Project Title – NBA Games & Players Data Analysis

Description:

- By using NBA data, I have performed analysis and visualized some of the results. I have answered following Questions by my analysis.

Questions:

1. From NBA Champions between 1947 and 2020, Find out how many times each team won NBA Championship.
2. Which teams are still to win their first NBA Championship? By checking their foundation year, we can imagine for how long their first championship is due.
3. How many years each team took to win its first NBA championship since it was founded? Also, I have sorted them so that we could find the teams which were quickest to win NBA Championship.
4. Between 2003 and 2020, list out teams from each conference which finished first in their conference and won NBA championship that year.
5. Find out Correlation between Free Throw Percentage and Players' Height for the players who played at least 300 matches during 2003-2020.
6. LeBron James played for 3 teams during 2003-2020: Miami Heat, Cleveland Cavaliers & Los Angeles Lakers and he won NBA championship with all the teams. So, I compared his performance for these teams and also made a plot of comparison.
7. Between 2003 and 2020, List out top 10 players with most no. of Triple Doubles.
8. Comparison between 2 selected players for all the statistical categories. This is per game's performance comparison.
9. Between 2003 and 2020 which team had most no. of Home Wins? Also, which team had most no. of Away Wins? List out Top 10 teams from each conference.

10. List out teams from every year's points table from each conference which qualified for playoffs during 2004-2020.
11. Between 2004 and 2020, List out top 10 teams from each conference which qualified most for the play-offs.
12. Plot the comparison of all teams' average (per game) performance in any statistical category for the selected year between 2004 and 2020.
13. Plot the side-by-side yearly comparison for home and away performance in any statistical category for the selected team between 2004 and 2020.
14. Yearwise comparison for selected team in these 7 categories: Total Wins, Total Loses, Winning Percentage, Home Wins, Away Wins, Home Loses, Away Loses, Team Position. Plot a comparison bar chart.
15. Between 1947 and 2020, List out NBA MVPs whose team won the championship.

Conclusion: Although, I faced difficulties in some of the questions, I have successfully answered all 15 questions.

Learning from Project:

- The main thing I learned from the project is interactive python in jupyter notebook for which I used ipywidgets module. I just linked my function with widget and it makes the function interactive. Also, I learned more about numpy, pandas, data pre-processing, data cleaning and data visualization.

Instructions:

- I have used interactive python in jupyter notebook, so after opening ipynb file, please click on Run All Cells. (Just opening file normally won't run some of the cells in the code)
- Please open these csv files with any text editor, opening these files with Microsoft Excel changes the format of some columns and affects the overall analysis.

Data:

I have uploaded data files on Google Drive. So, please use this link to download data files:

https://drive.google.com/drive/folders/1Yd3mBcvNovst2iac-fVh-GJ8rzyFspxL?usp=sharing

- I have used 7 dataset files for the analysis. I have downloaded this data files from Kaggle. The link of the original dataset is as below:

Original Data Link: https://www.kaggle.com/nathanlauga/nba-games

1. teams_info.csv:

- It has 9 columns. It shows information about all 30 NBA teams.

1. TEAM_ID: Unique ID of each team.
2. ABBREVIATION: 3 Letter Abbreviation code for each team.
3. Team Name: Name of each team.
4. Year Founded: Founding Year of each team.
5. City: City that each team represents.
6. ARENA: Each team's home Stadium/Arena Name.
7. OWNER: Owner of each team.
8. GENERALMANAGER: General Manager of each team.

2. ranking.csv:

- It has 13 columns. It shows daily points table for NBA Regular Season and Pre-Season for 2003-2020.

1. Team_ID: Unique ID of each team.
2. LEAGUE_ID: League ID. This column is not significant, so I just dropped it.
3. Season_ID: It represents Season. For Exp: if Season_ID=22019 means it represents 2019-20 season (Year=2020). So, in my code I used Year = Season_ID – 20000 + 1 formula many times.
4. Standings Date: Date for which ranking is given.
5. Conference: Conference Name.
6. Team Name: Name of Team.
7. G: No. of Games played by team at particular date.
8. W: No. of games Won by team at particular date.
9. L: No. of games Lost by team at particular date.
10. W_PCT: Winning percentage for each team at particular date.

11. HOME_RECORD: The value 20-16 means that team won 20 Home games and lost 16 Home games.
12. ROAD_RECORD: The value 20-16 means that team won 20 Road/Away games and lost 16 Road/Away games.
13. RETURTOPLAY: I have not used this column to analyze.

3. players.csv:

- It has 4 columns. It shows players data.

1. Player: Player Name.
2. Team_ID: Team's ID for which player played that season.
3. Player_ID: Player's unique ID.
4. Season: Year for which player played for particular team.

4. players_bios.csv:

- It has 7 columns. It includes players bios like their height, weight and birth data.

1. Player: Player Name.
2. Height: Player's height in cm.
3. Weight: Player's weight in kg.
4. College: College/University Name.
5. Birth_Year: Player's birth year.
6. Birth_City: Player's birth city.
7. Birth_State: Player's birth state.

5. NBA Finals and MVP.xlsx:

- It has 15 columns. It represents NBA finals data between 1947-2020. It also has data about MVPs.

1. Year: NBA final's year.
2. Western Champion: That year's western champion.
3. Eastern Champion: That year's eastern champion.
4. Result: Finals 7 game series result.
5. NBA Champion: That year's NBA champion's full name.
6. Champion Name: That year's NBA champion's nickname.
7. NBA Vice-Champion: That year's NBA vice-champion.
8. Final Sweep: If the winning team won NBA by 4-0 then it is called sweep. So, if the column Result has value 4-0 then it is TRUE otherwise FALSE.
9. MVP Name: That year's MVP Name.
10. MVP Height (m): MVP's Height in meters.

11. MVP Height (ft): MVP's Height in ft.
12. MVP Position: MVP's playing position.
13. MVP Team: MVP's team name.
14. MVP Nationality: MVP's nationality.
15. MVP status: MVP's team status means his team's final position.

6. games.csv:

- It has 21 columns. It represents each and every games data which were played between 2003 to 21st March,2021.

1. GAME_DATE_EST: Game date
2. GAME_ID: Game's unique ID
3. GAME_STATUS_TEXT: Game's Status. All rows have same value - Final.
4. HOME_TEAM_ID: Home Team's unique ID
5. VISITOR_TEAM_ID: Away Team's unique ID
6. SEASON: Season/Year
7. TEAM_ID_home: Home Team's unique ID
8. PTS_home: Home team's Points in that game
9. FG_PCT_home: Home team's Field Goal Percentage in that game
10. FT_PCT_home: Home team's Free Throw Percentage in that game
11. FG3_PCT_home: Home team's Field Goal 3-Point Percentage in that game
12. AST_home: Home team's Assists in that game
13. REB_home: Home team's Rebounds in that game
14. TEAM_ID_away: Away Team's unique ID
15. PTS_away: Away team's Points in that game
16. FG_PCT_away: Away team's Field Goal Percentage in that game
17. FT_PCT_away: Away team's Free Throw Percentage in that game
18. FG3_PCT_away: Away team's Field Goal 3-Point Percentage in that game
19. AST_away: Away team's Assists in that game
20. REB_away: Away team's Rebounds in that game
21. HOME_TEAM_WINS: If home team won the game, then the value is 1 otherwise 0.

7. games_details.csv:

- It has 28 columns. It represents each player's performance in each and every NBA 2003-2020 regular season's game.

1. Game_ID: Each game's unique ID
2. Team_ID: Each team's unique ID
3. Team_Abbreviation: Team's 3 letter abbreviation
4. Team City: City that each team represents.

5. Player_ID: Player's unique ID
6. Player Name: Player's name
7. Start_Position: Starting position of each player
8. Comment: Extra comment about player mostly about if that player did not play that match
9. MIN: Minutes played by player
10. FGM: Field Goal Made by player
11. FGA: Field Goal Attempted by player
12. FG_PCT: Field Goal Percentage of player
13. FG3M: Field Goal 3-Point Made by player
14. FG3A: Field Goal 3-Point Attempted by player
15. FG3_PCT: Field Goal 3-Point Percentage of player
16. FTM: Free Throw Made by player
17. FTA: Free Throw Attempted by player
18. FT_PCT: Free Throw Percentage of player
19. OREB: Offensive Rebounds by player
20. DREB: Defensive Rebounds by player
21. REB: Total Rebounds by player
22. AST: Assists of player
23. STL: Steals by player
24. BLK: Blocks of player
25. TO: Turn Overs by player
26. PF: Personal Fouls of player
27. PTS: Points made by player
28. PLUS_MINUS: It is not used for analysis.