NBA

Starting 5 Predictions

Sheldon Yen

TABLE OF CONTENTS

Project Context and The Why?

Group of Interest:

Starting 5 Lineups
- Individual Statistics
- Productivity with other players in lineup

Factor to consider:

Usage Rate
- Team Statistic
- Low vs. Medium vs. High Usage Rate Players

Why:

Long fans of Basketball
- Watch Basketball
    - Diehard Los Angeles Clippers Fan

The Data and The Data Cleaning

The Data:

Balldontlie.io API

- Player Data
- Season Averages

Data Cleaning:

- Inner Merged datasets together by unique player ID
- Removed all players that did not play in 2022
- Calculated unique team statistics for each player
- Calculated unique usage for each player

Usage Calculation:

100 * ((Field Goals Attempted + 0.44 * Free Throws Attempted + Turnover Rate) * (Total Team Minutes Played / 5)) / (Minutes Played * (Total Team Field Goals Attempted + 0.44 * Total Team Free Throws Attempted +Total Team Turnovers)).

Usage percentage is an estimate of the percentage of team plays used by a player while he was on the floor.

Machine Learning		MSE
Linear Regression	Simple Linear Regression using Usage and Points/Rebounds/Assists/Steals	21.24
K-Nearest	Simple K-Nearest Regression using Usage and Points/Rebounds/Assists/Steals	28.80
Multivariable Linear Regression	Multivariable Linear Regression using Usage/Weight/Height/Position and Points/Rebounds/Assists/Steals	21.97
Multivariable K-Nearest	Multivariable K-Nearest Regression using Usage/Weight/Height/Position and Points/Rebounds/Assists/Steals	26.35
Ensemble	Ensemble method using Voting Regressor of both the linear and k-nearest regressions	22.26

Regression Analysis

Variables of Interest:

Explanatory Variables
- Usage Rate
- New Usage

Response Variables
- Points
- Assists
- Rebounds
- Steals

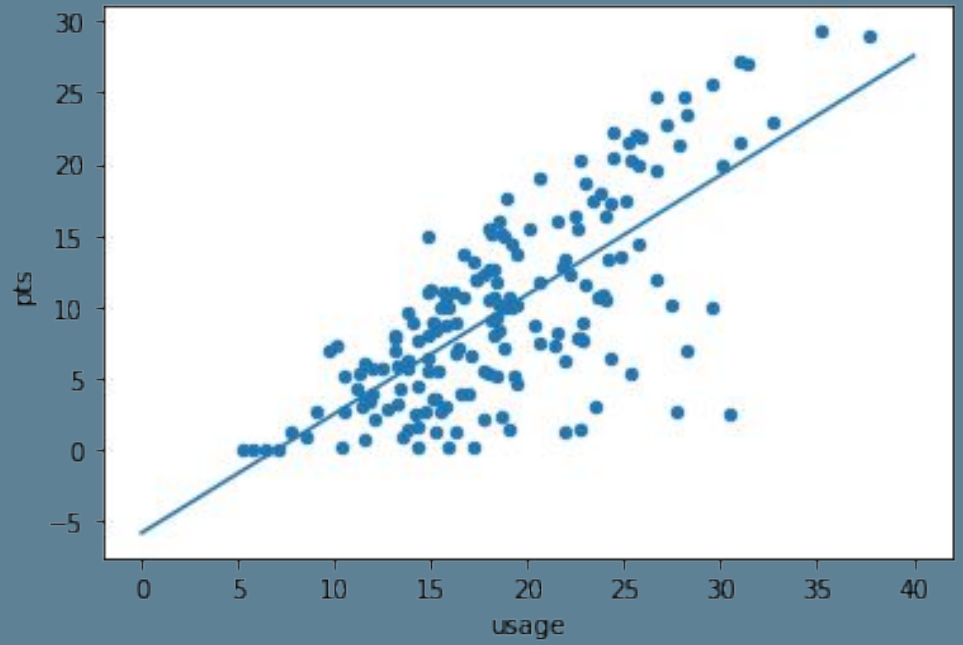
```
new['pts'] = 0.83476156 * new['new_usage'] - 5.556495932455569

new['reb'] = 0.15413808 * new['new_usage'] + 0.9763377142292433

new['ast'] = 0.17393668 * new['new_usage'] - 0.3528002733591336

new['stl'] = 0.02194919 * new['new_usage'] + 0.22063825643570223
```

Regression Plot

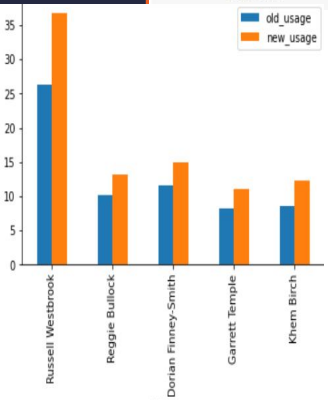


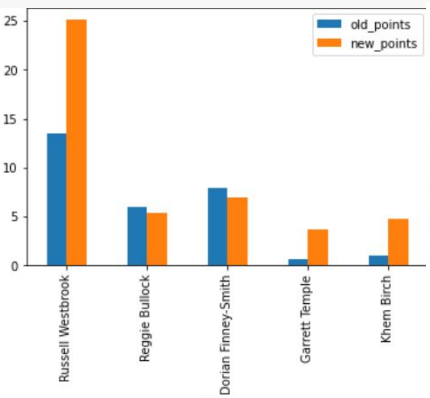
Simulator

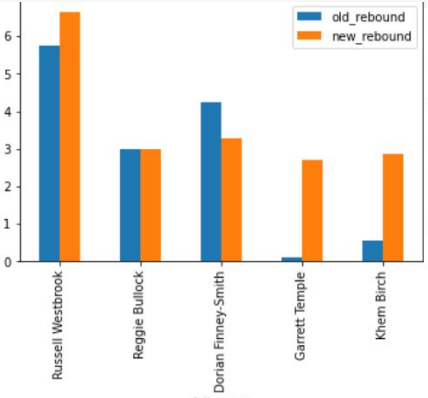
https://colab.research.google.com/drive/1DJP3kBH-CYCz-6JGP11btvMZ0NJu0ZjS?usp=sharing

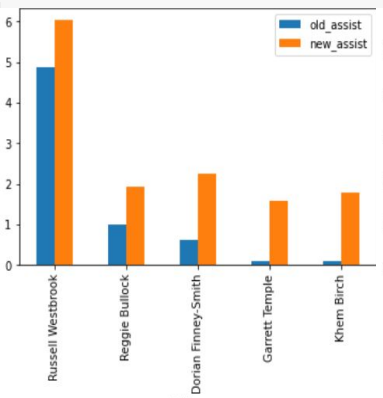
Russell Westbrook with Low Usage

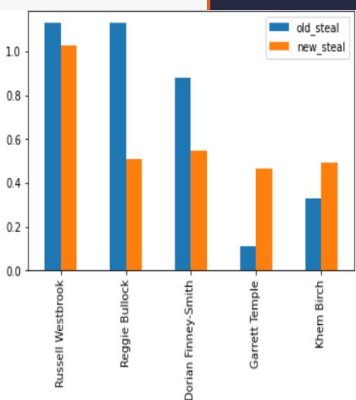
full_name	old_usage	new_usage	old_points	new_points	old_rebound	new_rebound	old_assist	new_assist	old_steal	new_steal
Russell Westbrook	26.344613	36.722932	13.50	25.098396	5.75	6.636740	4.88	6.034665	1.13	1.026677
Reggie Bullock	10.184853	13.161959	6.00	5.430602	3.00	3.005097	1.00	1.936547	1.13	0.509533
Dorian Finney-Smith	11.583336	14.969229	7.88	6.939241	4.25	3.283666	0.63	2.250898	0.88	0.549201
Garrett Temple	8.140678	11.109822	0.67	3.717556	0.11	2.688784	0.11	1.579605	0.11	0.464490
Khem Birch	8.500035	12.320884	1.00	4.728504	0.56	2.875455	0.11	1.790253	0.33	0.491072





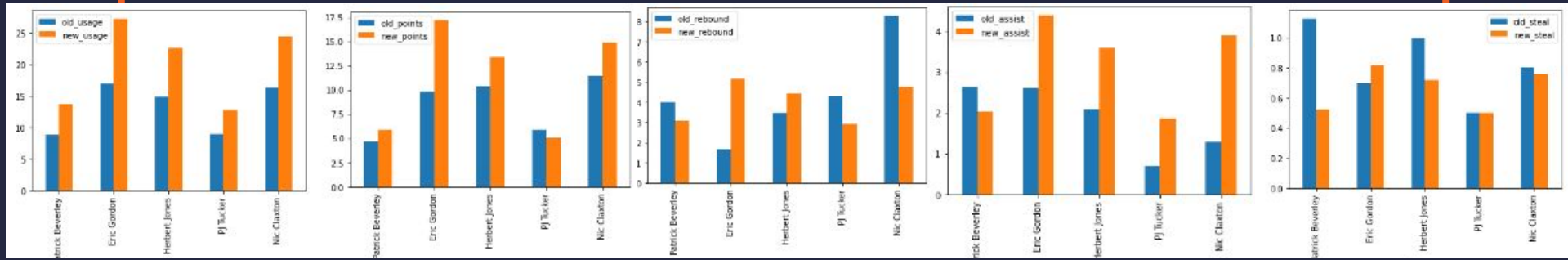






All Low Usage Player Lineup

full_name	old_usage	new_usage	old_points	new_points	old_rebound	new_rebound	old_assist	new_assist	old_steal	new_steal
Patrick Beverley	8.781712	13.712485	4.63	5.890160	4.0	3.089954	2.63	2.032304	1.13	0.521616
Eric Gordon	16.986393	27.278662	9.80	17.214682	1.7	5.181018	2.60	4.391960	0.70	0.819383
Herbert Jones	14.824726	22.663378	10.40	13.362021	3.5	4.469627	2.10	3.589192	1.00	0.718081
PJ Tucker	9.021693	12.770257	5.90	5.103624	4.3	2.944721	0.70	1.868416	0.50	0.500935
Nic Claxton	16.369622	24.487754	11.50	14.884940	8.3	4.750833	1.30	3.906518	0.80	0.758125



Conclusion

Limitations:

- Lack of Access to Advanced Statistics
  - Plus/Minus
  - Per 48 Minute Statistics
- Limited Definition for Player Positions
- Players with abnormally high usage rates
  - Lack of games

Observations:

- Good predictions overall from model with given resources
- High usage rate players favored
  - Highest usage rate player has best individual statistics
  - High usage rate players not optimal with other high usage rate players
- Low usage rate players see a jump with players with low/medium usage rates

Thank You and Happy Holidays!