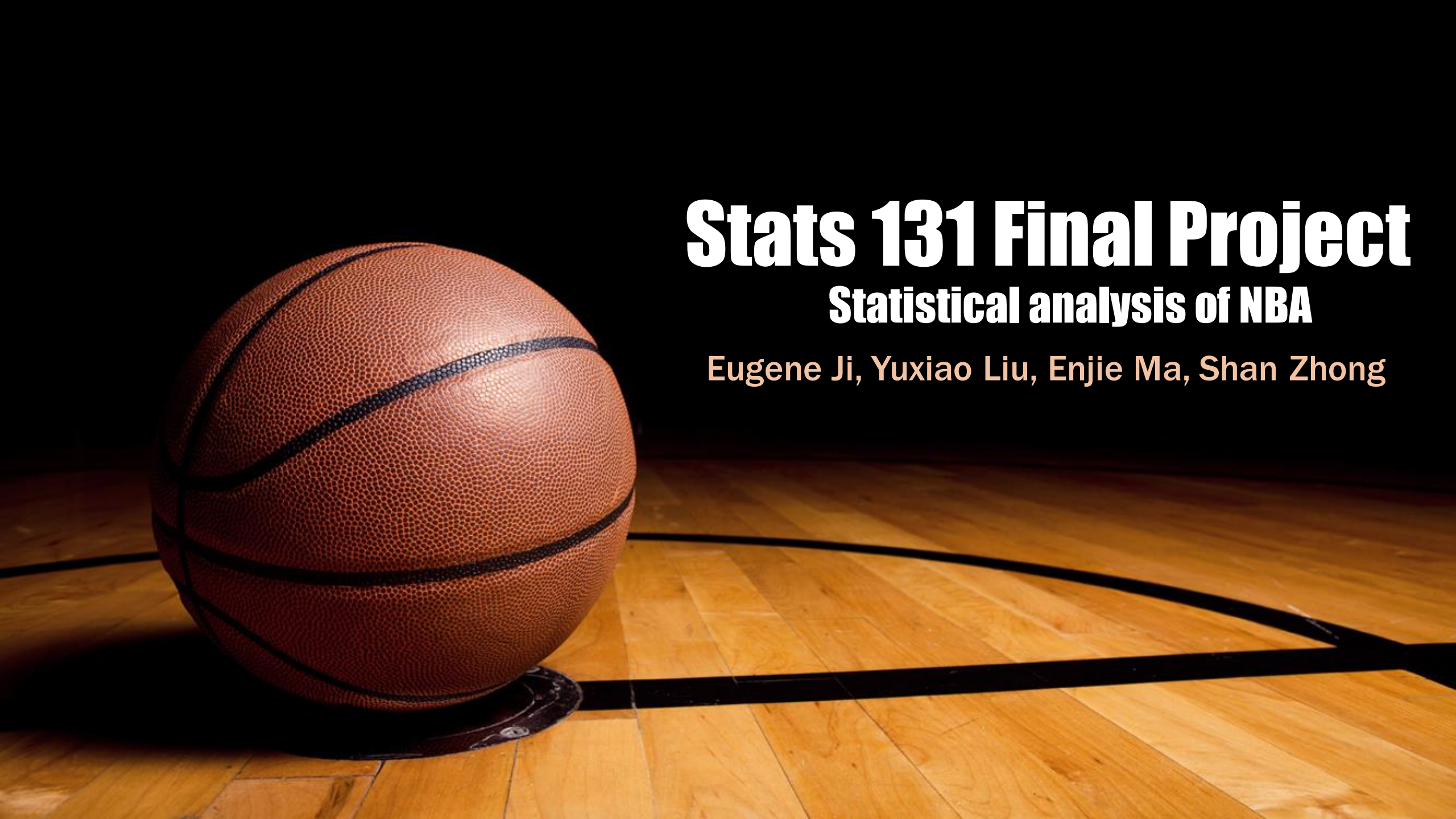


Stats 131 Final Project

Statistical analysis of NBA

Eugene Ji, Yuxiao Liu, Enjie Ma, Shan Zhong

Background

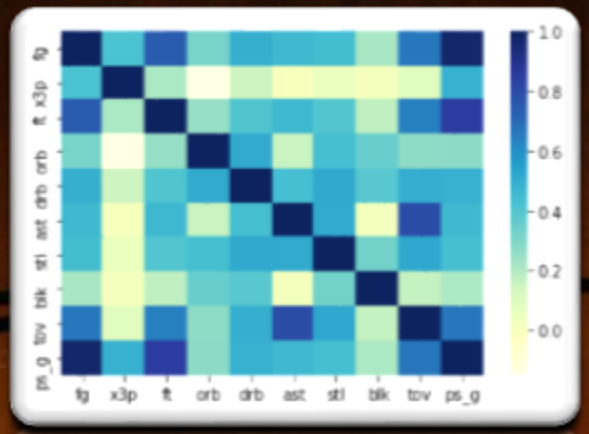
Motivation

- Point guard (PG) / Shooting guard (SG)/ Small forward (SF)/ Power forward (PF)/ Center (C)
- What statistical attributes are special for each basketball position, in the past ten years?

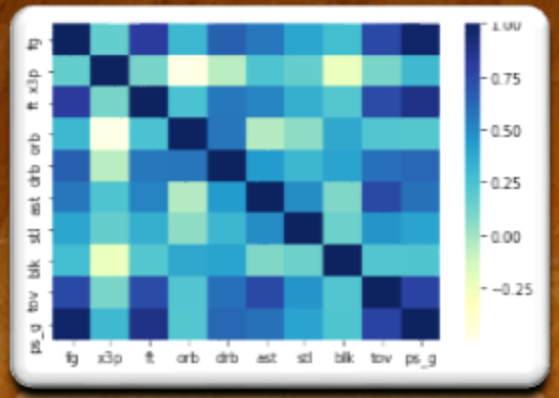
Data Information:

- - 19647 observations
- - 48 variables (offensive or defensive statistics)
- This data is the data of NBA players on regular seasons.
- Each observation (row) includes the statistics per game of a player during one particular NBA regular season.

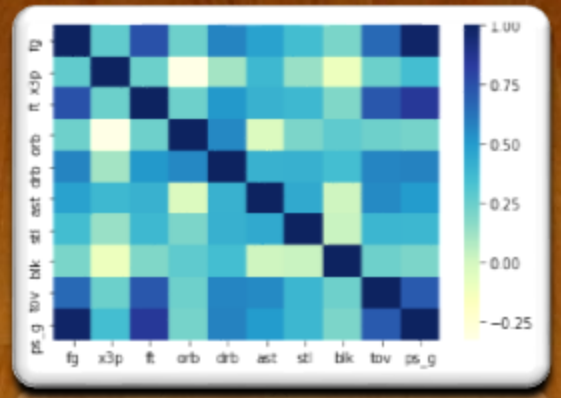
EDA — Guards, Forwards, Center



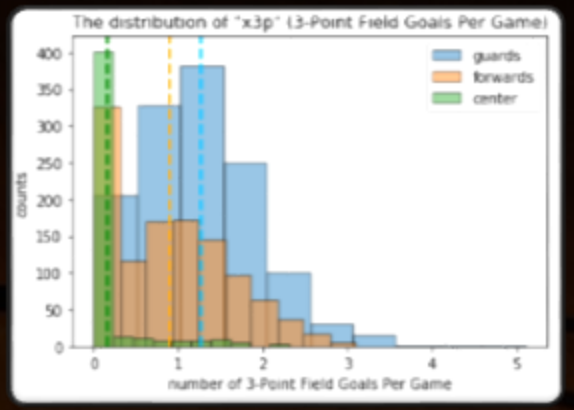
Guards

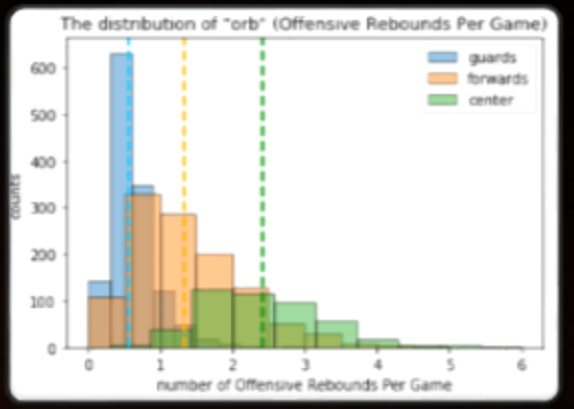


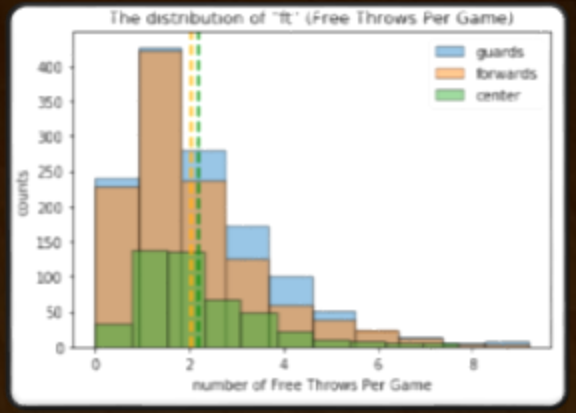
Forwards

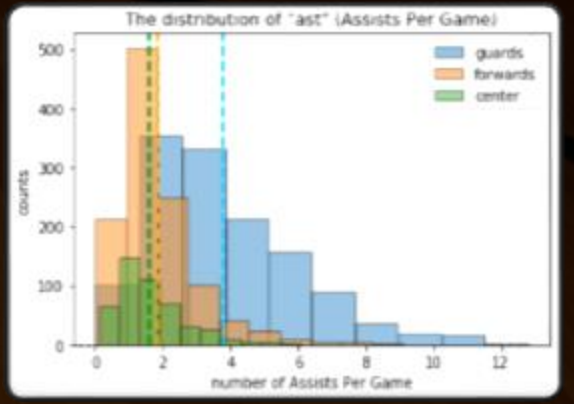


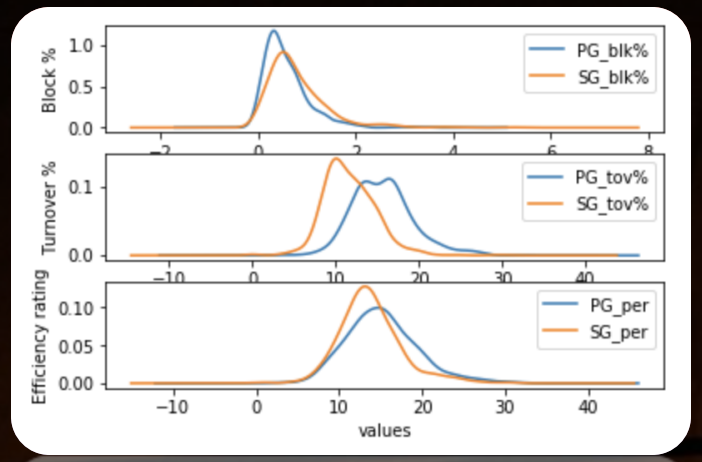
Center

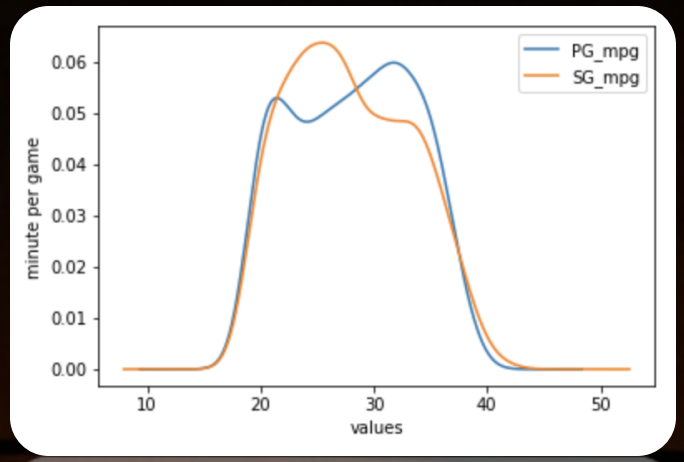


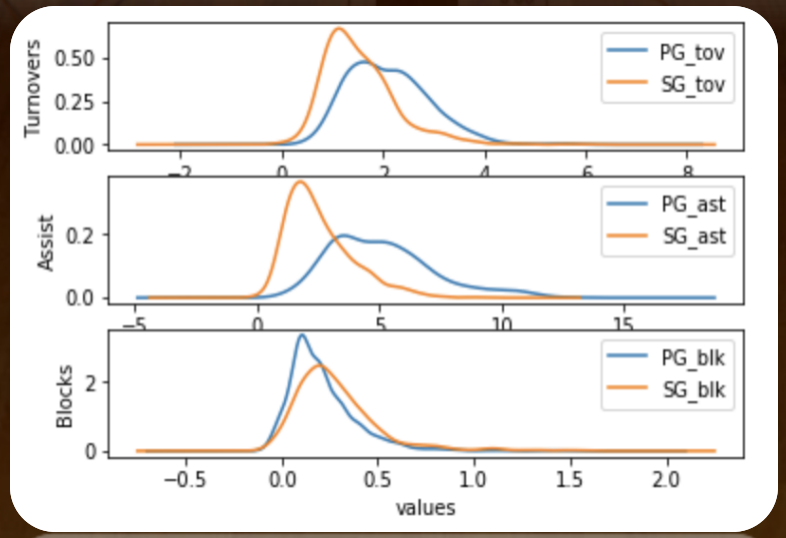












EDA — PG vs SG

After exploring entire features related to player's positions, we found 7 attributes that separate point guards from shooting guards: turnovers, turnover percentage, blocks, blocks percentage, minutes per game, assists and player efficiency rating.

1) Turnovers: Number of turnovers for point guards is generally higher than that of shooting guards

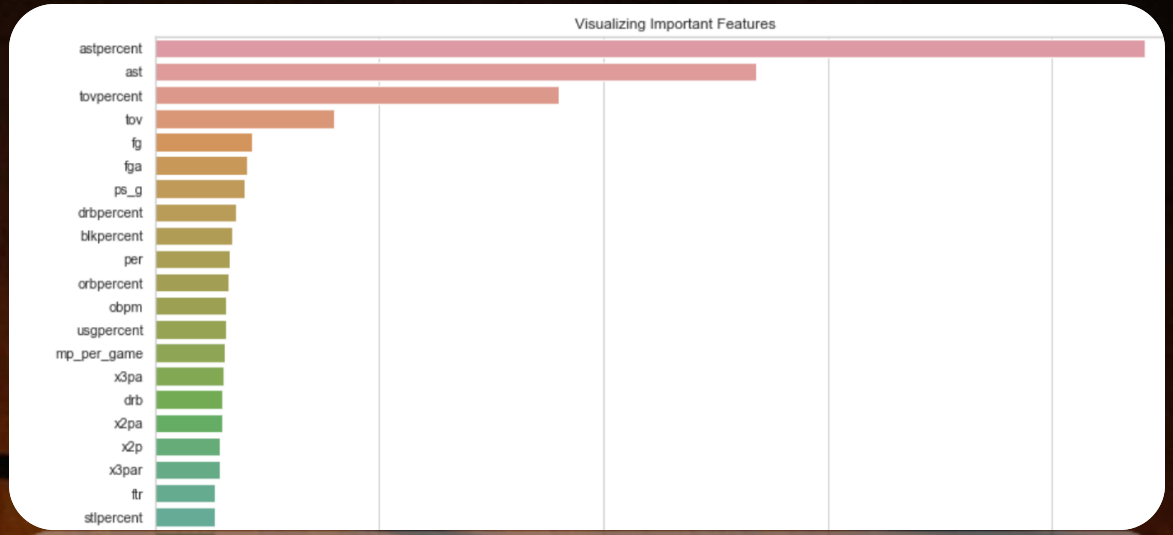
2) Blocks: Number of blocks per game for shooting guards is generally higher than that of point guards

3) Minutes per game: Minutes per game for point guards is generally higher than that of shooting guards

4) Assists: Number of assists for point guards is much higher than that of shooting guards

5) Player efficiency rating: PER for point guards is slightly higher than that of shooting guards

- Model category: Predictive Binary Classification Model (PG/ SG)
- Split data into "Russell" (case study) and "Others" (training model)
- KNN Method
  - The number of training observations: K = 1~10
  - Cross Validation: n_splits = 5, test_size = 0.25
  - CV Accuracy: 0.83
- Random Forest
  - test_size = 0.25, n_estimators=250
  - Feature Selection
  - Accuracy: 0.70
- Logistic Regression (final model):
  - Feature Selection (p-value)
  - Good Interpretability (coefficients)
  - CV Accuracy: 0.87



	Coef.	Std.Err.	z	P>|z|
age	-0.1262	0.0251	-5.0258	0.0000
gs	0.0181	0.0051	3.5489	0.0004
fga	-0.2942	0.0714	-4.1232	0.0000
x2pa	-0.1299	0.0737	-1.7611	0.0782
x2ppercent	-5.5782	2.1299	-2.6190	0.0088
mp_per_game	-0.0701	0.0488	-1.4353	0.1512
orb	3.3130	1.5977	2.0735	0.0381
tov	-0.4836	0.2584	-1.8716	0.0613
pf	0.3999	0.2225	1.7973	0.0723
orbpercent	-0.7600	0.3862	-1.9676	0.0491
drbpercent	-0.3078	0.0433	-7.1104	0.0000
astpercent	0.3669	0.0230	15.9765	0.0000
stlpercent	0.3501	0.1806	1.9387	0.0525
intercept	5.9802	1.6353	3.6569	0.0003

Model exploration

Case Analysis & Conclusions

Apply Logistic Model on Russell Westbrook:
The play style of today's point guards may not benefit the team

Generalization from EDA and Data Modelling:

- Compare and Contrast significant attributes derived from EDA and data modelling

- Summarize the point guard's distinctive feature in terms of offense and defense:
  - Control of the offense
  - Inactive Defensive
  - 3 pointers instead of mid range

	player	prediction	pos	i
	Russell Westbrook	PG	PG	2009
	Russell Westbrook	PG	PG	2010
	Russell Westbrook	PG	PG	2011
	Russell Westbrook	SG	PG	2012
	Russell Westbrook	PG	PG	2013
	Russell Westbrook	PG	PG	2014
	Russell Westbrook	PG	PG	2015
	Russell Westbrook	PG	PG	2016
	Russell Westbrook	PG	PG	2017
	Russell Westbrook	PG	PG	2018