CMSE 202 Final Project: NBA Predictions

Gabriel, Amaan, Vibhu, Vinayak

Overall Question and Goal

The goal of our project is to create a model to predict the next years stats for each player, then based off those stats predict the upcoming years MVP as well as the first and second all NBA teams.

Methodology

- **Using OLS and multiple linear regression modeling we are able to predict the stats of every player for the next season by measuring regression from year to year using different ages as the dependent variable.**
- **To make the MVP prediction we used a separate dataset with all of the past MVPs to make a model using the same tactics as before that can predict the MVP and all NBA teams.**

Results Up to This Point

- **This is one of the OLS summaries from the model we created to help predict next seasons stats. It has a very low r-squared value so we can definitely improve on that.**

Dep. Variable:	First	R-squared:	0.167
Model:	OLS	Adj. R-squared:	0.140
Method:	Least Squares	F-statistic:	6.239
Date:	Sun, 26 Nov 2023	Prob (F-statistic):	2.18e-09
Time:	22:48:28	Log-Likelihood:	-1588.5
No. Observations:	355	AIC:	3201.
Df Residuals:	343	BIC:	3247.
Df Model:	11		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-63.7009	32.986	-1.931	0.054	-128.582	1.180
Age	0.2353	0.322	0.731	0.465	-0.398	0.869
G	0.1944	0.122	1.591	0.113	-0.046	0.435
MP	-0.8127	0.507	-1.604	0.110	-1.809	0.184
PTS	1.6570	0.279	5.938	0.000	1.108	2.206
TRB	0.5411	0.594	0.910	0.363	-0.628	1.710
AST	1.6424	0.597	2.753	0.006	0.469	2.816
STL	2.5138	2.684	0.937	0.350	-2.765	7.793
BLK	2.9293	2.018	1.452	0.147	-1.039	6.898
FG%	50.6885	34.213	1.482	0.139	-16.606	117.983
3P%	11.5889	13.342	0.869	0.386	-14.653	37.831
FT%	-4.7131	18.023	-0.262	0.794	-40.162	30.736

Omnibus:	232.243	Durbin-Watson:	1.891
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1635.337
Skew:	2.847	Prob(JB):	0.00
Kurtosis:	11.840	Cond. No.	3.50e+03

Results Up to This Point

- **We then have our model for MVP which we tested on last years data and from our tests it seems to work fairly well, however it also has a fairly low r-squared value that we can improve on.**

```
[86]: nba_per_const = sm.add_constant(nba_per_new)

      predictions = result.predict(nba_per_const)
      result_df = pd.DataFrame({'player': nba_per['player'], 'predicted_rank': predictions})

      result_df.sort_values(by = 'predicted_rank', ascending = False).dropna()[0:10]
```

Out[86]:

	player	predicted_rank
1694	Nikola Jokić	315.193214
1596	LeBron James	314.436776
1481	Joel Embiid	313.085782
1365	Giannis Antetokounmpo	309.823229
1716	Paul George	293.982790
1605	Luka Dončić	292.300112
1558	Kevin Durant	282.795982
1276	Dejounte Murray	270.366737
1581	Kyrie Irving	269.349061
1407	Ja Morant	266.683401

Conclusions Drawn

The main conclusion we have come to is that it is very possible to model the NBA MVP and predict next years stats, however we have found that with our current knowledge it is hard to create an accurate model.

Plans for Finishing

We need to finish the predictions for next years stats and improve on the accuracy of our models.