Predicting NBA Player FanDuel Score

By: Dong Zhen

About this project.

What is FanDuel?

A sports focused gambling website that offers sportsbook, daily fantasy, and horse race betting

Objective

Help daily fantasy basketball betting group place in the top 23% **by predicting NBA players' score in FanDuel**

FanDuel's point system.

FD Basketball Scoring as of 5/6/2021

FGM	FTM	3PM	REB	AST	STL	BLK	TO
2	1	1	1.2	1.5	3	3	-1

	FGM	FTM	3PM	REB	AST	STL	BLK	TO
Russell Westbrook	13	6	0	11	4	4	0	5
FanDuel Score	26	6	0	13.2	6	12	0	-5

Total Points: 58.2

2017 - 19
NBA Data.

Players

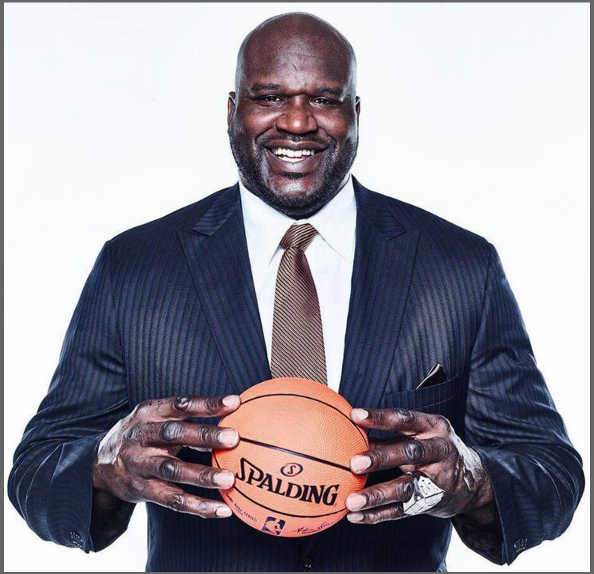
- List of players
- Minimum of 30 games a season
- Minimum 15 minutes per game

Teams

- NBA team rosters
- Team statistics

Why
2017-19?

Centers Dominate



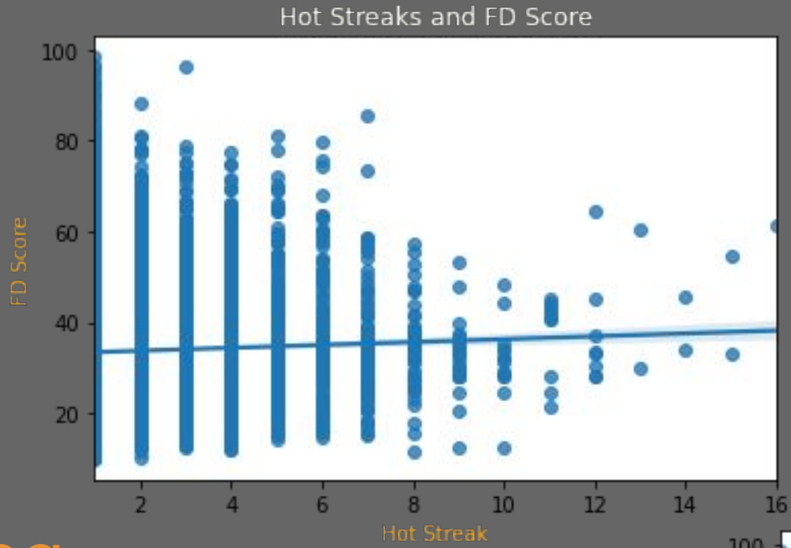
Guards Dominate

Generalizing FD Score.

Player tendencies + Player attributes + Momentum + Opponent’s defensive tendencies = **FD Score**

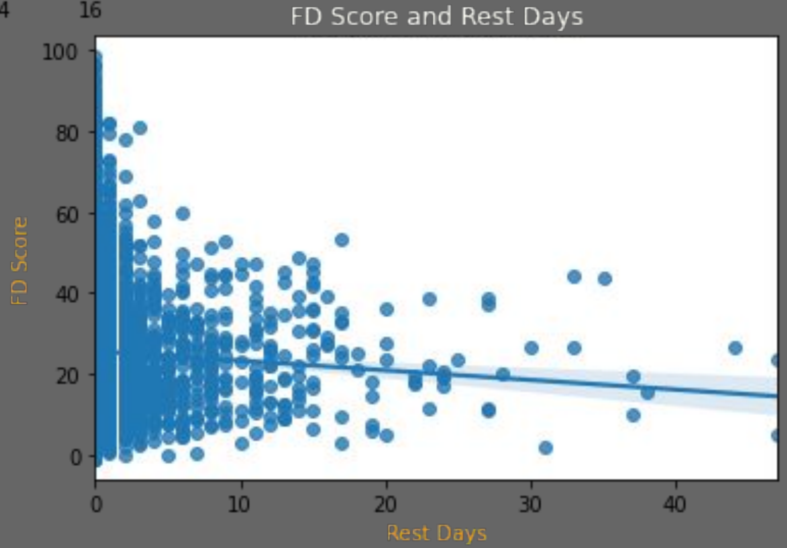
- ❑ Offensive activity example: FG
- ❑ Defensive activity example: fouls

Feature Designing.

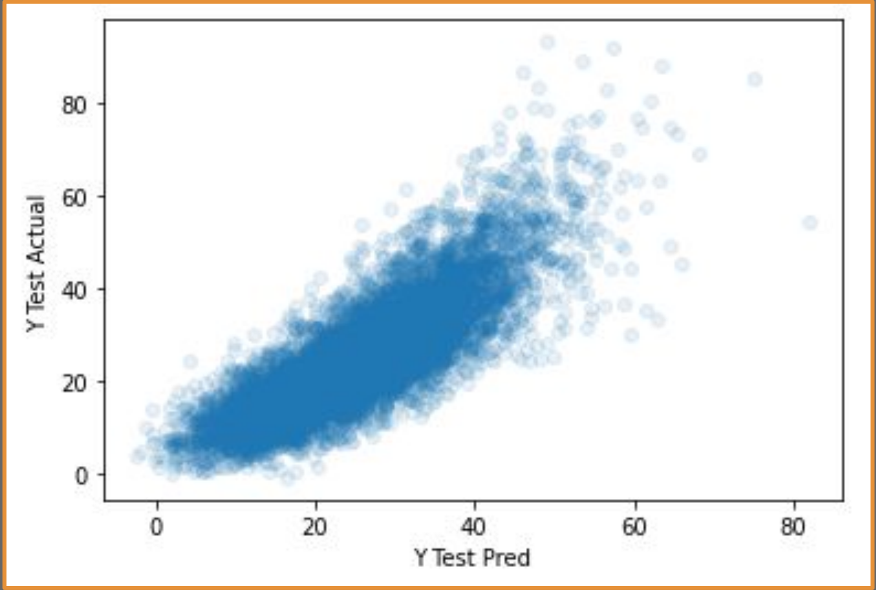


Hot streaks indicate momentum

Rest days are measured by number of games between previous and current

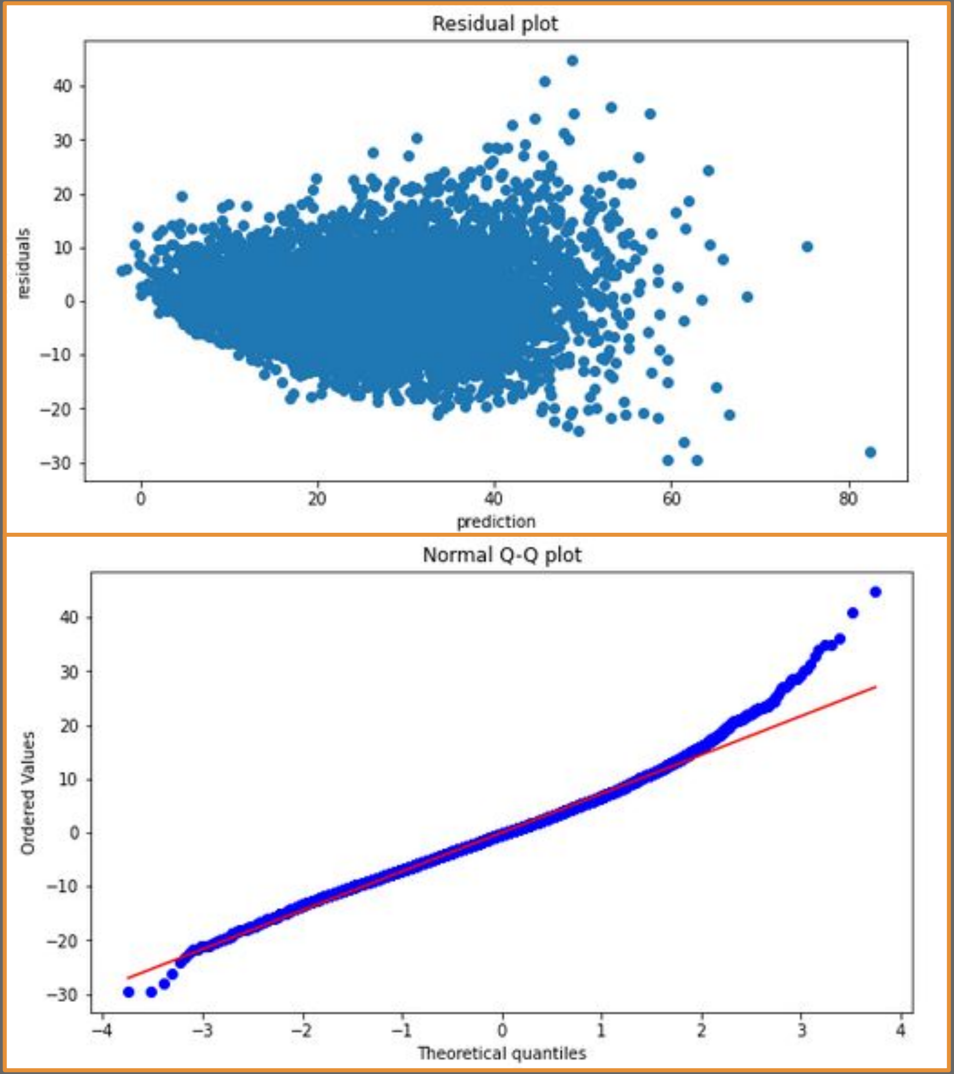


Using Linear Regression.

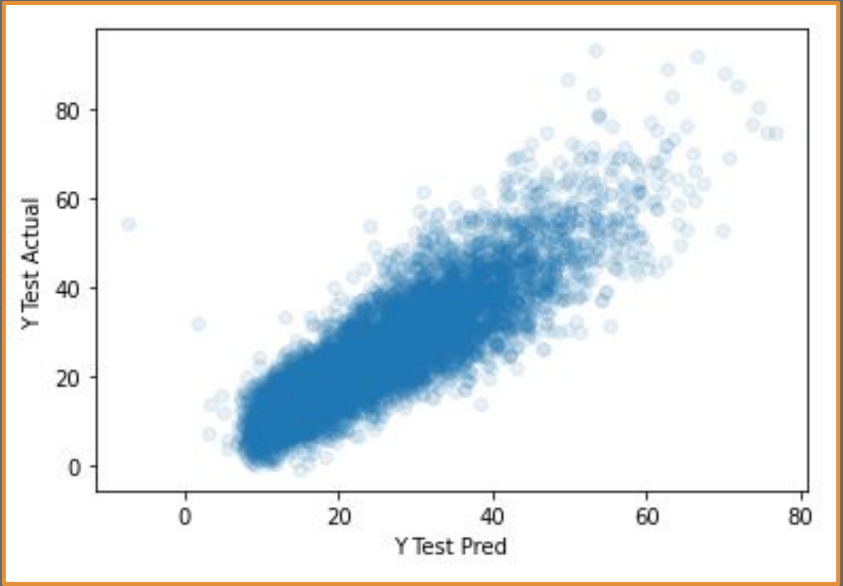


Training R^2 Score: 0.6850205173145631
Training RMSE: 7.256913673779181
Training MAE: 5.590420023489153
Test R^2 Score: 0.6844943075416979
Test RMSE: 7.256913673779181
Test MAE: 5.614603579545828
Cross validation scores : [0.68404225 0.68644522 0.67033296 0.68906962 0.69267176]
Cross validation scores : 0.6845123609680444

Checking for OLS Assumptions

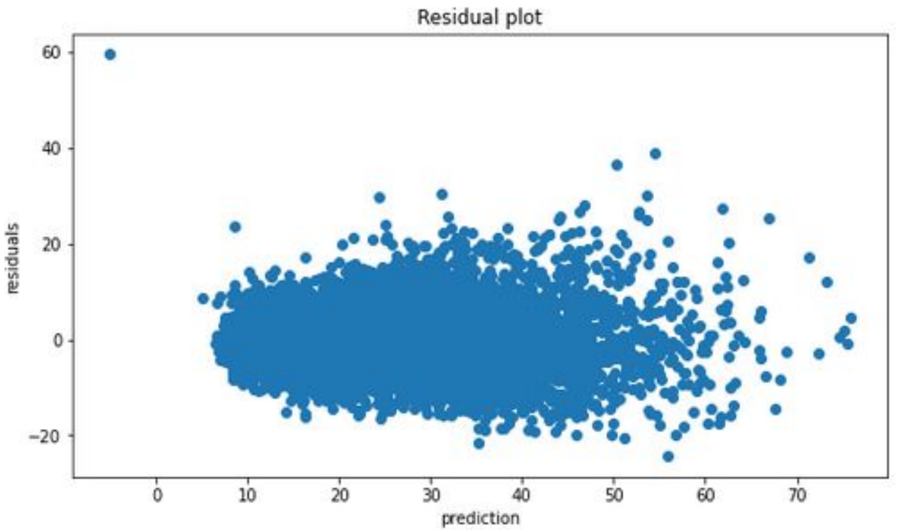


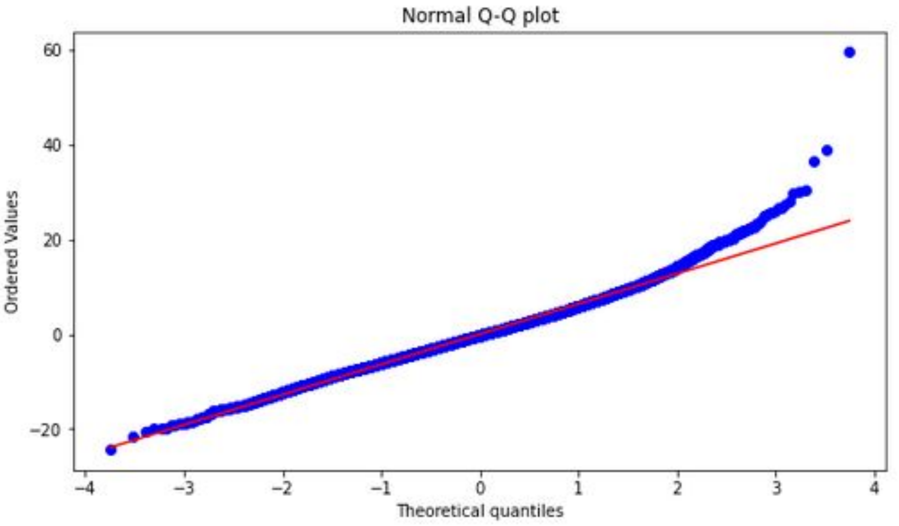
Polynomial to the second degree



Training R^2 Score: 0.751815672294646
Training RMSE: 6.441663002232507
Training MAE: 4.909333544347451
Test R^2 Score: 0.7436806607768427
Test RMSE: 6.441663002232507
Test MAE: 5.031599774824723
Cross validation scores: [0.74790621 0.74734498
0.74166187 0.74022639 0.75743132]
Cross validation scores: 0.7469141550672334

Residuals
Improvement.





Prediction.

```
This was a really close prediction without any of the changes that I made
```

```python
test_df2 = fds_df.sample(random_state = 7490)
test_df2.head()
```

	Date	Age	Name	Team	Opp	Minutes Played	Personal Fouls	Rest Days	FD Score	FGA/M	...	Weight	Exp	Opp_Avg_Pace	Opp_Avg_Off_Rtg	Opp_Avg_Def_Rtg	Games Started_Starter	Position_PF	Posit
18639	2018-01-19	28	Markieff Morris	WAS	DET	31.55	4	0	31.8	0.44374	...	245.0	6.0	96.2	107.2	107.3	1	1	

1 rows × 25 columns

```python
test_df_f2 = test_df2[['Age', 'Minutes Played', 'Personal Fouls', 'Rest Days', 'FGA/M',
                    'Hot Streak', 'Prev FD Score', 'Weight', 'Exp', 'Opp_Avg_Pace',
                    'Opp_Avg_Off_Rtg', 'Opp_Avg_Def_Rtg', 'Games Started_Starter',
                    'Position_PF', 'Position_PG', 'Position_SF', 'Position_SG']]
```

```python
test_df_f2.head()
```

	Age	Minutes Played	Personal Fouls	Rest Days	FGA/M	Hot Streak	Prev FD Score	Weight	Exp	Opp_Avg_Pace	Opp_Avg_Off_Rtg	Opp_Avg_Def_Rtg	Games Started_Starter	Position_PF	Position_PG	Position_SF	Positior
18639	28	31.55	4	0	0.44374	1	8.5	245.0	6.0	96.2	107.2	107.3	1	1	0	0	

```python
lm.predict(test_df_f2)
```

```
array([[30.8207839]])
```

```python
test_df_t2 = test_df2['FD Score']
test_df_t2
```

```
18639    31.8
Name: FD Score, dtype: float64
```

Future Work.

Add more **complexity**, use Selenium to get important per game player and opponent tendencies

Look into what is causing the **outliers**

Incorporate **FD salary** to help client make the best budget picks