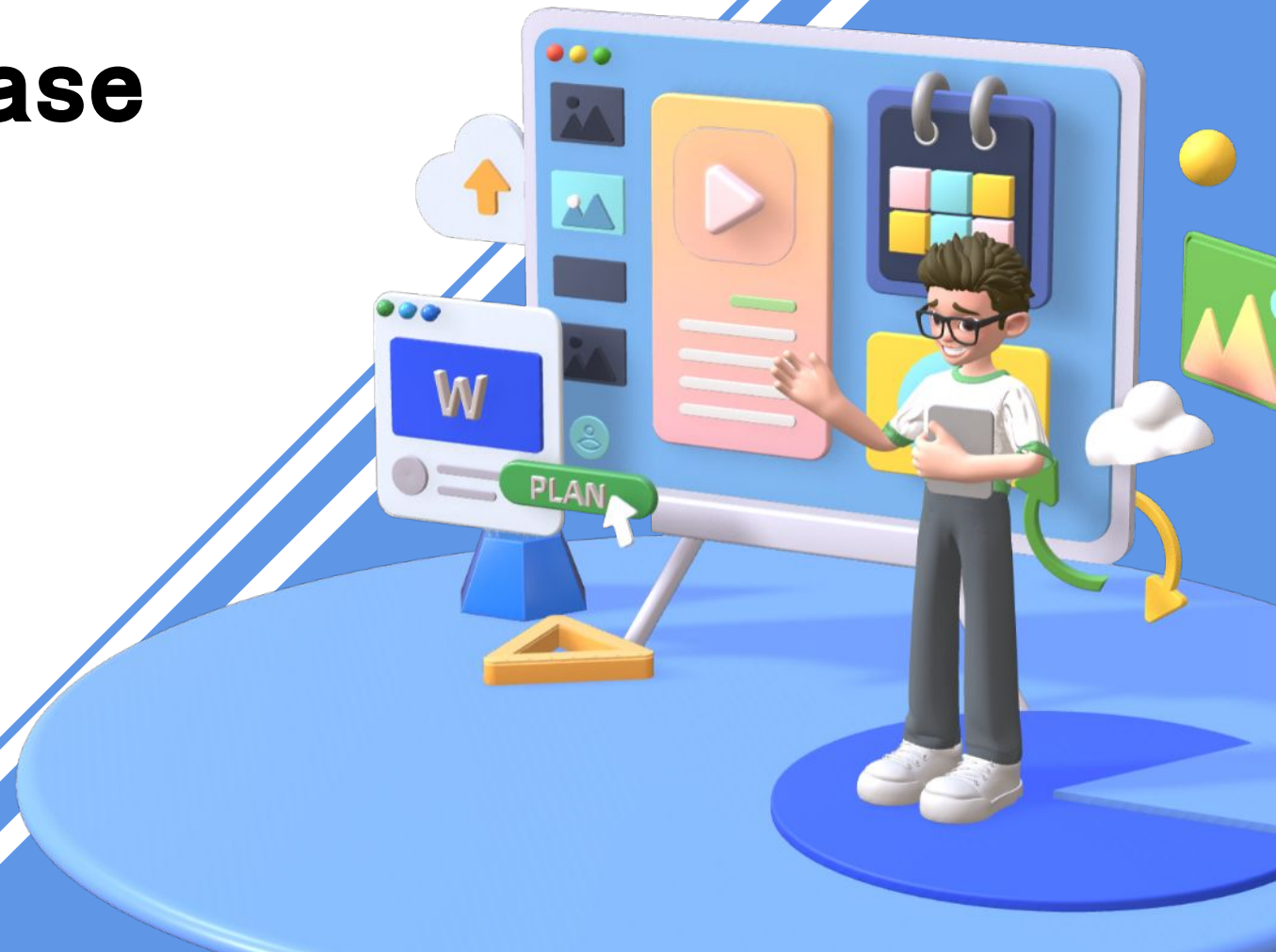
Cleaning and Showcase

what decides NBA players' salary

Ezzedine Ben Hadj Yahya, Edbert Faustine, Xin Li



01 Cleaning

Cleaning

Drop salary rows that don't have a player uuid

2019 - 2020:

before: 515 samples

after: 333 samples

2020 - 2021:

before: 578 samples

after: 367 samples

2021_2022:

before: 653 samples

after: 371 samples

2022_2023:

before: 574 sampels

after: 328 samples

Drop stats rows that don't have a player uuid or team uuid

2019 - 2020:

before: 216 samples

after: 203 samples

2020 - 2021:

before: 239 samples

after: 182 samples

2021_2022:

before: 217 samples

after: 165 samples

2022_2023:

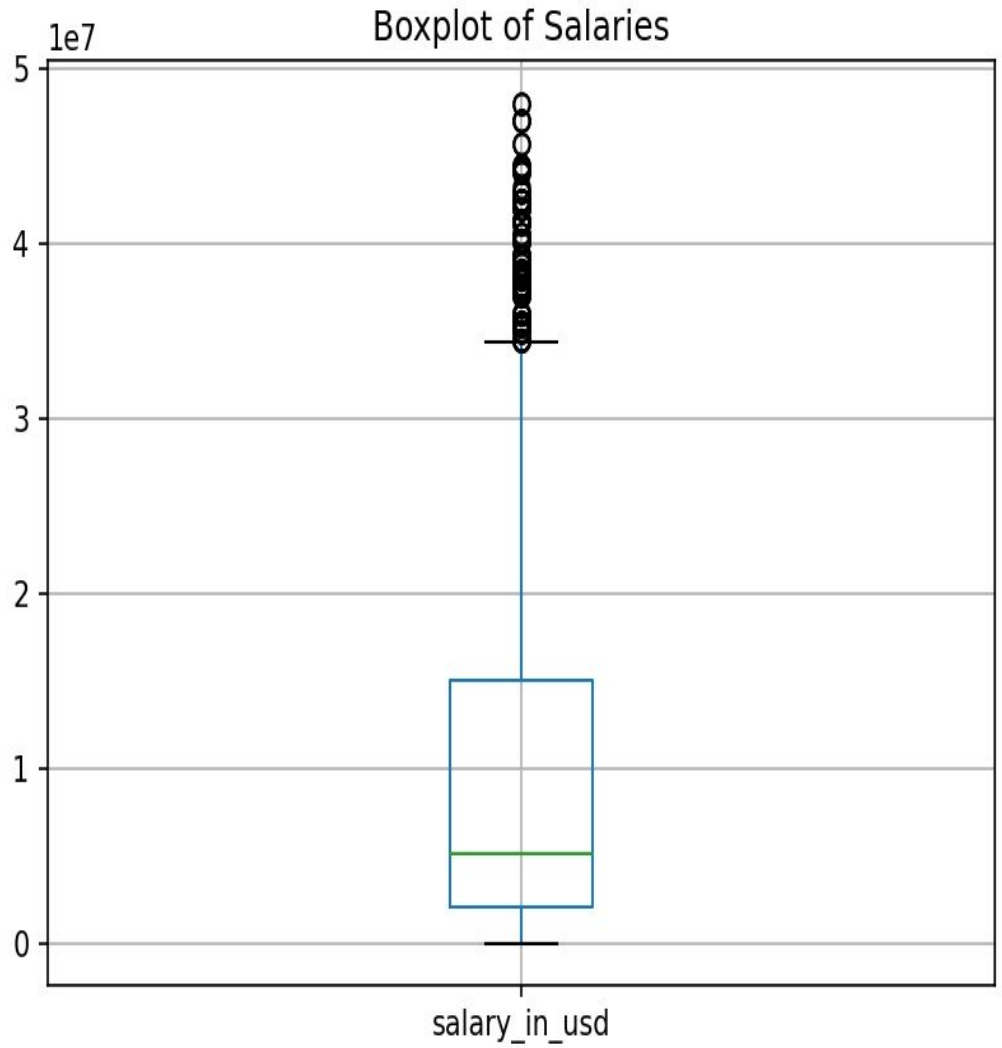
before: 216 samples

after: 163 samples

02 Find reletive values

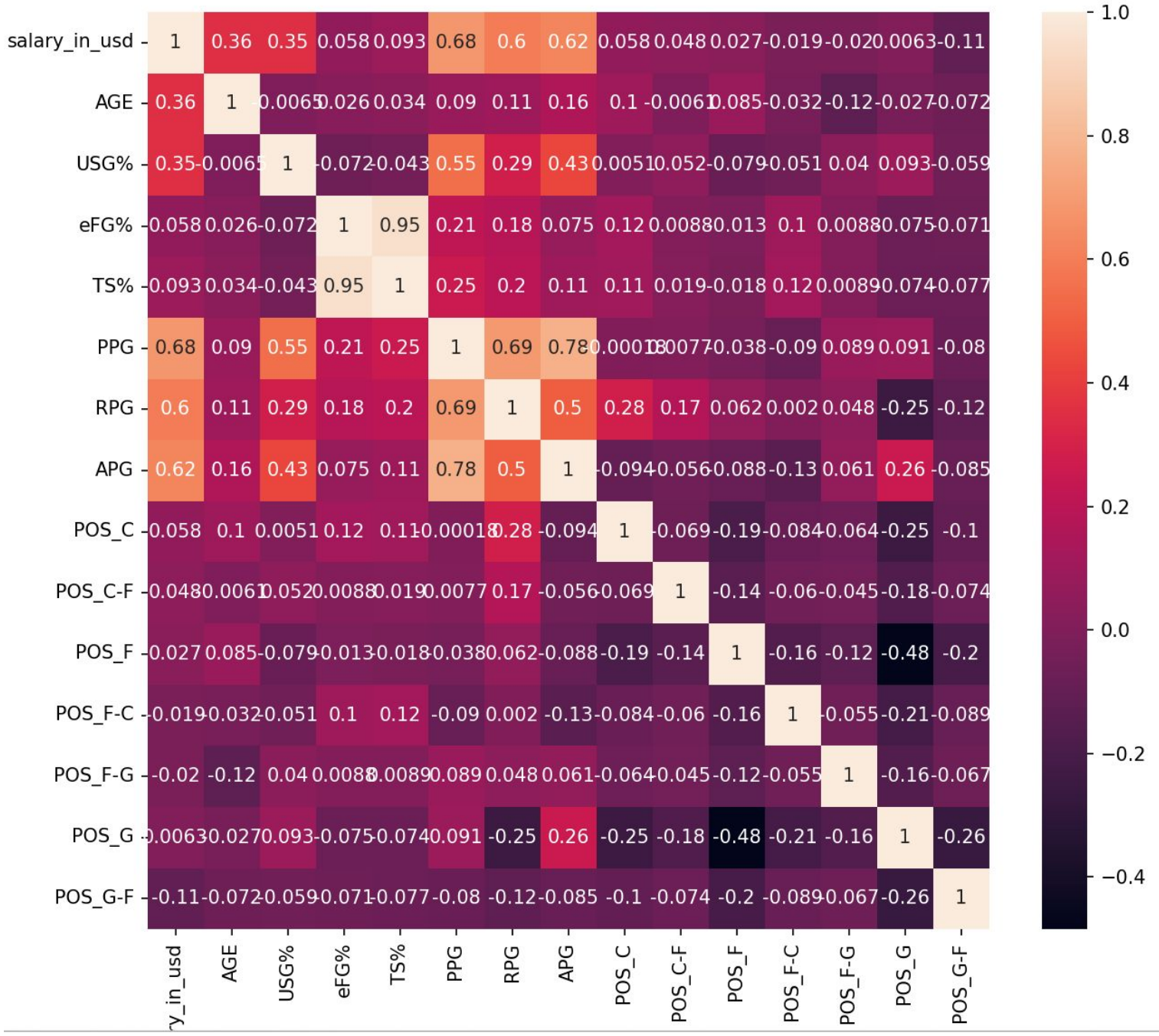
General Looking

- The median salary is approximately $5 million
- The middle 50% of the data is approximately from $2 to $15 million.
- Some people have extremely high salaries,



Find relative Keys

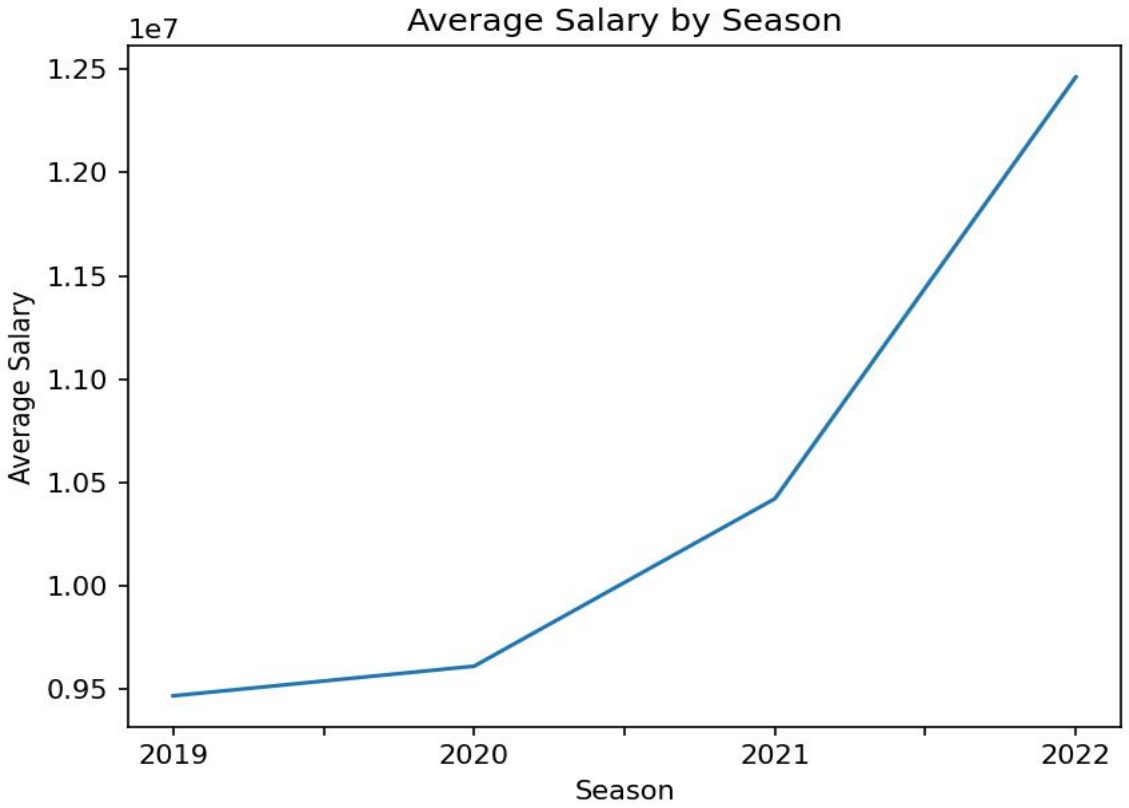
We have identified 5 keys with correlation coefficients greater than 0.15 in relation to salary: AGE, USG%, PPG, RPG, APG.

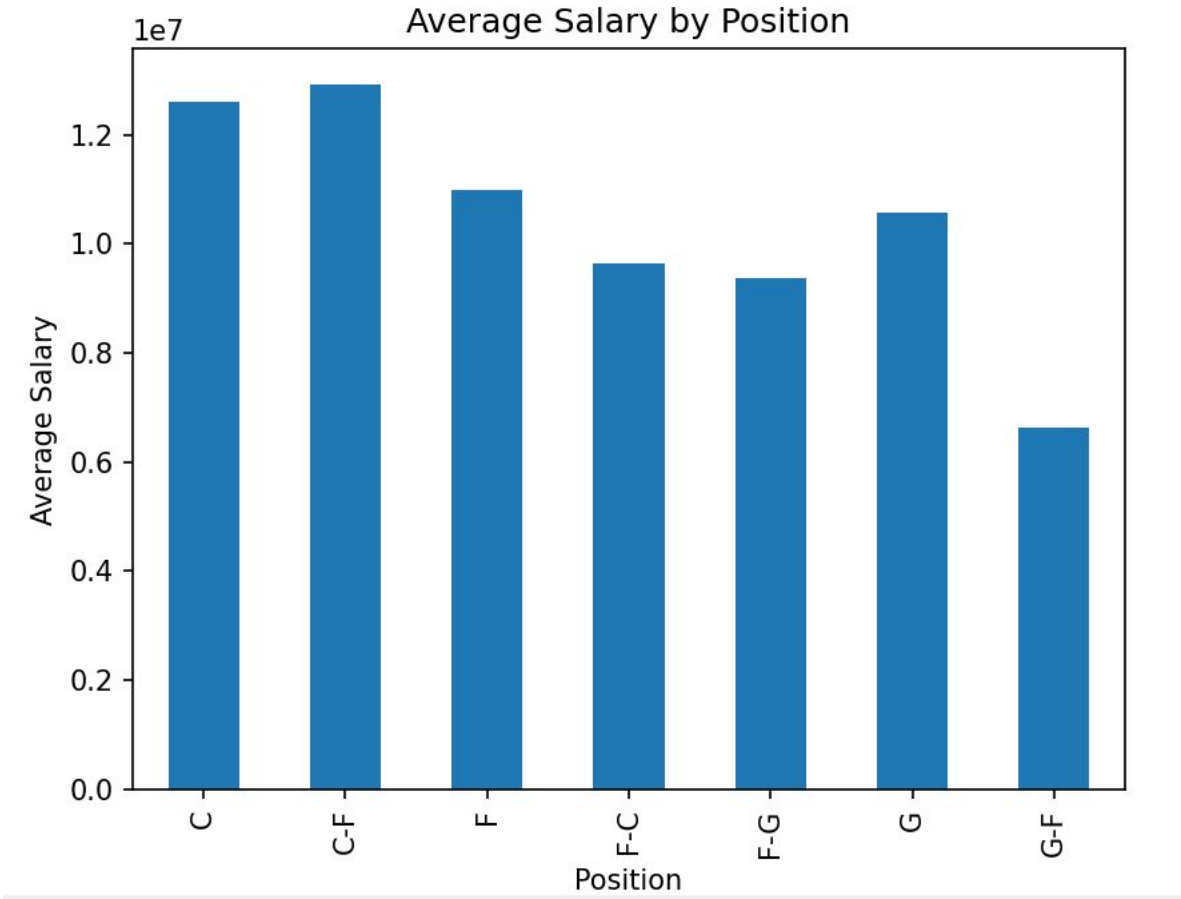


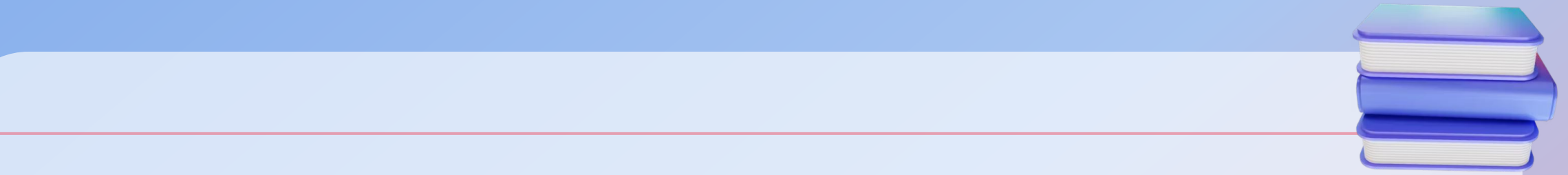
Find relative Keys

1、Different positions lead to varying salaries.

2、Salaries experience an exponential growth each season.







03

Guess the formula and obtain coefficients.

A Difficult

The five keys we initially found have relatively low correlations with each other (except between USG% and PPG). However, it is evident that both the season and age have synchronized growth and mutually influence each other. So, what should we do next?

Solution

Since salary, age, and season do not appear to have a linear relationship, we can try fitting them using other formulas. Do you remember that we previously discovered a relationship between season and salary that resembled an exponential function?

Guess Fromular:

position * (b1 * ppg + b2 * rpg + b3 * apg + b4 * age + c) * k ** (season - 2018)

Reach formular

Finally we got:

F ,G_F ,G ,C ,C_F ,FG ,FC

408.08, 309.39, 355.07, 363.98, 427.97, 264.02, 450.27

Position*(1080 * ppg + 1379 * rpg + 2542 * apg + 1367 * age -36328) * 1.1 ^ (season - 2018)

For example:

If a forward position player get a more ppg,he will get more 408.08 * 1080 = 440,726 dollors a year.

If he becomes elder 1 year, he will get more 408.08*1367 = 557,845 dollars a year.

And every year his salary will increase 10%.

We also calculate correlation between salary and number of championships

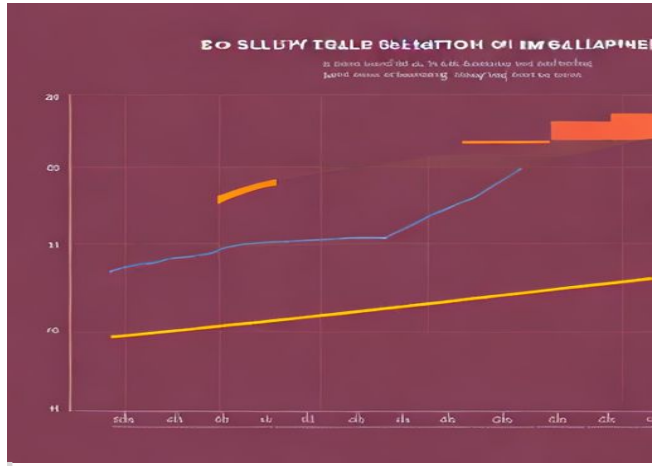
```
# Now count the number of championships for each team.

team_championships = champs['champion_uuid'].value_counts()

# Merge these two dataframes

team_data = pd.merge(team_salaries, team_championships, left_index=True, right_index=True)

# Now calculate correlation between salary and number of championships

correlation = team_data.corr()

print(correlation)
```

And we found the correlation is 0.43.

Conclusions

Salary is correlated with AGE, USG%, PPG, RPG, APG, season. Teams that already won a title tend to pay more.

Players in different positions have varying salary increases based on their performance. Forwards/centers (F-C and C-F) experience the highest increases, while guards/forwards (G-F) and guards (G) have the lowest increases. The difference between them can be as much as 50%.



Regardless of performance improvement, salary tends to increase with age. Additionally, players' overall income increases by 10% each year.

THANKS