



RAPPORT SAE SID



Jad MAOULIDA, Youssef ALTIT, Justin WALME

Table des matières

1. Présentation du projet.....	2
Contexte :	2
Problématique :	2
Objectifs :	2
2. Modélisation du SID	2
Choix du processus métier à modéliser :	2
Définition de la granularité :	3
Choix des dimensions :	3
Identification des faits (Mesures) :	3
3. Intégration de données (Processus ETL)	3
Description du processus d'extraction :	3
Transformation des données :	3
Gestion des erreurs et des données manquantes :	4
Création de colonnes personnalisées et conditionnelles :	4
Analyse de la qualité des données :	4
Création et utilisation de tables de temps :	4
Création des mesures avec DAX :	4
Description des modèles de données utilisés :	4
4. Restitution.....	5
Reporting :	5
Tableaux de bord :	5
Analyse équipe vs équipe :	5
Page dédiée aux joueurs :	5
Répartition des postes NBA :	5
Étude de la progression des équipes :	6

1. Présentation du projet

Contexte :

Le projet vise à analyser les performances des équipes et des joueurs de la NBA en utilisant des données variées, telles que les statistiques des matchs, les informations individuelles des joueurs, et les données de la draft. L'objectif est de créer un système d'aide à la décision (SID) qui permette de visualiser les tendances et d'identifier les facteurs clés de succès. Ce système sera conçu pour aider les analystes, les entraîneurs, et les fans à mieux comprendre les dynamiques de la NBA, en fournissant des insights basés sur des données concrètes.

Problématique :

L'analyse des performances en NBA est complexe en raison de la multitude de facteurs à prendre en compte. Comment peut-on comparer efficacement les performances des équipes et des joueurs ? Quels sont les éléments qui influencent le plus les résultats des matchs ? Par exemple, est-ce que la taille moyenne d'une équipe ou l'expérience des joueurs a un impact significatif sur leur taux de victoires ? Ces questions nécessitent une modélisation précise des données et une analyse approfondie pour y répondre.

Objectifs :

Le projet a pour but de concevoir un modèle de données robuste qui intègre plusieurs sources d'informations, comme les statistiques des matchs, les données des joueurs, et les mesures physiques de la draft. Ensuite, un processus ETL (Extract, Transform, Load) sera mis en place pour transformer et charger ces données dans Power BI, où des visualisations interactives seront créées. Des mesures DAX (Data Analysis Expressions) seront développées pour permettre des analyses approfondies, comme le calcul du taux de victoire (Win Rate) ou la comparaison des performances entre équipes. Enfin, le système permettra de filtrer les performances des joueurs par équipe, offrant ainsi une analyse personnalisée.

2. Modélisation du SID

Choix du processus métier à modéliser :

Le système se concentre sur trois processus métiers principaux : l'analyse des performances des équipes sur plusieurs saisons, la comparaison directe entre deux équipes (Équipe vs Équipe), et l'étude des statistiques individuelles des joueurs avec la possibilité de filtrer par équipe. Ces processus ont été choisis car ils couvrent les besoins principaux des utilisateurs, qu'il s'agisse d'analyser les tendances générales ou de se concentrer sur des aspects spécifiques comme les confrontations entre équipes ou les performances individuelles.

Définition de la granularité :

La granularité des données est définie au niveau du match individuel et du joueur individuel. Cela permet une analyse fine des performances, en tenant compte des détails de chaque match et des contributions spécifiques de chaque joueur. Par exemple, on pourra analyser combien de points un joueur a marqué lors d'un match précis, ou comment une équipe a performé sur une série de matchs.

Choix des dimensions :

Les dimensions retenues pour le modèle incluent le calendrier (année, mois, jour, saison NBA), les équipes (nom, identifiant, statistiques globales), les joueurs (nom, identifiant, âge, taille, expérience en NBA), et les données de la draft (attributs physiques des joueurs). Ces dimensions permettent de structurer les données de manière à faciliter les analyses temporelles, les comparaisons entre équipes, et l'évaluation des joueurs.

Identification des faits (Mesures) :

Les mesures incluent des indicateurs tels que la moyenne de points par équipe, le nombre total de victoires et défaites, le taux de victoire (Win Rate), et des comparaisons entre équipes (points marqués et encaissés, ratio victoires/défaites). Pour les joueurs, on mesurera l'âge moyen, la taille moyenne, et la répartition des postes sur le terrain. Ces mesures permettront de répondre à des questions comme : "Quelle équipe a le meilleur taux de victoire sur les 5 dernières saisons ?" ou "Quel est l'âge moyen des joueurs de cette équipe ?"

3. Intégration de données (Processus ETL)

Description du processus d'extraction :

Les données sont extraites à partir de fichiers CSV contenant les statistiques des équipes, les informations des joueurs, et les mesures physiques de la draft combine. Ces fichiers sont ensuite intégrés dans Power BI pour être traités. L'extraction est une étape cruciale car elle détermine la qualité et la quantité des données disponibles pour l'analyse.

Transformation des données :

Une fois les données extraites, elles sont transformées pour être utilisables. Cela inclut la suppression des colonnes non pertinentes, la conversion des types de données (par exemple, convertir des chaînes de caractères en nombres), et la fusion ou la séparation de colonnes pour normaliser les données. Par exemple, la taille des joueurs, souvent donnée en pieds et pouces, est convertie en centimètres pour une analyse plus standardisée.

Gestion des erreurs et des données manquantes :

Les données manquantes ou erronées sont gérées soit par imputation (remplacement par une valeur moyenne ou médiane), soit par suppression si elles ne sont pas pertinentes. Les doublons sont également supprimés pour éviter les biais dans l'analyse. Par exemple, si un joueur apparaît deux fois dans les données avec des statistiques différentes, une vérification est effectuée pour conserver la version la plus récente ou la plus fiable.

Création de colonnes personnalisées et conditionnelles :

Des colonnes personnalisées sont créées pour faciliter l'analyse. Par exemple, une colonne "Expérience en NBA" est ajoutée en fonction du nombre d'années de carrière d'un joueur. De même, des indicateurs de performance sont créés pour classer les équipes en fonction de leur taux de victoire ou de leur nombre de points marqués.

Analyse de la qualité des données :

La qualité des données est vérifiée en analysant leur distribution et en identifiant les valeurs aberrantes. Par exemple, si un joueur a une taille de 3 mètres, cela est probablement une erreur qui doit être corrigée. Cette étape est essentielle pour garantir la fiabilité des analyses.

Création et utilisation de tables de temps :

Une table calendrier est générée en utilisant la fonction `CALENDAR()` en DAX. Cette table permet de faciliter l'analyse temporelle, comme l'évolution des performances d'une équipe au fil des saisons ou des mois.

Création des mesures avec DAX :

Des mesures DAX sont créées pour calculer des indicateurs clés, comme le Win Rate (taux de victoire), les points moyens par équipe, ou l'âge moyen des joueurs. Ces mesures permettent de répondre à des questions spécifiques et de créer des visualisations interactives dans Power BI.

Description des modèles de données utilisés :

Le modèle de données est structuré en étoile, avec une table de faits principale (joueurs et équipes) et des tables de dimensions (calendrier, équipes, joueurs, draft stats). Cette structure permet une navigation fluide et une agrégation efficace des mesures, facilitant ainsi l'analyse.

4. Restitution

Reporting :

Le reporting inclut plusieurs analyses clés, comme l'évolution des performances des équipes sur plusieurs saisons, la comparaison des statistiques entre deux équipes, et l'étude des performances individuelles des joueurs avec la possibilité de filtrer par équipe. Par exemple, un rapport pourrait montrer comment une équipe a progressé au fil des années, ou comment deux équipes rivales se comparent en termes de points marqués et encaissés.

Tableaux de bord :

Les tableaux de bord sont conçus pour offrir une vue globale et détaillée des performances. Un dashboard général montre les performances globales des équipes, tandis qu'un autre se concentre sur les comparaisons entre deux équipes spécifiques. Une page dédiée aux joueurs permet d'analyser leurs caractéristiques, comme leur taille ou leur expérience, avec la possibilité de filtrer par équipe. Enfin, une visualisation de la répartition des postes sur le terrain montre les tendances de recrutement et les stratégies des équipes.

Analyse équipe vs équipe :

Cette section permet de comparer deux équipes en détail, en analysant leurs performances passées, leur ratio victoires/défaites, et leurs forces et faiblesses respectives. Par exemple, on pourrait voir que l'équipe A a un meilleur taux de victoire contre l'équipe B, mais que l'équipe B marque plus de points en moyenne.

Page dédiée aux joueurs :

Cette page permet d'explorer les caractéristiques des joueurs, comme leur âge, leur taille, ou leur expérience, avec la possibilité de filtrer par équipe. Par exemple, on pourrait voir que les joueurs de l'équipe X sont en moyenne plus jeunes que ceux de l'équipe Y, ce qui pourrait expliquer des différences de performance.

Répartition des postes NBA :

Cette visualisation montre comment les joueurs sont répartis sur le terrain (meneur, ailier, pivot, etc.) et permet d'identifier des tendances de recrutement. Par exemple, on pourrait voir que certaines équipes privilégient les joueurs de grande taille, tandis que d'autres misent sur la vitesse et l'agilité.

Étude de la progression des équipes :

Enfin, une analyse de la progression des équipes en fonction de leur composition et de leurs évolutions permet de comprendre comment les changements de joueurs ou de stratégies ont impacté leurs performances au fil du temps. Par exemple, on pourrait voir qu'une équipe a amélioré son taux de victoire après avoir recruté de jeunes joueurs talentueux.

Ce projet offre une vision complète et détaillée des performances en NBA, en combinant des analyses globales et spécifiques pour répondre aux besoins variés des utilisateurs.