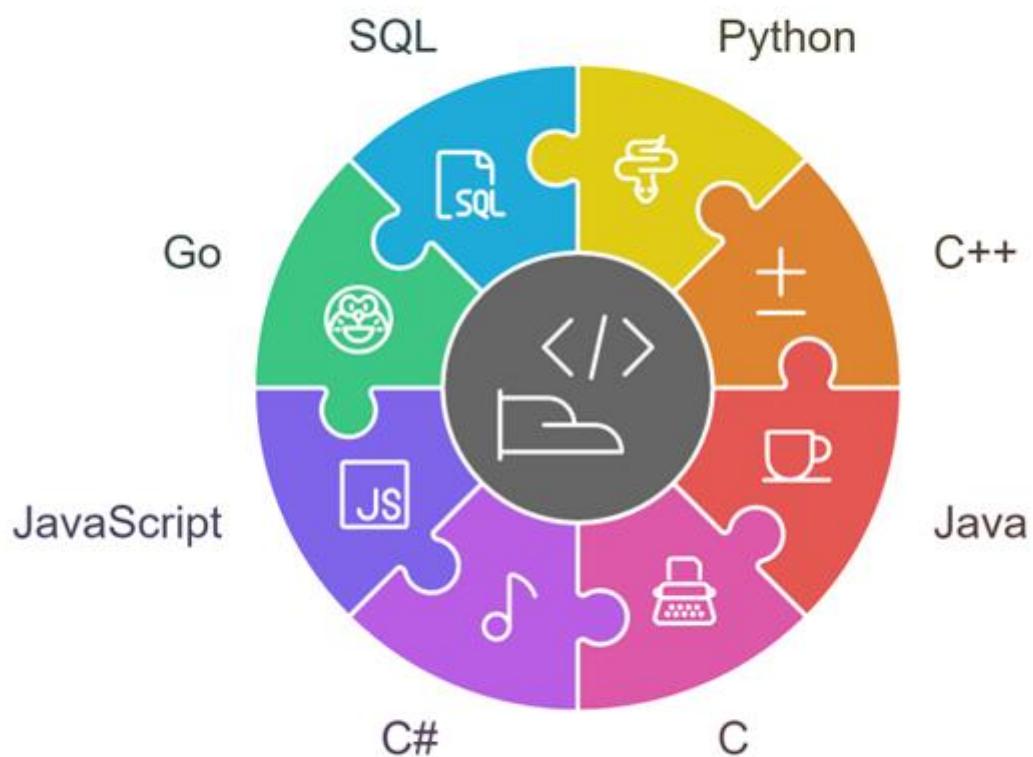


Rapport page web python



Résumé de la base de données

La base de données utilisée regroupe des **statistiques NBA au niveau équipe par match**. Elle contient **31 340 observations et 32 variables**, chaque ligne correspondant à la performance d'une équipe lors d'un match donné (un match apparaît donc deux fois, une fois par équipe).

Les variables couvrent plusieurs dimensions :

- **Identification du match et de l'équipe** (ID, nom, ville, saison),
- **Contexte du match** (équipe à domicile, minutes jouées, période COVID),
- **Statistiques offensives** (tirs à 2 points, tirs à 3 points, lancers francs et leurs pourcentages),
- **Jeu collectif et rebonds** (passes décisives, rebonds offensifs, défensifs et totaux),
- **Aspects défensifs et discipline** (interceptions, contres, balles perdues, fautes),
- **Indicateurs de performance avancés** (points, plus/minus, EFG%, PIE, pourcentage de victoires),
- Ainsi que le **résultat du match** (victoire ou défaite).

Le jeu de données est **complet, sans valeurs manquantes**, et présente des variables quantitatives et qualitatives bien structurées. Il est donc particulièrement adapté à des **analyses statistiques avancées**, telles que l'analyse exploratoire, l'ACP, le clustering ou la modélisation prédictive des performances et des victoires.

Aperçu de la base de données

	TEAM_ID	TEAM_NAME	TEAM_ABBREVIATION	TEAM_CITY	HOME_TEAM	MIN	FGM	FGA	FG_PCT	FG3M	FG3A
0	1610612764	Wizards	WAS	Washington	Cavaliers	240:00	32	90	0.356	8	32
1	1610612739	Cavaliers	CLE	Cleveland	Cavaliers	240:00	36	79	0.456	7	20
2	1610612748	Heat	MIA	Miami	Heat	240:00	43	79	0.544	8	16
3	1610612738	Celtics	BOS	Boston	Heat	240:00	39	75	0.52	6	13
4	1610612747	Lakers	LAL	Los Angeles	Lakers	240:00	38	77	0.494	3	13
5	1610612742	Mavericks	DAL	Dallas	Lakers	240:00	40	85	0.471	5	15
6	1610612761	Raptors	TOR	Toronto	Raptors	240:00	33	91	0.363	6	17
7	1610612754	Pacers	IND	Indiana	Raptors	240:00	37	78	0.474	5	13
8	1610612743	Nuggets	DEN	Denver	76ers	240:00	33	88	0.375	4	18
9	1610612755	76ers	PHI	Philadelphia	76ers	240:00	30	85	0.353	7	25

Description du script app.py – Application d’analyse NBA

Description détaillée du script *app.py*

Le script *app.py* implémente une application web interactive développée en Python à l'aide du framework Streamlit. Cette application prend la forme d'un dashboard NBA multi-pages, permettant l'exploration, l'analyse statistique avancée et la visualisation interactive des performances des équipes NBA sur plusieurs saisons.

L'objectif principal de ce dashboard est de fournir un outil d'aide à l'analyse décisionnelle, combinant statistiques descriptives, indicateurs avancés, méthodes multivariées, classification et modélisation prédictive, tout en restant accessible à un utilisateur non spécialiste en programmation.

1. Environnement et bibliothèques utilisées

Le script mobilise plusieurs bibliothèques majeures de l'écosystème Python :

- Streamlit pour la création de l'interface web interactive et la gestion de la navigation multi-pages.
- Pandas et NumPy pour la manipulation, le nettoyage et la transformation des données.
- Plotly (Express et Graph Objects) pour les visualisations interactives dynamiques.
- Scikit-learn pour les méthodes d'analyse avancée :
 - Standardisation des données,
 - Analyse en composantes principales (ACP),
 - Clustering par K-means,
 - Régression logistique pour la prédiction des résultats.
- Matplotlib et Seaborn pour certaines visualisations statistiques spécifiques (matrice de confusion).

L'ensemble de ces bibliothèques permet de couvrir un large spectre d'analyses, allant de l'exploration simple à la modélisation statistique.

2. Configuration générale et mise en forme

Le script commence par la configuration globale de l'application via `st.set_page_config`, définissant :

- Le titre de l'application,
- Une mise en page large adaptée aux tableaux et graphiques.

Un style CSS personnalisé est ensuite appliqué afin d'améliorer l'ergonomie et la lisibilité de l'interface. Cette personnalisation renforce l'aspect professionnel du dashboard en uniformisant les couleurs, les titres et les indicateurs clés.

3. Chargement et préparation des données

Les données sont chargées à partir d'un fichier final_data.csv, avec un mécanisme robuste permettant de tester plusieurs chemins possibles. Le chargement est mis en cache afin d'optimiser les performances de l'application.

Un travail de nettoyage et de préparation est ensuite réalisé :

- Conversion de la variable MIN (initialement au format *minutes: secondes*) en une variable numérique continue MIN_CLEAN,
- Suppression des doublons,
- Gestion des valeurs manquantes,
- Identification et extraction des variables numériques exploitables pour les analyses statistiques.

Cette phase garantit la fiabilité des analyses ultérieures.

4. Filtres interactifs et export des données

Une barre latérale permet à l'utilisateur de filtrer dynamiquement les données selon :

- La saison,
- L'équipe,
- Le résultat du match (victoire ou défaite).

Les données complètes ou filtrées peuvent être exportées au format CSV, facilitant leur réutilisation dans un cadre académique ou professionnel.

5. Architecture multi-pages du dashboard

L'application est structurée autour de plusieurs pages thématiques, accessibles via un menu de navigation :

Page Home

Elle fournit une vue d'ensemble du jeu de données :

- Indicateurs clés (nombre de matchs, équipes, points moyens, taux de victoire),
- Classement des équipes selon les points moyens,
- Aperçu tabulaire des données.

Page Profil équipe

Cette page propose une analyse détaillée d'une équipe donnée :

- Statistiques moyennes,

- Évolution match par match de plusieurs indicateurs (points, rebonds, passes).

Page Analyse multi-saisons

Elle permet d'étudier l'évolution temporelle des performances d'une équipe sur plusieurs saisons pour différentes statistiques clés.

Page Analyse avancée

Cette section calcule et exploite des statistiques avancées telles que :

- Le True Shooting Percentage (TS%),
- L'Effective Field Goal Percentage (FG%),
- Le taux de lancers francs.

Ces indicateurs offrent une vision plus fine de l'efficacité offensive des équipes.

Page Clustering équipes

Une méthode de clustering non supervisé (K-means) est appliquée après standardisation des variables et réduction de dimension par ACP. Cette approche permet d'identifier des profils d'équipes aux caractéristiques similaires et de les visualiser dans un plan factoriel.

Page Prévision simple

Une régression logistique est utilisée pour prédire le résultat d'un match (victoire/défaite) à partir de statistiques sélectionnées par l'utilisateur. Les performances du modèle sont évaluées à l'aide d'une matrice de confusion et de l'interprétation des coefficients.

Page Exploration libre

Cette page offre une liberté totale de visualisation en permettant à l'utilisateur de choisir les variables à représenter sur les axes, la coloration et la taille des points.

Page Comparateur équipes

Deux équipes peuvent être comparées à l'aide :

- De tableaux récapitulatifs,
- D'un graphique radar normalisé mettant en évidence leurs forces et faiblesses relatives.

Page Stats & Insights

Cette section génère automatiquement :

- Une matrice de corrélation,
- Des classements automatiques,
- Un résumé analytique prêt à être intégré dans un rapport.

Page Analyse complète par équipe

Cette page combine visualisations et interprétations automatiques des graphiques. L'utilisateur peut cliquer sur un graphique pour afficher une analyse textuelle générée dynamiquement et l'exporter au format texte.

Page Conclusion générale

Elle synthétise l'ensemble des résultats obtenus et met en évidence les principaux enseignements statistiques et analytiques de l'étude.

6. Génération automatique d'interprétations

Le script intègre une fonction dédiée à l'interprétation automatique des graphiques, qui calcule des indicateurs statistiques (moyenne, dispersion, corrélation, tendance) et les traduit en texte analytique clair. Cette fonctionnalité renforce l'aspect pédagogique et directement exploitable du dashboard.

Conclusion

Cette étude s'est appuyée sur un jeu de données NBA comprenant plus de **31 000 observations**, décrivant les performances des équipes match par match sur plusieurs saisons. La richesse de la base, combinant statistiques traditionnelles et indicateurs avancés, a permis d'analyser la performance collective sous plusieurs angles complémentaires.

L'analyse descriptive a mis en évidence une **variabilité marquée des performances entre les équipes**, soulignant que la victoire ne dépend pas uniquement du nombre de points marqués, mais également de l'efficacité offensive, de l'organisation collective et de la discipline de jeu. Les statistiques avancées, telles que l'**Effective Field Goal Percentage**, le **Player Impact Estimate** et le **plus/minus**, se sont révélées particulièrement pertinentes pour résumer l'impact réel des équipes sur le jeu.

Les analyses multivariées, notamment l'Analyse en Composantes Principales et le clustering par K-means, ont permis d'identifier **différents profils d'équipes**, mettant en évidence des styles de jeu distincts et des niveaux de performance contrastés. Par ailleurs, l'approche prédictive basée sur une régression logistique a montré qu'un ensemble restreint de variables clés permettait déjà d'expliquer une part significative du résultat des matchs.

Enfin, le développement du dashboard interactif a constitué un apport majeur de ce travail. Il offre un outil d'exploration dynamique, capable de générer automatiquement des visualisations et des interprétations exploitables dans un cadre académique ou professionnel. Ce projet confirme ainsi que l'analyse des performances en NBA repose sur une approche **multidimensionnelle**, combinant données, méthodes statistiques et visualisation interactive, et ouvre des perspectives intéressantes pour des analyses plus approfondies et des modèles prédictifs plus avancés.