163 Final Project

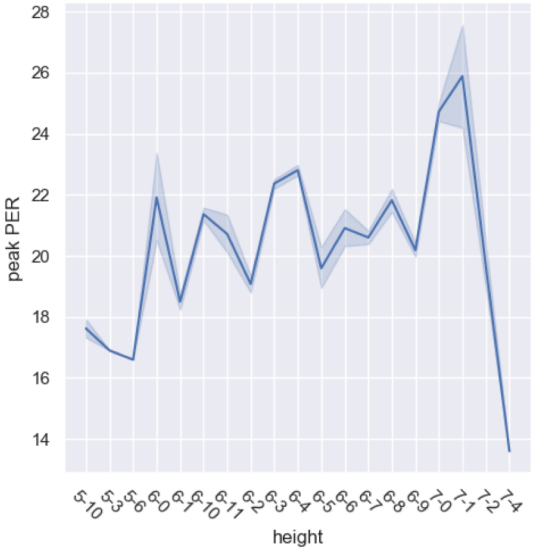
Predicting basketball players' peak age and decline

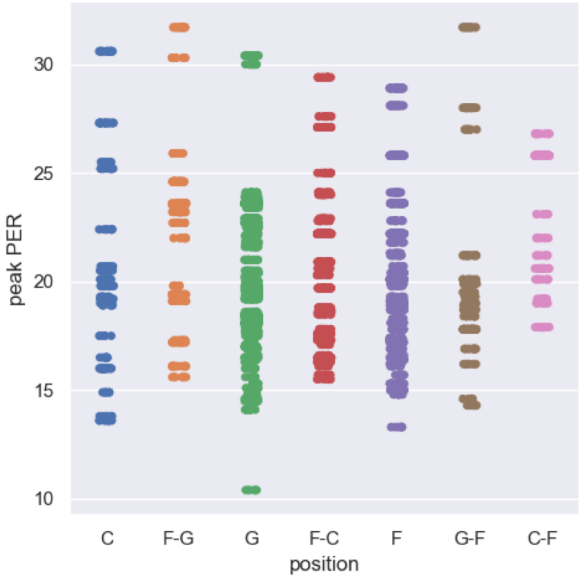
Preliminary Computations

Before we began the project, we computed average MVP and all star age, and plotted different graphs. We have computed values and graphs below.

Average age of All-Star: 26.5 Years

Average age of MVP: 27.9 Years





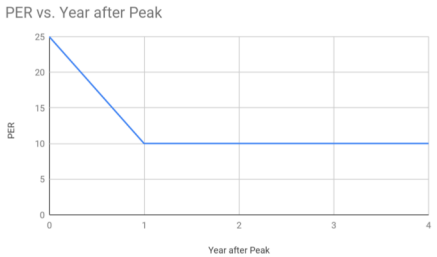
Introduction

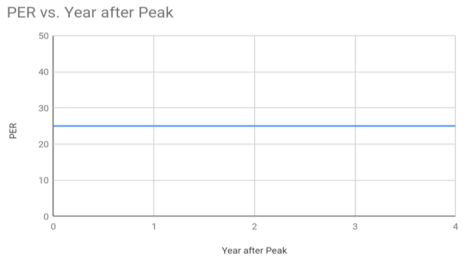
Used indicators:

- Used Players Efficiency Rating for performance, and to determine peak age

- Used the area of the in order to determine the magnitude of players' decline

Worst Case(52.5 unit of decline) vs Best case(0 unit of decline); area of the rectangle with height of max PER and width with 4 - area of the PER graph





Part 1 execution

Investigated the relationship between the players' peak age and variables: height, weight, position, and age when entering the league.

Used decision tree regressor

Usd Mean Squared Error to measure accuracy

```
uPER = (1 / MP) *
     [ 3P
     + (2/3) * AST
     + (2 - factor * (team_AST / team_FG)) * FG
     + (FT *0.5 * (1 + (1 - (team_AST / team_FG)) + (2/3) * (team_AST / team_FG)))
     - VOP * TOV
     - VOP * DRB% * (FGA - FG)
     - VOP * 0.44 * (0.44 + (0.56 * DRB%)) * (FTA - FT)
     + VOP * (1 - DRB%) * (TRB - ORB)
     + VOP * DRB% * ORB
     + VOP * STL
     + VOP * DRB% * BLK
     - PF * ((lg_FT / lg_PF) - 0.44 * (lg_FTA / lg_PF) * VOP) ]
```

Part 2 execution

Project Part 2

Investigated the relationship between the players’ decline and variables: 'position', 'beginning_age', 'peak age', 'peak PER', 'decline measure', 'USG%', 'FG%', 'FG', '3P', 'AST', 'STL', 'BLK', 'VORP', '3PA','WS' and 'BPM.'

Used decision tree regressor

Usd Mean Squared Error to measure accuracy

Conclusion

MSE for peak age: approximately 8

MSE for decline measure: approximately 36

Speculation: It seems like there is no strong correlation between height, weight, ect. And the age a player peaks at. It seems like there are a lot more variables we are not taking into account, and if we had access to a player's high-school data, we may get more accurate results. But this also shows that players are still moldable after they enter the league, and development can matter a lot.

Used sources

https://www.kaggle.com/drgilermo/nba-players-stats

(originally from basketball-reference.com)

Contains data of players' bodies' data and their performance for every season after 1950.

https://www.kaggle.com/open-source-sports/mens-professional-basketball

Contains data of the awards for each player.

Python libraries:

-Pandas

-Scikit learn

-matplotplib