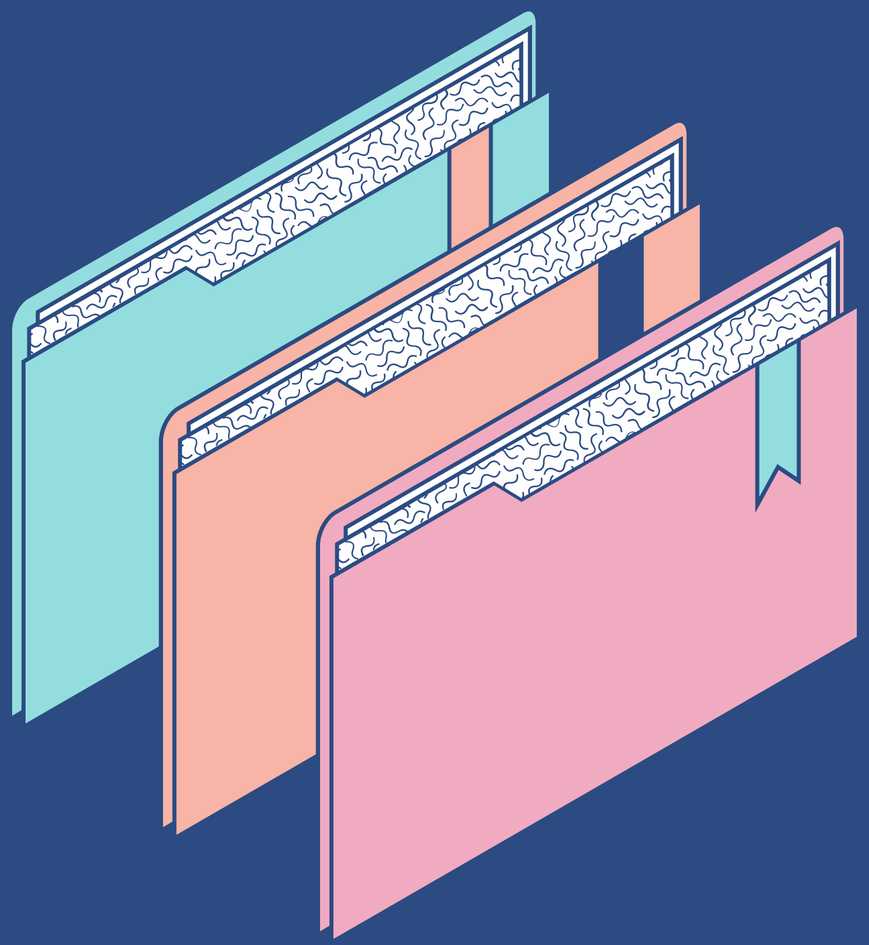
Practice Case NBA Player Analysis 2017

Gempar T. Asmara



Content List

- Data Understanding
- Data Pre-processing
- Data Analytics and Insight

Data Understanding

Background

Use dataset NBA Player Stats from 1950 until 2017

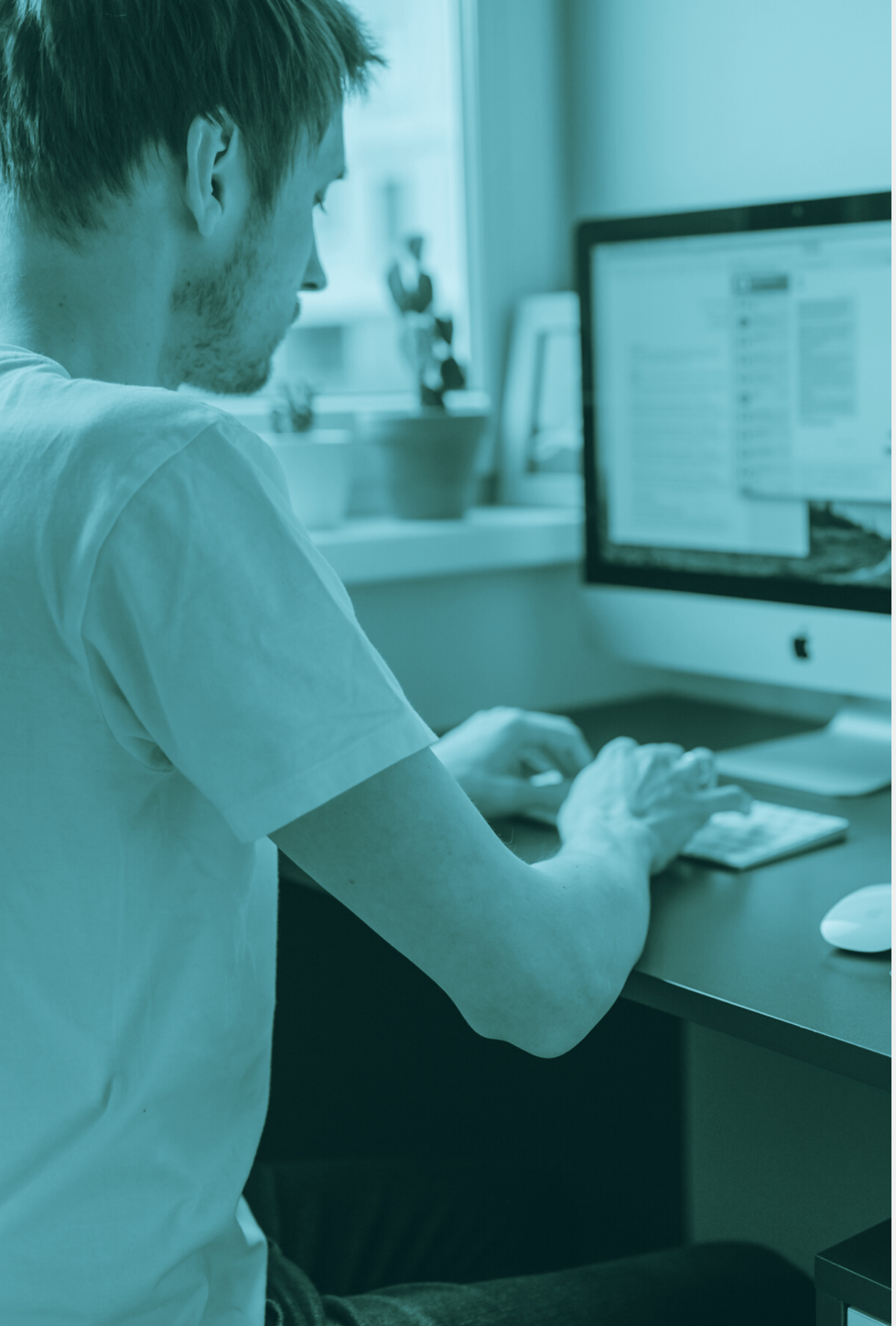
Objectives

We need only last stat (which means 2017) for get some insight

Condition of data

```
dtypes: float64(49), object(3)
```

There are information about the count of non- null (non missing value) and also the data type.



Data Pre-Processing

CHECK DUPLICATE VALUES

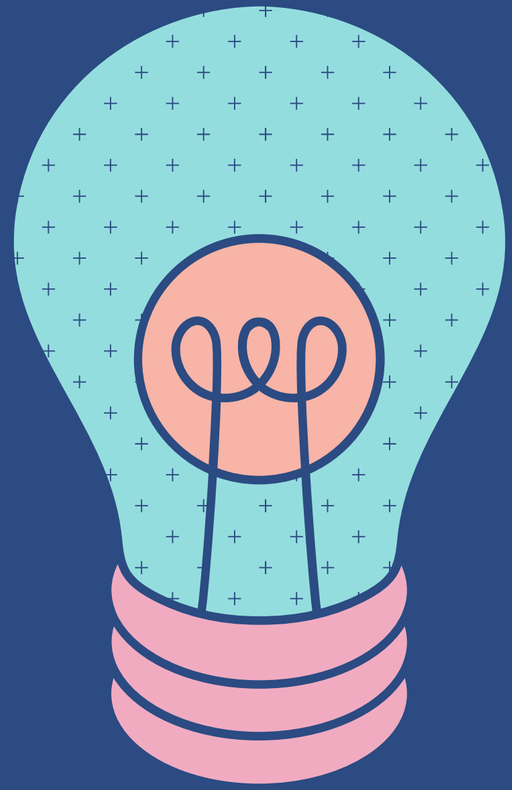
CHECK MISSING VALUES

CHECK NULL ROW/COLUMN

DATA CLEANSING

Data Analysis and Insight

1. Find the Youngest and Oldest Player for each Team
2. Find the most minutes played in each position
3. The highest average percentage score by team for total **rebound**, total **assist**, total **steal**, and block.
4. Assign who the best NBA player 2017
5. Assign who the best team 2017



Young Players

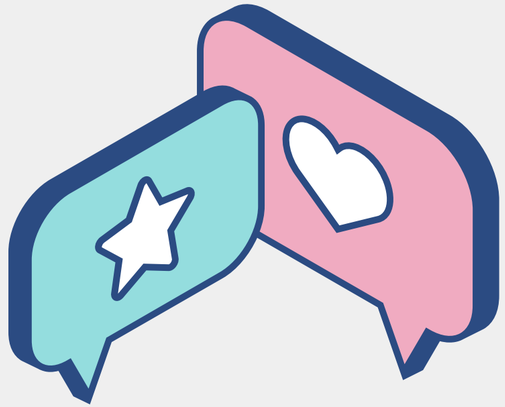
Tm	Player	Age
ATL	DeAndre' Bembry	22.0
BOS	Jaylen Brown	20.0
BRK	Isaiah Whitehead	21.0
CHI	Bobby Portis	21.0
CHO	Christian Wood	21.0
CLE	Kay Felder	21.0
DAL	Ben Bentil	21.0
DEN	Jamal Murray	19.0
DET	Henry Ellenson	20.0
GSW	Kevon Looney	20.0
HOU	Chinanu Onuaku	20.0
IND	Myles Turner	20.0
LAC	Diamond Stone	19.0
LAL	Brandon Ingram	19.0

Tm	Player	Age
MEM	Wade Baldwin	20.0
MIA	Justise Winslow	20.0
MIL	Thon Maker	19.0
MIN	Tyus Jones	20.0
NOP	Cheick Diallo	20.0
NYK	Kristaps Porzingis	21.0
OKC	Domantas Sabonis	20.0
ORL	Stephen Zimmerman	20.0
PHI	Jahlil Okafor	21.0
PHO	Dragan Bender	19.0
POR	Noah Vonleh	21.0
SAC	Georgios Papagiannis	19.0
SAS	Dejounte Murray	20.0
TOR	Jakob Poeltl	21.0
UTA	Trey Lyles	21.0
WAS	Chris McCullough	21.0

Oldest Players

Tm	Player	Age
ATL	Mike Dunleavy	36.0
BOS	Gerald Green	31.0
BRK	Luis Scola	36.0
CHI	Dwyane Wade	35.0
CHO	Brian Roberts	31.0
CLE	Chris Andersen	38.0
DAL	Dirk Nowitzki	38.0
DEN	Mike Miller	36.0
DET	Beno Udrih	34.0
GSW	Matt Barnes	36.0
HOU	Nene Hilario	34.0
IND	Aaron Brooks	32.0
LAC	Paul Pierce	39.0
LAL	Metta World	37.0
MEM	Vince Carter	40.0

Tm	Player	Age
MIA	Udonis Haslem	36.0
MIL	Jason Terry	39.0
MIN	John Lucas	34.0
NOP	Jarrett Jack	33.0
NYK	Carmelo Anthony	32.0
OKC	Nick Collison	36.0
ORL	C.J. Watson	32.0
PHI	Tiago Splitter	32.0
PHO	Tyson Chandler	34.0
POR	Evan Turner	28.0
SAC	Arron Afflalo	31.0
SAS	Manu Ginobili	39.0
TOR	P.J. Tucker	31.0
UTA	Joe Johnson	35.0
WAS	Marcin Gortat	32.0



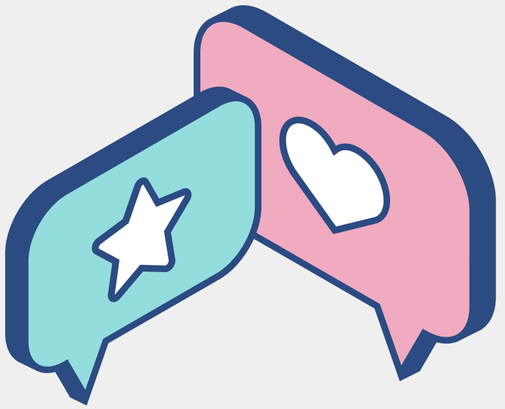
Youngest and Oldest

To get the values, I selected the only equired columns, Tm, Player, and Age.

After that, I group the data by 'Tm' and find the smallest and largest values in the data by using transform() function. Then, I sort the results by Tm.

Most Minutes Play each Position

	Pos	Player	MP
0	SF	Andrew Wiggins	3048.0
1	C	Karl-Anthony Towns	3030.0
2	PG	James Harden	2947.0
3	PF	Harrison Barnes	2803.0
4	SG	C.J. McCollum	2796.0



The calculation is the same as the previous case, the difference is in this case I use the **Post**, **Player** and **MP** columns and grouped according to position (Post)

Highest Average Score

TOTAL REBOUND %

Tm	TRB%
WAS	12.735294

TOTAL ASSIST %

Tm	AST%
DEN	15.723529

TOTAL STEAL %

Tm	STL%
MIN	2.413333

TOTAL BLOCK%

Tm	BLK%
MIL	2.741176

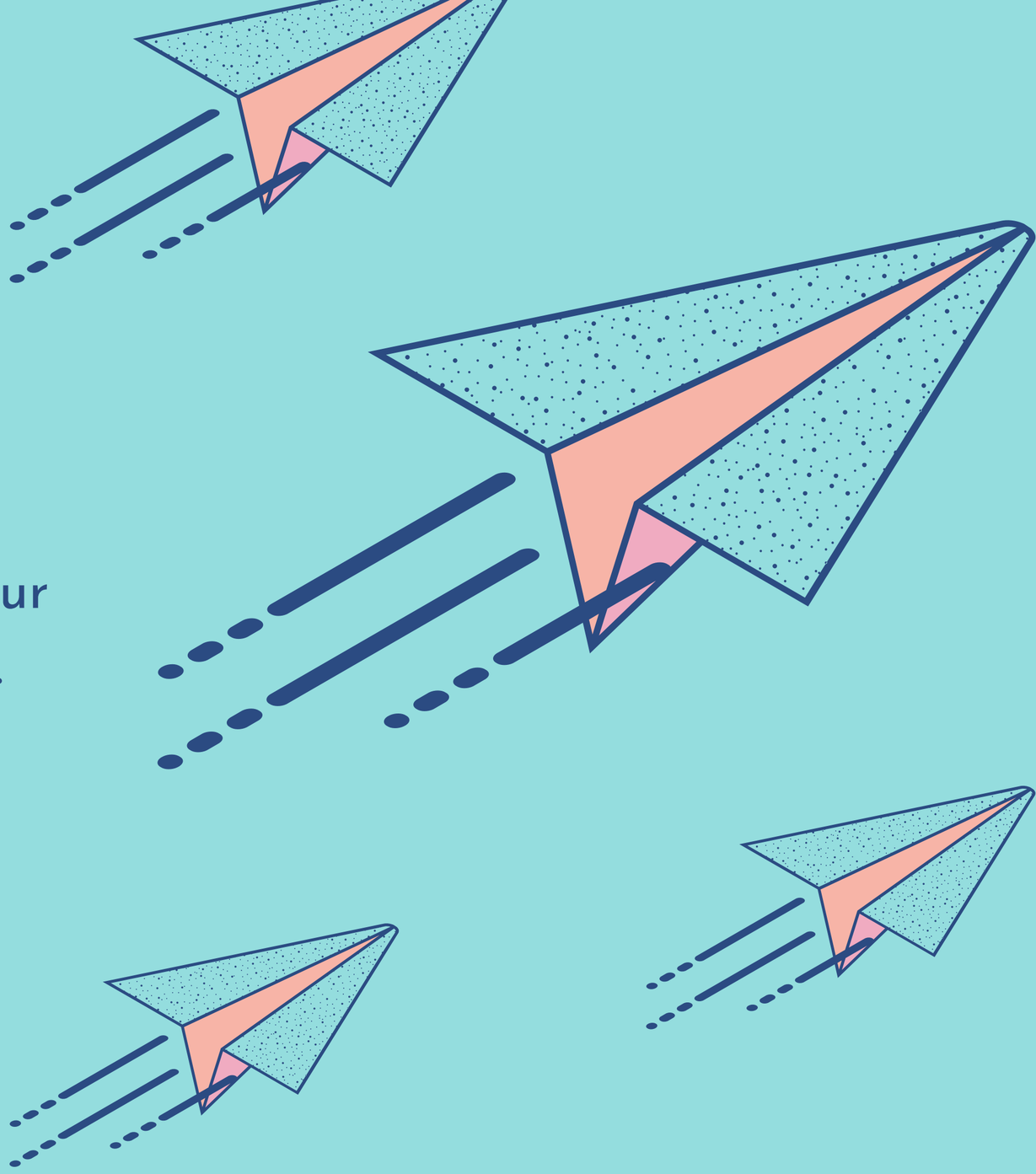
To get that value, I select the only required column from the 4 parameters above, then sum and group by team. After that descending ther data to find the highest value.

Best Player

Russell Westbrook	2831.576

To find this name, I use column from Four Factor by Basketball reference website.

Then I sum all value and sort until we have the highest score.



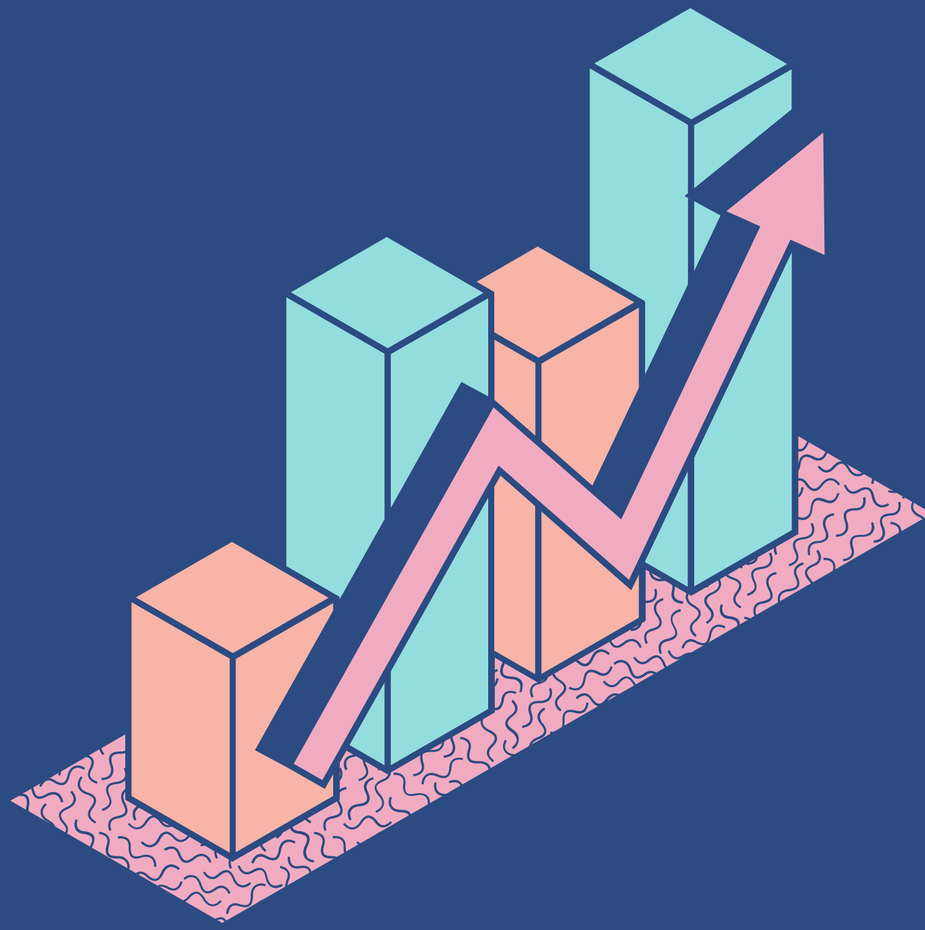
Source : Basketball reference website.

Best Team

Tm	Total
GSW	1575.9278

To find this team, I sum all score column with lambda function `df['Total'] = df.iloc[:,5:].apply(lambda x: x.mean(),axis=1)` to make a new column as **Total**.

Then I group by team and sort until have the highest **Total** score.



Thank you!

Software

SQL *

JUPITER NOTEBOOK *

PYTHON *

Library : Numpy, Pandas, Scikit-learn

BI TOOLS *

Power BI, Tableau, Redash

“We need technology in every classroom and in every student and teacher’s hand, because it is the pen and paper of our time, and it is the lens through which we experience much of our world."

DAVID WARLICK