DATA ANALITYCS

SPORT - NBA Icon

Description automatically generated  Graphical user interface

Description automatically generated

TEAM/PLAYERS PERFORMANCE : MILWAUKEE BUCKS AND GIANNIS ANTETOKOUNMPO

RNCP

BOUBACAR TRAORE

DECEMBER 2022

**Table of contents**

**Introduction1**

**Data and data sources 4**

**Data cleaning and Exploratory data analysis4**

Python cleaning process5

Data visualization5

**Database4**

Type of database5

Entities. ER Model5

SQL database 5

**Conclusion4**

**Introduction**

**Performances of the Milwaukee bucks and its superstar Giannis Antetokounmpo**

The NBA (National Basketball Association) is a professional basketball league in North America and the biggest basketball league in the world. The league is composed of 30 teams (29 in the United States and 1 in Canada)

Passionate about basketball I am following the NBA and most particularly the Bucks, the team of Milwaukee city. Giannis Antetokounmpo is a Greek player who was drafted by the bucks in 2013, at the age of 18. He is now 28 and one of the best players in the NBA with two MVP (Most Valuable Trophies) in 2019 and 2020.

Having learnt the basic tools of data analysis I now want to do my own approach on analyzing data about the evolution of the Bucks

The goal of this project was to analyze the weight of Giannis Antetokounmpo performances on the results of the Milwaukee Bucks team by comparing the player individuals’ statistics and the team statistics. I also want to show the evolution of the Milwaukee Bucks across the years team since Giannis started being in the team in 2013.

The plan of the project was to choose a data source that was able to provide all the game results and statistics of the Milwaukee bucks since 2013.

After export of the data into CSV files, imported the files into a new script of Python, to produce data cleaning (by managing with outliers, missing values, and features that were not relevant for the study) and visualizations, to perform a primary data analysis.

**DATA AND DATA SOURCES**

The data sources are from (<https://www.kaggle.com/datasets/nathanlauga/nba-games?resource=download>)

It is composed of 5 CSV files: games.csv, games\_details.csv, players.csv, ranking.csv, teams.csv.

I have only used these three files: games.csv, games\_details.csv and ranking.csv.

|  |  |  |
| --- | --- | --- |
| **file** | Games.csv | Games\_details.csv |
| **Content** | Every NBA game from 2003 to 2022 with stats of the two teams | Every NBA game from 2003 to 2022 with stats for each player of the two teams |
| **shape** | 25 796 rows,  21 columns | 645 953 rows,  29 columns |

**Description of some not-explicit columns**:

*Games.csv*

'PTS\_home/away': Total points scored by the home/away team

'AST\_ home/away': Total assists (pass before a shot made) by the home/away team

'REB\_ home/away': Total rebound by the home/away team

'FG\_PCT\_ home/away': percentage (accuracy) of shot made by the home/away team

'FT\_PCT\_ home/away': percentage (accuracy) of free throws made by the home/away team

'FG3\_PCT\_ home/away': percentage (accuracy) of 3 points shots made by the home/away team

Games\_details.csv

‘START\_POSITION': role of the player when he started the game

'MIN': total minutes played by a player during a game

'FGA': (Field goal attempts) number of shots attempts by a player

'FGM': (Field goal made) number of shots made by a player

'FG\_PCT':(Field goal percentage) = FGM / FGA

'OREB': offensive rebounds by a player

'DREB': defensive rebounds by a player

'REB': total rebounds by a player

'AST': total assists by a player

'STL': total steal by a player

'BLK': total block by a player

'TO': total turnover by a player

'PTS': total points by a player

'PLUS\_MINUS'**:** used to measure a player's impact on the game, represented by the difference between their team's total scoring versus their opponent's when the player is in the game.