

NBA_Data_Cleaning_and_Feature_Engineering

December 12, 2022

##Data Cleaning

```
[ ]: import pandas as pd
import numpy as np
```

```
[ ]: pd.set_option("display.max_columns", None)
```

```
[ ]: games = pd.read_csv("/content/drive/MyDrive/Machine Learning NBA /Final Data/
↳games.csv", index_col="id")
```

```
[ ]: games
```

```
[ ]:
      date  home_team_score  period  postseason  season \
id
47179  2019-01-30T00:00:00.000Z      126      4      False    2018
48751  2019-02-09T00:00:00.000Z      112      4      False    2018
48739  2019-02-08T00:00:00.000Z      117      4      False    2018
48740  2019-02-08T00:00:00.000Z      119      4      False    2018
48746  2019-02-08T00:00:00.000Z      102      4      False    2018
...
473801  2021-12-11T00:00:00.000Z      112      4      False    2021
473863  2021-12-19T00:00:00.000Z      137      4      False    2021
473890  2021-12-23T00:00:00.000Z      107      4      False    2021
473895  2021-12-23T00:00:00.000Z      110      4      False    2021
473818  2021-12-13T00:00:00.000Z      117      4      False    2021
```

```
      status  time  visitor_team_score  home_team.id  home_team.abbreviation \
id
47179  Final      94      2      BOS
48751  Final     123      2      BOS
48739  Final     110     23      PHI
48740  Final     106     30      WAS
48746  Final      96     26      SAC
...
473801  Final  NaN     127     27      SAS
473863  Final  NaN     106     24      PHX
473890  Final  NaN     115      8      DEN
473895  Final  NaN     138     14      LAL
```

473818 Final NaN 103 2 BOS

| | home_team.city | home_team.conference | home_team.division | \ |
|--------|----------------|----------------------|--------------------|---|
| id | | | | |
| 47179 | Boston | East | Atlantic | |
| 48751 | Boston | East | Atlantic | |
| 48739 | Philadelphia | East | Atlantic | |
| 48740 | Washington | East | Southeast | |
| 48746 | Sacramento | West | Pacific | |
| ... | ... | ... | ... | |
| 473801 | San Antonio | West | Southwest | |
| 473863 | Phoenix | West | Pacific | |
| 473890 | Denver | West | Northwest | |
| 473895 | Los Angeles | West | Pacific | |
| 473818 | Boston | East | Atlantic | |

| | home_team.full_name | home_team.name | visitor_team.id | \ |
|--------|---------------------|----------------|-----------------|---|
| id | | | | |
| 47179 | Boston Celtics | Celtics | 4 | |
| 48751 | Boston Celtics | Celtics | 13 | |
| 48739 | Philadelphia 76ers | 76ers | 8 | |
| 48740 | Washington Wizards | Wizards | 6 | |
| 48746 | Sacramento Kings | Kings | 16 | |
| ... | ... | ... | ... | |
| 473801 | San Antonio Spurs | Spurs | 8 | |
| 473863 | Phoenix Suns | Suns | 4 | |
| 473890 | Denver Nuggets | Nuggets | 4 | |
| 473895 | Los Angeles Lakers | Lakers | 27 | |
| 473818 | Boston Celtics | Celtics | 17 | |

| | visitor_team.abbreviation | visitor_team.city | visitor_team.conference | \ |
|--------|---------------------------|-------------------|-------------------------|---|
| id | | | | |
| 47179 | CHA | Charlotte | East | |
| 48751 | LAC | LA | West | |
| 48739 | DEN | Denver | West | |
| 48740 | CLE | Cleveland | East | |
| 48746 | MIA | Miami | East | |
| ... | ... | ... | ... | |
| 473801 | DEN | Denver | West | |
| 473863 | CHA | Charlotte | East | |
| 473890 | CHA | Charlotte | East | |
| 473895 | SAS | San Antonio | West | |
| 473818 | MIL | Milwaukee | East | |

| | visitor_team.division | visitor_team.full_name | visitor_team.name |
|-------|-----------------------|------------------------|-------------------|
| id | | | |
| 47179 | Southeast | Charlotte Hornets | Hornets |

| | | | |
|--------|-----------|---------------------|-----------|
| 48751 | Pacific | LA Clippers | Clippers |
| 48739 | Northwest | Denver Nuggets | Nuggets |
| 48740 | Central | Cleveland Cavaliers | Cavaliers |
| 48746 | Southeast | Miami Heat | Heat |
| ... | ... | ... | ... |
| 473801 | Northwest | Denver Nuggets | Nuggets |
| 473863 | Southeast | Charlotte Hornets | Hornets |
| 473890 | Southeast | Charlotte Hornets | Hornets |
| 473895 | Southwest | San Antonio Spurs | Spurs |
| 473818 | Central | Milwaukee Bucks | Bucks |

[50000 rows x 22 columns]

```
[ ]: # time column is empty
games.drop("time", axis=1, inplace=True)
```

```
[ ]: # dropping a random unfinished game
games.drop(games[games["status"].eq("8:30 PM ET")].index, axis=0, inplace=True)
```

```
[ ]: # dropping unplayed games
games.drop(games[games["home_team_score"].eq(0)].index, axis=0, inplace=True)
```

```
[ ]: # dropping unnecessary name info
games.drop(["home_team.city", "visitor_team.city", "home_team.name", "visitor_team.name"], axis=1, inplace=True)
```

```
[ ]: games
```

```
[ ]:
      date  home_team_score  period  postseason  season \
id
47179  2019-01-30T00:00:00.000Z      126      4      False    2018
48751  2019-02-09T00:00:00.000Z      112      4      False    2018
48739  2019-02-08T00:00:00.000Z      117      4      False    2018
48740  2019-02-08T00:00:00.000Z      119      4      False    2018
48746  2019-02-08T00:00:00.000Z      102      4      False    2018
...
473801  2021-12-11T00:00:00.000Z      112      4      False    2021
473863  2021-12-19T00:00:00.000Z      137      4      False    2021
473890  2021-12-23T00:00:00.000Z      107      4      False    2021
473895  2021-12-23T00:00:00.000Z      110      4      False    2021
473818  2021-12-13T00:00:00.000Z      117      4      False    2021

      status  visitor_team_score  home_team.id  home_team.abbreviation \
id
47179  Final          94          2          BOS
48751  Final          123          2          BOS
48739  Final          110         23          PHI
```

| | | | | | |
|--------|-------|-----|-----|-----|-----|
| 48740 | Final | | 106 | 30 | WAS |
| 48746 | Final | | 96 | 26 | SAC |
| ... | ... | ... | ... | ... | ... |
| 473801 | Final | | 127 | 27 | SAS |
| 473863 | Final | | 106 | 24 | PHX |
| 473890 | Final | | 115 | 8 | DEN |
| 473895 | Final | | 138 | 14 | LAL |
| 473818 | Final | | 103 | 2 | BOS |

| | home_team.conference | home_team.division | home_team.full_name | \ |
|--------|----------------------|--------------------|---------------------|---|
| id | | | | |
| 47179 | East | Atlantic | Boston Celtics | |
| 48751 | East | Atlantic | Boston Celtics | |
| 48739 | East | Atlantic | Philadelphia 76ers | |
| 48740 | East | Southeast | Washington Wizards | |
| 48746 | West | Pacific | Sacramento Kings | |
| ... | ... | ... | ... | |
| 473801 | West | Southwest | San Antonio Spurs | |
| 473863 | West | Pacific | Phoenix Suns | |
| 473890 | West | Northwest | Denver Nuggets | |
| 473895 | West | Pacific | Los Angeles Lakers | |
| 473818 | East | Atlantic | Boston Celtics | |

| | visitor_team.id | visitor_team.abbreviation | visitor_team.conference | \ |
|--------|-----------------|---------------------------|-------------------------|---|
| id | | | | |
| 47179 | 4 | CHA | East | |
| 48751 | 13 | LAC | West | |
| 48739 | 8 | DEN | West | |
| 48740 | 6 | CLE | East | |
| 48746 | 16 | MIA | East | |
| ... | ... | ... | ... | |
| 473801 | 8 | DEN | West | |
| 473863 | 4 | CHA | East | |
| 473890 | 4 | CHA | East | |
| 473895 | 27 | SAS | West | |
| 473818 | 17 | MIL | East | |

| | visitor_team.division | visitor_team.full_name |
|--------|-----------------------|------------------------|
| id | | |
| 47179 | Southeast | Charlotte Hornets |
| 48751 | Pacific | LA Clippers |
| 48739 | Northwest | Denver Nuggets |
| 48740 | Central | Cleveland Cavaliers |
| 48746 | Southeast | Miami Heat |
| ... | ... | ... |
| 473801 | Northwest | Denver Nuggets |
| 473863 | Southeast | Charlotte Hornets |

| | | |
|--------|-----------|-------------------|
| 473890 | Southeast | Charlotte Hornets |
| 473895 | Southwest | San Antonio Spurs |
| 473818 | Central | Milwaukee Bucks |

[49995 rows x 17 columns]

```
[ ]: games.to_csv("/content/drive/MyDrive/Machine Learning NBA /Final Data/
↳games_clean.csv")
```

##Stats Data

This is used to build the model

```
[ ]: stats = pd.read_csv("/content/drive/MyDrive/Machine Learning NBA /Final Data/
↳stats_raw.csv", index_col="id")
```

/usr/local/lib/python3.7/dist-packages/IPython/core/interactiveshell.py:3326:
DtypeWarning: Columns (13,28,36) have mixed types.Specify dtype option on import
or set low_memory=False.

```
exec(code_obj, self.user_global_ns, self.user_ns)
```

```
[ ]: stats
```

```
[ ]:
      ast  blk  dreb  fg3_pct  fg3a  fg3m  fg_pct  fga  fgm  ft_pct  \
id
1069106  0.0  0.0  0.0  0.000000  0.0  0.0  0.000000  0.0  0.0  0.000
1069107  NaN  NaN  NaN         NaN  NaN  NaN         NaN  NaN  NaN    NaN
1069108  0.0  0.0  4.0  0.000000  1.0  0.0  0.667000  9.0  6.0  0.000
1069109  1.0  0.0  9.0  0.500000  10.0  5.0  0.556000  18.0  10.0  0.667
1069110  4.0  0.0  6.0  0.500000  2.0  1.0  0.308000  13.0  4.0  0.500
...
12399585  4.0  1.0  3.0  0.000000  2.0  0.0  0.166667  6.0  1.0  0.000
12399582  0.0  1.0  1.0  0.000000  0.0  0.0  0.000000  0.0  0.0  0.000
12399586  1.0  2.0  3.0  0.571429  7.0  4.0  0.500000  8.0  4.0  0.000
12399580  1.0  0.0  1.0  0.000000  1.0  0.0  0.000000  2.0  0.0  0.000
12399579  1.0  0.0  0.0  0.000000  0.0  0.0  0.000000  0.0  0.0  0.000

      fta  ftm   min  oreb  pf  pts  reb  stl  turnover  game.id  \
id
1069106  0.0  0.0  1:09  0.0  0.0  0.0  0.0  0.0         0.0  45167
1069107  NaN  NaN   NaN  NaN  NaN  NaN  NaN  NaN         NaN  45167
1069108  0.0  0.0  33:00  1.0  2.0  12.0  5.0  1.0         0.0  45206
1069109  3.0  2.0  25:09  1.0  3.0  27.0  10.0  0.0         2.0  45206
1069110  2.0  1.0  27:49  7.0  2.0  10.0  13.0  0.0         1.0  45206
...
12399585  0.0  0.0   13  0.0  1.0  2.0  3.0  1.0         0.0  857660
12399582  0.0  0.0   04  0.0  2.0  0.0  1.0  0.0         1.0  857660
12399586  0.0  0.0   29  1.0  4.0  12.0  4.0  1.0         2.0  857660
```

| | | | | | | | | | | |
|----------|-----|-----|----|-----|-----|-----|-----|-----|-----|--------|
| 12399580 | 0.0 | 0.0 | 08 | 1.0 | 1.0 | 0.0 | 2.0 | 0.0 | 0.0 | 857660 |
| 12399579 | 0.0 | 0.0 | 03 | 0.0 | 1.0 | 0.0 | 0.0 | 0.0 | 1.0 | 857660 |

| | game.date | game.home_team_id | game.home_team_score | \ |
|----------|--------------------------|-------------------|----------------------|---|
| id | | | | |
| 1069106 | 2019-01-17T00:00:00.000Z | 12 | 96 | |
| 1069107 | 2019-01-17T00:00:00.000Z | 12 | 96 | |
| 1069108 | 2019-01-17T00:00:00.000Z | 8 | 135 | |
| 1069109 | 2019-01-17T00:00:00.000Z | 8 | 135 | |
| 1069110 | 2019-01-17T00:00:00.000Z | 8 | 135 | |
| ... | ... | ... | ... | |
| 12399585 | 2022-11-28T00:00:00.000Z | 14 | 115 | |
| 12399582 | 2022-11-28T00:00:00.000Z | 14 | 115 | |
| 12399586 | 2022-11-28T00:00:00.000Z | 14 | 115 | |
| 12399580 | 2022-11-28T00:00:00.000Z | 14 | 115 | |
| 12399579 | 2022-11-28T00:00:00.000Z | 14 | 115 | |

| | game.period | game.postseason | game.season | game.status | game.time | \ |
|----------|-------------|-----------------|-------------|-------------|-----------|---|
| id | | | | | | |
| 1069106 | 4 | False | 2018 | Final | | |
| 1069107 | 4 | False | 2018 | Final | | |
| 1069108 | 4 | False | 2018 | Final | | |
| 1069109 | 4 | False | 2018 | Final | | |
| 1069110 | 4 | False | 2018 | Final | | |
| ... | ... | ... | ... | ... | ... | |
| 12399585 | 4 | False | 2022 | Final | Final | |
| 12399582 | 4 | False | 2022 | Final | Final | |
| 12399586 | 4 | False | 2022 | Final | Final | |
| 12399580 | 4 | False | 2022 | Final | Final | |
| 12399579 | 4 | False | 2022 | Final | Final | |

| | game.visitor_team_id | game.visitor_team_score | player.id | \ |
|----------|----------------------|-------------------------|------------|---|
| id | | | | |
| 1069106 | 23 | 120 | 240.0 | |
| 1069107 | 23 | 120 | 352.0 | |
| 1069108 | 5 | 105 | 222.0 | |
| 1069109 | 5 | 105 | 297.0 | |
| 1069110 | 5 | 105 | 377.0 | |
| ... | ... | ... | ... | |
| 12399585 | 12 | 116 | 304.0 | |
| 12399582 | 12 | 116 | 17896035.0 | |
| 12399586 | 12 | 116 | 38017507.0 | |
| 12399580 | 12 | 116 | 666468.0 | |
| 12399579 | 12 | 116 | 666442.0 | |

| | player.first_name | player.height_feet | player.height_inches | \ |
|----|-------------------|--------------------|----------------------|---|
| id | | | | |

| | | | |
|----------|----------|-----|------|
| 1069106 | Alize | 6.0 | 9.0 |
| 1069107 | Kyle | 6.0 | 10.0 |
| 1069108 | Chandler | 6.0 | 7.0 |
| 1069109 | Lauri | 7.0 | 0.0 |
| 1069110 | Bobby | 6.0 | 11.0 |
| ... | ... | ... | ... |
| 12399585 | T.J. | 6.0 | 2.0 |
| 12399582 | Isaiah | NaN | NaN |
| 12399586 | Andrew | NaN | NaN |
| 12399580 | Oshae | NaN | NaN |
| 12399579 | Goga | NaN | NaN |

| | player.last_name | player.position | player.team_id | \ |
|----------|------------------|-----------------|----------------|---|
| id | | | | |
| 1069106 | Johnson | F | 19.0 | |
| 1069107 | O'Quinn | C | 23.0 | |
| 1069108 | Hutchison | F | 24.0 | |
| 1069109 | Markkanen | F | 29.0 | |
| 1069110 | Portis | F | 17.0 | |
| ... | ... | ... | ... | |
| 12399585 | McConnell | G | 12.0 | |
| 12399582 | Jackson | F | 12.0 | |
| 12399586 | Nembhard | G-F | 12.0 | |
| 12399580 | Brissett | F | 12.0 | |
| 12399579 | Bitadze | C | 12.0 | |

| | player.weight_pounds | team.id | team.abbreviation | team.city | \ |
|----------|----------------------|---------|-------------------|-----------|---|
| id | | | | | |
| 1069106 | 212.0 | 12 | IND | Indiana | |
| 1069107 | 250.0 | 12 | IND | Indiana | |
| 1069108 | 197.0 | 5 | CHI | Chicago | |
| 1069109 | 240.0 | 5 | CHI | Chicago | |
| 1069110 | 250.0 | 5 | CHI | Chicago | |
| ... | ... | ... | ... | ... | |
| 12399585 | 190.0 | 12 | IND | Indiana | |
| 12399582 | NaN | 12 | IND | Indiana | |
| 12399586 | NaN | 12 | IND | Indiana | |
| 12399580 | NaN | 12 | IND | Indiana | |
| 12399579 | NaN | 12 | IND | Indiana | |

| | team.conference | team.division | team.full_name | team.name | player |
|---------|-----------------|---------------|----------------|-----------|--------|
| id | | | | | |
| 1069106 | East | Central | Indiana Pacers | Pacers | NaN |
| 1069107 | East | Central | Indiana Pacers | Pacers | NaN |
| 1069108 | East | Central | Chicago Bulls | Bulls | NaN |
| 1069109 | East | Central | Chicago Bulls | Bulls | NaN |
| 1069110 | East | Central | Chicago Bulls | Bulls | NaN |

| | | | | | |
|----------|------|---------|----------------|--------|-----|
| ... | ... | ... | ... | ... | ... |
| 12399585 | East | Central | Indiana Pacers | Pacers | NaN |
| 12399582 | East | Central | Indiana Pacers | Pacers | NaN |
| 12399586 | East | Central | Indiana Pacers | Pacers | NaN |
| 12399580 | East | Central | Indiana Pacers | Pacers | NaN |
| 12399579 | East | Central | Indiana Pacers | Pacers | NaN |

[1192900 rows x 46 columns]

```
[ ]: # drop unneeded columns
stats.drop(["game.period", "game.postseason", "game.status", "game.time",
           ↪ "player.height_feet", "player.height_inches",
           "player.weight_pounds", "team.abbreviation", "team.city", "team.
           ↪ conference", "team.division", "team.name",
           "player", "player.first_name", "player.last_name", "player.
           ↪ position", "team.full_name", "player.team_id"],
           axis=1, inplace=True)
```

```
[ ]: # a null value generally indicates that the player did not play in that game
stats.dropna(axis=0, how="any", inplace=True)
```

```
[ ]: # clean time column to get a consistent format. ("mm:ss" or "m:ss")
stats["min"] = stats["min"].astype(str)

# drop the row if the player didn't play in the game
played_0min = stats[stats["min"].eq("0:00") | stats["min"].str.startswith("0")
           ↪ | stats["min"].eq("")].index
stats.drop(played_0min, axis=0, inplace=True)

# Convert times like "27.0" to "27:0"
stats["min"] = stats["min"].str.replace(".", ":")

# convert times like "27" to "27:00"
minutes_only_times = stats["min"][~stats["min"].str.contains(":").index
stats["min"].loc[minutes_only_times] += ":00"

minutes = [time[0] for time in stats["min"].str.split(":").values]
seconds = [time[1] for time in stats["min"].str.split(":").values]

# convert times like "27:0" to "27:00"
for i, second in enumerate(seconds):
    if len(second) == 1:
        seconds[i] = second + "0"

# convert times like "8:60" to "9:00"
for i, second in enumerate(seconds):
```



```

    if second == "60":
        seconds[i] = "00"
        minutes[i] = str(int(minutes[i]) + 1) # increment minutes by 1

times = [":".join(list(item)) for item in list(zip(minutes,seconds))]

stats["min"] = times

```

/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:9: FutureWarning:
The default value of regex will change from True to False in a future version.
In addition, single character regular expressions will *not* be treated as
literal strings when regex=True.

```

    if __name__ == '__main__':
/usr/local/lib/python3.7/dist-packages/pandas/core/indexing.py:1732:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
 self._setitem_single_block(indexer, value, name)

```
[ ]: stats
```

```

[ ]:
      ast  blk  dreb  fg3_pct  fg3a  fg3m  fg_pct  fga  fgm  \
id
1069106  0.0  0.0  0.0  0.000000  0.0  0.0  0.000000  0.0  0.0
1069108  0.0  0.0  4.0  0.000000  1.0  0.0  0.667000  9.0  6.0
1069109  1.0  0.0  9.0  0.500000  10.0  5.0  0.556000  18.0  10.0
1069110  4.0  0.0  6.0  0.500000  2.0  1.0  0.308000  13.0  4.0
1069111  6.0  0.0  3.0  0.000000  3.0  0.0  0.267000  15.0  4.0
...
12398943  1.0  0.0  3.0  0.428571  7.0  3.0  0.545455  11.0  6.0
12399578  14.0  1.0  5.0  0.428571  7.0  3.0  0.454545  22.0  10.0
12399584  1.0  0.0  6.0  0.375000  8.0  3.0  0.454545  22.0  10.0
12399585  4.0  1.0  3.0  0.000000  2.0  0.0  0.166667  6.0  1.0
12399586  1.0  2.0  3.0  0.571429  7.0  4.0  0.500000  8.0  4.0

      ft_pct  fta  ftm  min  oreb  pf  pts  reb  stl  turnover  \
id
1069106  0.000000  0.0  0.0  1:09  0.0  0.0  0.0  0.0  0.0  0.0
1069108  0.000000  0.0  0.0  33:00  1.0  2.0  12.0  5.0  1.0  0.0
1069109  0.667000  3.0  2.0  25:09  1.0  3.0  27.0  10.0  0.0  2.0
1069110  0.500000  2.0  1.0  27:49  7.0  2.0  10.0  13.0  0.0  1.0
1069111  0.000000  0.0  0.0  28:48  2.0  4.0  8.0  5.0  0.0  2.0
...
12398943  0.000000  0.0  0.0  25:00  2.0  2.0  15.0  5.0  0.0  2.0
12399578  0.333333  3.0  1.0  35:00  2.0  1.0  24.0  7.0  1.0  0.0

```

| | | | | | | | | | | |
|----------|----------|-----|-----|-------|-----|-----|------|-----|-----|-----|
| 12399584 | 0.000000 | 0.0 | 0.0 | 32:00 | 2.0 | 0.0 | 23.0 | 8.0 | 1.0 | 2.0 |
| 12399585 | 0.000000 | 0.0 | 0.0 | 13:00 | 0.0 | 1.0 | 2.0 | 3.0 | 1.0 | 0.0 |
| 12399586 | 0.000000 | 0.0 | 0.0 | 29:00 | 1.0 | 4.0 | 12.0 | 4.0 | 1.0 | 2.0 |

| | game.id | game.date | game.home_team_id | \ |
|----------|---------|--------------------------|-------------------|---|
| id | | | | |
| 1069106 | 45167 | 2019-01-17T00:00:00.000Z | 12 | |
| 1069108 | 45206 | 2019-01-17T00:00:00.000Z | 8 | |
| 1069109 | 45206 | 2019-01-17T00:00:00.000Z | 8 | |
| 1069110 | 45206 | 2019-01-17T00:00:00.000Z | 8 | |
| 1069111 | 45206 | 2019-01-17T00:00:00.000Z | 8 | |
| ... | ... | ... | ... | |
| 12398943 | 857658 | 2022-11-28T00:00:00.000Z | 26 | |
| 12399578 | 857660 | 2022-11-28T00:00:00.000Z | 14 | |
| 12399584 | 857660 | 2022-11-28T00:00:00.000Z | 14 | |
| 12399585 | 857660 | 2022-11-28T00:00:00.000Z | 14 | |
| 12399586 | 857660 | 2022-11-28T00:00:00.000Z | 14 | |

| | game.home_team_score | game.season | game.visitor_team_id | \ |
|----------|----------------------|-------------|----------------------|---|
| id | | | | |
| 1069106 | 96 | 2018 | 23 | |
| 1069108 | 135 | 2018 | 5 | |
| 1069109 | 135 | 2018 | 5 | |
| 1069110 | 135 | 2018 | 5 | |
| 1069111 | 135 | 2018 | 5 | |
| ... | ... | ... | ... | |
| 12398943 | 117 | 2022 | 24 | |
| 12399578 | 115 | 2022 | 12 | |
| 12399584 | 115 | 2022 | 12 | |
| 12399585 | 115 | 2022 | 12 | |
| 12399586 | 115 | 2022 | 12 | |

| | game.visitor_team_score | player.id | team.id |
|----------|-------------------------|------------|---------|
| id | | | |
| 1069106 | 120 | 240.0 | 12 |
| 1069108 | 105 | 222.0 | 5 |
| 1069109 | 105 | 297.0 | 5 |
| 1069110 | 105 | 377.0 | 5 |
| 1069111 | 105 | 268.0 | 5 |
| ... | ... | ... | ... |
| 12398943 | 122 | 272.0 | 24 |
| 12399578 | 116 | 3547245.0 | 12 |
| 12399584 | 116 | 38017686.0 | 12 |
| 12399585 | 116 | 304.0 | 12 |
| 12399586 | 116 | 38017507.0 | 12 |

[968619 rows x 28 columns]

```
[ ]: stats.to_csv("/content/drive/MyDrive/Machine Learning NBA /Final Data/
↳stats_clean.csv")

#Feature Engineering
##Games Data

[ ]: all_games = pd.read_csv("/content/drive/MyDrive/Machine Learning NBA /Final_
↳Data/games_clean.csv", index_col="id")

[ ]: all_games.shape

[ ]: (49995, 17)

[ ]: # Feature engineering

g = all_games.copy()

g["date"] = pd.to_datetime(all_games["date"]).dt.tz_localize(None)

# creating target variable
g["winner"] = np.where(g["home_team_score"] > g["visitor_team_score"], 1, 0)

# creating avg points by team across all years (not sure how useful this will_
↳be)
home_avg_pts_map = g[["home_team.id", "home_team_score"]].groupby("home_team.
↳id").mean().squeeze()
g["home_team_avg_score_historical"] = g["home_team.id"].map(home_avg_pts_map).
↳round(1)

visitor_avg_pts_map = g[["visitor_team.id", "visitor_team_score"]].
↳groupby("visitor_team.id").mean().squeeze()
g["visitor_team_avg_score_historical"] = g["visitor_team.id"].
↳map(visitor_avg_pts_map).round(1)

[ ]: # creating an identifier for the team in a specific season
g["home_team_id_year"] = (g["home_team.id"].astype(str) + " " + g["season"].
↳astype(str)).values
g["visitor_team_id_year"] = (g["visitor_team.id"].astype(str) + " " +
↳g["season"].astype(str)).values

[ ]: # creating average pts feature
# average pts scored per game in that season
# based on if they are home or away because there is a statistically_
↳significant difference
home_avg_score_map = g[["home_team_id_year", "home_team_score"]].
↳groupby(["home_team_id_year"]).mean().squeeze()
```

```

visitor_avg_score_map = g[["visitor_team_id_year", "visitor_team_score"]].
    ↪groupby(["visitor_team_id_year"]).mean().squeeze()

g["home_team_avg_score"] = g["home_team_id_year"].map(home_avg_score_map).
    ↪round(1)
g["visitor_team_avg_score"] = g["visitor_team_id_year"].
    ↪map(visitor_avg_score_map).round(1)

```

```

[ ]: # create an average difference in pts feature
# will calculate on average how much a team wins or loses by
# this will hopefully benefit lower scoring teams that also play defensively
avg_score_diff = g[["home_team_id_year", "home_team_avg_score",
    ↪"visitor_team_avg_score"]].groupby("home_team_id_year").mean()
avg_score_diff = avg_score_diff["home_team_avg_score"] -
    ↪avg_score_diff["visitor_team_avg_score"]
g["home_avg_score_diff"] = g["home_team_id_year"].map(avg_score_diff)

avg_score_diff = g[["visitor_team_id_year", "visitor_team_avg_score",
    ↪"home_team_avg_score"]].groupby("visitor_team_id_year").mean()
avg_score_diff = avg_score_diff["visitor_team_avg_score"] -
    ↪avg_score_diff["home_team_avg_score"]
g["visitor_avg_score_diff"] = g["visitor_team_id_year"].map(avg_score_diff)

```

```

[ ]: g

```

```

[ ]:
      date  home_team_score  period  postseason  season  status  \
id
47179  2019-01-30           126        4         False    2018   Final
48751  2019-02-09           112        4         False    2018   Final
48739  2019-02-08           117        4         False    2018   Final
48740  2019-02-08           119        4         False    2018   Final
48746  2019-02-08           102        4         False    2018   Final
...      ...              ...      ...      ...      ...
473801 2021-12-11           112        4         False    2021   Final
473863 2021-12-19           137        4         False    2021   Final
473890 2021-12-23           107        4         False    2021   Final
473895 2021-12-23           110        4         False    2021   Final
473818 2021-12-13           117        4         False    2021   Final

      visitor_team_score  home_team.id  home_team.abbreviation  \
id
47179                94                2                   BOS
48751               123                2                   BOS
48739               110               23                   PHI
48740               106               30                   WAS
48746                96               26                   SAC

```

| | | | |
|--------|-----|-----|-----|
| ... | ... | ... | ... |
| 473801 | 127 | 27 | SAS |
| 473863 | 106 | 24 | PHX |
| 473890 | 115 | 8 | DEN |
| 473895 | 138 | 14 | LAL |
| 473818 | 103 | 2 | BOS |

| | home_team.conference | home_team.division | home_team.full_name | \ |
|--------|----------------------|--------------------|---------------------|---|
| id | | | | |
| 47179 | East | Atlantic | Boston Celtics | |
| 48751 | East | Atlantic | Boston Celtics | |
| 48739 | East | Atlantic | Philadelphia 76ers | |
| 48740 | East | Southeast | Washington Wizards | |
| 48746 | West | Pacific | Sacramento Kings | |
| ... | ... | ... | ... | |
| 473801 | West | Southwest | San Antonio Spurs | |
| 473863 | West | Pacific | Phoenix Suns | |
| 473890 | West | Northwest | Denver Nuggets | |
| 473895 | West | Pacific | Los Angeles Lakers | |
| 473818 | East | Atlantic | Boston Celtics | |

| | visitor_team.id | visitor_team.abbreviation | visitor_team.conference | \ |
|--------|-----------------|---------------------------|-------------------------|---|
| id | | | | |
| 47179 | 4 | CHA | East | |
| 48751 | 13 | LAC | West | |
| 48739 | 8 | DEN | West | |
| 48740 | 6 | CLE | East | |
| 48746 | 16 | MIA | East | |
| ... | ... | ... | ... | |
| 473801 | 8 | DEN | West | |
| 473863 | 4 | CHA | East | |
| 473890 | 4 | CHA | East | |
| 473895 | 27 | SAS | West | |
| 473818 | 17 | MIL | East | |

| | visitor_team.division | visitor_team.full_name | winner | \ |
|--------|-----------------------|------------------------|--------|---|
| id | | | | |
| 47179 | Southeast | Charlotte Hornets | 1 | |
| 48751 | Pacific | LA Clippers | 0 | |
| 48739 | Northwest | Denver Nuggets | 1 | |
| 48740 | Central | Cleveland Cavaliers | 1 | |
| 48746 | Southeast | Miami Heat | 1 | |
| ... | ... | ... | ... | |
| 473801 | Northwest | Denver Nuggets | 0 | |
| 473863 | Southeast | Charlotte Hornets | 1 | |
| 473890 | Southeast | Charlotte Hornets | 0 | |
| 473895 | Southwest | San Antonio Spurs | 0 | |

473818 Central Milwaukee Bucks 1

| | home_team_avg_score_historical | visitor_team_avg_score_historical | \ |
|--------|--------------------------------|-----------------------------------|---|
| id | | | |
| 47179 | 105.7 | 98.4 | |
| 48751 | 105.7 | 100.6 | |
| 48739 | 103.2 | 104.2 | |
| 48740 | 103.4 | 98.4 | |
| 48746 | 105.5 | 96.1 | |
| ... | ... | ... | |
| 473801 | 105.8 | 104.2 | |
| 473863 | 108.4 | 98.4 | |
| 473890 | 109.5 | 98.4 | |
| 473895 | 107.8 | 101.4 | |
| 473818 | 105.7 | 101.2 | |

| | home_team_id_year | visitor_team_id_year | home_team_avg_score | \ |
|--------|-------------------|----------------------|---------------------|---|
| id | | | | |
| 47179 | 2 2018 | 4 2018 | 112.8 | |
| 48751 | 2 2018 | 13 2018 | 112.8 | |
| 48739 | 23 2018 | 8 2018 | 117.9 | |
| 48740 | 30 2018 | 6 2018 | 116.4 | |
| 48746 | 26 2018 | 16 2018 | 114.9 | |
| ... | ... | ... | ... | |
| 473801 | 27 2021 | 8 2021 | 113.0 | |
| 473863 | 24 2021 | 4 2021 | 111.5 | |
| 473890 | 8 2021 | 4 2021 | 111.5 | |
| 473895 | 14 2021 | 27 2021 | 106.6 | |
| 473818 | 2 2021 | 17 2021 | 108.8 | |

| | visitor_team_avg_score | home_avg_score_diff | visitor_avg_score_diff |
|--------|------------------------|---------------------|------------------------|
| id | | | |
| 47179 | 108.3 | 3.620000 | -3.831707 |
| 48751 | 113.1 | 3.620000 | 0.581818 |
| 48739 | 108.2 | 8.725532 | -4.670213 |
| 48740 | 103.8 | 7.429268 | -8.419512 |
| 48746 | 105.4 | 5.129268 | -6.670732 |
| ... | ... | ... | ... |
| 473801 | 108.1 | 5.469231 | -0.170370 |
| 473863 | 113.6 | 4.622727 | 4.932258 |
| 473890 | 113.6 | 3.431034 | 4.932258 |
| 473895 | 107.6 | -0.912500 | -0.908000 |
| 473818 | 113.8 | 2.200000 | 5.013793 |

[49995 rows x 26 columns]

```
[ ]: g.to_csv("/content/drive/MyDrive/Machine Learning NBA /Final Data/
↳games_with_features.csv")
```

##Stats Data

```
[ ]: stats = pd.read_csv("/content/drive/MyDrive/Machine Learning NBA /Final Data/
↳stats_clean.csv", index_col="id")
```

Convert the date column to datetime format so the data can be sorted by date.

Convert the minutes column to timedelta so that they can be added together. (The rows of data are stats put up by a specific player in a specific game. These stats from these rows will be aggregated... grouping by team and game. This will result in the stats that the TEAM put up for that game, which will be useful because we are predicting which TEAM will win the game.)

```
[ ]: # Convert game date to datetime
stats["game.date"] = pd.to_datetime(stats["game.date"]).dt.tz_localize(None)

# Convert string to timedelta
stats["min"] = [pd.Timedelta(minutes=int(time[0]), seconds=int(time[1])) for
↳time in stats["min"].str.split(":").values]

# Create target variable & a "labels" dataframe
stats["winner"] = np.where(stats["game.home_team_score"].values > stats["game.
↳visitor_team_score"].values, 1,0)
labels = stats[["game.id", "game.date", "game.season", "winner"]].groupby("game.
↳id").first()
```

This code block aggregates the individual player statistics into team statistics

```
[ ]: # split into 2 dataframes so that grouping data by game id doesn't
# group players of opposing teams
home_stats = stats[stats["team.id"].eq(stats["game.home_team_id"])]
away_stats = stats[stats["team.id"].eq(stats["game.visitor_team_id"])]

# define how to aggregate statistics when grouping player stats
agg_map = {"ast": "sum",
           "blk": "sum",
           "dreb": "sum",
           "fg3_pct": "mean",
           "fg3a": "sum",
           "fg3m": "sum",
           "fg_pct": "mean",
           "fga": "sum",
           "fgm": "sum",
           "ft_pct": "mean",
           "fta": "sum",
           "ftm": "sum",
```

```

        "min": "sum",
        "oreb": "sum",
        "pf": "sum",
        "pts": "sum",
        "reb": "sum",
        "stl": "sum",
        "turnover": "sum",
        "game.id": "first",
        "game.date": "first",
        "game.season": "first",
        "game.home_team_id": "first",
        "game.home_team_score": "first",
        "game.visitor_team_id": "first",
        "game.visitor_team_score": "first",
        "player.id": "first",
        "team.id": "first",}

# aggregate player stats to team stats
home_games = home_stats.groupby("game.id").agg(agg_map)
away_games = away_stats.groupby("game.id").agg(agg_map)

# player stats were aggregated to team stats, so player id doesn't make sense
↳ anymore
# team id isn't needed either
home_games.drop(["player.id", "team.id"], axis=1, inplace=True)
away_games.drop(["player.id", "team.id"], axis=1, inplace=True)

```

To predict which team will win a game, the model will look at the average stats that a team put up in their 20 most recent games.* (called the rolling average from here on out)

Their 20 most recent home games if the game being predicted is a home game
 Their 20 most recent away games if the game being predicted is an away game

```

[ ]: # sort rows by team and date to prepare for the rolling average
home_games = home_games.sort_values(["game.home_team_id", "game.date"])
away_games = away_games.sort_values(["game.visitor_team_id", "game.date"])

```

```

[ ]: Fro# Only use the columns that refer to stats when calculating the rolling
↳ average
stats_cols =
↳ ["ast", "blk", "dreb", "fg3_pct", "fg3a", "fg3m", "fg_pct", "fga", "fgm", "ft_pct", "fta", "ftm", "oreb",
    "pf", "pts", "reb", "stl", "turnover"]

rolling_stats_home = pd.DataFrame()
for team in home_games["game.home_team_id"].unique():
    ↳ for each team

```



```

    games = home_games[home_games["game.home_team_id"].eq(team)] #
    ↪get only home games for that team
    games_stats = games[stats_cols] #
    ↪Only use the columns that refer to stats when calculating the rolling average
    rolling_stats_home = rolling_stats_home.append(games_stats.
    ↪rolling(window=20).mean()) # calculate the rolling average (20 game average)
    rolling_stats_home = rolling_stats_home.shift() #
    ↪make the rolling average not include the current game
rolling_stats_home.dropna(inplace=True)

rolling_stats_away = pd.DataFrame()
for team in away_games["game.visitor_team_id"].unique():
    games = away_games[away_games["game.visitor_team_id"].eq(team)]
    games_stats = games[stats_cols]
    rolling_stats_away = rolling_stats_away.append(games_stats.
    ↪rolling(window=20).mean())
    rolling_stats_away = rolling_stats_away.shift() # make the rolling average
    ↪not include the current game
rolling_stats_away.dropna(inplace=True)

```

```

[ ]: # Rename columns in the home and away dataframes
rolling_stats_home.columns = ["home_" + col_name for col_name in
    ↪rolling_stats_home.columns]
rolling_stats_away.columns = ["away_" + col_name for col_name in
    ↪rolling_stats_away.columns]
# Now that columns are renamed they can be merged into the same dataframe
rolling_stats = pd.merge(rolling_stats_home, rolling_stats_away, on="game.id")

```

```

[ ]: # NOTE: adding columns to the start of the dataframe will mess up this code
# adding columns to the end will not
rolling_stats_diff = pd.DataFrame(index=rolling_stats.index)
home_cols = rolling_stats.columns[0:18]
away_cols = rolling_stats.columns[18:36]

```

```

[ ]: # Create a dataframe where each column is equal to the home value minus the
    ↪away value
##### This should be altered so that percentages are divided by each other
    ↪instead of subtracted
for col in zip(stats_cols, home_cols, away_cols):
    rolling_stats_diff[col[0]] = rolling_stats[col[1]] - rolling_stats[col[2]]

```

```

[ ]: # Rename the columns
rolling_stats_diff.columns = ["diff_" + col_name for col_name in
    ↪rolling_stats_away.columns]

```

```
# putting it all together in 1 dataframe
rolling_stats = pd.merge(rolling_stats, rolling_stats_diff, on="game.id") #_
    ↳adding diff to the rolling stats
rolling_stats = pd.merge(labels, rolling_stats, on="game.id") #_
    ↳adding labels
```

```
[ ]: rolling_stats
```

```
[ ]:
      game.date  game.season  winner  home_ast  home_blk  home_dreb  \
game.id
1      2018-10-16          2018        1      23.40      5.35      36.05
2      2018-10-16          2018        1      31.55      7.60      34.45
3      2018-10-17          2018        0      21.35      5.15      35.85
4      2018-10-17          2018        1      23.15      3.40      33.55
5      2018-10-17          2018        1      22.55      4.25      33.20
...      ...      ...      ...      ...      ...      ...
857658  2022-11-28          2022        0      25.85      4.40      32.05
857659  2022-11-28          2022        0      24.95      4.80      34.45
857660  2022-11-28          2022        0      22.90      6.00      33.20
857661  2022-11-28          2022        1      25.35      3.95      32.25
857662  2022-11-28          2022        1      29.55      3.05      34.65

      home_fg3_pct  home_fg3a  home_fg3m  home_fg_pct  home_fga  home_fgm  \
game.id
1           0.283508      32.10      11.60      0.407498      85.30      38.10
2           0.186014      27.60      11.35      0.511330      83.35      43.50
3           0.211261      26.70       9.45      0.415333      87.05      38.70
4           0.239457      28.45      11.50      0.431416      84.55      38.95
5           0.291932      25.80      10.00      0.453450      84.70      41.25
...      ...      ...      ...      ...      ...      ...
857658      20.284341      33.80      12.50      32.655897      84.20      40.30
857659      13.977462      39.25      14.65      28.952156      85.15      41.65
857660      24.626431      33.05      10.85      42.398149      86.40      40.20
857661      20.856653      32.15      10.60      42.469124      87.40      43.05
857662      33.899859      37.75      14.35      45.480293      87.85      42.80

      home_ft_pct  home_fta  home_ftm  home_oreb  home_pf  home_pts  \
game.id
1           0.392525      21.45      16.35         8.80      20.20      104.15
2           0.385137      20.15      16.35         8.60      18.85      114.70
3           0.515664      27.40      21.25         9.75      17.05      108.10
4           0.400628      20.05      14.95         9.30      17.85      104.35
5           0.432828      20.20      15.30         9.40      18.80      107.80
...      ...      ...      ...      ...      ...      ...
857658      39.248915      25.65      20.35         6.95      19.10      113.45
857659      29.378268      27.50      20.55        10.90      20.70      118.50
857660      43.942857      22.15      16.05         9.90      19.20      107.30
```

| | | | | | | |
|--------|-----------|-------|-------|-------|-------|--------|
| 857661 | 44.212607 | 22.00 | 17.35 | 11.85 | 20.40 | 114.05 |
| 857662 | 46.114460 | 21.25 | 17.05 | 8.75 | 19.65 | 117.00 |

| | home_reb | home_stl | home_turnover | away_ast | away_blk | away_dreb | \ |
|---------|----------|----------|---------------|----------|----------|-----------|---|
| game.id | | | | | | | |
| 1 | 44.85 | 7.05 | 12.90 | 25.60 | 4.75 | 34.20 | |
| 2 | 43.05 | 7.65 | 15.25 | 22.55 | 5.60 | 32.80 | |
| 3 | 45.60 | 6.45 | 11.35 | 21.30 | 4.65 | 32.00 | |
| 4 | 42.85 | 7.90 | 13.30 | 21.60 | 4.10 | 35.45 | |
| 5 | 42.60 | 7.90 | 13.00 | 20.55 | 4.25 | 30.00 | |
| ... | ... | ... | ... | ... | ... | ... | |
| 857658 | 39.00 | 7.15 | 13.65 | 27.15 | 3.40 | 32.40 | |
| 857659 | 45.35 | 7.00 | 13.60 | 24.30 | 4.80 | 33.65 | |
| 857660 | 43.10 | 7.85 | 12.45 | 24.15 | 5.25 | 33.80 | |
| 857661 | 44.10 | 8.10 | 12.95 | 23.00 | 4.00 | 35.10 | |
| 857662 | 43.40 | 6.80 | 11.85 | 22.50 | 4.00 | 32.00 | |

| | away_fg3_pct | away_fg3a | away_fg3m | away_fg_pct | away_fga | away_fgm | \ |
|---------|--------------|-----------|-----------|-------------|----------|----------|---|
| game.id | | | | | | | |
| 1 | 0.242414 | 29.45 | 10.85 | 0.451425 | 83.50 | 39.20 | |
| 2 | 0.235524 | 31.20 | 11.20 | 0.406164 | 90.10 | 41.50 | |
| 3 | 0.226351 | 25.30 | 8.60 | 0.398101 | 82.05 | 37.80 | |
| 4 | 0.249849 | 31.60 | 10.25 | 0.425655 | 85.15 | 37.45 | |
| 5 | 0.276580 | 25.85 | 9.40 | 0.433132 | 81.05 | 35.85 | |
| ... | ... | ... | ... | ... | ... | ... | |
| 857658 | 23.005717 | 29.45 | 10.50 | 42.006387 | 88.85 | 43.20 | |
| 857659 | 26.124760 | 31.30 | 11.75 | 44.094360 | 87.80 | 41.45 | |
| 857660 | 26.426683 | 35.45 | 12.15 | 39.111668 | 88.60 | 38.70 | |
| 857661 | 26.378944 | 36.95 | 12.50 | 38.601197 | 89.50 | 39.10 | |
| 857662 | 28.075487 | 39.15 | 13.50 | 42.805649 | 87.50 | 39.50 | |

| | away_ft_pct | away_fta | away_ftm | away_oreb | away_pf | away_pts | \ |
|---------|-------------|----------|----------|-----------|---------|----------|---|
| game.id | | | | | | | |
| 1 | 0.477155 | 24.25 | 18.20 | 11.10 | 21.15 | 107.45 | |
| 2 | 0.353259 | 22.60 | 16.20 | 12.50 | 20.05 | 110.40 | |
| 3 | 0.433145 | 23.85 | 18.70 | 9.50 | 21.55 | 102.90 | |
| 4 | 0.506771 | 25.15 | 19.70 | 9.15 | 20.65 | 104.85 | |
| 5 | 0.429151 | 20.30 | 16.55 | 9.60 | 23.85 | 97.65 | |
| ... | ... | ... | ... | ... | ... | ... | |
| 857658 | 39.141589 | 19.15 | 15.50 | 9.50 | 22.40 | 112.40 | |
| 857659 | 40.042889 | 20.40 | 16.50 | 8.60 | 19.00 | 111.15 | |
| 857660 | 41.549211 | 23.35 | 17.60 | 11.75 | 20.00 | 107.15 | |
| 857661 | 37.552899 | 20.45 | 15.00 | 10.70 | 18.95 | 105.70 | |
| 857662 | 46.193101 | 25.35 | 18.45 | 10.50 | 20.15 | 110.95 | |

| | away_reb | away_stl | away_turnover | diff_away_ast | diff_away_blk | \ |
|---------|----------|----------|---------------|---------------|---------------|---|
| game.id | | | | | | |

| | | | | | |
|--------|-------|------|-------|-------|-------|
| 1 | 45.30 | 8.00 | 15.55 | -2.20 | 0.60 |
| 2 | 45.30 | 7.80 | 12.10 | 9.00 | 2.00 |
| 3 | 41.50 | 8.80 | 14.30 | 0.05 | 0.50 |
| 4 | 44.60 | 7.30 | 16.60 | 1.55 | -0.70 |
| 5 | 39.60 | 6.75 | 14.50 | 2.00 | 0.00 |
| ... | ... | ... | ... | ... | ... |
| 857658 | 41.90 | 7.95 | 13.05 | -1.30 | 1.00 |
| 857659 | 42.25 | 6.65 | 12.20 | 0.65 | 0.00 |
| 857660 | 45.55 | 6.80 | 14.25 | -1.25 | 0.75 |
| 857661 | 45.80 | 7.95 | 13.30 | 2.35 | -0.05 |
| 857662 | 42.50 | 7.65 | 15.40 | 7.05 | -0.95 |

| | diff_away_dreb | diff_away_fg3_pct | diff_away_fg3a | diff_away_fg3m | \ |
|---------|----------------|-------------------|----------------|----------------|---|
| game.id | | | | | |
| 1 | 1.85 | 0.041094 | 2.65 | 0.75 | |
| 2 | 1.65 | -0.049510 | -3.60 | 0.15 | |
| 3 | 3.85 | -0.015089 | 1.40 | 0.85 | |
| 4 | -1.90 | -0.010392 | -3.15 | 1.25 | |
| 5 | 3.20 | 0.015352 | -0.05 | 0.60 | |
| ... | ... | ... | ... | ... | |
| 857658 | -0.35 | -2.721376 | 4.35 | 2.00 | |
| 857659 | 0.80 | -12.147298 | 7.95 | 2.90 | |
| 857660 | -0.60 | -1.800252 | -2.40 | -1.30 | |
| 857661 | -2.85 | -5.522291 | -4.80 | -1.90 | |
| 857662 | 2.65 | 5.824371 | -1.40 | 0.85 | |

| | diff_away_fg_pct | diff_away_fga | diff_away_fgm | diff_away_ft_pct | \ |
|---------|------------------|---------------|---------------|------------------|---|
| game.id | | | | | |
| 1 | -0.043927 | 1.80 | -1.10 | -0.084629 | |
| 2 | 0.105166 | -6.75 | 2.00 | 0.031878 | |
| 3 | 0.017233 | 5.00 | 0.90 | 0.082519 | |
| 4 | 0.005761 | -0.60 | 1.50 | -0.106143 | |
| 5 | 0.020318 | 3.65 | 5.40 | 0.003677 | |
| ... | ... | ... | ... | ... | |
| 857658 | -9.350490 | -4.65 | -2.90 | 0.107326 | |
| 857659 | -15.142204 | -2.65 | 0.20 | -10.664621 | |
| 857660 | 3.286482 | -2.20 | 1.50 | 2.393646 | |
| 857661 | 3.867926 | -2.10 | 3.95 | 6.659708 | |
| 857662 | 2.674644 | 0.35 | 3.30 | -0.078641 | |

| | diff_away_fta | diff_away_ftm | diff_away_oreb | diff_away_pf | \ |
|---------|---------------|---------------|----------------|--------------|---|
| game.id | | | | | |
| 1 | -2.80 | -1.85 | -2.30 | -0.95 | |
| 2 | -2.45 | 0.15 | -3.90 | -1.20 | |
| 3 | 3.55 | 2.55 | 0.25 | -4.50 | |
| 4 | -5.10 | -4.75 | 0.15 | -2.80 | |
| 5 | -0.10 | -1.25 | -0.20 | -5.05 | |

| | | | | |
|--------|-------|-------|-------|-------|
| ... | ... | ... | ... | ... |
| 857658 | 6.50 | 4.85 | -2.55 | -3.30 |
| 857659 | 7.10 | 4.05 | 2.30 | 1.70 |
| 857660 | -1.20 | -1.55 | -1.85 | -0.80 |
| 857661 | 1.55 | 2.35 | 1.15 | 1.45 |
| 857662 | -4.10 | -1.40 | -1.75 | -0.50 |

| | diff_away_pts | diff_away_reb | diff_away_stl | diff_away_turnover |
|---------|---------------|---------------|---------------|--------------------|
| game.id | | | | |
| 1 | -3.30 | -0.45 | -0.95 | -2.65 |
| 2 | 4.30 | -2.25 | -0.15 | 3.15 |
| 3 | 5.20 | 4.10 | -2.35 | -2.95 |
| 4 | -0.50 | -1.75 | 0.60 | -3.30 |
| 5 | 10.15 | 3.00 | 1.15 | -1.50 |
| ... | ... | ... | ... | ... |
| 857658 | 1.05 | -2.90 | -0.80 | 0.60 |
| 857659 | 7.35 | 3.10 | 0.35 | 1.40 |
| 857660 | 0.15 | -2.45 | 1.05 | -1.80 |
| 857661 | 8.35 | -1.70 | 0.15 | -0.35 |
| 857662 | 6.05 | 0.90 | -0.85 | -3.55 |

[46673 rows x 57 columns]

```
[ ]: rolling_stats.to_csv("/content/drive/MyDrive/Machine Learning NBA /Final Data/
↳stats_feats.csv")
```