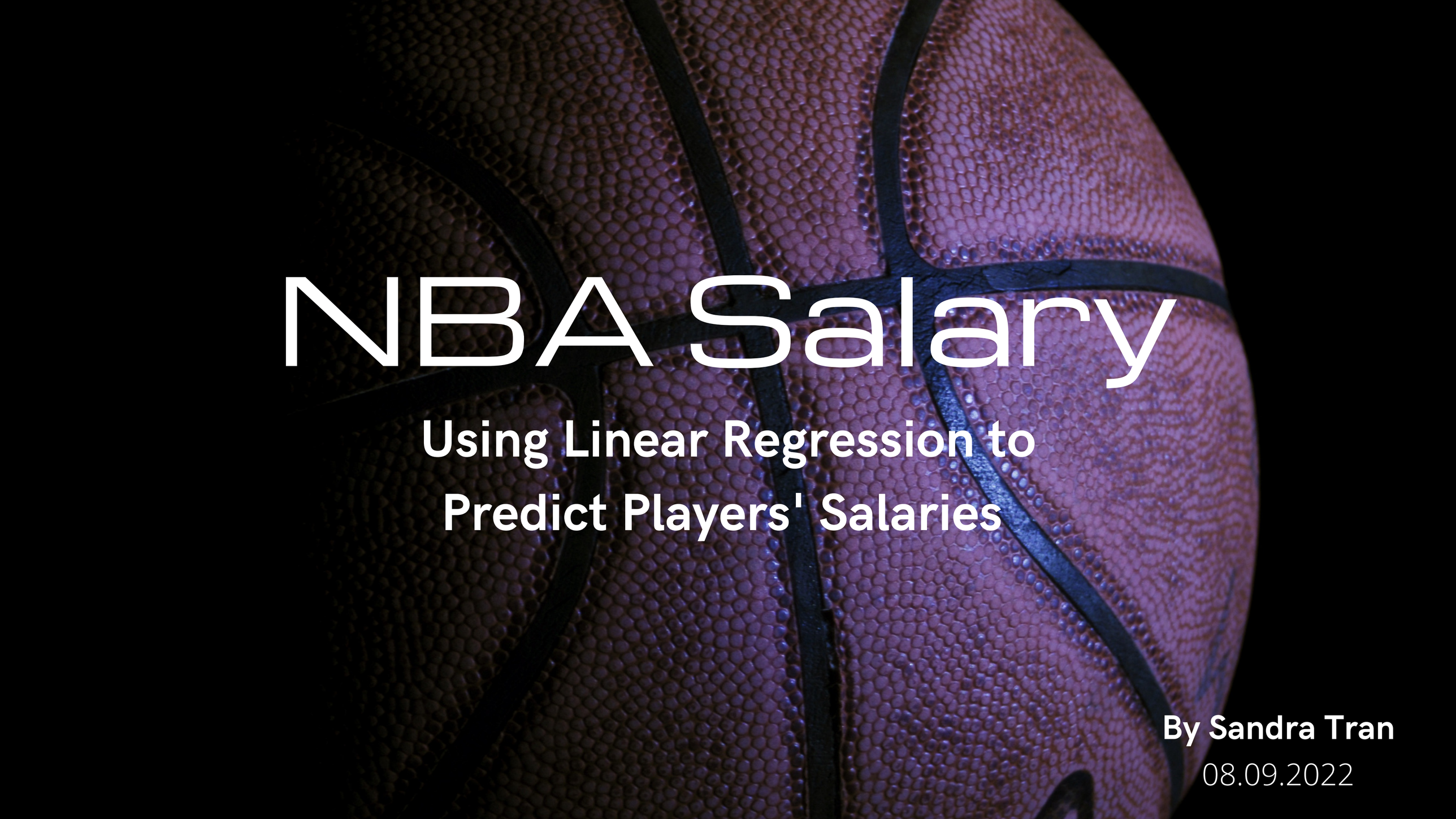


NBA Salary

Using Linear Regression to Predict Players' Salaries

By Sandra Tran

08.09.2022

Design

Objective:

- Use linear regression modeling to build a predictive model for an NBA player's salary based on their game stats
- The model with the best R^2 value will represent the model with the most accurate predictions

Goal:

- Build a model with the best R^2 value

big_3.jpg (1920×1080) (moddingway.com)

Data

- NBA Players + Game Stats: www.nba.com
  - Regular Seasons
    - 1996-97 to 2021-22
  - Player
    - Name
    - Age
    - Games Played (Wins, Losses)
    - Minutes Played
    - Points made
    - FG%, 3P%, FT%
    - Etc.
- NBA Players + Salary: www.hoopshype.com
  - Regular Seasons
    - 1996-97 to 2021-22
  - Player Salaries*
- Merged Tables - over 11K rows of data
  - Each row represents a player in a given season

nba player stats - Bing images

*Current US Dollars adjusted for inflation

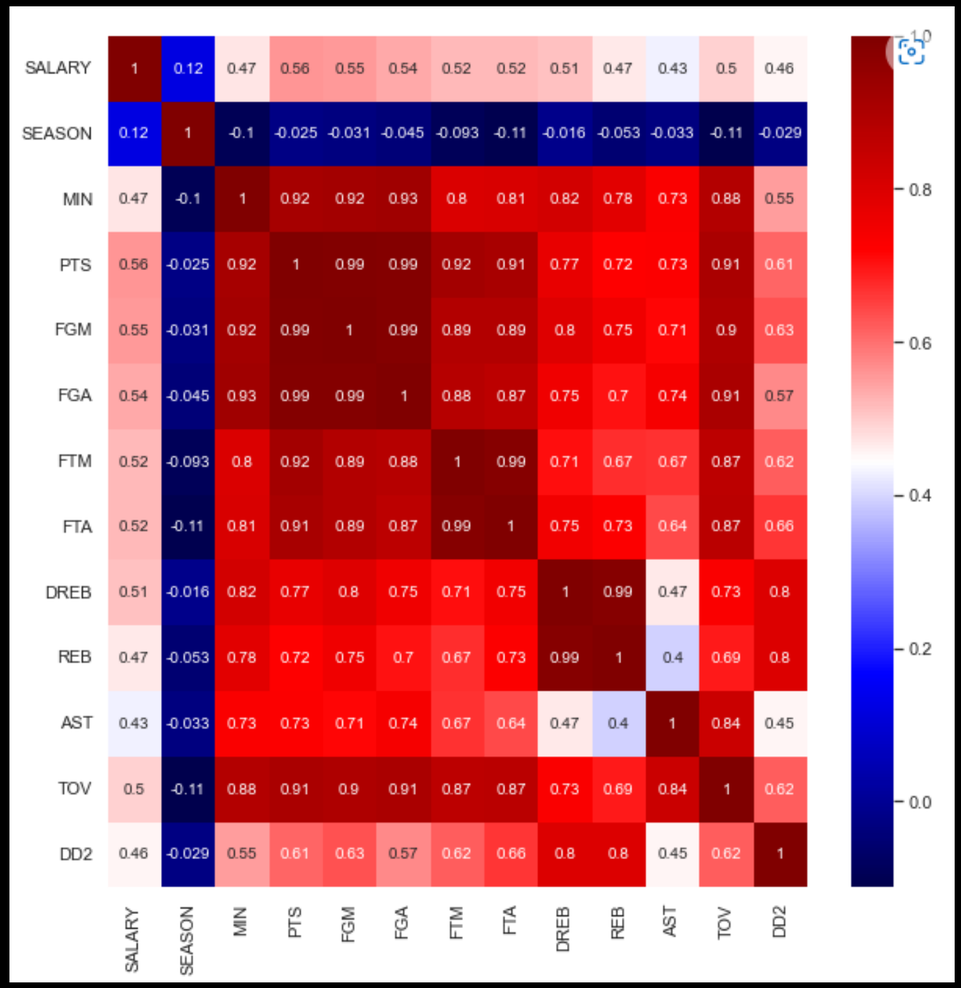
Tools

- Webscraping
  - BeautifulSoup
  - Selenium
- Regression Modeling
  - Python
  - Numpy
  - Pandas
  - Scikit-learn
  - Statsmodels
  - Matplotlib
  - Seaborn
- Presentation
  - Canva

selenium webscraping - Bing images

Data Cleaning & EDA

- Data Cleaning
  - Clean up column names
  - Change datatypes to numeric
  - Drop all rows with null values in 'SALARY'
  - Update players' names (modified or misspelled)
  - Merge stats + salaries tables
- Exploratory Data Analysis (EDA)
  - Pairplot, heatmap, correlation map for feature correlations
  - Distribution plots of target and features
  - VIF for multicollinearity check



Modeling

Regularization

Base Model (Simple Linear)

- **R^2 (train) = 0.504**
- **R^2 (test) = 0.525**

Ridge Regression

- R^2 (train) = 0.504
- R^2 (test) = 0.525

LASSO Regression

- R^2 (train) = 0.504
- R^2 (test) = 0.525

All scores are equal

- **a sign that it may be underfit**

WhirlwindAmpleGemsbok-size_restricted.gif (435×250) (gfycat.com)

Feature Engineering

Polynomial Transformation (PT)

Simple Linear + PT

- R^2 (train) = 0.637
- R^2 (test) = 0.493

Ridge Regression + PT

- R^2 (train) = 0.617
- R^2 (test) = 0.514

LASSO Regression + PT

- R^2 (train) = 0.599
- R^2 (test) = 0.58

nba warriors dance gif - Bing images

- Higher test score!
- Smaller gap between train/test!

Feature Engineering

Log Transformation (Log) of Target

Log (Target) + PT (Features)

- R^2 (train) = 0.548
- R^2 (test) = 0.539

LASSO Regression + PT (Features)

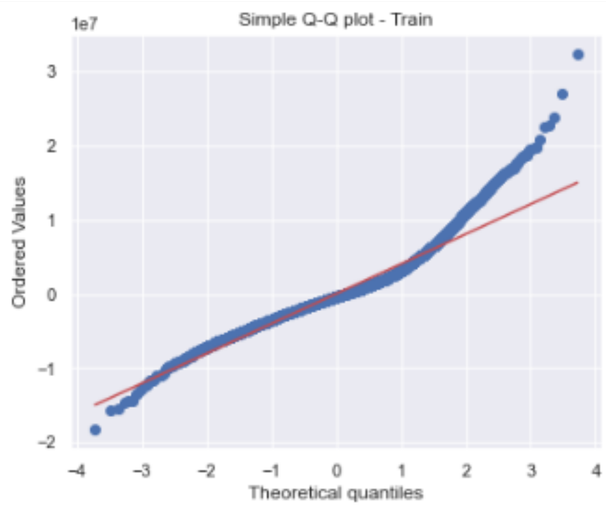
- R^2 (train) = 0.599
- R^2 (test) = 0.58

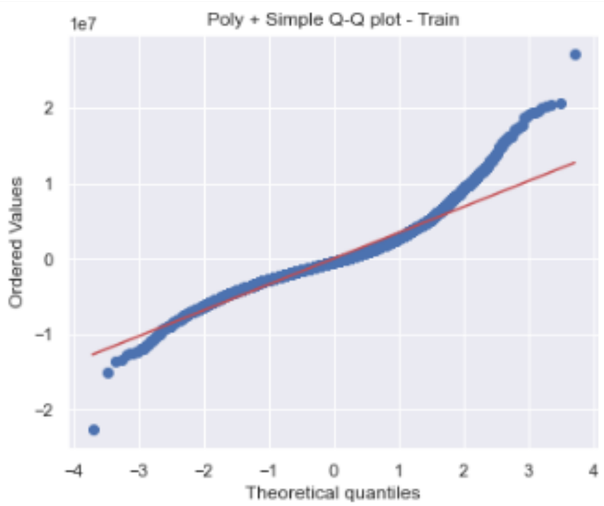
Feature Interpretation

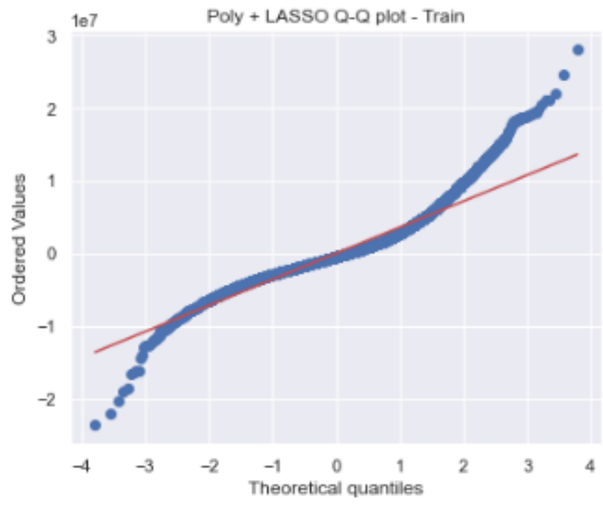
	coef	features	abs_val_coef
85	-4.510335e+06	GP PTS	4.510335e+06
55	-3.968606e+06	AGE^2	3.968606e+06
2	3.927981e+06	AGE	3.927981e+06
60	3.741859e+06	AGE PTS	3.741859e+06
56	-2.332580e+06	AGE GP	2.332580e+06
61	1.811059e+06	AGE FGM	1.811059e+06
76	1.662335e+06	AGE BLK	1.662335e+06
196	1.544900e+06	FGM^2	1.544900e+06
30	1.453157e+06	SEASON GP	1.453157e+06
276	1.134559e+06	3PA DREB	1.134559e+06

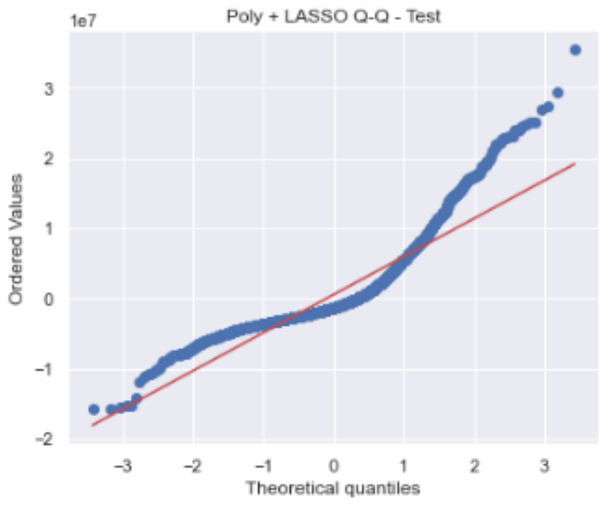
- Standardized coefficients
  - Measured in units of standard deviation
- Sorted in (descending) order of absolute value
- MOST important variable = **product of GP (Games Played) and PTS (Points)**
  - Negatively correlated
- **AGE** = top variable affecting SALARY positively
- 109 total features in the end
  - Started with 27

Diagnostic Plots









- Model will make poor predictions at low and high ends
- Model may still be underfit
- High outliers, esp right tail

Future Works

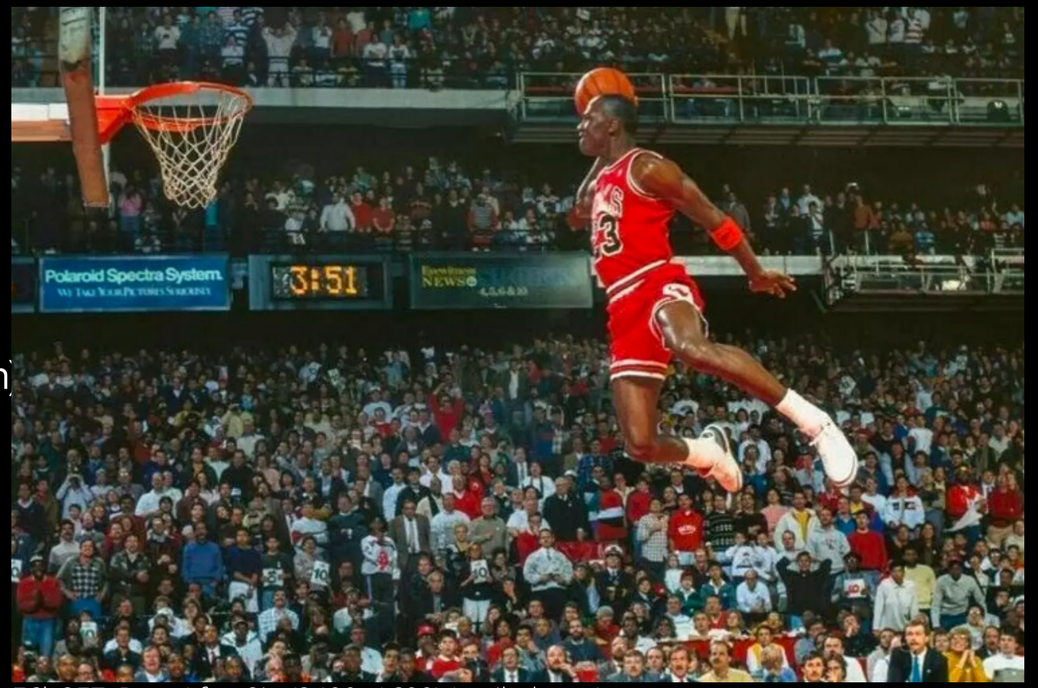
Model could be improved by...

- Adding more features
  - Team
    - Create dummy variables and see how being on different teams affect salaries
  - Teammates
    - Having chemistry with other players may help stats, in turn, salary
  - Injuries
  - Salary caps
    - Layered equation
    - Luxury Tax
    - '84-'85 salary cap = $3.6 mil (~$8 mil, adjusted for inflation for 2015)
    - '22-'23 salary cap = $122 mil
    - Changes every year
- Dividing player contracts into subcategories (i.e. Rookie 1-5 yrs, Non-Rookie 5+ yrs)

References

Images

1. ranking-the-top-15-best-players-in-nba-history-with-combined-stats.jpg (1200×900) (fadeawayworld.net)
2. NBA-Salaries-Then-vs-Now.jpg (1000×1267) (wp.com)
3. Typical-NBA-Salary.jpg (1000×1309) (wp.com)
4. 5956b22ea3630f62588b67a2 (1201×900) (insider.com
5. selenium webscraping - Bing images
6. big_3.jpg (1920×1080) (moddingway.com)
7. 67-court-10.gif (527×280) (nba.com)
8. WhirlwindAmpleGemsbok-size_restricted.gif (435×250) (gfycat.com)
9. E6ly2EZzRyemLfzgs2jp (2400×1600) (grailed.com)



E6ly2EZzRyemLfzgs2jp (2400×1600) (grailed.com)

Data

1. www.nba.com
2. www.hoopshype.com/salaries/players

Resources

1. Standardized and Unstandardized Regression Coefficients (analyticsvidhya.com)