trouve dans la partie V.Résultats - E.radar des joueurs par saison - 2. Par rapport aux autres équipes à la page 31)

season	team_api_id	player_id
2008/2009	8635	34480
2008/2009	8635	106013
2008/2009	8635	38388
2008/2009	8635	26458
2008/2009	8635	13423
2008/2009	8635	38389
2008/2009	8635	38798
2008/2009	8635	30949
2008/2009	8635	38253
2008/2009	8635	38383
2008/2009	8635	46552
2008/2009	9987	37937
2008/2009	9987	38780
2008/2009	9987	38293
2008/2009	9987	148313
2008/2009	9987	104411
2008/2009	9987	148314
2008/2009	9987	37202
2008/2009	9987	43158
2008/2009	9987	9307
2008/2009	9987	42153
2008/2009	9987	32690
2008/2009	9987	38782

Figure : Fichier Excel qui associe à chaque saison un joueur et son équipe

Ensuite, avec Talend, on a réalisé une jointure avec les dimensions team et player pour obtenir les noms.

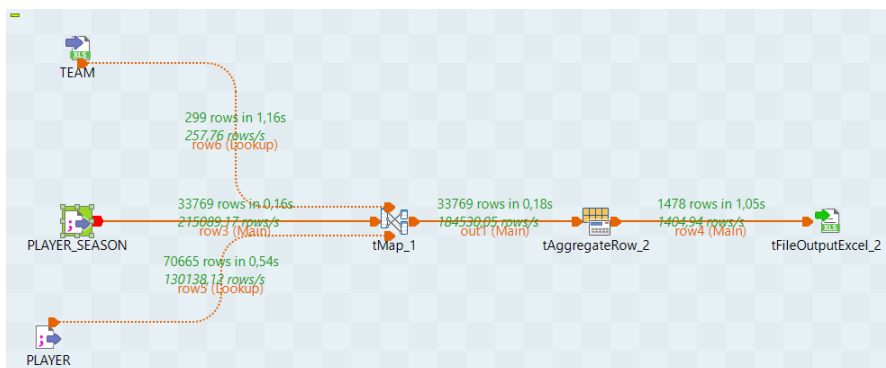


Figure : Traitement du fichier Excel des joueurs par saison dans Talend

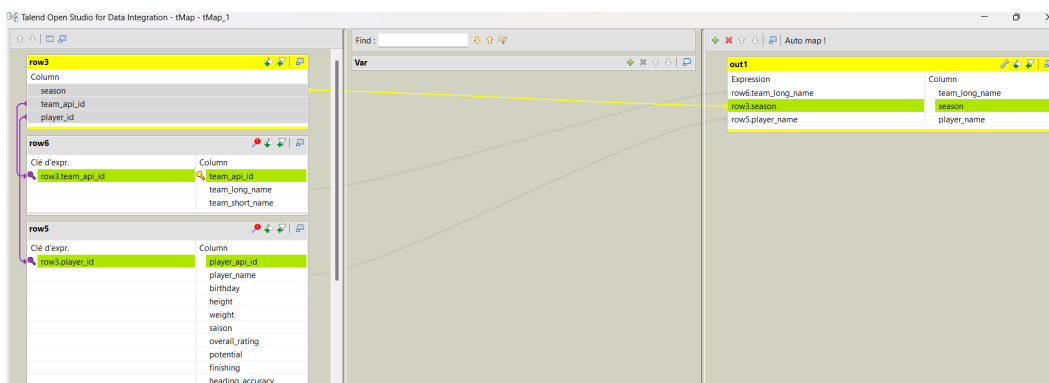


Figure : Récupération des noms des joueurs et des équipes

On souhaitait obtenir tous les joueurs d'une équipe pour une saison donnée sur une seule ligne. Donc on a utilisé tAggregateRow pour regrouper par équipe et par saison. Comme fonction d'agrégation, on a choisi "liste", qui donne la liste de tous les joueurs dans une seule variable.

On a donc obtenu, le fichier Excel suivant où pour chaque ligne, on a le nom de l'équipe, la saison et la liste de tous les joueurs de cette saison pour cette équipe. Comme ce format n'était pas exploitable, on a séparé la colonne des joueurs en plusieurs colonnes. Pour cela, on a utilisé la fonctionnalité d'Excel permettant de scinder une colonne en plusieurs grâce à un séparateur.

	A	B	C
1	team_long_name	season	player_name
2	Livorno	2009/2010	Francesco Benussi,Alessandro Bernardini,Marcus Diniz,Dario Knezevic,Leonardo Martin Miglionico,Cristian Raimondi,Davide Moro,Nico Pulzetti,Antonio Candreva,Mirko Pieri,Francesco Tavano,Cristiano Lucarelli,Alfi
3	AC Ajaccio	2011/2012	Guillermo Ochoa,Jackson Mendy,Anthony Lippini,Arnaud Maire,Yohann Poulard,Samuel Bouhours,Carl Medjani,Mehdi Mostefa,Paul Bastien Lasne,Johan Cavalli,Frederic Sammaritano,Richard Socrier,Leyti N'Diaye,Je
4	Málaga CF	2014/2015	Carlos Kameni,Juanpi,Roberto Rosales,Sergio Sanchez,Welington,Antunes,Ignacio Camacho,Sergi Darder,Samuel Garcia,Juanmi,Samuel Castillejo,Nordin Amrabat,Miguel Torres,Roque Santa Cruz,Marcos Angeleri,Dudi
5	FC Twente	2011/2012	Sander Boschker,Bart Buysse,Tim Cornelisse,Peter Wisgerhof,Rasmus Bengtsson,Dwight Tiendalli,Wout Brama,Luuk de Jong,Willem Janssen,Bryan Ruiz,Marc Janko,Steven Berghuis,Nikolai Mihalov,Roberto Rosales,
6	Montpellier Hérault SC	2012/2013	Geoffrey Jourden,Karim Ait Fana,Garry Bocaly,Hilton,Henri Bedimo,Daniel Congre,Jamel Saihi,Remy Cabella,Younes Belhanda,John Utaka,Gaetan Charbonnier,Souleymane Camara,Abdelhamid El Kaoutari,Mapou Ya
7	Borussia Dortmund	2013/2014	Mitchell Langerak,Milos Jojic,Kevin Grosskreutz,Neven Subotic,Mats Hummels,Marcel Schmelzer,Sven Bender,Nuri Sahin,Pierre-Emerick Aubameyang,Illkay Guendogan,Marco Reus,Robert Lewandowski,Roman Weide
8	West Bromwich Albion	2014/2015	Ben Foster,Boaz Myhill,Andre Wisdom,Craig Dawson,Joleon Lescott,Sebastien Pocognoli,James Morrison,Craig Gardner,Graham Dorrans,Stephane Sessegnon,Chris Brunt,Saido Berahino,Chris Baird,Claudio Yacob,Ge
9	Stade Rennais FC	2010/2011	Nicolas Douchez,Georges Mandjeck,Romain Danze,Rod Fanni,Jean-Armel Kana-Biyik,Samuel Souprayen,Jires Kembo-Ekoko,Stephane Dalmat,Yann M'Vila,Abdoul Razzaqui Camara,Victor Hugo Montano,Kader Manga
10	AS Saint-Étienne	2013/2014	Stephane Ruffier,Banel Nicolita,Francois Clerc,Loic Perrin,Kurt Zouma,Jonathan Brison,Jeremy Clement,Fabien Lemoine,Renaud Cohade,Romain Hamouma,Franck Tabanou,Brandao,Faouzi Ghoulam,Jamael Diomand
11	Hannover 96	2009/2010	Robert Enke,Altin Lala,Steven Cherundolo,Karim Haggui,Christian Schulz,Constant Djakpa,Hanno Balitsch,Jan Rosenthal,Jacek Krzynowek,Arnold Bruggink,Jiri Stajner,Mikael Forssell,Konstantin Rausch,Sergio Pinto,Did
12	Napoli	2012/2013	Morgan De Sanctis,Federico Fernandez,Hugo Campagnaro,Paolo Cannavaro,Miguel Britos,Christian Maggio,Valon Behrami,Gokhan Inler,Marek Hamsik,Salvatore Aronica,Lorenzo Insigne,Edison Cavani,Alessandro C
13	Roda JC Kerkrade	2008/2009	Bram Castro,Anour Hadouir,Davy de Fauw,Pa Modou Kah,Kris De Wree,Marcel De Jong,Ruud Vormer,Marcel Meeuwis,Laurent Delorge,Boldizar Bodor,Sekou Cisse,Willem Janssen,Vincent Lachambre
14	Rayo Vallecano	2014/2015	Cristian Alvarez,Jorge Garcia Morcillo,Tito,Ze Castro,Abdoulaye Ba,Emiliano Insua,Roberto Trashorras,Lica,Alberto Bueno,Gael Kakuta,Leo Baptistao,Quini,Antonio Amaya,Jose Raul Beena,Jozabed Javier Aquino,Ton
15	Bolton Wanderers	2008/2009	Jussi Jaaskelainen,Euzebiusz Smolarek,Gretar Rafn Steinsson,Gary Cahill,Andrew O'Brien,Illyod Samuel,Kevin Davies,Gavin McCann,Fabrice Muamba,Matthew Taylor,Ricardo Gardner,Johan Elmander,Kevin Nolan,Joe
16	Roma	2008/2009	Doni,Souleymane Adama Diamoutene,Christian Panucci,Philippe Mexes,Juan,John Arne Riise,Daniele De Rossi,David Pizarro,Rodrigo Taddei,Simone Perrotta,Mirko Vucinic,Julio Baptista,Cicinho,Matteo Brighi,France
17	St. Mirren	2012/2013	Craig Samson,Esmael Goncalves,David van Zanten,Marc McCausland,Lee Mair,Paul Dummett,Thomas Reilly,Dougie Imrie,Jim Goodwin,Kenny McLean,Steven Thompson,Sam Parkin,Gary Teale,Graham Carey,John McC
18	Lazio	2008/2009	Juan Pablo Carrizo,Francois Lallou,Stephan Lichtsteiner,Sebastiano Siviglia,David Rozehnal,Aleksandar Kolarov,Christian Brocchi,Christian Daniel Ledesma,Stefano Mauri,Goran Pandev,Mauro Zarate,Stefan De
19	Sporting Lokeren	2014/2015	Davino Verhulst,Cyriel Dessers,Arthur,Mijat Maric,Alexander Scholz,Denis Odo,Koen Persoons,Kilian Overmeire,Hans Vanaken,Hamdi Harbaoui,Jordan Remacle,Boubacar Barry,Copa,Gregory Mertens,Nill De Pauw,M
20	Cesena	2011/2012	Francesco Antonioli,Jorge Martinez,Luca Ceccarelli,Yohan Benalouane,Guillermo Rodriguez,Marco Rossi,Roberto Guana,Marco Parolo,Abdelkader Ghezzal,Adnan Muta,Antonio Candreva,Maurizio Lauro,Gianluca Co
21	Karlsruher SC	2008/2009	Markus Miller,Mahir Saglik,Andreas Goeftitz,Tim Sebastian,Martin Stoll,Christian Eichner,Godfried Aduobe,Michael Mutzel,Sebastian Freis,Alexander Iashvili,Joshua Kennedy,Stefano Celozzi,Antonio da Silva,Timo Staf
22	Jagiellonia Bialystok	2010/2011	Grzegorz Sandomierski,Igor Lewczuk,Alexis Norambuena Ruz,Maciej Makuszewski,Rafal Grzyb,Neves Soares Hermes,Tomasz Kupisz,Kamil Grosicki,Przemyslaw Trytko,Thiago Clonек
23	Stade Rennais FC	2009/2010	Nicolas Douchez,Lucien Aubey,Kader Mangane,Rod Fanni,Petter Hansson,Romain Danze,Jerome Leroy,Sylvain Marveaux,Bruno Cheyrou,Yann M'Vila,Jires Kembo-Ekoko,Moussa Sow,Carlos Bocanegra,Alexander Tet
24	Heracles Almelo	2011/2012	Remko Pasveer,Joey Belterman,Tim Breukers,Mike Te Wierik,Antoine van der Linden,Marko Vejinovic,Willy Overtoom,Kwame Quansah,Darl Douglas,Samuel Armenteros,Everton,Mark Looms,Ben Rienstra,Glynor Ple
25	BSC Young Boys	2015/2016	Yvon Mvogo,Samuel Afum,Florent Hadergionaj,Milan Vilotic,Steve von Bergen,Jan Lecjaks,Leonardo Bertone,Milan Gajic,Renato Steffen,Alexander Gerndt,Miralem Sulejmani,Guillaume Hoarau,Scott Lee Sutter,Denis
26	Bayern Munich	2013/2014	Manuel Neuer,Philipp Lahm,Thomas Muller,Arjen Robben,Bastian Schweinsteiger,David Alaba,Robert Lewandowski,Jerome Boateng,Sebastian Rudy,David Alaba,Thomas Muller,Arjen Robben,Bastian Schweinsteiger,David Alaba,Robert Lewandowski,Jerome Boateng,Sebastian Rudy,David Alaba

Figure : Fichier Excel donnant la liste de tous les joueurs dans une équipe pendant une saison

Finalement, on a affiché le résultat avec PowerBI sous la forme d'un tableau. Il y a aussi un filtre sur les saisons pour sélectionner une saison en particulier.

season	team	player1	player2	player3	player4
<input type="checkbox"/> 2008/2009	Aberdeen	Jamie Langfield	Peter Pawlett	Lee Mair	Andrew Considine
<input type="checkbox"/> 2009/2010	AC Bellinzona	Carlo Zotti	Genc Mehmeti	Angelo Raso	Iacopo La Rocca
<input type="checkbox"/> 2010/2011	Académica de Coimbra	Boris Peskovic	Jose Luis Garces	Pedrinho	Orlando
<input type="checkbox"/> 2011/2012	ADO Den Haag	Robert Zwinkels	Bogdan Milic	Ahmed Ammi	Csaba Horvath
<input type="checkbox"/> 2012/2013	AJ Auxerre	Remy Riou	Valter Birsa	Cedric Hengbart	Moussa Narry
<input type="checkbox"/> 2013/2014	Ajax	Maarten Stekelenburg	Santiago Leonardo	Bruno Silva	Oleguer
<input type="checkbox"/> 2014/2015	Amadora	Nelson	Hugo Carreira	Hugo Gomes	Nuno Andre Coelho
<input type="checkbox"/> 2015/2016	Arka Gdynia	Norbert Witkowski	Marcin Budzinski	Blazej Telichowski	Bartosz Lawa
	Arsenal	Manuel Almunia	Vito Mannone	Bacary Sagna	Kolo Toure
	AS Monaco	Stephane Ruffier	Thomas Mangani	Dario Simic	Leandro Cufre
	AS Nancy-Lorraine	Gennaro Bracigliano	Joel Sami	Andre Luiz	Damian Macaluso Rojas
	AS Saint-Etienne	Jody Viviani	Sebastien Grax	Efstathios Tavliridis	Mouhamadou Dabo
	Aston Villa	Brad Friedel	Emile Heskey	Luke Young	Carlos Cuellar
	Atalanta	Ferdinando Coppola	Claudio Rivalta	Gyorgy Garics	Daniele Capelli
	Athletic Club de Bilbao	Gorka Iraizoz Moreno	Eneko Boveda	Andoni Iraola	Aitor Ocio
	Atlético Madrid	Gregory Coupet	Luis Garcia	Luis Perea	Johnny Heitinga
	AZ	Joey Didulica	Brett Holman	Gill Syerts	Kew Jaliens
	Bayer 04 Leverkusen	Rene Adler	Karim Haggui	Gonzalo Castro	Manuel Friedrich
	Beerschot AC	Silvio Proto	Henri Muryaneza	Pieterjan Monteyne	Kurt van Dooren
	Blackburn Rovers	Paul Robinson	Keith Treacy	Andre Ooijer	Christopher Samba
	Bologna	Francesco Antonioli	Luigi Lavecchia	Cristian Zenoni	Vangelis Moras
	Bolton Wanderers	Jussi Jaaskelainen	Euzebiusz Smolarek	Gretar Rafn Steinsson	Gary Cahill
	Borussia Dortmund	Roman Weidenfeller	Robert Kovac	Antonio Rukavina	Neven Subotic
	Borussia Mönchengladbach	Fredric Loehe	Thomas Kleine	Patrick Paauwe	Filip Daems
	BSC Young Boys	Marco Woelfli	Guillermo Ariel Pereyra	Miguel Portillo	Saif Ghezal
	CA Osasuna	Roberto Fernandez	Javier Flano	Cesar Azpilicueta	Miguel Flano
	Cagliari	Federico Marchetti	Cristiano Lupatelli	Carlos Matheu	Diego Luis Breijo Lopez

Figure : Tableau de la composition des équipes filtrées par saison dans PowerBI

B. Caractéristiques des joueurs au fil des saisons

Traitement de la table PLAYER_ATTRIBUTE :

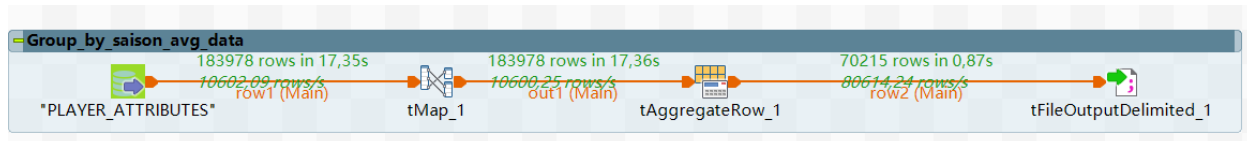


Figure : Traitement de la table PLAYER_ATTRIBUTE dans Talend

Dans la table PLAYER_ATTRIBUTE, chaque joueur a plusieurs enregistrements de statistiques de match, et chaque enregistrement a une date. En l'absence d'informations explicites, nous supposons que cela représente les performances d'un joueur lors du match à la date de l'enregistrement. Etant donné que toutes les demandes auxquelles nous devons répondre sont basées sur la saison, lors du traitement des données, nous avons choisi de convertir toutes les dates en saisons.

Nous devons d'abord savoir la date de début et la date de fin pour une saison. En consultant les données de la table MATCH, nous pouvons trouver la date du dernier match de chaque saison :

2008/2009	2008-07-18 00:00:00	2012/2013	2012-07-13 00:00:00
2008/2009	2009-05-31 00:00:00	2012/2013	2013-06-02 00:00:00
2009/2010	2009-07-11 00:00:00	2013/2014	2013-07-13 00:00:00
2009/2010	2010-05-16 00:00:00	2013/2014	2014-05-18 00:00:00
2010/2011	2010-07-17 00:00:00	2014/2015	2014-07-18 00:00:00
2010/2011	2011-05-29 00:00:00	2014/2015	2015-05-31 00:00:00
2011/2012	2011-07-16 00:00:00	2015/2016	2015-07-17 00:00:00
2011/2012	2012-05-23 00:00:00	2015/2016	2016-05-25 00:00:00

Figure : Premier match et dernier match enregistrés pour chaque saison

Nous savons que la date du premier match d'une saison peut différer de la date de début de la saison. Dans la table de dimension SAISON, nous spécifions que la date de début de chaque saison est le 1er juillet de l'année en cours et que la date de fin est le 30 juin de l'année suivante. De cette manière, nous pouvons convertir la date de PLAYER_ATTRIBUTE en saison par une exécution suivante :

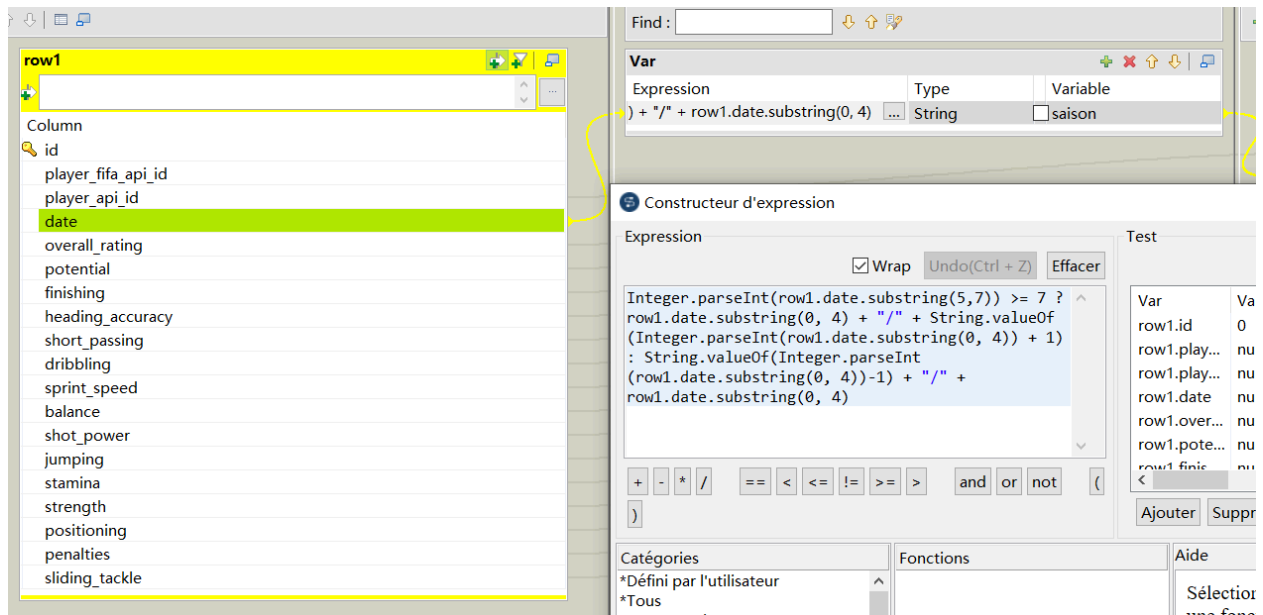


Figure : Convertissement de la date à la saison

C'est-à-dire que nous déterminons à quelle saison appartient un enregistrement en vérifiant si la date associée est après juillet.

Ensuite, nous connectons les données produites par le composant tMap au composant tAggregateRow, en regroupant par player_api_id et la saison, afin de calculer la moyenne de chaque performance :

tAggregateRow_1

Paramètres simples

Paramètres avancés

Paramètres dynamiques

Vue

Documentation

Schéma

Built-in

Modifier le schéma

Sync colonnes

Group by

Colonne de sortie	Position de la colonne d'entrée
player_api_id	player_api_id
saison	saison

Opérations

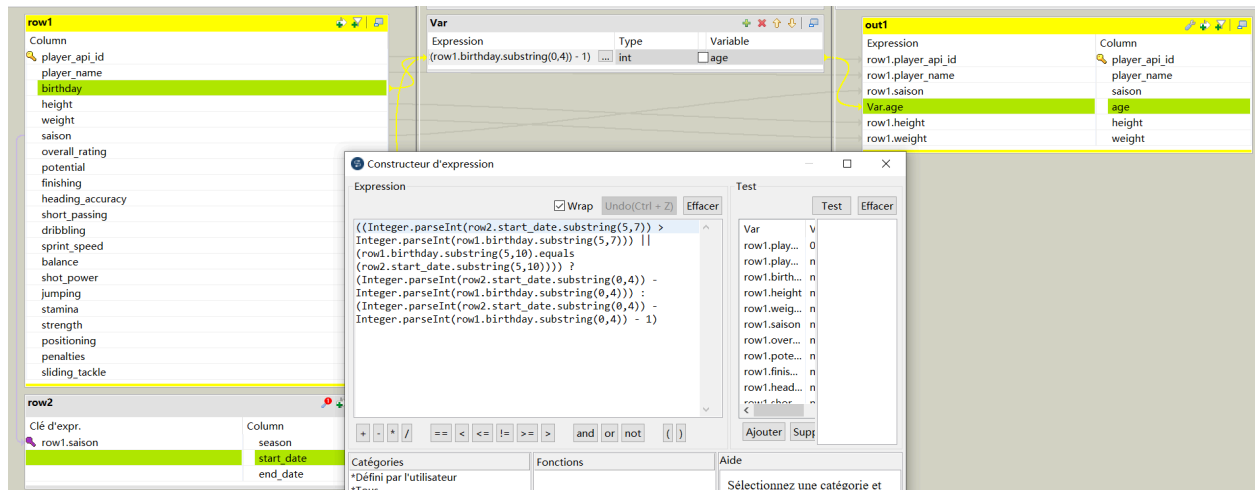
Colonne de sortie	Fonction	Position de la colonne d'e...	<input type="checkbox"/> Ignorer les valeurs nulles
overall_rating	avg	overall_rating	<input checked="" type="checkbox"/>
potential	avg	potential	<input checked="" type="checkbox"/>
finishing	avg	finishing	<input checked="" type="checkbox"/>
heading_accuracy	avg	heading_accuracy	<input checked="" type="checkbox"/>
short_passing	avg	short_passing	<input checked="" type="checkbox"/>
dribbling	avg	dribbling	<input checked="" type="checkbox"/>
sprint_speed	avg	sprint_speed	<input checked="" type="checkbox"/>
balance	avg	balance	<input checked="" type="checkbox"/>
shot_power	avg	shot_power	<input checked="" type="checkbox"/>
jumping	avg	jumping	<input checked="" type="checkbox"/>
stamina	avg	stamina	<input checked="" type="checkbox"/>
strength	avg	strength	<input checked="" type="checkbox"/>
positioning	avg	positioning	<input checked="" type="checkbox"/>
penalties	avg	penalties	<input checked="" type="checkbox"/>
sliding_tackle	avg	sliding_tackle	<input checked="" type="checkbox"/>

Figure : Exécution du composant tAggregateRow

Nous appelons la table PLAYER_ATTRIBUTE_ALL_TIME qui prend les données sorties de tAggregateRow. En joignant cette table et la table PLAYER par player_api_id, nous obtenons enfin la dimension PLAYER.

Dans la base de données d'origine, la taille et le poids des joueurs sont stockés dans la table PLAYER et ne varient pas avec la saison, seul l'âge des joueurs change au fil des saisons. Nous spécifions que l'âge d'un joueur pour une saison donnée est celui qu'il a le premier jour de cette saison. Par exemple, si la date de naissance d'un joueur est le 01-01-2000, alors son âge pour la saison 2018/2019 est de 18 ans; si la date de naissance d'un joueur est le 2 juillet 2000, alors son âge pour la saison 2018/2019 est de 17 ans.

Nous connectons la table de dimension PLAYER et la table de dimension TIME via une jointure à l'aide de PLAYER.saison et TIME.season. Nous calculons l'âge d'un joueur pour une saison en fonction de sa date de naissance et de la date de début de la saison :



Nous avons donc une table qui contient les caractères (taille, poids et l'âge pour la saison) des joueurs.

player_name	height	weight	saison	age
<input type="checkbox"/> Aaron Appindangoye	170.18	146	2006/2007	16
<input type="checkbox"/> Aaron Appindangoye	170.18	146	2008/2009	18
<input checked="" type="checkbox"/> Aaron Cresswell	170.18	146	2009/2010	19
<input checked="" type="checkbox"/> Aaron Cresswell	170.18	146	2010/2011	20
<input type="checkbox"/> Aaron Doran	170.18	146	2011/2012	21
<input type="checkbox"/> Aaron Galindo	170.18	146	2012/2013	22
<input type="checkbox"/> Aaron Hughes	170.18	146	2013/2014	23
<input type="checkbox"/> Aaron Hughes	170.18	146	2014/2015	24
	170.18	146	2015/2016	25

Figure : Exemple des caractéristiques du joueur

C. Matchs gagnés par équipes

1. Par saison

Pour déterminer le nombre de matchs gagnés par équipe par saison, on a récupéré, le nom de l'équipe, la saison et les victoires pour away_team et home_team. On a utilisé les 2 équipes car une même équipe peut être parfois home_team et parfois away_team. Il fallait donc additionner les 2 victoires.

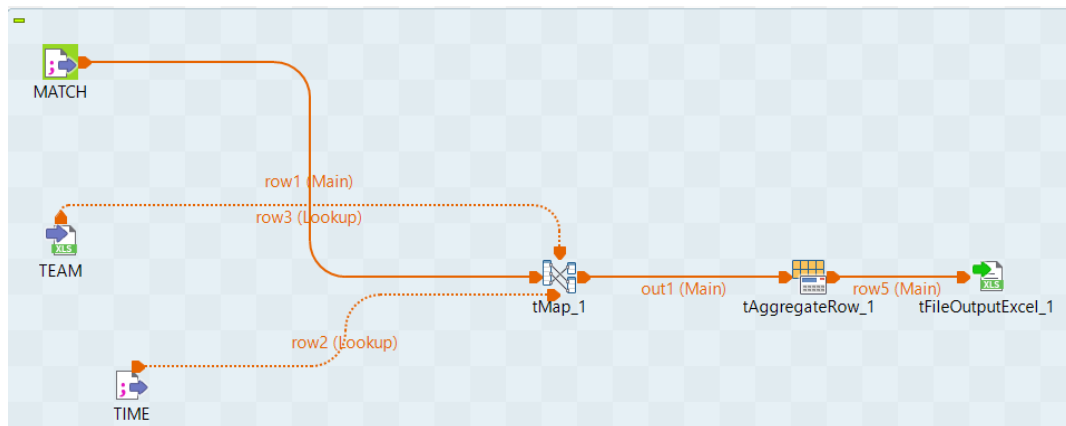


Figure : Création d'un fichier pour obtenir toutes les victoires par saison par équipes

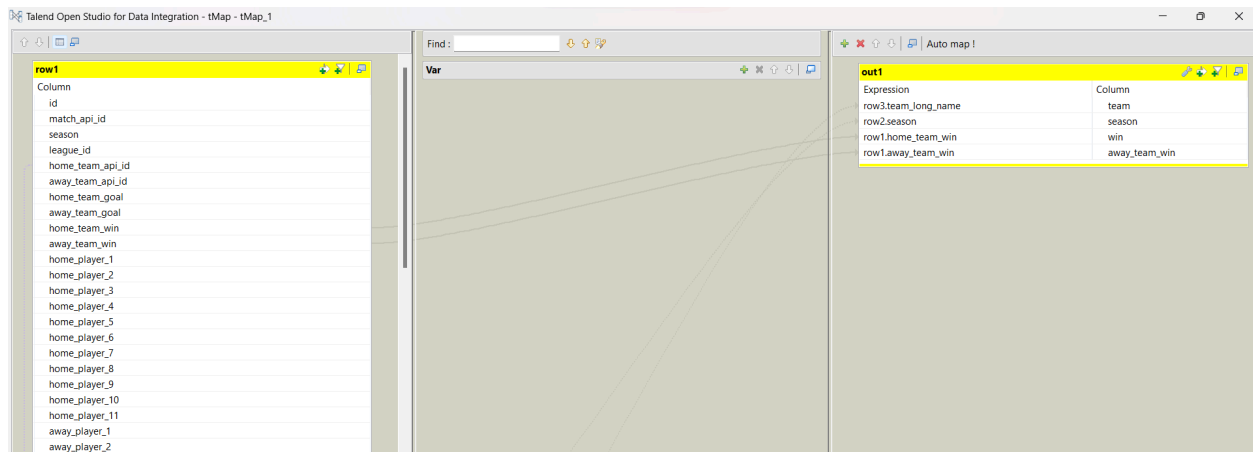


Figure : Récupération de la saison, des victoires des équipes et de leur nom

On a ensuite fait une agrégation par saison et équipes avec pour fonction d'agrégation la somme des 2 victoires.

Sur le fichier Excel obtenu, on a sommé home_team_win et away_team_win pour avoir le nombre total de victoires.

	A	B	C	D	E
1	team	season	home_team_win	away_team_win	wins
2	Aberdeen	2008/2009	9	5	14
3	Aberdeen	2009/2010	6	10	16
4	Aberdeen	2015/2016	12	3	15
5	Aberdeen	2012/2013	6	4	10
6	Aberdeen	2011/2012	6	5	11
7	Aberdeen	2013/2014	10	5	15
8	Aberdeen	2010/2011	6	11	17
9	Aberdeen	2014/2015	12	4	16
10	AC Ajaccio	2011/2012	6	5	11
11	AC Ajaccio	2013/2014	3	10	13
12	AC Ajaccio	2012/2013	7	7	14
13	AC Arles-Avignon	2010/2011	2	11	13
14	AC Bellinzona	2010/2011	4	8	12
15	AC Bellinzona	2009/2010	5	10	15
16	AC Bellinzona	2008/2009	8	6	14
17	Académica de Coimbra	2008/2009	8	2	10
18	Académica de Coimbra	2011/2012	4	7	11
19	Académica de Coimbra	2009/2010	4	6	10
20	Académica de Coimbra	2012/2013	4	6	10
21	Académica de Coimbra	2015/2016	5	7	12
22	Académica de Coimbra	2013/2014	5	4	9
23	Académica de Coimbra	2014/2015	1	4	5
24	Académica de Coimbra	2010/2011	4	5	9
25	ADO Den Haag	2011/2012	5	6	11
26	ADO Den Haag	2010/2011	10	4	14

Figure : Fichier résultat

On a donc construit le graphe suivant des victoires en fonction des saison pour une équipe. Il suffit de sélectionner une équipe sur le filtre pour avoir les données de cette équipe.

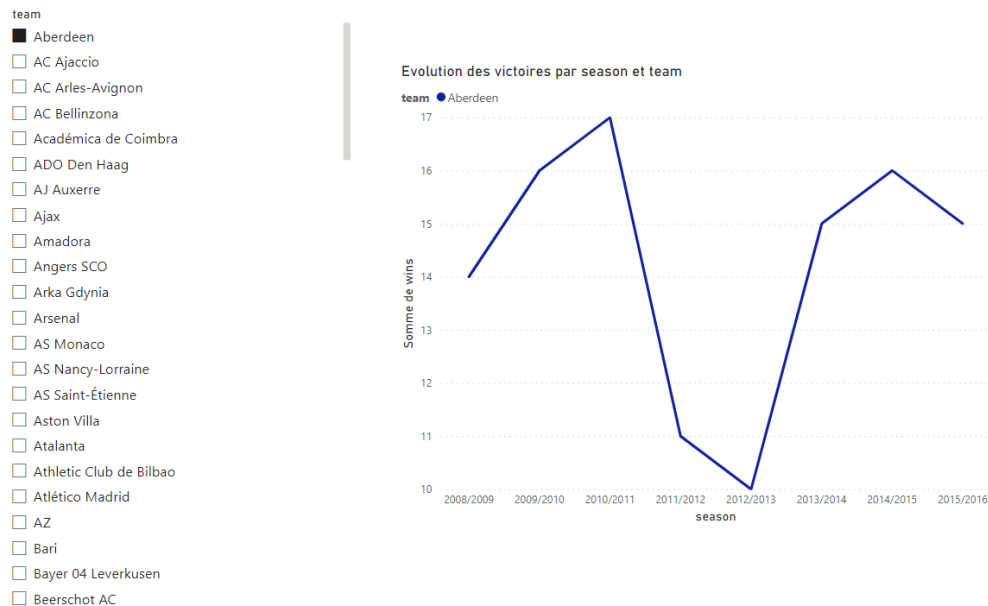


Figure : Courbe d'évolution des victoires par saison et équipes

2. Par ligue

La méthode pour déterminer les victoires par ligues est exactement la même que celle pour déterminer les victoires par saisons: on a récupéré, le nom de l'équipe, la ligue et les victoires pour away_team et home_team.

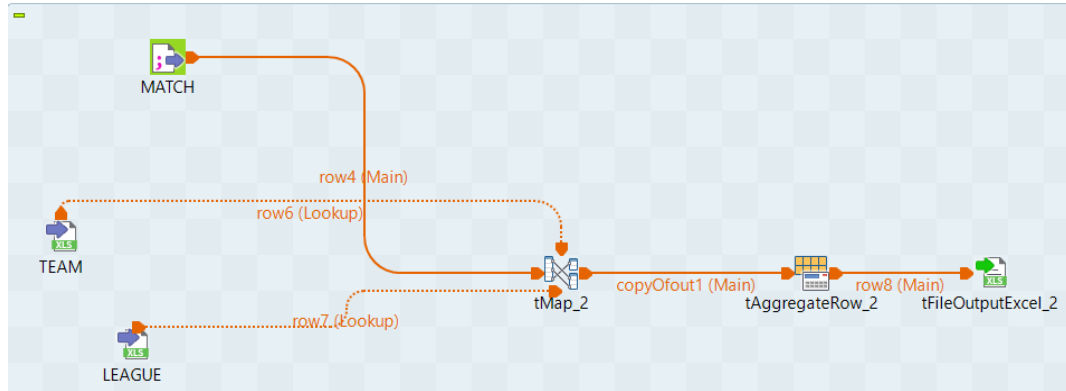


Figure : Création d'un fichier pour obtenir toutes les victoires par ligues par équipes

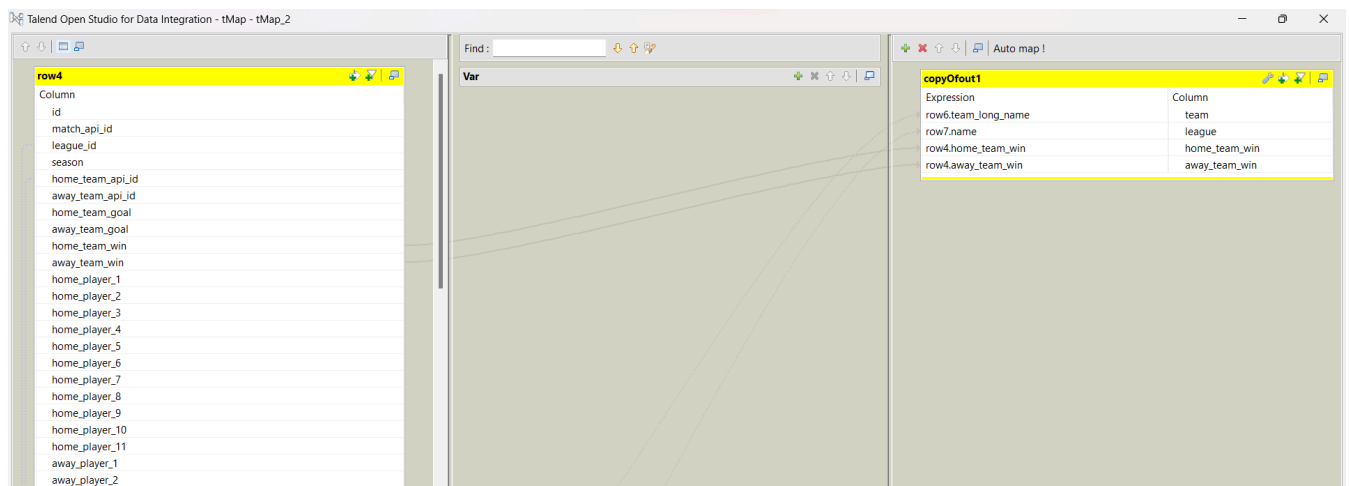


Figure : Récupération des victoires des équipes, de leur ligue et de leur nom

On a ensuite fait une agrégation par saison et équipes avec pour fonction d'agrégation la somme des 2 victoires.

Sur le fichier Excel obtenu, on a sommé home_team_win et away_team_win pour avoir le nombre total de victoires.

	A	B	C	D	E
1	team	league	home_team_win	away_team_win	wins
2	Aberdeen	Scotland Premier League	67	47	114
3	AC Ajaccio	France Ligue 1	16	22	38
4	AC Arles-Avignon	France Ligue 1	2	11	13
5	AC Bellinzona	Switzerland Super League	17	24	41
6	Académica de Coimbra	Portugal Liga ZON Sagres	35	41	76
7	ADO Den Haag	Netherlands Eredivisie	50	47	97
8	AJ Auxerre	France Ligue 1	31	18	49
9	Ajax	Netherlands Eredivisie	103	9	112
10	Amadora	Portugal Liga ZON Sagres	7	4	11
11	Angers SCO	France Ligue 1	6	5	11
12	Arka Gdynia	Poland Ekstraklasa	16	17	33
13	Arsenal	England Premier League	97	21	118
14	AS Monaco	France Ligue 1	53	22	75
15	AS Nancy-Lorraine	France Ligue 1	30	36	66
16	AS Saint-Étienne	France Ligue 1	75	37	112
17	Aston Villa	England Premier League	45	57	102
18	Atalanta	Italy Serie A	55	42	97
19	Athletic Club de Bilbao	Spain LIGA BBVA	80	41	121
20	Atlético Madrid	Spain LIGA BBVA	103	25	128
21	AZ	Netherlands Eredivisie	82	28	110
22	Bari	Italy Serie A	13	15	28
23	Bayer 04 Leverkusen	Germany 1. Bundesliga	76	26	102
24	Beerschot AC	Belgium Jupiler League	30	21	51
25	Birmingham City	England Premier League	14	7	21
26	Blackburn Rovers	England Premier League	20	26	46

Figure : Fichier résultat

On a donc construit 2 graphes:

- celui des victoires en fonction des équipes pour une ligue. Il suffit de sélectionner une ligue sur le filtre pour avoir les données de cette ligue. Nous voulions initialement uniquement tracer l'évolution des victoires en fonction des ligues mais chaque équipe n'était associée qu'à une seule ligue. Le graphe ne contenait juste qu'un point, ce qui ne semblait pas pertinent. On a donc aussi tracé ce graphe.

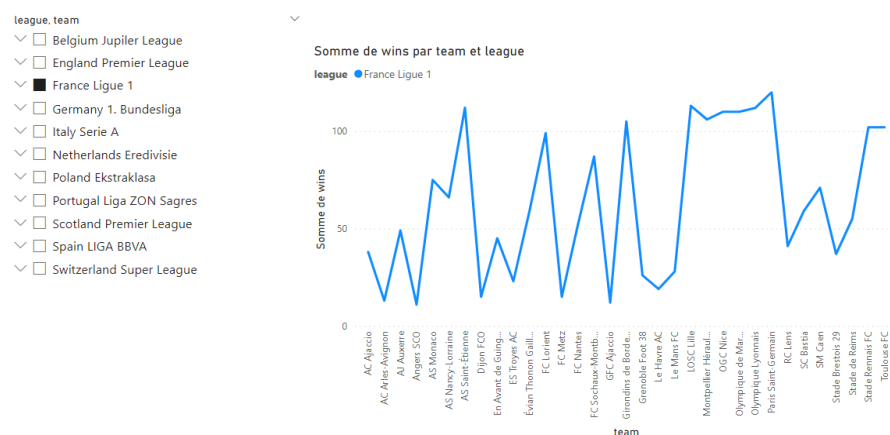


Figure : Courbe d'évolution des victoires par équipe sur une ligue

- celui des victoires par équipes par ligues. Le filtre permet de filtrer par équipe. Chaque graphe filtré contient donc un seul point.

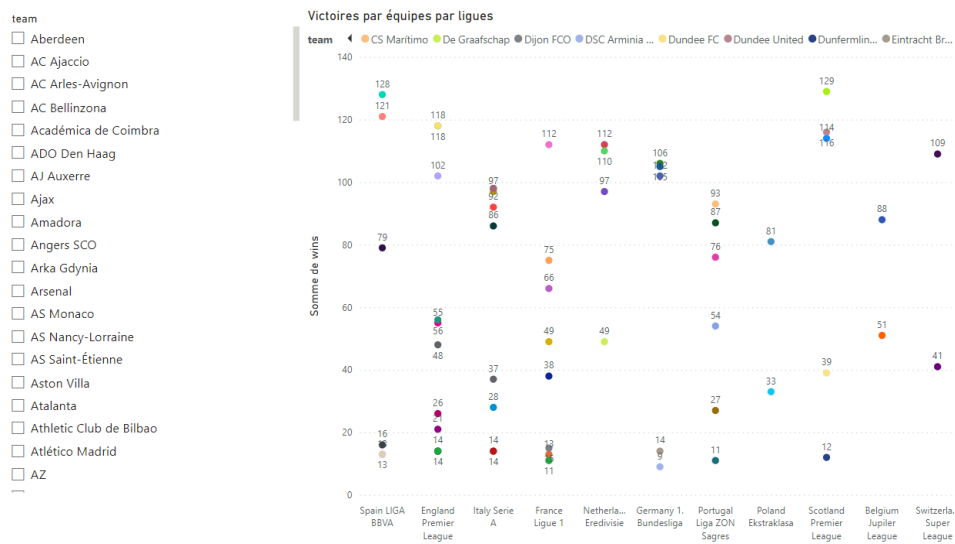


Figure : Nuage de points des victoires par ligues et par équipes

3. Par équipes

La méthode pour déterminer les victoires par équipes est presque la même que celles pour déterminer les victoires par saisons et ligues: on a créé 2 output. Dans le 1er, on a le nom de l'équipe home, la ligue et les victoires pour away_team et home_team. Dans le 2ème, on a le nom de l'équipe away, la ligue et les victoires pour away_team et home_team. Pour le 2ème, on a aussi échangé le nom des variables entre home_team et away_team pour avoir le même nom de colonnes

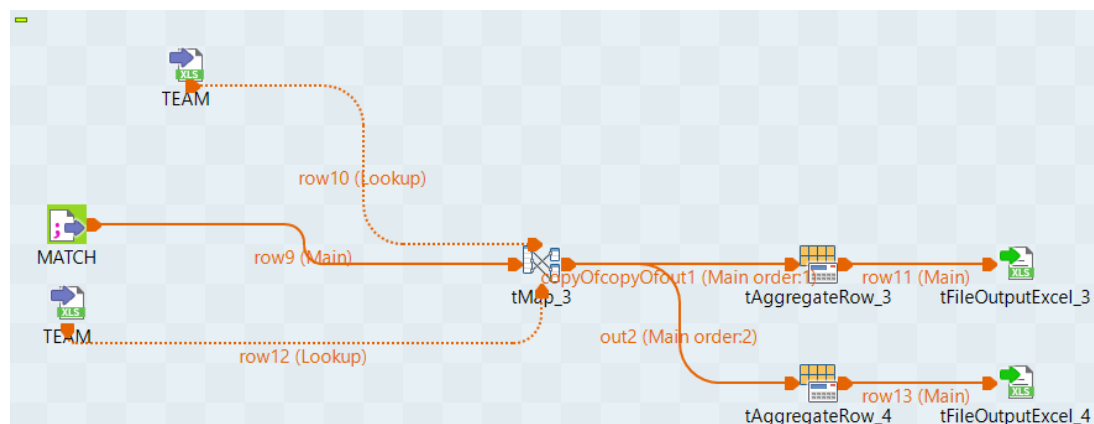


Figure : Création d'un fichier pour obtenir toutes les victoires par équipes par rapport aux autres équipes

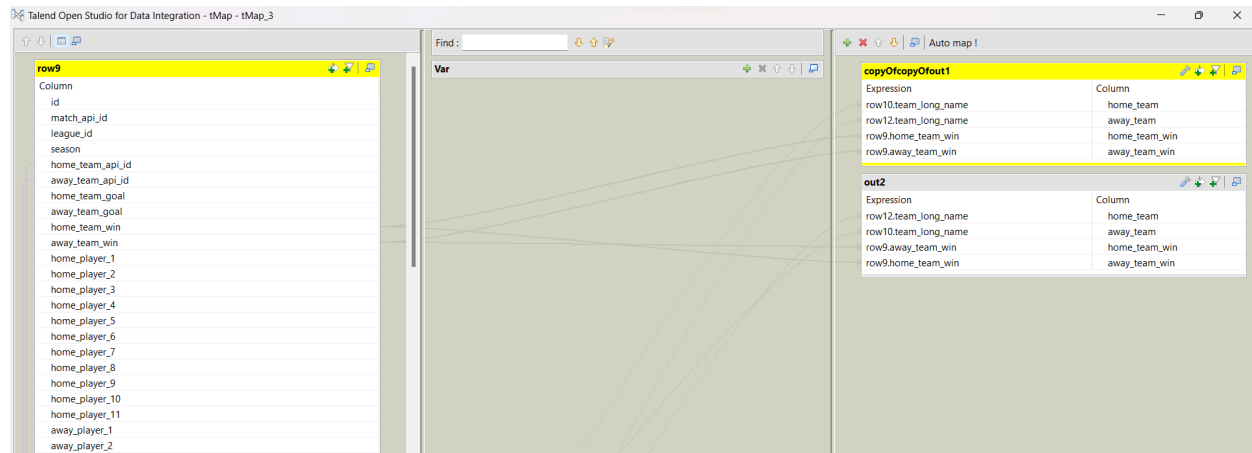


Figure : Récupération des victoires et des noms des 2 équipes

On a donc obtenu 2 fichiers Excel pour chaque output du tmap après regroupement par équipes. Les fichiers ont été fusionnés en gardant seulement les colonnes home_team_win. Puis les 2 colonnes restantes ont été additionnées pour avoir le nombre de victoires total.

	A	B	C	D
1	home_team	away_team	home_team_win	away_team_win
2	Liverpool	Chelsea	3	3
3	Wolverhampton Wanderers	Portsmouth	0	1
4	FC Zürich	FC Basel	1	0
5	FC Köln	Karlsruher SC	0	0
6	Stade Rennais FC	OGC Nice	4	2
7	Manchester City	Swansea City	5	0
8	Sunderland	Leicester City	0	1
9	Borussia Dortmund	FC Ingolstadt 04	1	0
10	Sparta Rotterdam	De Graafschap	0	0
11	Dundee FC	Inverness Caledonian Thistle	0	3
12	Legia Warszawa	Piast Gliwice	5	0
13	Hertha BSC Berlin	VfB Stuttgart	4	2
14	Sporting CP	FC Paços de Ferreira	4	2
15	Hull City	Manchester City	1	2
16	FC Augsburg	Hertha BSC Berlin	2	1
17	Angers SCO	AS Monaco	1	0
18	Tubize	KSV Cercle Brugge	1	0
19	Waasland-Beveren	Royal Excel Mouscron	1	0
20	Real Valladolid	Xerez Club Deportivo	0	0
21	CD Numancia	Getafe CF	1	0
22	Olympique Lyonnais	Girondins de Bordeaux	3	2
23	FC Basel	AC Bellinzona	5	0
24	Sunderland	Swansea City	1	1
25	Stade Rennais FC	Olympique Lyonnais	2	3

Figure : 1er fichier résultat

	A	B	C	D
1	home_team	away_team	home_team_win	away_team_win
2	Liverpool	Chelsea	4	2
3	Wolverhampton Wanderers	Portsmouth	0	1
4	FC Zürich	FC Basel	2	10
5	FC Köln	Karlsruher SC	1	0
6	Stade Rennais FC	OGC Nice	3	4
7	Manchester City	Swansea City	2	1
8	Sunderland	Leicester City	0	1
9	Borussia Dortmund	FC Ingolstadt 04	1	0
10	Sparta Rotterdam	De Graafschap	0	0
11	Dundee FC	Inverness Caledonian Thistle	0	2
12	Legia Warszawa	Piast Gliwice	2	2
13	Hertha BSC Berlin	VfB Stuttgart	1	3
14	Sporting CP	FC Paços de Ferreira	3	2
15	Hull City	Manchester City	0	2
16	FC Augsburg	Hertha BSC Berlin	0	1
17	Angers SCO	AS Monaco	0	1
18	Tubize	KSV Cercle Brugge	0	1
19	Waasland-Beveren	Royal Excel Mouscron	1	1
20	Real Valladolid	Xerez Club Deportivo	0	1
21	CD Numancia	Getafe CF	0	1
22	Olympique Lyonnais	Girondins de Bordeaux	3	4
23	FC Basel	AC Bellinzona	4	1
24	Sunderland	Swansea City	1	1
25	Stade Rennais FC	Olympique Lyonnais	2	2

Figure : 2ème fichier résultat

	A	B	C	D	E
1	home_team	away_team	home_team_win	home_team_win	wins
2	Aberdeen	Dundee FC	4	3	7
3	Aberdeen	Hamilton Academical FC	6	3	9
4	Aberdeen	Dundee United	5	7	12
5	Aberdeen	Kilmarnock	9	6	15
6	Aberdeen	St. Johnstone FC	4	5	9
7	Aberdeen	Heart of Midlothian	3	2	5
8	Aberdeen	Motherwell	4	3	7
9	Aberdeen	Celtic	4	5	9
10	Aberdeen	St. Mirren	6	0	6
11	Aberdeen	Dunfermline Athletic	2	1	3
12	Aberdeen	Partick Thistle F.C.	2	4	6
13	Aberdeen	Ross County FC	4	2	6
14	Aberdeen	Rangers	1	0	1
15	Aberdeen	Falkirk	2	1	3
16	Aberdeen	Inverness Caledonian Thistle	6	2	8
17	Aberdeen	Hibernian	5	0	5
18	AC Ajaccio	ES Troyes AC	0	0	0
19	AC Ajaccio	Dijon FCO	1	0	1
20	AC Ajaccio	Paris Saint-Germain	0	0	0
21	AC Ajaccio	Toulouse FC	0	2	2
22	AC Ajaccio	FC Sochaux-Montbéliard	1	1	2
23	AC Ajaccio	Stade Brestois 29	1	0	1
24	AC Ajaccio	Girondins de Bordeaux	1	0	1
25	AC Ajaccio	Stade Rennais FC	2	0	2

Figure : Fichier final résultat de la fusion des 2 fichiers obtenus

On a donc obtenu le graphe suivant des victoires en fonction des autres équipes pour une équipe. Il suffit de sélectionner une équipe sur le filtre pour avoir les données de cette équipe

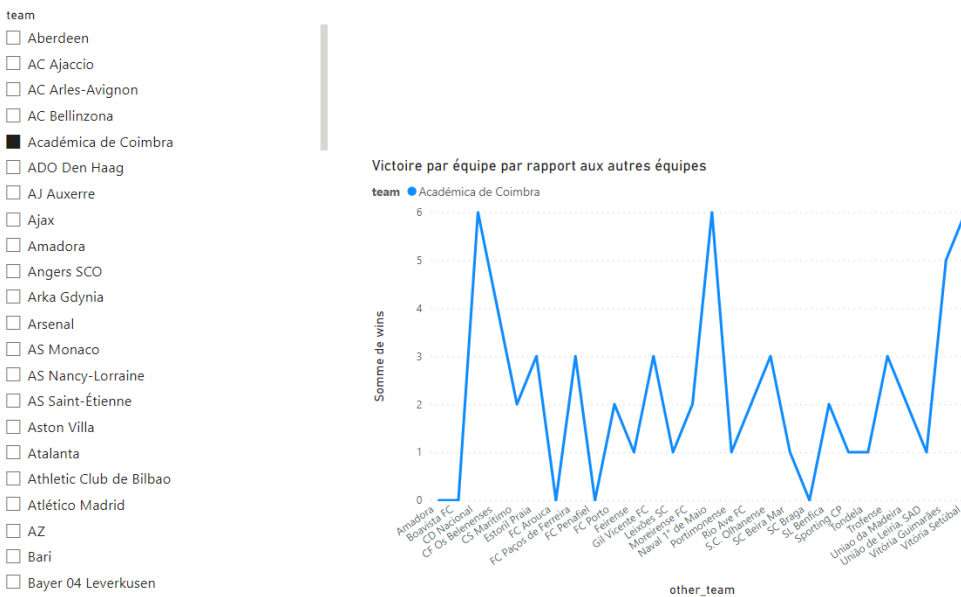


Figure : Courbe d'évolution du nombre de victoires par équipes par rapport aux autres équipes

D. Buts marqués par équipes

1. Par saison

Pour déterminer le nombre de buts marqués par équipe par saison, on a récupéré, le nom de l'équipe, la saison et les goals pour away_team et home_team. On a utilisé les 2 équipes car une même équipe peut être parfois home_team et parfois away_team. Il fallait donc additionner les 2 goals.

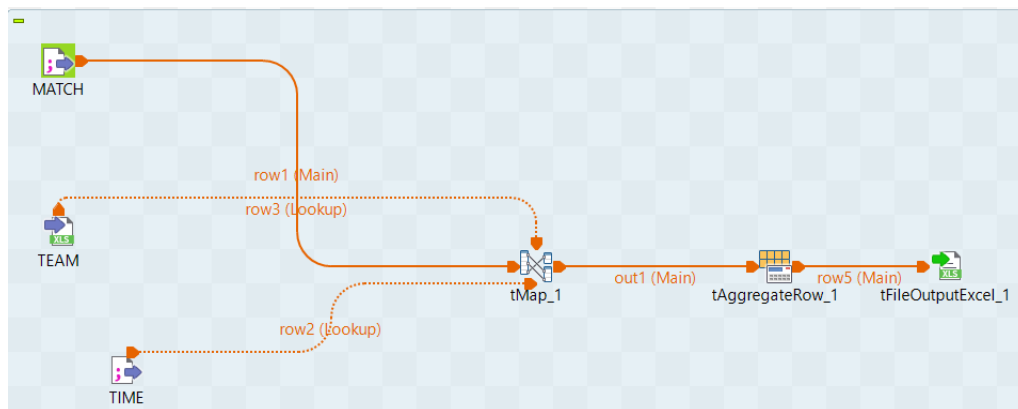


Figure : Création d'un fichier pour obtenir tous les goals par saison par équipes

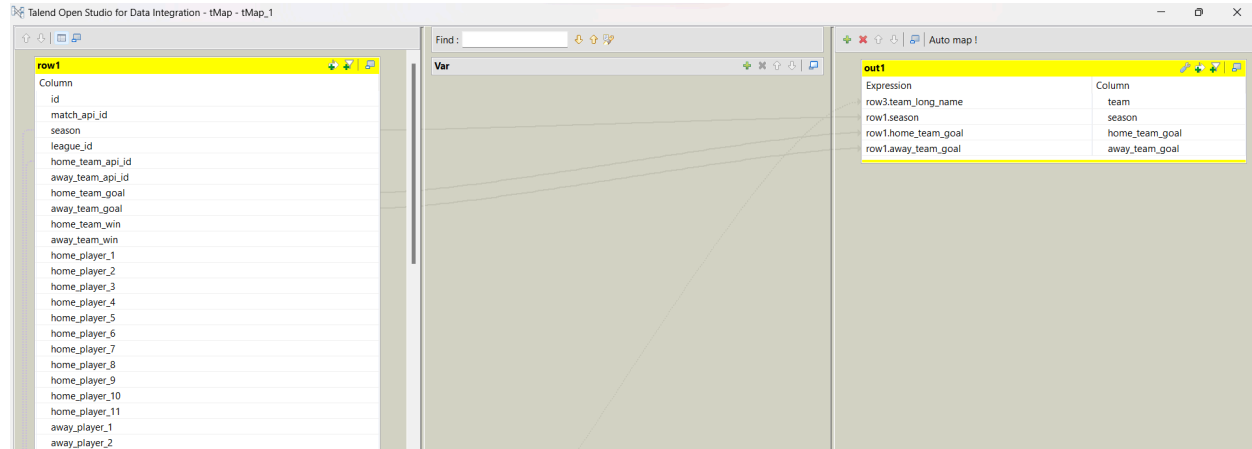


Figure : Récupération des buts des équipes et de leur nom

On a ensuite fait une agrégation par saison et équipes avec pour fonction d'agrégation la somme des 2 goals.

Sur le fichier Excel obtenu, on a sommé home_team_goal et away_team_goal pour avoir le nombre total de goals.

	A	B	C	D	E
1	team	season	home_team_goal	away_team_goal	goals
2	Aberdeen	2008/2009	22	17	39
3	Aberdeen	2009/2010	20	31	51
4	Aberdeen	2015/2016	30	19	49
5	Aberdeen	2012/2013	18	15	33
6	Aberdeen	2011/2012	22	16	38
7	Aberdeen	2013/2014	20	13	33
8	Aberdeen	2010/2011	22	21	43
9	Aberdeen	2014/2015	32	15	47
10	AC Ajaccio	2011/2012	22	24	46
11	AC Ajaccio	2013/2014	23	33	56
12	AC Ajaccio	2012/2013	19	20	39
13	AC Arles-Avignon	2010/2011	14	31	45
14	AC Bellinzona	2010/2011	21	34	55
15	AC Bellinzona	2009/2010	24	40	64
16	AC Bellinzona	2008/2009	29	23	52
17	Académica de Coimbra	2008/2009	18	13	31
18	Académica de Coimbra	2011/2012	11	12	23
19	Académica de Coimbra	2009/2010	20	20	40
20	Académica de Coimbra	2012/2013	22	27	49
21	Académica de Coimbra	2015/2016	20	28	48
22	Académica de Coimbra	2013/2014	13	16	29
23	Académica de Coimbra	2014/2015	16	23	39
24	Académica de Coimbra	2010/2011	16	17	33
25	ADO Den Haag	2011/2012	22	26	48
26	ADO Den Haag	2010/2011	26	25	51

Figure : Fichier résultat

On a donc construit le graphe suivant du nombre de buts en fonction des saisons pour une équipe. Il suffit de sélectionner une équipe sur le filtre pour avoir les données de cette équipe.

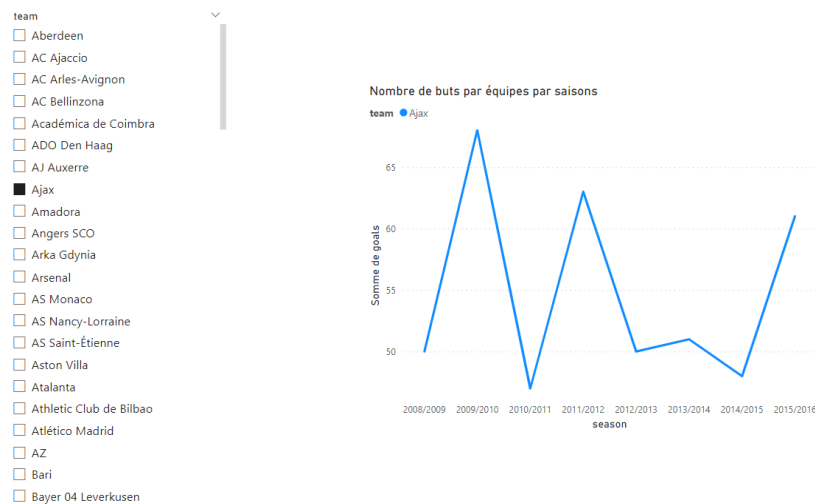


Figure : Courbe d'évolution des buts par saison et équipes

2. Par ligue

Pour déterminer les goals par ligue, on a récupéré, le nom de l'équipe, la ligue et les goals pour away_team et home_team.

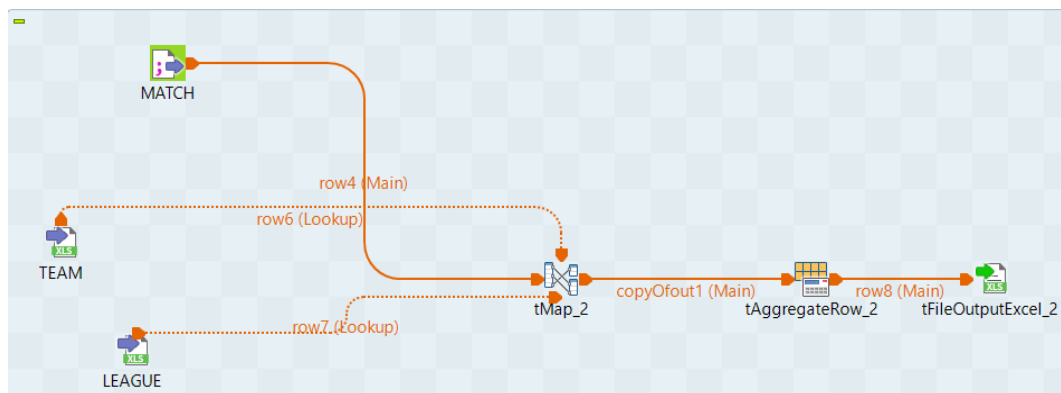


Figure : Création d'un fichier pour obtenir tous les buts par ligues par équipes

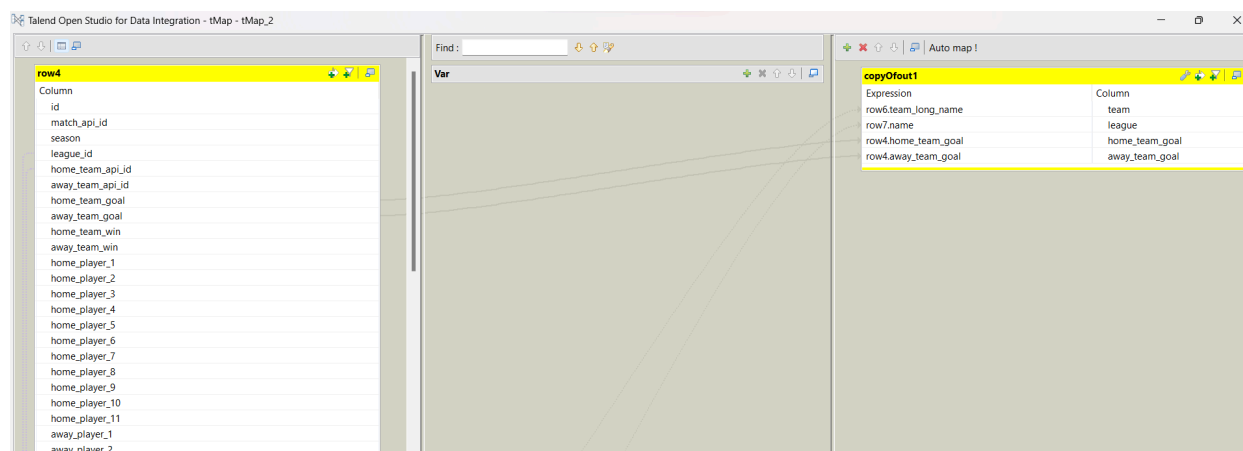


Figure : Récupération des goals des équipes, de leur ligue et de leur nom

On a ensuite fait une agrégation par saison et équipes avec pour fonction d'agrégation la somme des 2 goals.

Sur le fichier Excel obtenu, on a sommé home_team_goal et away_team_goal pour avoir le nombre total de goals.

	A	B	C	D	E
1	team	league	home_team_goal	away_team_goal	goals
2	Aberdeen	Scotland Premier League	186	147	333
3	AC Ajaccio	France Ligue 1	64	77	141
4	AC Arles-Avignon	France Ligue 1	14	31	45
5	AC Bellinzona	Switzerland Super League	74	97	171
6	Académica de Coimbra	Portugal Liga ZON Sagres	136	156	292
7	ADO Den Haag	Netherlands Eredivisie	204	200	404
8	AJ Auxerre	France Ligue 1	98	72	170
9	Ajax	Netherlands Eredivisie	360	78	438
10	Amadora	Portugal Liga ZON Sagres	16	15	31
11	Angers SCO	France Ligue 1	20	15	35
12	Arka Gdynia	Poland Ekstraklasa	52	53	105
13	Arsenal	England Premier League	306	122	428
14	AS Monaco	France Ligue 1	153	97	250
15	AS Nancy-Lorraine	France Ligue 1	103	116	219
16	AS Saint-Étienne	France Ligue 1	223	138	361
17	Aston Villa	England Premier League	179	198	377
18	Atalanta	Italy Serie A	170	151	321
19	Athletic Club de Bilbao	Spain LIGA BBVA	250	176	426
20	Atlético Madrid	Spain LIGA BBVA	321	124	445
21	AZ	Netherlands Eredivisie	279	145	424
22	Bari	Italy Serie A	42	43	85
23	Bayer 04 Leverkusen	Germany 1. Bundesliga	270	153	423
24	Beerschot AC	Belgium Jupiler League	112	98	210
25	Birmingham City	England Premier League	38	35	73
26	Blackburn Rovers	England Premier League	98	90	188

Figure : Fichier résultat

On a donc construit 2 graphes:

- celui des buts en fonction des équipes pour une ligue. Il suffit de sélectionner une ligue sur le filtre pour avoir les données de cette ligue. Nous voulions initialement uniquement tracer l'évolution des buts en fonction des ligues mais chaque équipe n'était associée

qu'à une seule ligue. Le graphe ne contenait juste qu'un point, ce qui ne semblait pas pertinent. On a donc aussi tracé ce graphe.

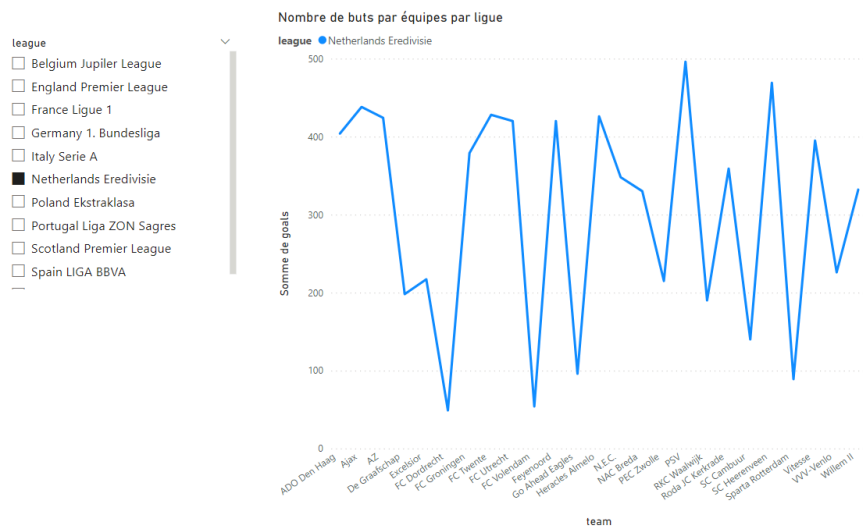


Figure : Courbe d'évolution des goals par équipe sur une ligue

- celui des buts par équipes par ligues. Le filtre permet de filtrer par équipe. Chaque graphe filtré contient donc un seul point.

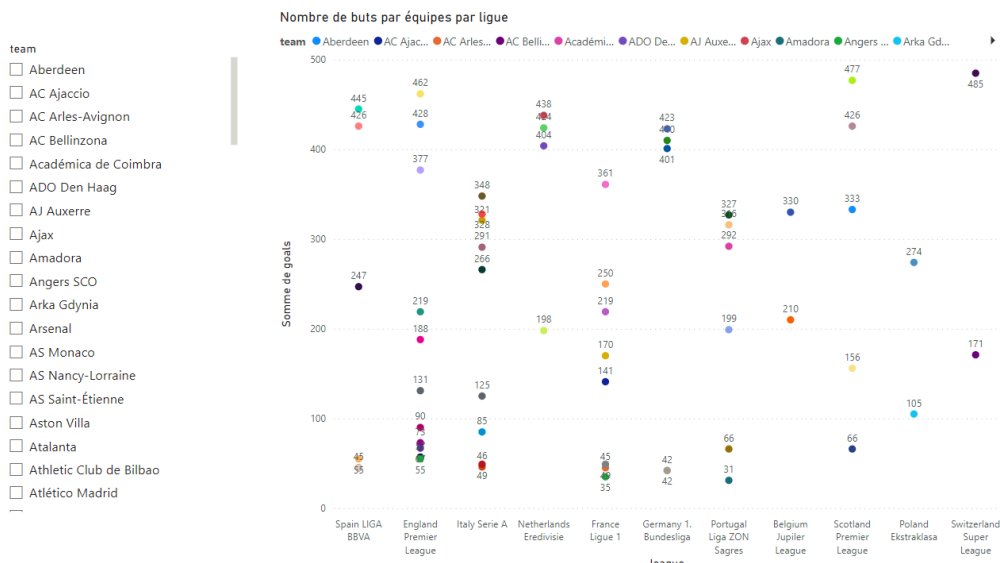


Figure : Nuage de points des goals par ligues et par équipes

3. Par équipes

Pour déterminer les goals par équipes, on a créé 2 output. Dans le 1er, on a le nom de l'équipe home, la ligue et les goals pour away_team et home_team. Dans le 2ème, on a le nom de l'équipe away, la ligue et les goals pour away_team et home_team. Pour le 2ème, on a aussi échangé le nom des variables entre home_team et away_team pour avoir le même nom de colonnes.

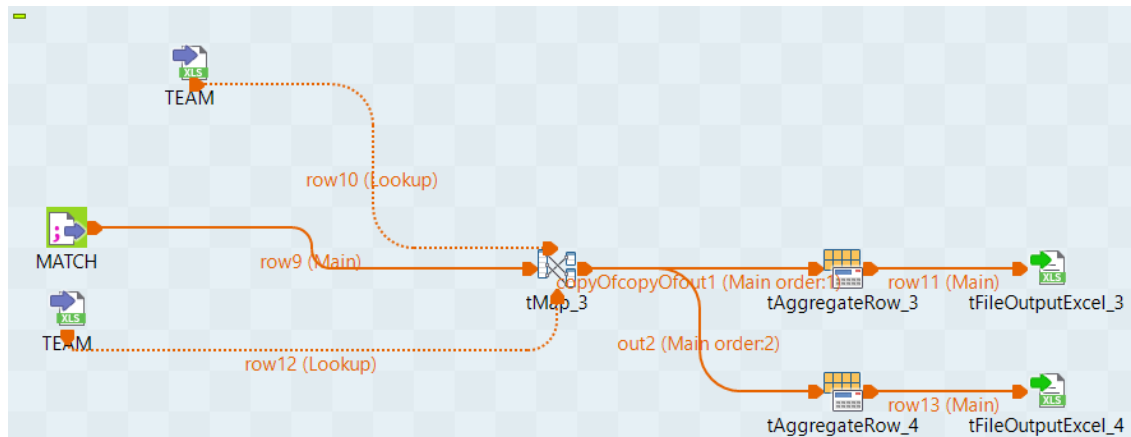


Figure : Création d'un fichier pour obtenir tous les goals par équipes par rapport aux autres équipes

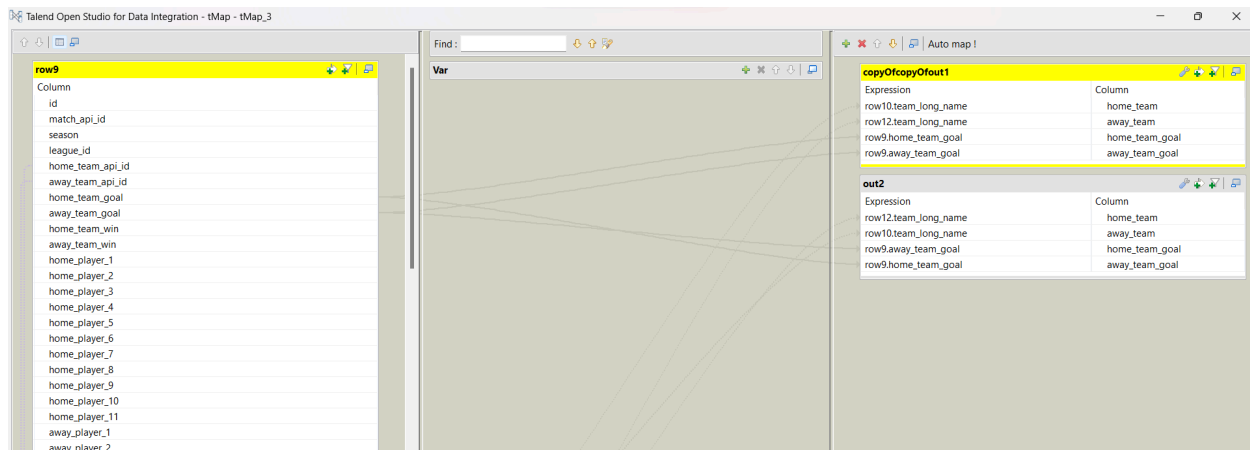


Figure : Récupération des goals et des noms des 2 équipes

On a donc obtenu 2 fichiers Excel pour chaque output du tmap après regroupement par équipes. Les fichiers ont été fusionnés en gardant seulement les colonnes home_team_goal. Puis les 2 colonnes restantes ont été additionnées pour avoir le nombre de goals total.

	A	B	C	D
1	home_team	away_team	home_team_goal	away_team_goal
2	Liverpool	Chelsea	12	10
3	Wolverhampton Wanderers	Portsmouth	0	1
4	FC Zürich	FC Basel	19	35
5	FC Köln	Karlsruher SC	0	0
6	Stade Rennais FC	OGC Nice	11	11
7	Manchester City	Swansea City	12	2
8	Sunderland	Leicester City	0	2
9	Borussia Dortmund	FC Ingolstadt 04	2	0
10	Sparta Rotterdam	De Graafschap	0	0
11	Dundee FC	Inverness Caledonian Thistle	5	10
12	Legia Warszawa	Piast Gliwice	16	5
13	Hertha BSC Berlin	VfB Stuttgart	8	6
14	Sporting CP	FC Paços de Ferreira	12	6
15	Hull City	Manchester City	6	9
16	FC Augsburg	Hertha BSC Berlin	4	1
17	Angers SCO	AS Monaco	3	0
18	Tubize	KSV Cercle Brugge	3	0
19	Waasland-Beveren	Royal Excel Mouscron	5	4
20	Real Valladolid	Xerez Club Deportivo	0	0
21	CD Numancia	Getafe CF	2	0
22	Olympique Lyonnais	Girondins de Bordeaux	10	7
23	FC Basel	AC Bellinzona	15	4
24	Sunderland	Swansea City	4	4
25	Stade Rennais FC	Olympique Lyonnais	10	8
26	Borussia Mönchengladbach	VfB Wolfsburg	13	16

Figure : 1er fichier résultat

	A	B	C	D
1	home_team	away_team	home_team_goal	away_team_goal
2	Liverpool	Chelsea	10	8
3	Wolverhampton Wanderers	Portsmouth	1	3
4	FC Zürich	FC Basel	19	38
5	FC Köln	Karlsruher SC	2	0
6	Stade Rennais FC	OGC Nice	7	11
7	Manchester City	Swansea City	8	6
8	Sunderland	Leicester City	2	4
9	Borussia Dortmund	FC Ingolstadt 04	4	0
10	Sparta Rotterdam	De Graafschap	3	3
11	Dundee FC	Inverness Caledonian Thistle	3	10
12	Legia Warszawa	Piast Gliwice	6	7
13	Hertha BSC Berlin	VfB Stuttgart	3	11
14	Sporting CP	FC Paços de Ferreira	10	7
15	Hull City	Manchester City	3	9
16	FC Augsburg	Hertha BSC Berlin	2	3
17	Angers SCO	AS Monaco	0	1
18	Tubize	KSV Cercle Brugge	1	4
19	Waasland-Beveren	Royal Excel Mouscron	1	1
20	Real Valladolid	Xerez Club Deportivo	0	3
21	CD Numancia	Getafe CF	0	1
22	Olympique Lyonnais	Girondins de Bordeaux	14	10
23	FC Basel	AC Bellinzona	13	6
24	Sunderland	Swansea City	7	9
25	Stade Rennais FC	Olympique Lyonnais	7	9
26	Borussia Mönchengladbach	VfB Wolfsburg	5	16

Figure : 2ème fichier résultat

	A	B	C	D	E
1	home_team	away_team	home_team_goal	home_team_goal	total
2	Aberdeen	Dundee FC	9	11	20
3	Aberdeen	Hamilton Academical FC	16	10	26
4	Aberdeen	Dundee United	18	17	35
5	Aberdeen	Kilmarnock	20	17	37
6	Aberdeen	St. Johnstone FC	11	14	25
7	Aberdeen	Heart of Midlothian	7	11	18
8	Aberdeen	Motherwell	15	12	27
9	Aberdeen	Celtic	18	17	35
10	Aberdeen	St. Mirren	16	1	17
11	Aberdeen	Dunfermline Athletic	5	3	8
12	Aberdeen	Partick Thistle F.C.	6	9	15
13	Aberdeen	Ross County FC	11	6	17
14	Aberdeen	Rangers	5	3	8
15	Aberdeen	Falkirk	3	2	5
16	Aberdeen	Inverness Caledonian Thistle	14	9	23
17	Aberdeen	Hibernian	12	0	12
18	AC Ajaccio	ES Troyes AC	0	2	2
19	AC Ajaccio	Dijon FCO	2	1	3
20	AC Ajaccio	Paris Saint-Germain	2	2	4
21	AC Ajaccio	Toulouse FC	4	7	11
22	AC Ajaccio	FC Sochaux-Montbéliard	3	2	5
23	AC Ajaccio	Stade Brestois 29	1	2	3
24	AC Ajaccio	Girondins de Bordeaux	2	3	5
25	AC Ajaccio	Stade Rennais FC	6	2	8

Figure : Fichier final résultat de la fusion des 2 fichiers obtenus

On a donc obtenu le graphe suivant des buts marqués en fonction des autres équipes pour une équipe. Il suffit de sélectionner une équipe sur le filtre pour avoir les données de cette équipe

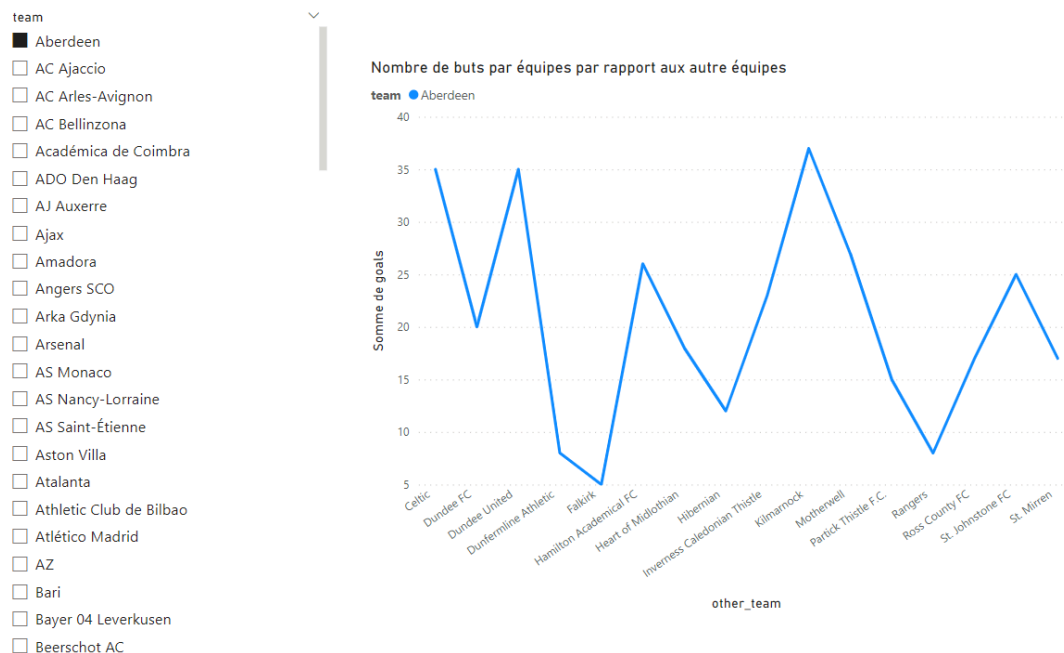


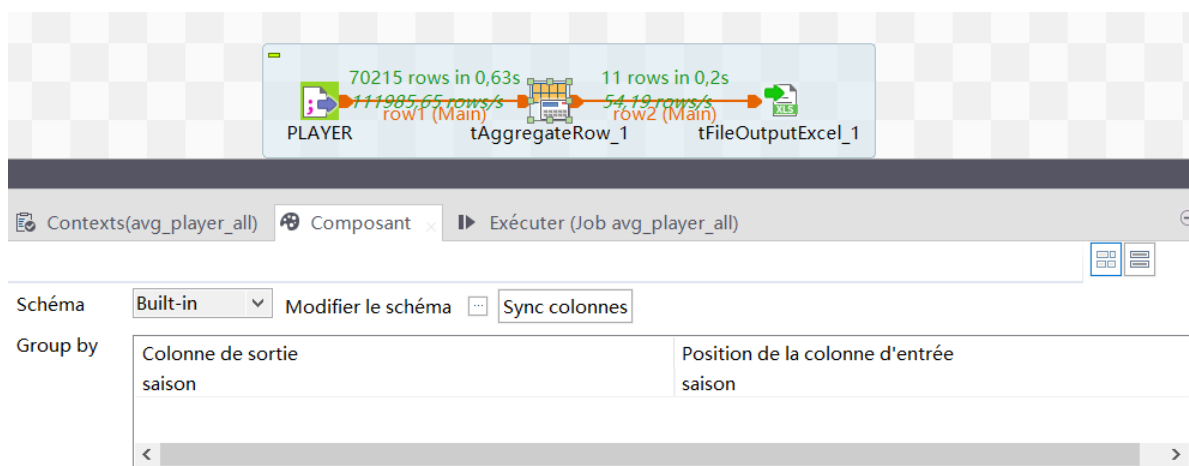
Figure : Courbe d'évolution du nombre de buts par équipes par rapport aux autres équipes

E. Radar des joueurs par saison

1. Par rapport à tous les autres joueurs

Nous utilisons la moyenne des autres joueurs comme la comparaison. En d'autres termes, si les données de performances de X joueurs sont enregistrées pour une saison, et que nous voulons visualiser le "radar" d'un joueur Y, nous souhaitons obtenir la moyenne des données de performances des X - 1 autres joueurs, qui sera affichée conjointement avec les données de Y sur le même radar. Donc nous ne pouvons pas calculer directement la moyenne des données des joueurs pour chaque saison, mais plutôt en soustrayant les données du joueur Y avant de calculer la moyenne.

Nous commençons d'abord par la table dimension PLAYER en effectuant les opérations suivantes : regrouper par saison à l'aide du composant tAggregateRow, calculer le nombre total de joueurs pour chaque saison, ainsi que la somme respective des autres données :



Contexts(avg_player_all) Composant Exécuter (Job avg_player_all)

Schéma Built-in Modifier le schéma Sync colonnes

Group by

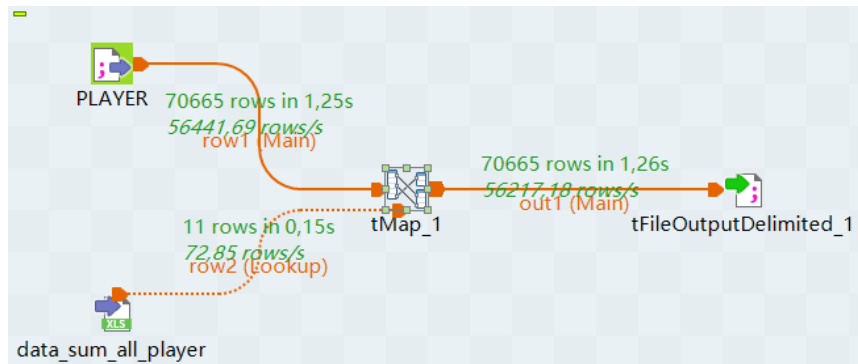
Colonne de sortie	Position de la colonne d'entrée
saison	saison

Opérations

Colonne de sortie	Fonction	Position de la colonne d'en...	<input type="checkbox"/> Ignorer les valeurs nulles
player_api_id	nombre	player_api_id	<input checked="" type="checkbox"/>
overall_rating	somme	overall_rating	<input checked="" type="checkbox"/>
potential	somme	potential	<input checked="" type="checkbox"/>
finishing	somme	finishing	<input checked="" type="checkbox"/>
heading_accuracy	somme	heading_accuracy	<input checked="" type="checkbox"/>
short_passing	somme	short_passing	<input checked="" type="checkbox"/>
dribbling	somme	dribbling	<input checked="" type="checkbox"/>
sprint_speed	somme	sprint_speed	<input checked="" type="checkbox"/>
balance	somme	balance	<input checked="" type="checkbox"/>
shot_power	somme	shot_power	<input checked="" type="checkbox"/>
jumping	somme	jumping	<input checked="" type="checkbox"/>
stamina	somme	stamina	<input checked="" type="checkbox"/>
strength	somme	strength	<input checked="" type="checkbox"/>
positioning	somme	positioning	<input checked="" type="checkbox"/>
penalties	somme	penalties	<input checked="" type="checkbox"/>
sliding_tackle	somme	sliding_tackle	<input checked="" type="checkbox"/>

Figure : Préparation pour le calcul de la moyenne des données

Ensuite, nous connectons la table résultante avec la table de dimension PLAYER par saison. Pour chaque donnée de performance, nous soustrayons les données d'un joueur de l'ensemble des données, puis nous divisons par le nombre total de joueurs moins un :



row1	
Column	
player_api_id	
player_name	
birthday	
height	
weight	
saison	
overall_rating	
potential	
finishing	
heading_accuracy	
short_passing	
dribbling	
sprint_speed	
balance	
shot_power	
jumping	
stamina	
strength	
positioning	
penalties	
sliding_tackle	

row2	
Clé d'expr.	Column
row1.saison	number
	saison
	overall_rating
	potential
	finishing

Figure : Jointure en saison

The screenshot displays a data processing tool interface. On the left, a list of columns for 'row1' is shown, including player attributes and 'overall_rating'. On the right, a list of columns for 'row2' is shown, with 'saison' highlighted. The 'Constructeur d'expression' (Expression Builder) pane is active, showing the formula:
$$\frac{(\text{row2.overall_rating} - \text{row1.overall_rating})}{(\text{row2.number} - 1)}$$
 This formula is used to calculate the average overall rating for each player across different seasons.

Figure : Calcul de la moyenne, exemple de "overall_rating"

out1			
Expression	Column		
row1.player_api_id	player_api_id	row1.sprint_speed	sprint_speed
row1.player_name	player_name	Var.sprint_speed_avg_others	sprint_speed_avg_others
row2.saison	saison	row1.balance	balance
row1.overall_rating	overall_rating	Var.balance_avg_others	balance_avg_others
Var.overall_rating_avg_others	overall_rating_avg_others	row1.shot_power	shot_power
row1.potential	potential	Var.shot_power_avg_others	shot_power_avg_others
Var.potential_avg_others	potential_avg_others	row1.jumping	jumping
row1.finishing	finishing	Var.jumping_avg_others	jumping_avg_others
Var.finishing_avg_others	finishing_avg_others	row1.stamina	stamina
row1.heading_accuracy	heading_accuracy	Var.stamina_avg_others	stamina_avg_others
Var.heading_accuracy_avg_others	heading_accuracy_avg_others	row1.strength	strength
row1.short_passing	short_passing	Var.strength_avg_others	strength_avg_others
Var.short_passing_avg_others	short_passing_avg_others	row1.positioning	positioning
row1.dribbling	dribbling	Var.positioning_avg_others	positioning_avg_others
Var.dribbling_avg_others	dribbling_avg_others	row1.penalties	penalties
row1.sprint_speed	sprint_speed	Var.penalties_avg_others	penalties_avg_others
		row1.sliding_tackle	sliding_tackle
		Var.sliding_tackle	sliding_tackle_1

Figure : output du traitement

Ainsi, nous avons une table enregistrant les données individuelles des joueurs ainsi que les moyennes des données des autres joueurs.

2. Par rapport à leur équipe

Nous avons besoin de savoir quels joueurs sont dans chaque équipe. Les informations sur les joueurs et les équipes sont stockées dans la table de fait MATCH, mais elle stocke à la fois les équipes à domicile (home_team_api_id) et à l'extérieur (away_team_api_id), ainsi que les joueurs à domicile (home_team) et à l'extérieur (away_team). Nous savons qu'une équipe peut apparaître en tant que home_team_api_id ou away_team_api_id dans différents matchs.

Pour intégrer les informations sur les équipes et les joueurs, nous les séparons pour chaque match. Les joueurs de chaque match sont convertis en une chaîne de caractères séparée par des point-virgules :

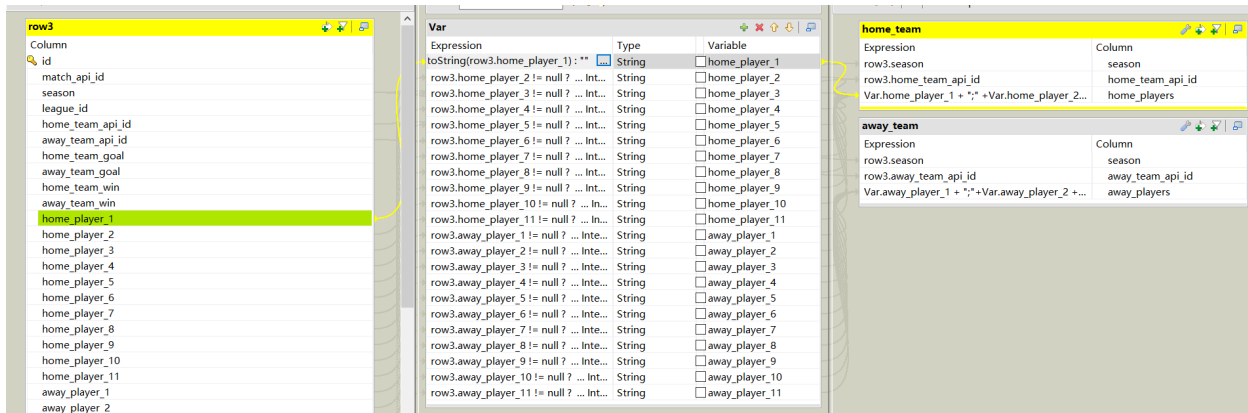


Figure : Séparation des équipes

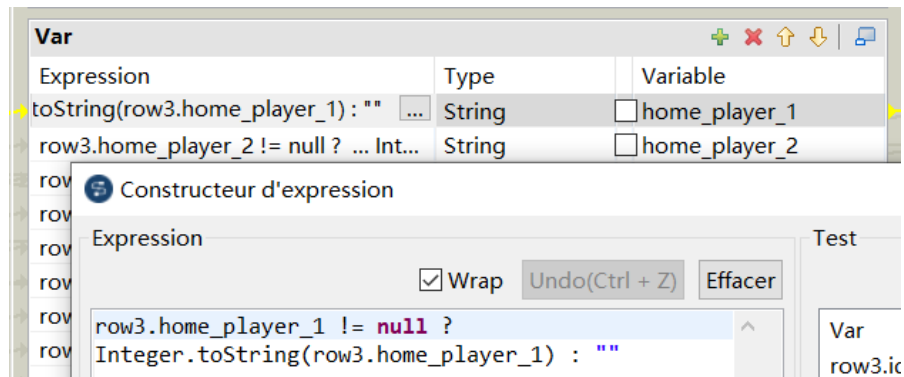


Figure : Convertissement de Integer en String pour les joueurs

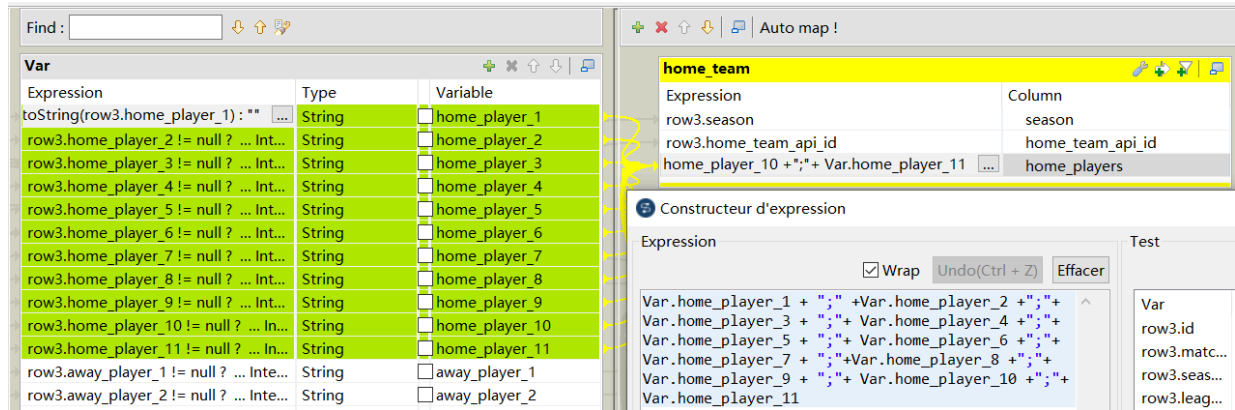


Figure : Concaténation des joueurs

Ensuite, nous normalisons les de sortie, avec le point-virgule comme séparateur d'éléments, afin que chaque identifiant de joueur corresponde à une saison et à un identifiant d'équipe respectifs :

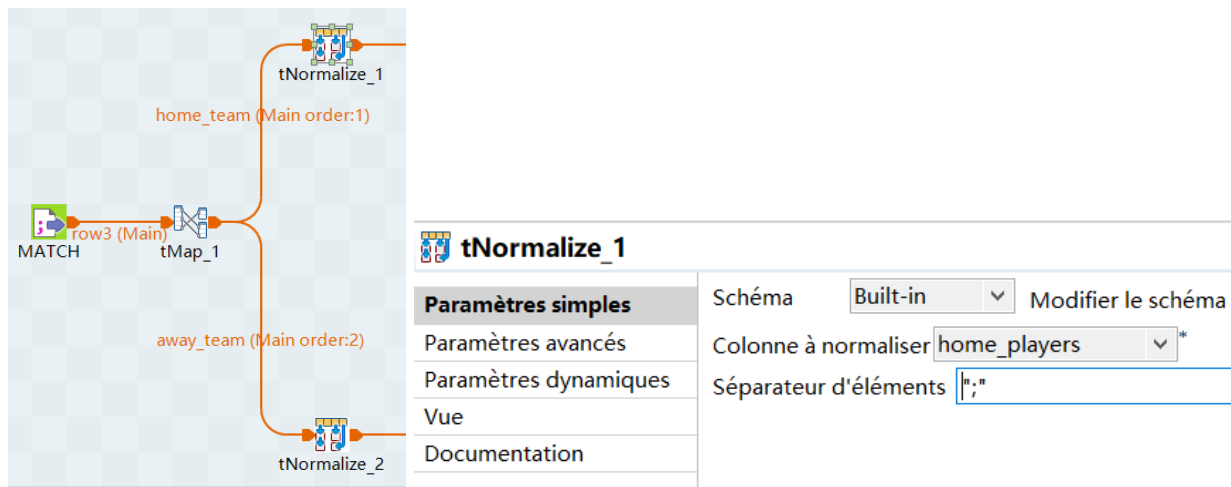


Figure : Normalisation des joueurs avec tNoemalize

En raison du fait que dans la table MATCH, les informations de certains joueurs pour certains matchs n'ont pas été enregistrées, certaines chaînes de caractères de joueurs sont vides. Nous devons filtrer ces lignes vides :

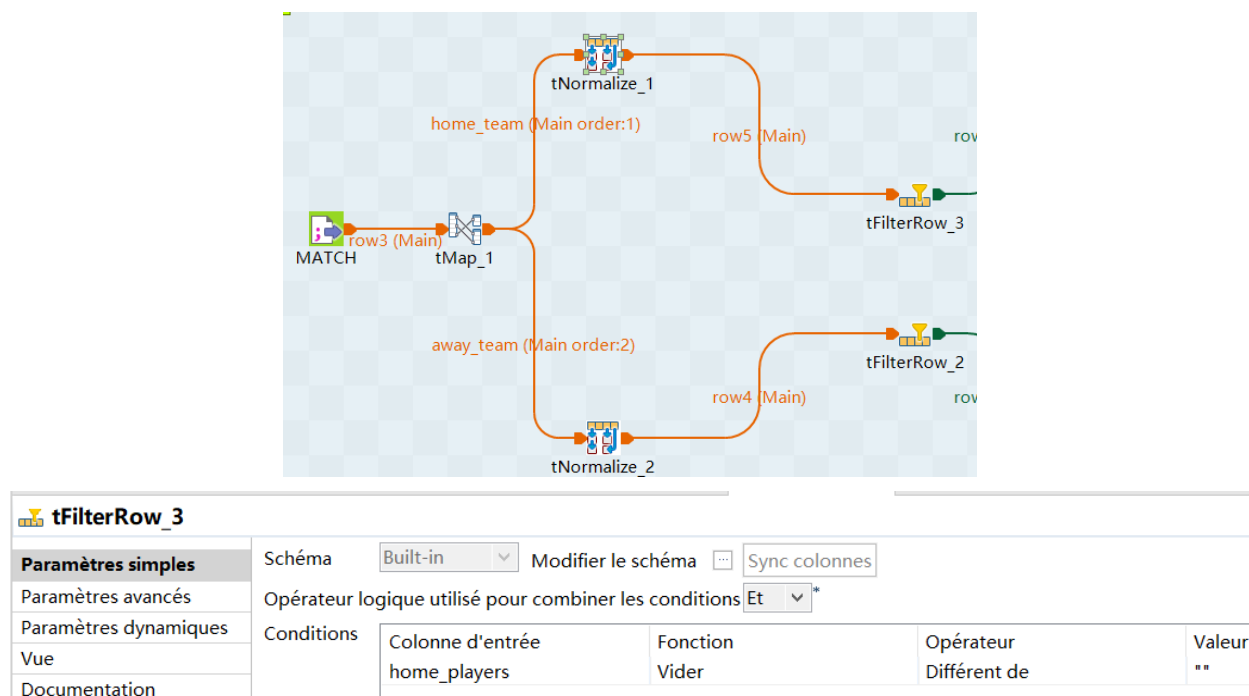
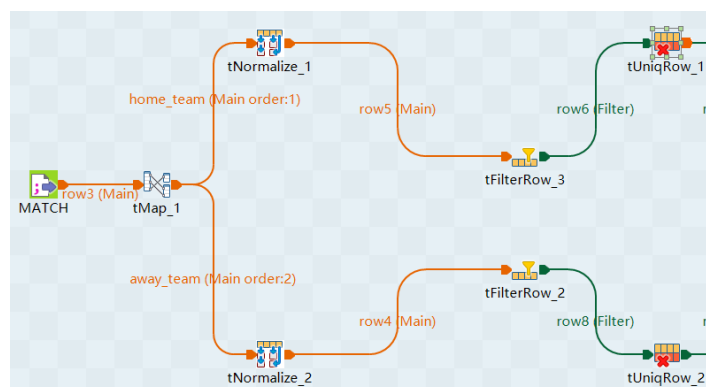


Figure : Implémentation de tFilterRow

Enfin, en supprimant les doublons de saison - ID d'équipe - ID de joueur, nous obtenons les identifiants d'équipe et de joueur des équipes à domicile et à l'extérieur pour chaque saison :

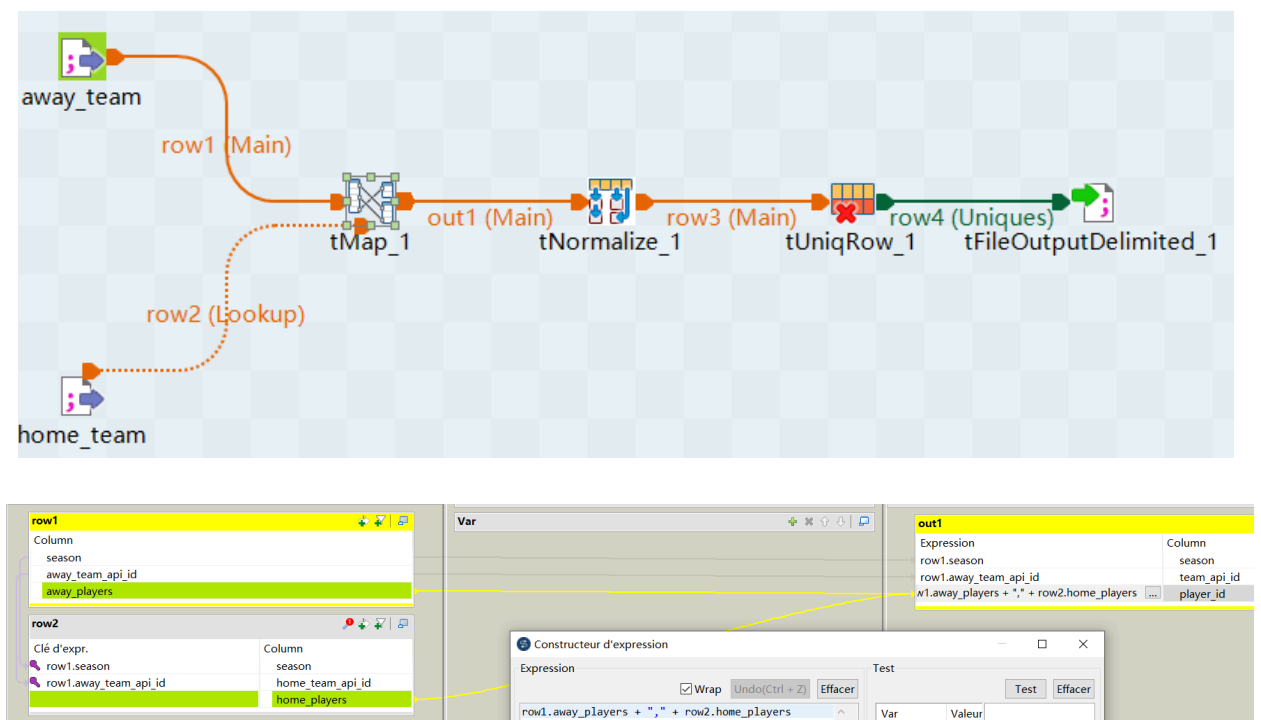


tUniqRow_1

Paramètres simples	Schéma	Built-in	Modifier le schéma	...	Sync colonnes
Paramètres avancés	Clé unique	Column		<input type="checkbox"/> Attribut de clé	<input type="checkbox"/> Sensible à la casse
Paramètres dynamiques		season		<input checked="" type="checkbox"/>	<input type="checkbox"/>
Vue		home_team_api_id		<input checked="" type="checkbox"/>	<input type="checkbox"/>
Documentation		home_players		<input checked="" type="checkbox"/>	<input type="checkbox"/>

Figure : Supprimer les doublons par tUniqRow

Répéter les exécutions de jointure, normalisation et suppression de doublons pour obtenir la table d'informations sur les joueurs de chaque équipe pour chaque saison :



tNormalize_1

Paramètres simples	Schéma	Built-in	Modifier le schéma	...	Sync colonnes
Paramètres avancés	Colonne à normaliser	player_id		*	
Paramètres dynamiques	Séparateur d'éléments	","			
Vue					
Documentation					

tUniqRow_1

Paramètres simples
 Paramètres avancés
 Paramètres dynamiques
 Vue
 Documentation

Schéma Built-in Modifier le schéma Sync colonnes

Clé unique	Column	Attribut de clé	Sensible à la casse
season		<input checked="" type="checkbox"/>	<input type="checkbox"/>
team_api_id		<input checked="" type="checkbox"/>	<input type="checkbox"/>
player_id		<input checked="" type="checkbox"/>	<input type="checkbox"/>

Figure : Opérations pour obtenir la table de joueurs de chaque équipe pour chaque saison

```
season;team_api_id;player_id
2008/2009;8635;34480
2008/2009;8635;106013
2008/2009;8635;38388
2008/2009;8635;26458
2008/2009;8635;13423
2008/2009;8635;38389
2008/2009;8635;38798
2008/2009;8635;30949
2008/2009;8635;38253
2008/2009;8635;38383
2008/2009;8635;46552
```

Figure : Table de sortie (partie)

Récupérer les données de performances par la jointure avec la table de fait PLAYER :

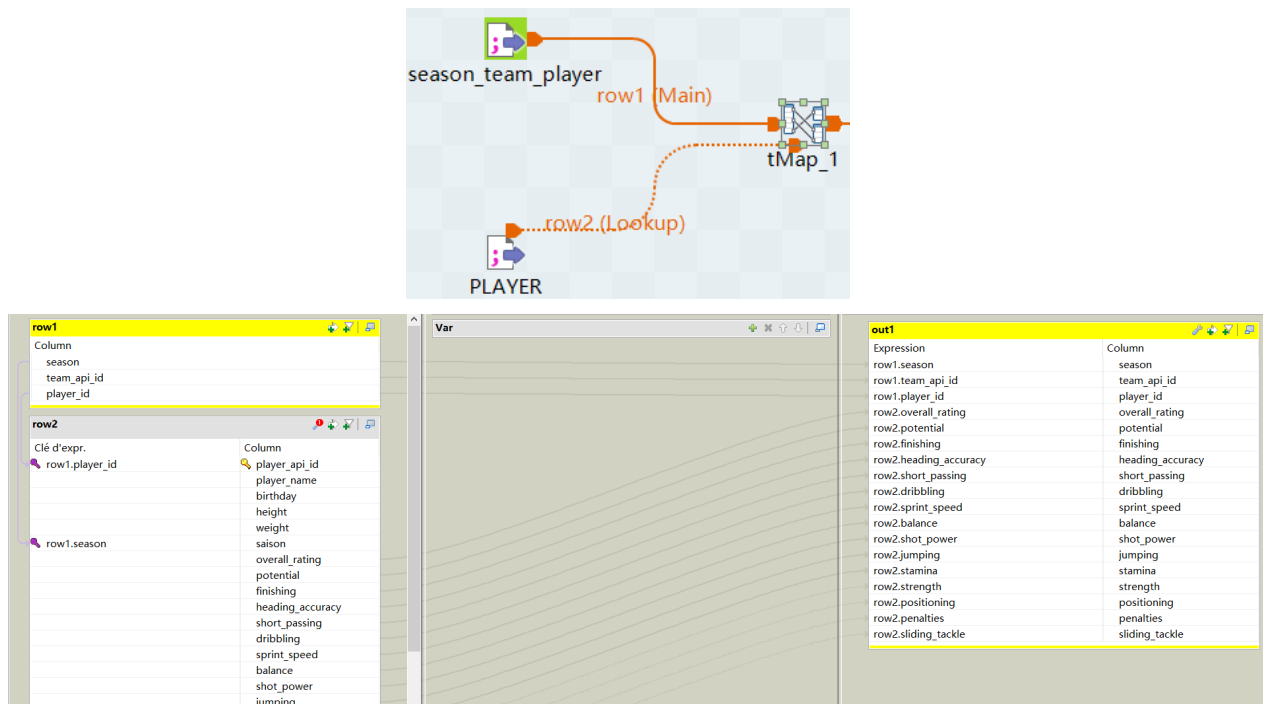
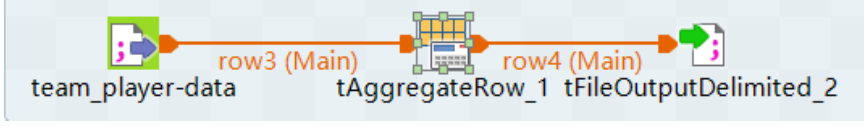


Figure : Jointure avec PLAYER

Ensuite, effectuer les opérations de la même manière que lors du calcul précédent de toutes les données des joueurs :



tAggregateRow_1

Schéma: Built-in | Modifier le schéma | Sync colonnes

Paramètres simples

Paramètres avancés

Paramètres dynamiques

Vue

Documentation

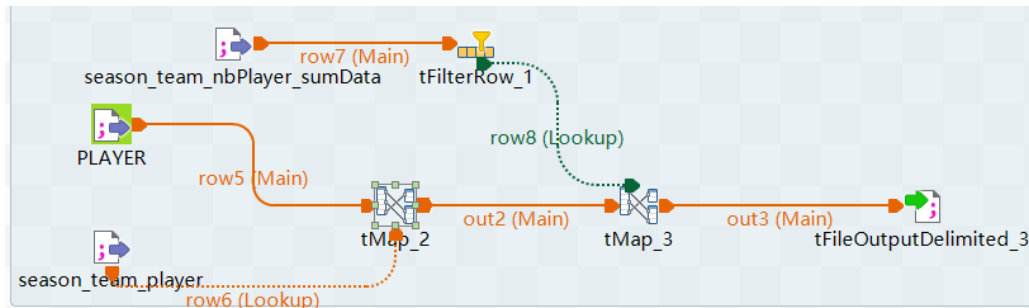
Group by

Colonne de sortie	Position de la colonne d'entrée
season	season
team_api_id	team_api_id

Opérations

Colonne de sortie	Fonction	Position de la colonne d'e...	<input type="checkbox"/> Ignorer les valeurs nulles
overall_rating	somme	overall_rating	<input checked="" type="checkbox"/>
potential	somme	potential	<input checked="" type="checkbox"/>
finishing	somme	finishing	<input checked="" type="checkbox"/>
heading_accuracy	somme	heading_accuracy	<input checked="" type="checkbox"/>
short_passing	somme	short_passing	<input checked="" type="checkbox"/>
dribbling	somme	dribbling	<input checked="" type="checkbox"/>
sprint_speed	somme	sprint_speed	<input checked="" type="checkbox"/>
balance	somme	balance	<input checked="" type="checkbox"/>
shot_power	somme	shot_power	<input checked="" type="checkbox"/>
jumping	somme	jumping	<input checked="" type="checkbox"/>
stamina	somme	stamina	<input checked="" type="checkbox"/>
strength	somme	strength	<input checked="" type="checkbox"/>
positioning	somme	positioning	<input checked="" type="checkbox"/>
penalties	somme	penalties	<input checked="" type="checkbox"/>
sliding_tackle	somme	sliding_tackle	<input checked="" type="checkbox"/>
player_id	nombre	player_id	<input checked="" type="checkbox"/>

Figure : Agrégation des données



row5	Var	out2
Column		Column
player_api_id		player_api_id
player_name		player_name
birthday		birthday
height		height
weight		weight
saïson		team_api_id
overall_rating		saïson
potential		overall_rating
finishing		potential
heading_accuracy		finishing
short_passing		heading_accuracy
dribbling		short_passing
sprint_speed		dribbling
balance		sprint_speed
shot_power		balance
jumping		shot_power
stamina		jumping
strength		stamina
positioning		strength
penalties		positioning
sliding_tackle		penalties
		sliding_tackle
row6		
Clé d'expr.	Column	
season	team_api_id	
team_api_id	player_id	
row5.player_api_id		

Figure : Ajout de l'équipe pour un joueur

out2	Var	out3
Column		Column
player_api_id		player_api_id
player_name		player_name
birthday		team_api_id
height		saïson
weight		overall_rating
team_api_id		potential
saïson		potential_team_avg
overall_rating		finishing
potential		heading_accuracy
finishing		heading_accuracy_team_avg
heading_accuracy		short_passing
short_passing		short_passing_team_avg
dribbling		dribbling
sprint_speed		dribbling_team_avg
balance		sprint_speed
shot_power		sprint_speed_team_avg
jumping		balance
stamina		balance_team_avg
strength		shot_power
positioning		shot_power_team_avg
penalties		jumping
sliding_tackle		jumping_team_avg
		stamina
		stamina_team_avg
		strength
		strength_team_avg
		positioning
		positioning_team_avg
		penalties
		penalties_team_avg
		sliding_tackle
		sliding_tackle
row8		
Property	Value	
Lookup Model	Charger une fois	
Match Model	Correspondance unique	
Join Model	Inner Join	
Store temp data	false	
Clé d'expr.	Column	
out2.saison	season	
out2.team_api_id	team_api_id	
	player_nb	

Figure : Table contenant les données de joueurs et les données moyennes de son équipe

Nous fusionnons cette table avec la table des données moyennes de tous les joueurs. Dans cette nouvelle table, nous avons player_api_id, player_name, team_api_id, saison, données de performances des joueurs, données de performances de tous les joueurs, données de performances de l'équipe des joueurs.

En donnant une saison et un nom (ou un player_api_id), nous pouvons tirer une ligne de données et gérer le diagramme radar en Excel. Par exemple, pour le joueur Bruno Pereira en saison 2012/2013 :

player_name	saison	overall_rating	potential	finishing	heading_accuracy	short_passing	dribbling	sprint_speed	balance	shot_power	jumping	stamina	strength	positioning	penalties	sliding_tackle
Bruno Pereira	2012/2013	71.14286	75.28571	60.42857	53.71429	69.57143	74.42857	76.57143	78.71429	66	73.14286	83.14286	62.85714	63	67.28571	70.28571
all_avg	2012/2013	68.10162	72.52689	47.78074	56.20935	61.51774	57.29245	66.63178	63.85452	61.08678	66.62674	66.07583	67.73607	52.72929	53.09422	46.87769
team_avg	2012/2013	74.1078	80.85075	51.50419	61.75926	68.15917	67.50309	74.34911	70.65093	68.43761	69.35212	70.19422	67.35714	61.47994	58.82275	56.67989

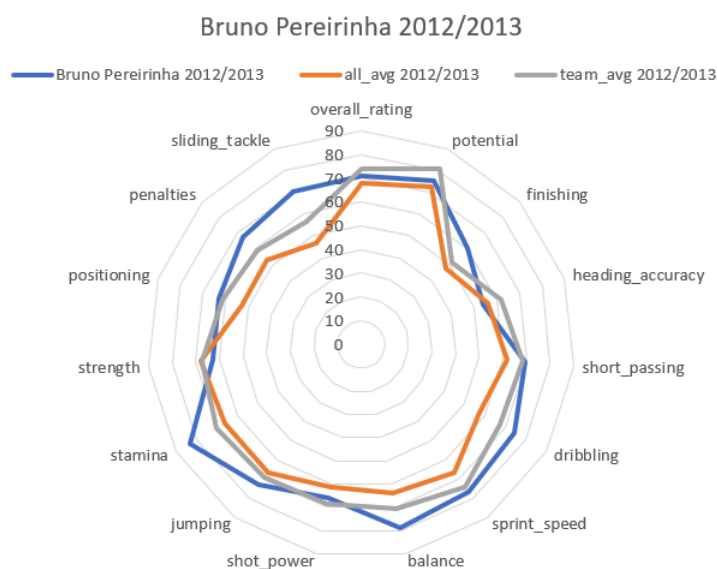


Figure : exemple de radar

V. Les difficultés rencontrées

Une première difficulté rencontrée concerne la création des radars. Nous voulions utiliser Power BI pour afficher les graphiques radar, mais comme nous n'avons pas trouvé de tutoriel adapté à la forme de nos données, nous avons opté pour Excel.

Une autre difficulté que nous avons rencontrée concerne les courbes d'évolutions des matchs gagnés/buts par équipes par rapport aux autres équipes. En utilisant une simple agrégation, on avait seulement le total quand l'équipe était en home_team ou le total quand l'équipe était en away_team. Pour résoudre ce problème, on a donc décidé de créer 2 fichiers Excel. Ainsi, on avait les 2 cas de figures et il suffisait de fusionner les 2 fichiers en récupérant les colonnes qui nous intéressaient.