

Problématique :

Quels sont les indicateurs statistiques clés, au niveau individuel et collectif, qui permettent à une équipe NBA de réussir en playoffs ? Existe-t-il un lien entre les performances, les salaires et la "valeur réelle" des joueurs ? Peut-on identifier des joueurs "surcotés", c'est-à-dire dont les performances ne justifient pas les salaires perçus ?

Cette problématique vise à :

- Identifier les facteurs de performance collective en playoffs : Première fois que Stephen Curry gagne le titre de champion NBA. Est-ce que ses stats reflètent ses performances ?



- Déterminer si payer cher un joueur garantit réellement un bon rendement
- Mettre en lumière les joueurs surévalués ou sous-évalués
- Comprendre l'impact du contexte d'équipe faible ou forte sur les statistiques individuelles

Démarche

1. Collecte et préparation des données

- Fichiers sources collectés depuis Kaggle, GitHub :
 - `players_stats.csv` : stats individuelles
 - `nba-salaries.csv` : salaires des joueurs
 - `NBA_Teams.csv` : stats d'équipe
 - `line_score.csv` : résultats de matchs (par quart-temps)
 - `team_history.csv` : infos historiques des équipes
 - `cities_unique.csv` : données géographiques
- Nettoyage des doublons, normalisation des noms (`Team`, `Name`, `TEAM_ABBV`)
- Filtrage de la saison 2014–2015, ciblant la saison régulière

2. Modélisation des données

a. Table de faits : `game_data`

- Scores par équipe et par match
- Ajout de colonnes `timestamp` pour jointure avec la dimension `dates`

b. Dimensions

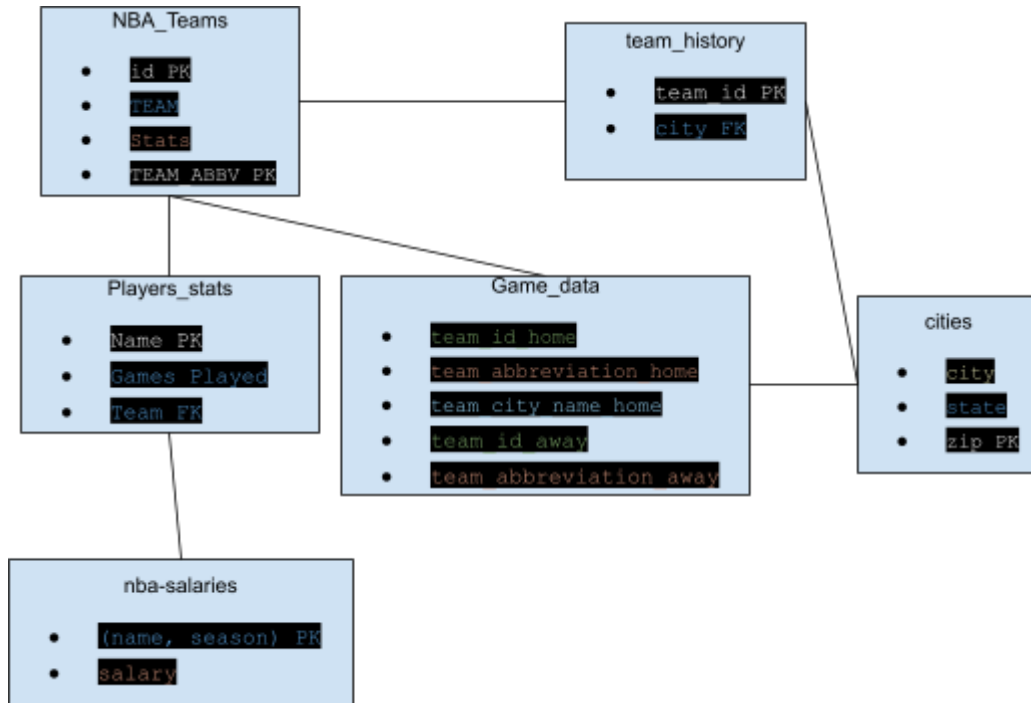
- `dim_players` : joueur, stats de saison, salaire, position, âge
- `dim_teams` : info géographique, historiques, stats agrégées
- `dates` : jour, mois, année pour analyses temporelles

3. Création d'une table dénormalisée (optionnelle pour Superset)

- Fusion des DataFrames avec `merge()` et non `concat()`

- Ajout des informations sur les équipes home/away et joueurs

Modèle relationnel



Dans la table NBA_Teams on a vérifié que TEAM_ABBV est unique pour chaque équipe.
Donc c'est la FK pour players

Modèle dénormalisé en étoile



Les features de team_history et cities sont agrégées dans la table NBA_Teams et le salaire des joueurs NBA est mis dans la table Players_stats renommée Players. On utilise la méthode pandas merge pour faire ces agrégations.

Résultats attendus / Visualisations ciblées

Joueurs

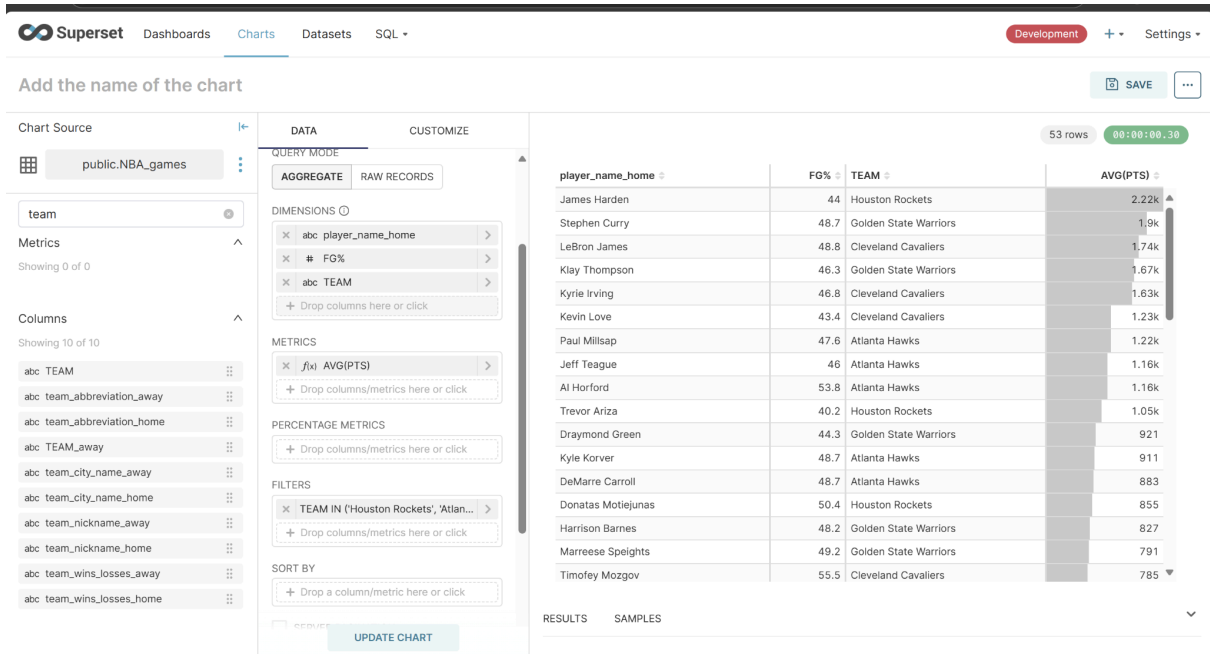
- **Corrélation entre salaire et performance réelle (EFF, PTS, AST, STL...)**
- **Indice de sur/sous-cotation : salaire vs contribution réelle. On voit que le joueur vétérán Kobe Bryant touche plus que les meilleurs joueurs de la ligue. Le MVP (most valuable player n'est même pas dans liste ci-dessous des joueurs les plus payés.**

player_name_home ↕	salary ▼	AVG(PTS) ↕
Kobe Bryant	23500000	782 ▲
Carmelo Anthony	22458401	966
Dwight Howard	21436271	646
Chris Bosh	20644400	928
LeBron James	20644400	1.74k
Chris Paul	20068563	1.56k
Rudy Gay	19317326	1.43k
Kevin Durant	18995624	686
Derrick Rose	18862876	904
Blake Griffin	17674613	1.47k
Zach Randolph	16500000	1.14k
LaMarcus Aldridge	16006000	1.66k
Paul George	15937290	53
Marc Gasol	15829688	1.41k
Kevin Love	15719063	1.23k
David Lee	15012000	388
Roy Hibbert	14898938	802 ▼

Équipes

- **Rôle du collectif vs individualités : Dans la visualisation Superset suivante on a les indicateurs individuelles de PTS marqué au cours de la saison avec le pourcentage de réussite pour les 4 équipes les plus prolifiques de la saison 2014/2015. Ainsi on voit que Golden State Warriors avec Stephen Curry n'ont pas le meilleur marqueur de la ligue mais le 2 ème avec une très bonne**

efficacité.



- En playoff, le match se joue dans le dernier quart-temps, est ce que les meilleurs équipes sont fortes dans ce dernier quart temps en saison régulière ?

team_abbreviation_home	AVG(pts_qtr4_home)
GSW	25.79
ATL	25.42
CLE	23.31
HOU	23.17

- L'équipe la plus clutch du BIG 4 est Golden State. Malheureusement, nous n'avons pas pris de csv pour connaître les stats des joueurs matchs par matchs et nous ne pouvons pas dire de quelle joueur vient cette contribution.

Difficultés rencontrées

Problème	Solution de contournement
Fichiers CSV avec tailles et structures différentes	Construction d'une modélisation en étoile, avec des jointures (merge) sur des clés cohérentes
Données incomplètes pour certains joueurs ou équipes J'ai dû rajouter Boston celtics (17 titres NBA) à un des fichiers	Utilisation de jointures left pour garder toutes les données principales même si partiellement liées
Incohérence dans les noms d'équipes/joueurs AJ Price et dans un autre fichier A.J. Price	Pas de stats joueurs pour chaque match (dommage ça aurait pu être notre table de fait)
Lien difficile entre performances individuelles et résultats d'équipe	