Analyze the value of NBA players and factor related

CMSE 202

Cong Fu

Questions

In NBA games, different players have different data and salaries.

My topic is to analyze what factors are most relevant to their salary.

It can also be said that what factors have the greatest impact on salary?

Computational techniques and Packages

Matplotlib - Visualizing data

Pandas - load the data frame

Seaborn - Provide more advanced and concise functions, and check the correlation by sns.heatmap

Statsmodels - show OLS result

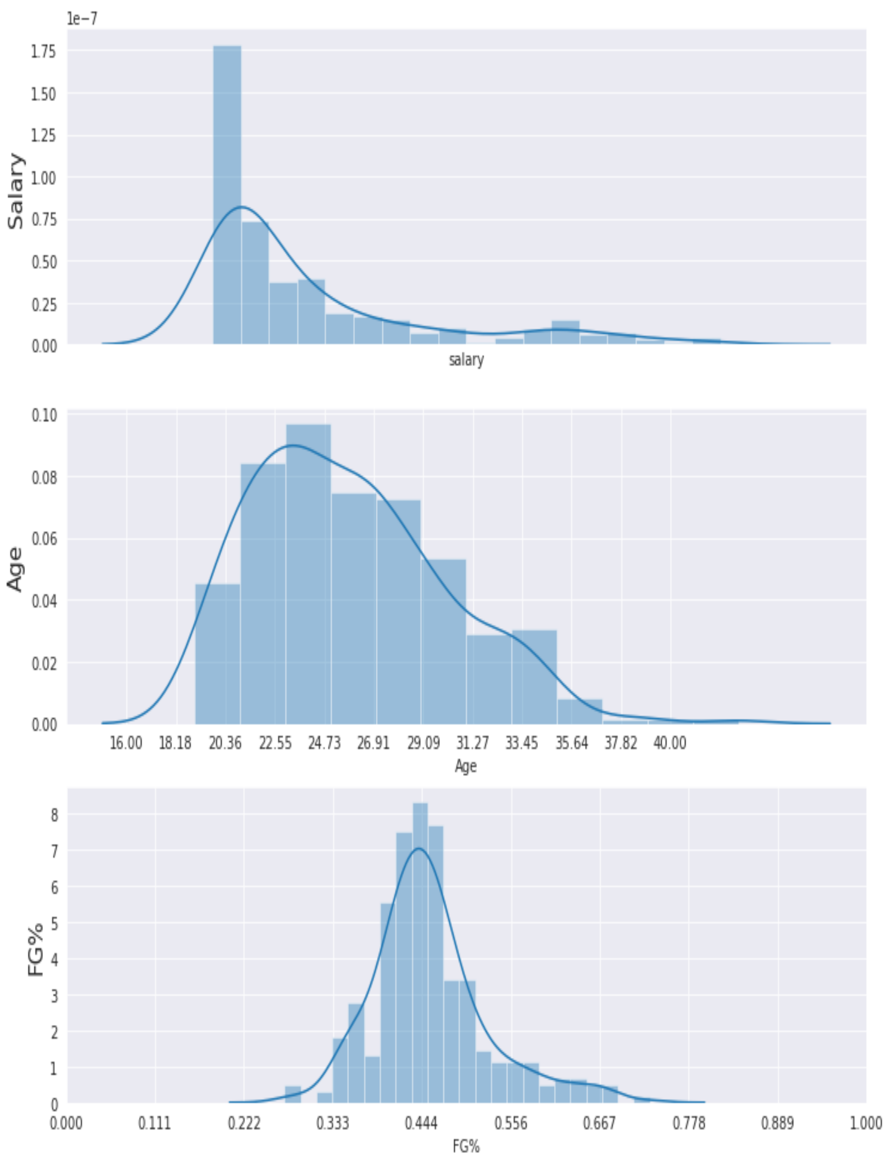
NumPy - get the mean and median

Overview NBA salary, age and FG%

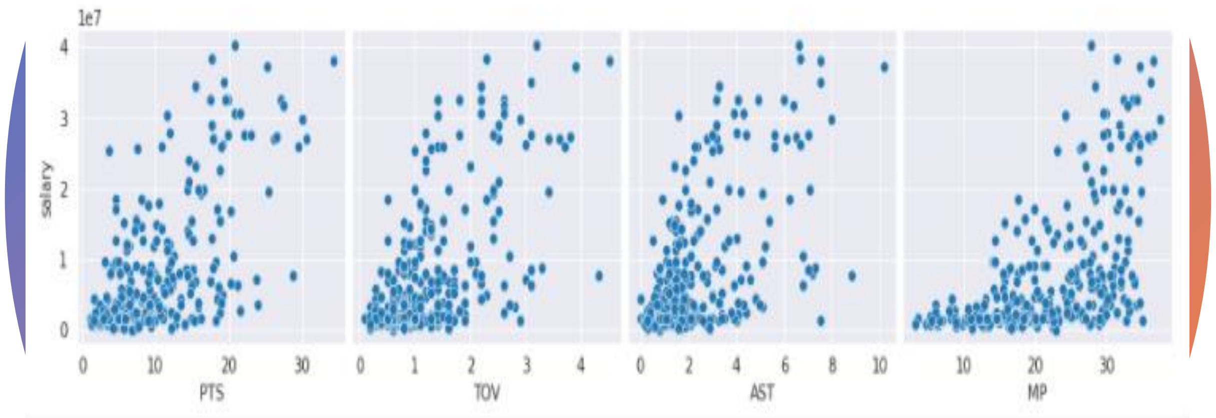
First, I use SNS to give an overview of salary, age and FG%.

After calculating the average and median, I found that these three figures are positive skewness distribution.

From this, we can see that the center of gravity of the data side is on the left and relatively concentrated at a relatively low level.



Overview factors related with the Salary



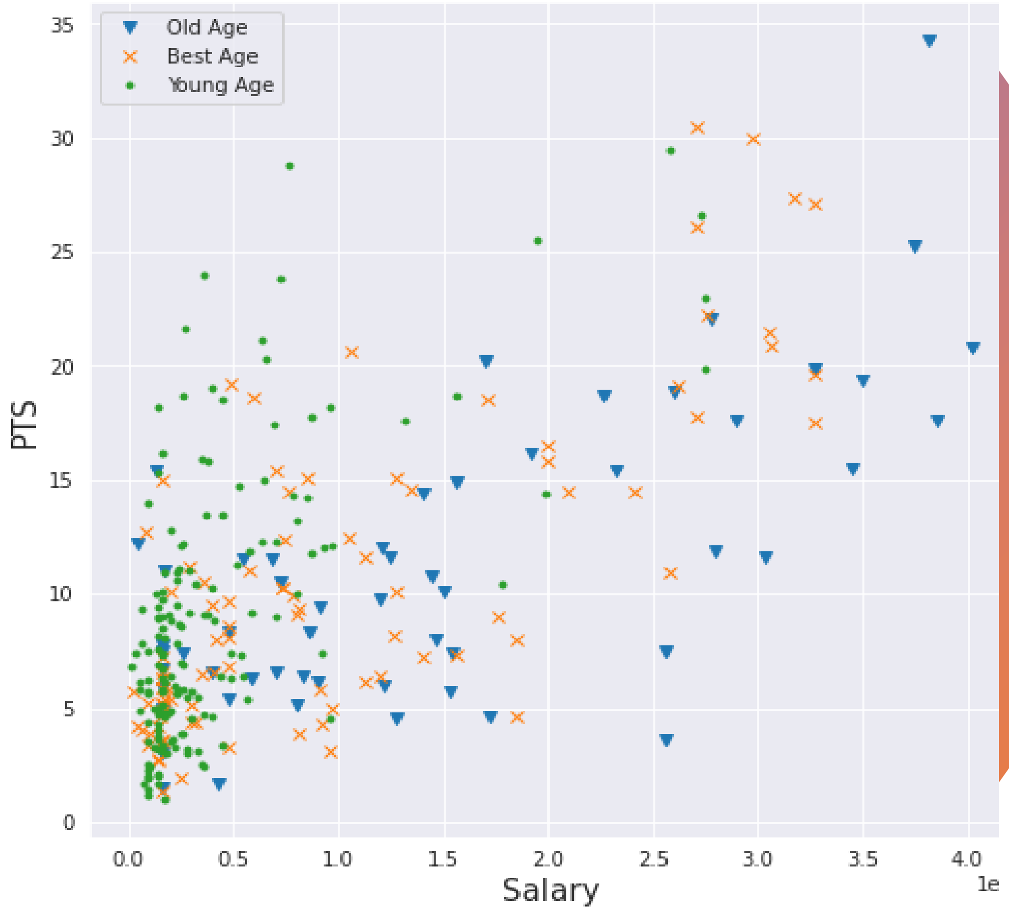
Through the correlation analysis of the overall data

PTS (0.6577), TOV (0.6198), AST (0.5969) and MP (0.5824) are the four values with relatively high correlation.

By analyzing the scatter diagram of the impact of these factors on salary, we can see that there is no particularly obvious trend here due to the degree of correlation.

However, we can still see that with the increase of data, wages will also increase.

Visualizing Data with Old-Age, Best Age, and Young Age



- I created a new column age based on the existing data called age_ type.

- Old age refers to players older than or equal to 30.
- Young age player refers to players
- younger than or equal to 25, and the player in the middle is called best age

- After data analysis and image analysis, I found that the number of young age is the largest (160). But on average, old_ Age players perform better in salary and data than the other two.

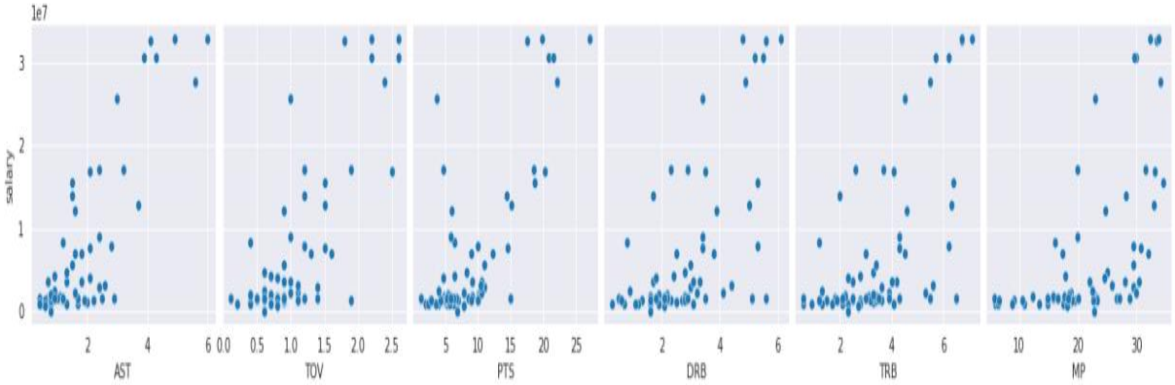
Spilt the Data frame to C, PG, PF, SF and SG

- Then I divided the data frame into five parts according to C, PF, PG, SG and SF.

- Then I analyze the most relevant factors of salary for players in different positions.

- Then through a series of analysis and fitting models, the best performance of the five is the SF data frame. So, I will use SF as the data for presentation.

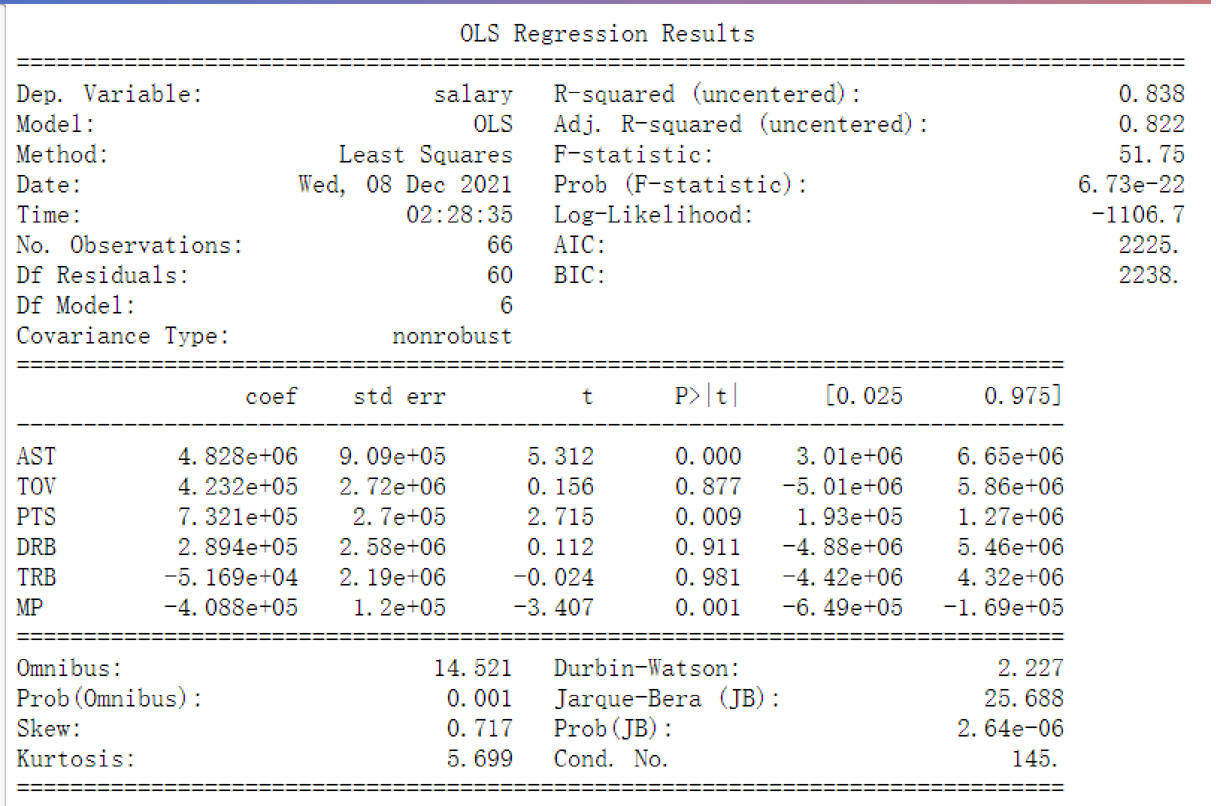
For SF data frame



- For SF, the factors most related to salary are
- AST (0.8385), TOV (0.7799), PTS (0.7597), DRB (0.6372), TRB (0.6232) and MP (0.5929).

- The first three factors have high positive correlation
- when analyzing the influence graph of independent variables on dependent variables, the three show a more obvious positive correlation scatter distribution.

OLS model and summary

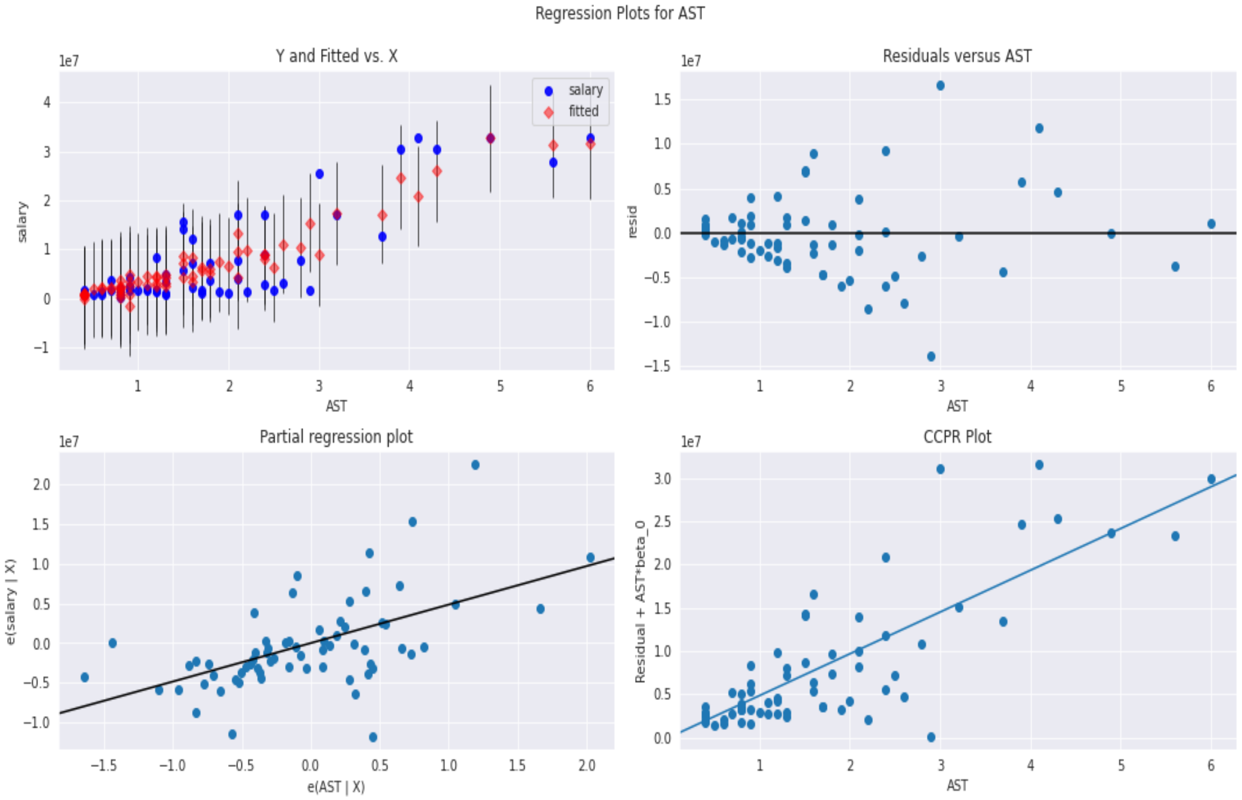


```
                            OLS Regression Results
=======================================================================================
Dep. Variable:                 salary   R-squared (uncentered):                   0.838
Model:                            OLS   Adj. R-squared (uncentered):              0.822
Method:                 Least Squares   F-statistic:                              51.75
Date:                Wed, 08 Dec 2021   Prob (F-statistic):                    6.73e-22
Time:                        02:28:35   Log-Likelihood:                         -1106.7
No. Observations:                  66   AIC:                                      2225.
Df Residuals:                      60   BIC:                                      2238.
Df Model:                           6
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
AST         4.828e+06   9.09e+05      5.312      0.000    3.01e+06    6.65e+06
TOV         4.232e+05   2.72e+06      0.156      0.877   -5.01e+06    5.86e+06
PTS         7.321e+05    2.7e+05      2.715      0.009    1.93e+05    1.27e+06
DRB         2.894e+05   2.58e+06      0.112      0.911   -4.88e+06    5.46e+06
TRB        -5.169e+04   2.19e+06     -0.024      0.981   -4.42e+06    4.32e+06
MP         -4.088e+05    1.2e+05     -3.407      0.001   -6.49e+05   -1.69e+05
==============================================================================
Omnibus:                       14.521   Durbin-Watson:                   2.227
Prob(Omnibus):                  0.001   Jarque-Bera (JB):               25.688
Skew:                           0.717   Prob(JB):                     2.64e-06
Kurtosis:                       5.699   Cond. No.                         145.
==============================================================================
```

- I set dependent value as salary, and then substitute the factors obtained before into the model.

- Then we get the summary of the model. Our model R-squared is 0.838, which proves that our model has an excellent degree of fitting.

- At the same time, the P values of AST, PTS and MP are lower than 0.05, which proves that they are significant.



OLS model and analysis plot

- I chose AST as the analysis factor because it has the lowest p value.

- I choose AST for the plot. From the figure, we can see that AST has an excellent performance in four different types of graphs.

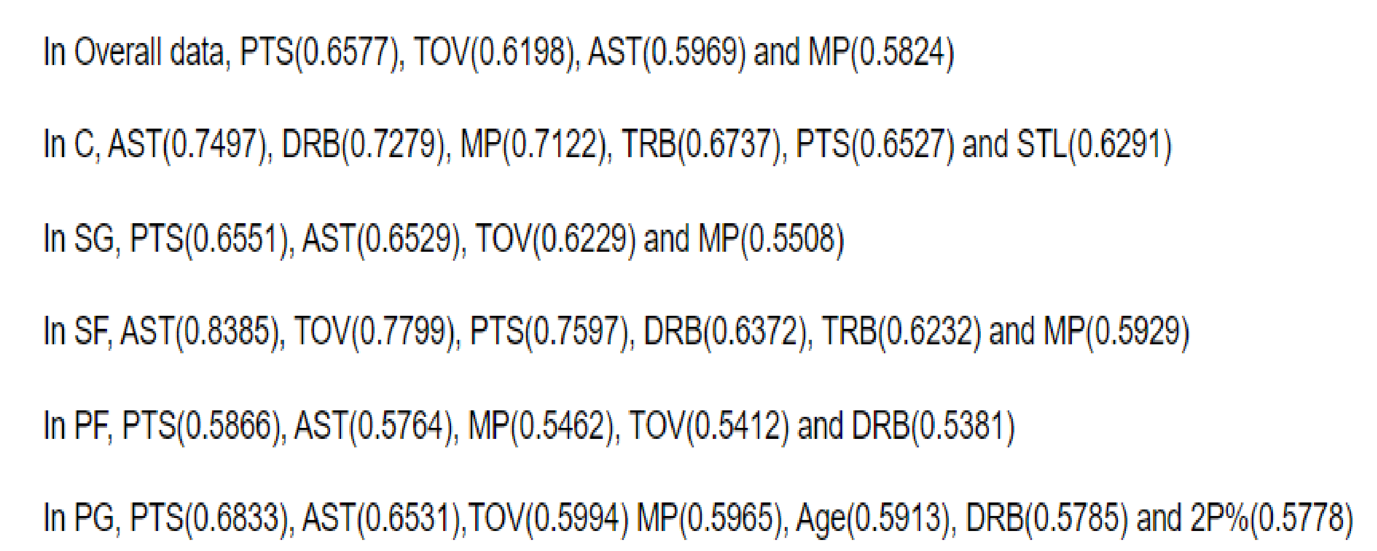
- It shows a good and obvious positive correlation in partial region and CCPR. Finally, it also has a good fitted degree in the fitting plot.

Answers

In my final summary, PTS has the most extensive impact on the salary of NBA players.

Among them, it has the greatest impact on SG, PF and PG, and has different effects on the other two positions.

In my initial assumptions, I thought the hit rate and efficiency would be the most influential. After analysis, it seems that this is not the case.



In Overall data, PTS(0.6577), TOV(0.6198), AST(0.5969) and MP(0.5824)

In C, AST(0.7497), DRB(0.7279), MP(0.7122), TRB(0.6737), PTS(0.6527) and STL(0.6291)

In SG, PTS(0.6551), AST(0.6529), TOV(0.6229) and MP(0.5508)

In SF, AST(0.8385), TOV(0.7799), PTS(0.7597), DRB(0.6372), TRB(0.6232) and MP(0.5929)

In PF, PTS(0.5866), AST(0.5764), MP(0.5462), TOV(0.5412) and DRB(0.5381)

In PG, PTS(0.6833), AST(0.6531),TOV(0.5994) MP(0.5965), Age(0.5913), DRB(0.5785) and 2P%(0.5778)

Difficulties or Complications

In this project, the most difficult point is to find a data frame with salary and player data at the same time.

Finally, I found two data frames and combined salary and player data one by one.

In this project, the most difficult point is to find a data frame with salary and player data at the same time.

Finally, I found two data frames and combined salary and player data one by one.

Reference

1.BasketballGlossary https://www.basketball-reference.com/about/glossary.html

2.Datasalary https://www.kaggle.com/junfenglim/nba-player-salaries-201920

3. Data stats https://www.basketball-reference.com/leagues/NBA_2020_per_game.html