Best NBA Players, Value for Money

Web Scraping and Linear Regression Project to Predict Salaries

Gulay Samatli-Pac

Introduction

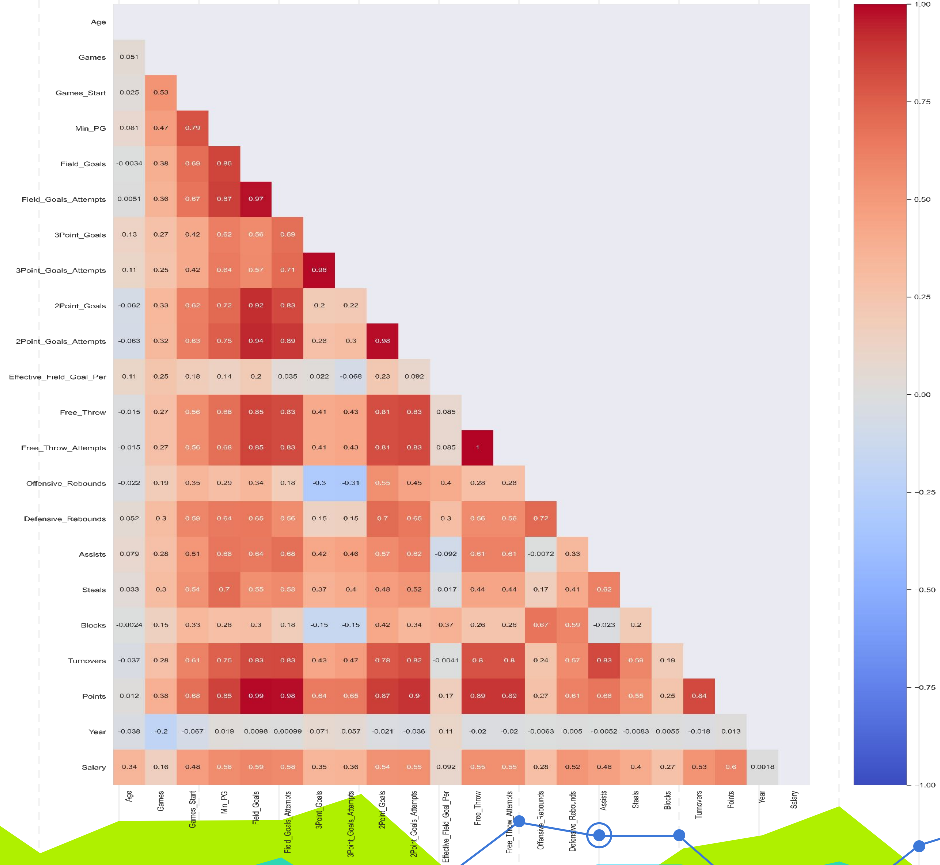
Question :

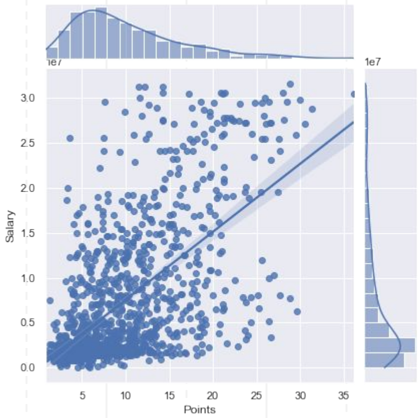
Discover 10 most underpaid NBA players

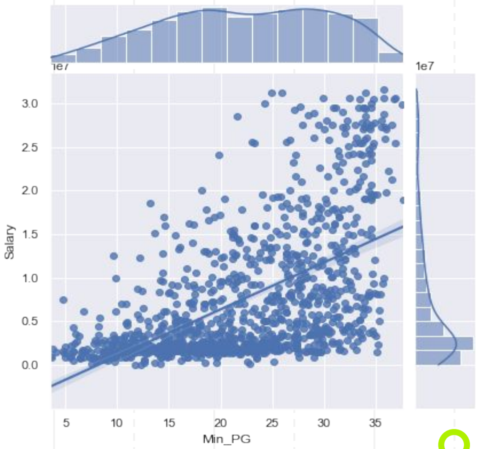
Data :

- ➔ **https://www.basketball-reference.com/**
- ➔ annual salaries and more than **20 per game statistics** for more than **400 players**
- ➔ Season 2019, 2020 and 2021

Data


Age
Games
Games_Start
Min_PG
Field_Goals
Field_Goals_Attempts
3Point_Goals
3Point_Goals_Attempts
2Point_Goals
2Point_Goals_Attempts
Effective_Field_Goal_Per
Free_Throw
Free_Throw_Attempts
Offensive_Rebounds
Defensive_Rebounds
Assists
Steals
Blocks
Turnovers
Points
Year
Salary
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00


1e7
Salary
3.0
2.5
2.0
1.5
1.0
0.5
0.0
5 10 15 20 25 30 35
Points


1e7
Salary
3.0
2.5
2.0
1.5
1.0
0.5
0.0
5 10 15 20 25 30 35
Min_PG

Linear Regression

Notes from initial model:

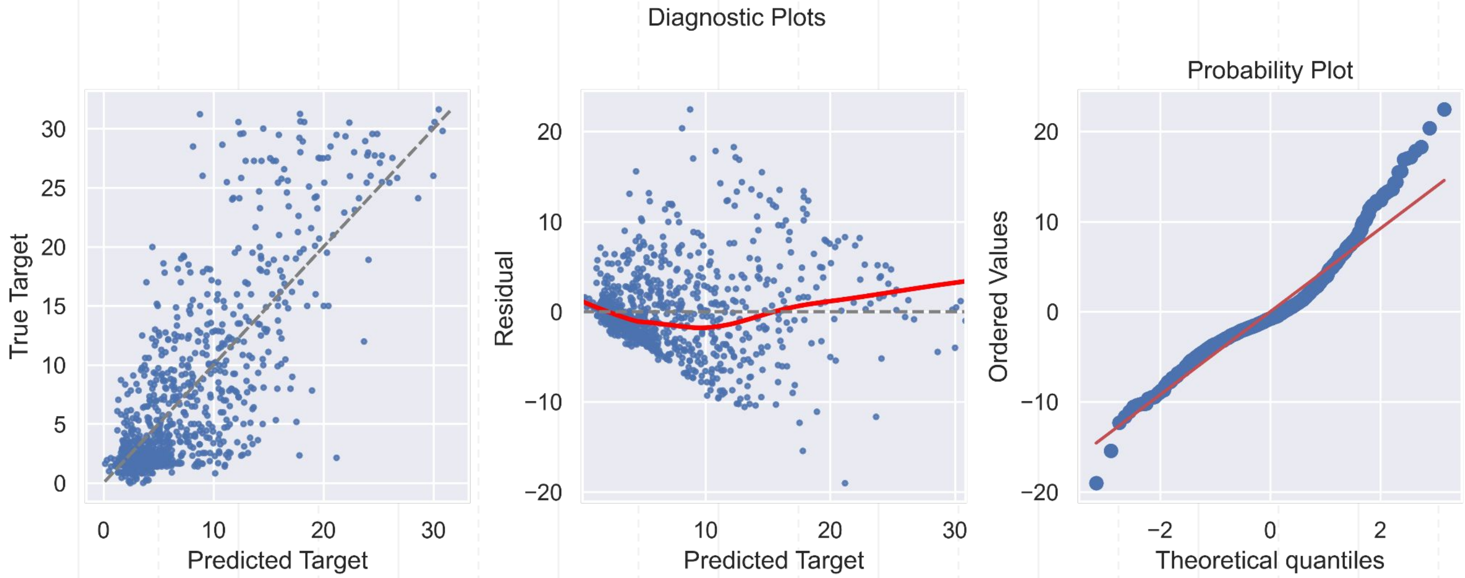
- Skewed target variable
- Outliers
- Under and overfitting
- Polynomial trend

Linear Regression	
R^2 Train	R^2 Test
0.53	0.56

0.7% improvement in MAE

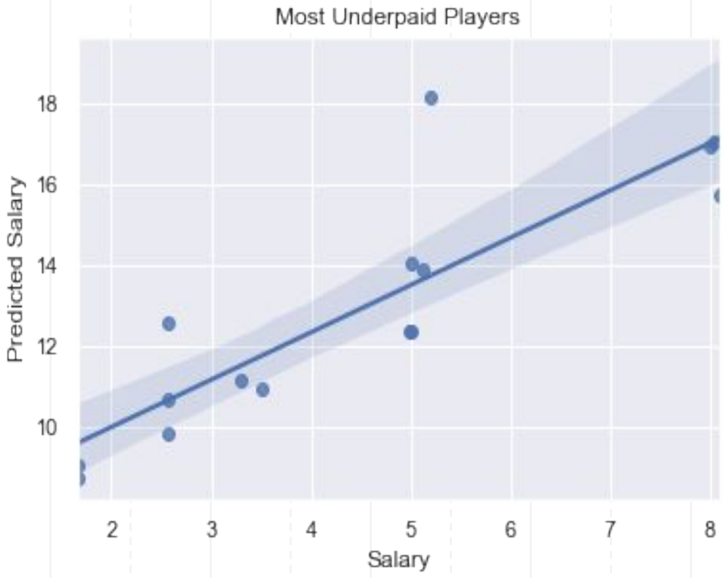
Poly(degree = 2), LASSO	
R^2 Train	R^2 Test
0.61	0.60

Polynomial transformation (degree = 2), LASSO



10 Top Underpaid Players

Name	Salary(in Million)	Predicted	95% CI Upper Limit	NBA Draft
Donovan Mitchell	$5.20	$18.18	$27.46	2017
Enes Kanter	***$5.01***	***$14.06***	***$23.34***	***2011***
Luka Dončić	$8.05	$17.07	$26.35	2018
Mason Plumlee	***$8.00***	***$16.97***	***$26.25***	***2013***
Bam Adebayo	$5.12	$13.90	$23.18	2017
Jeff Green	***$2.56***	***$10.72***	***$20.00***	***2007***
Michael Carter-Williams	***$3.30***	***$11.20***	***$20.48***	***2013***
De'Aaron Fox	$8.10	$15.74	$25.02	2017
Derrick White	$3.52	$10.98	$20.26	2017
Kendrick Nunn	$1.66	$9.09	$18.37	2018

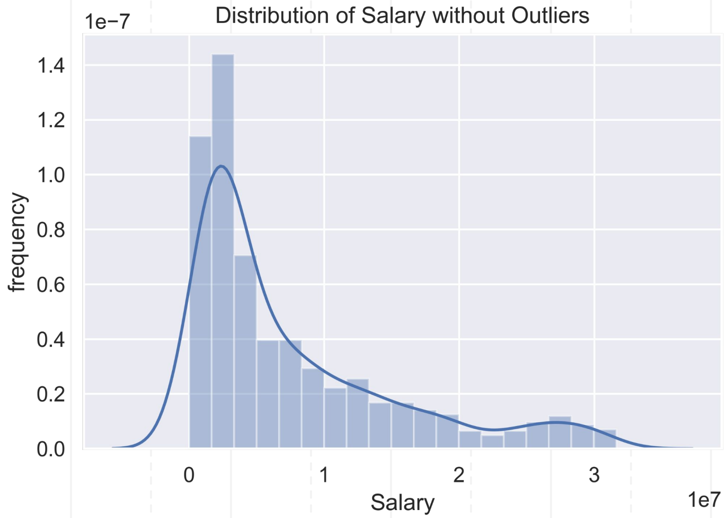


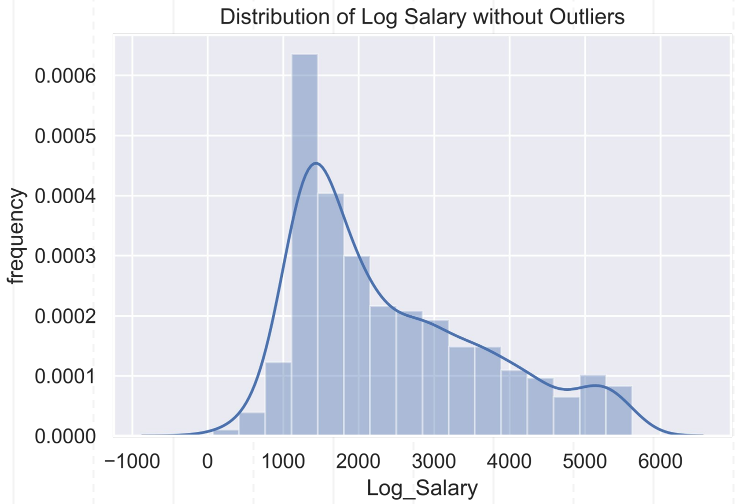
Future Works

- Add more metrics to describe a player’s performance: injuries, years of experience.
- Add more features that might indirectly affect salaries: team market size, expected fan-based revenue, merchandising and advertising.
- Apply different models: Decision Tree Regressor, Random Forest Regressor, Gradient Boosted Tree Regressor.
- Apply deeper feature selection to simplify the model without losing the prediction power.

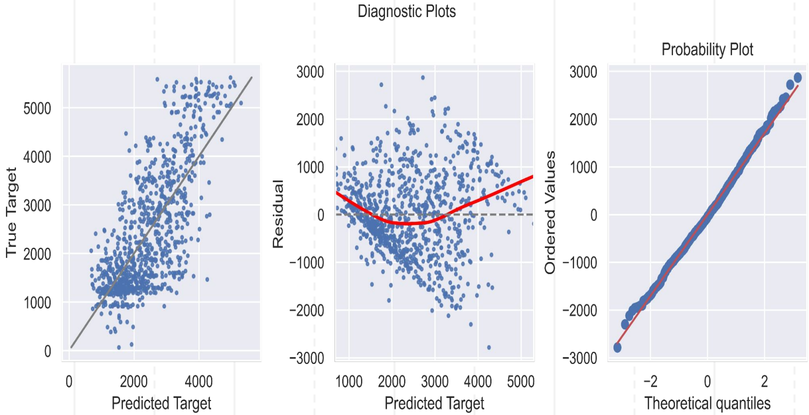
Thank you

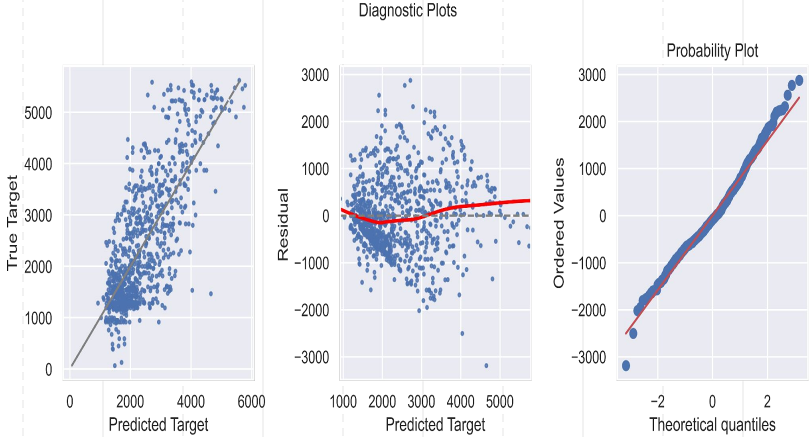
The Distribution of Salary





Linear Regression with Log Salary





Linear Reg-Log Salary	
R^2 Train	**R^2 Test**
0.55	0.57

Poly(degree = 2), LASSO- Log Salary	
R^2 Train	**R^2 Test**
0.61	0.58