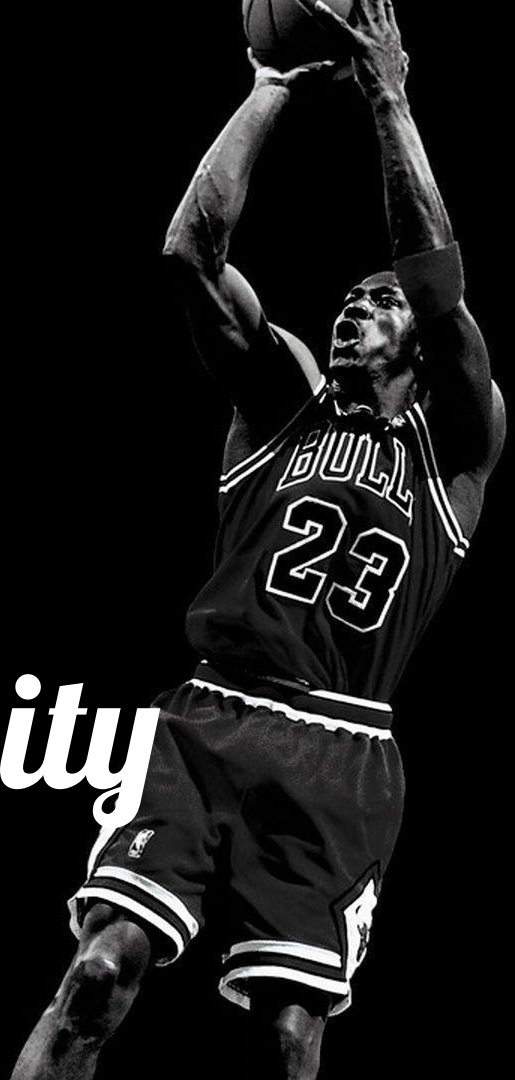
NBA Project

Predicting players longevity



By Marc INIZAN

Why did I chose this project ?

This project is part of a larger project that aims to predict best picks for **NBA Fantasy Leagues** lineups.

Knowing the longevity of a **new player** coming to the NBA will allow me to **extrapolate his performances** using similar NBA players's data.

Predicting NBA players longevity

The goal of this project is to predict new players longevity in the NBA.

Short: less than 8 seasons

Long: 8 or more seasons

What Data did I use?

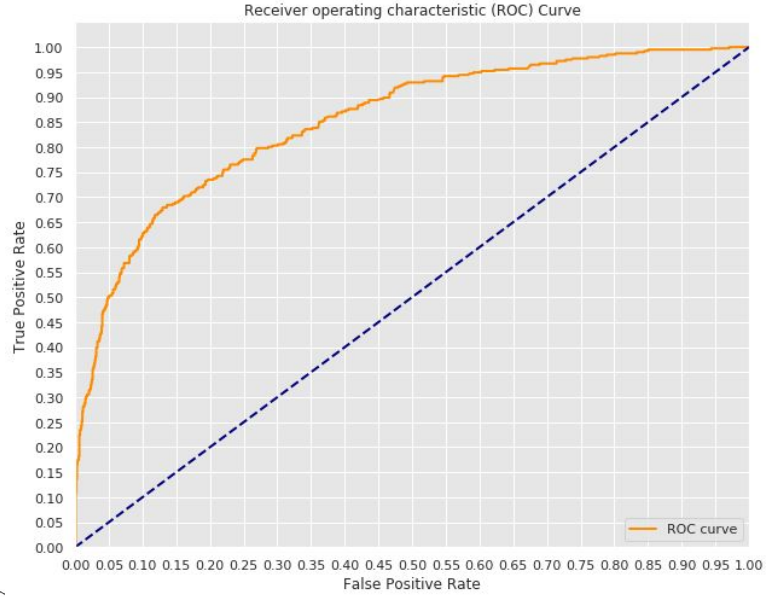
-Dataset from the 1996/1997 - 2016/2017 seasons.

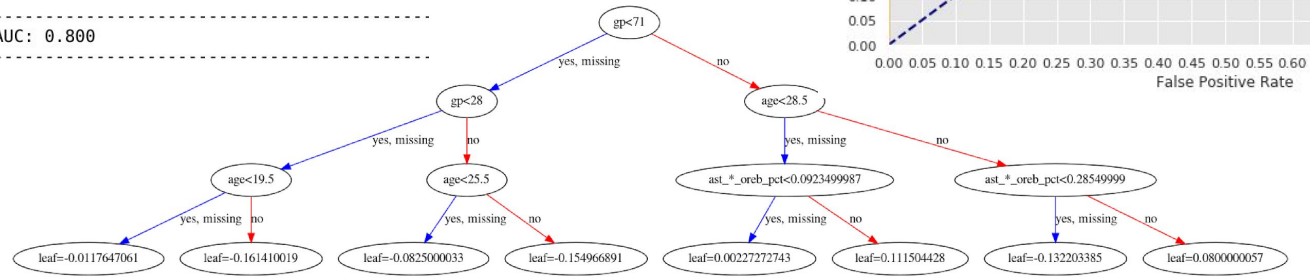
-Each row contains the stats of one player for one season.

-Variables used: age, game played, points, rebounds, assists, etc...

XGBoost Metrics

```
------------------------------------------------------------

Grid search XGBoost Classifier metrics on TRAIN DATASET
ROC AUC Score: 0.8508575007008311

------- Metrics for threshold 0.5 -------
- Precision Score: 0.7907949790794979
- Recall Score: 0.4772727272727273
- Accuracy Score: 0.8301387970918704
- F1 Score: 0.5952755905511812

------------------------------------------------------------

Grid search XGBoost Classifier metrics on TEST DATASET

ROC AUC Score: 0.8038312495430285

------- Metrics for threshold 0.5 -------
- Precision Score: 0.6666666666666666
- Recall Score: 0.4742268041237113
- Accuracy Score: 0.8047493403693932
- F1 Score: 0.5542168674698795

------------------------------------------------------------
Validation ROC AUC: 0.800
------------------------------------------------------------
```



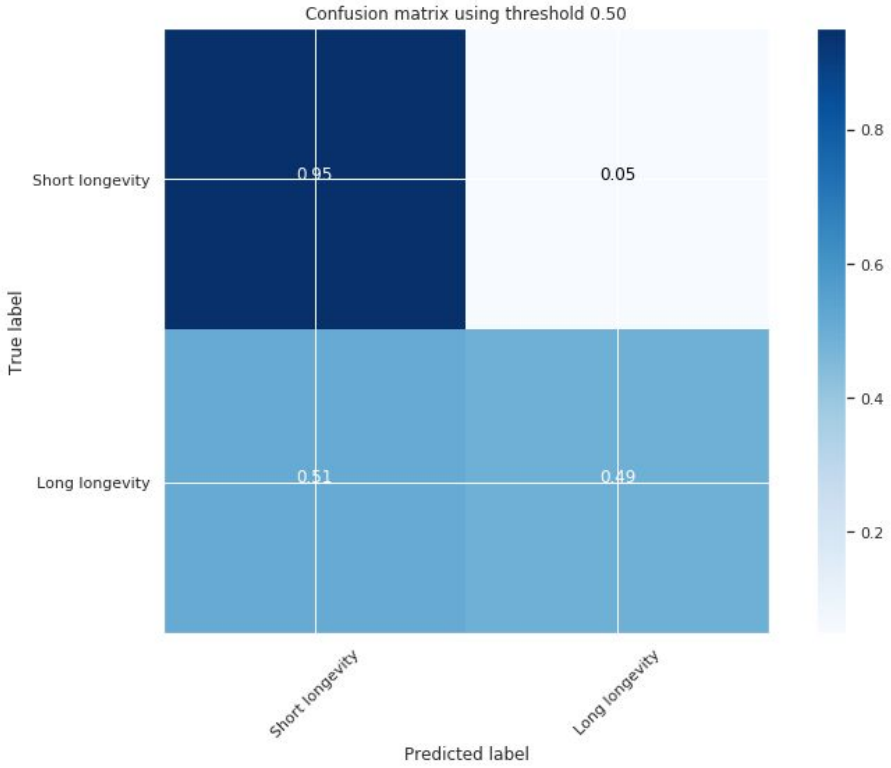


Confusion Matrix

For this project the impact of having a high recall and a low precision or the opposite is not critical.

FPR costs and FNR cost were set at the same value.

Giving a threshold of 0.496



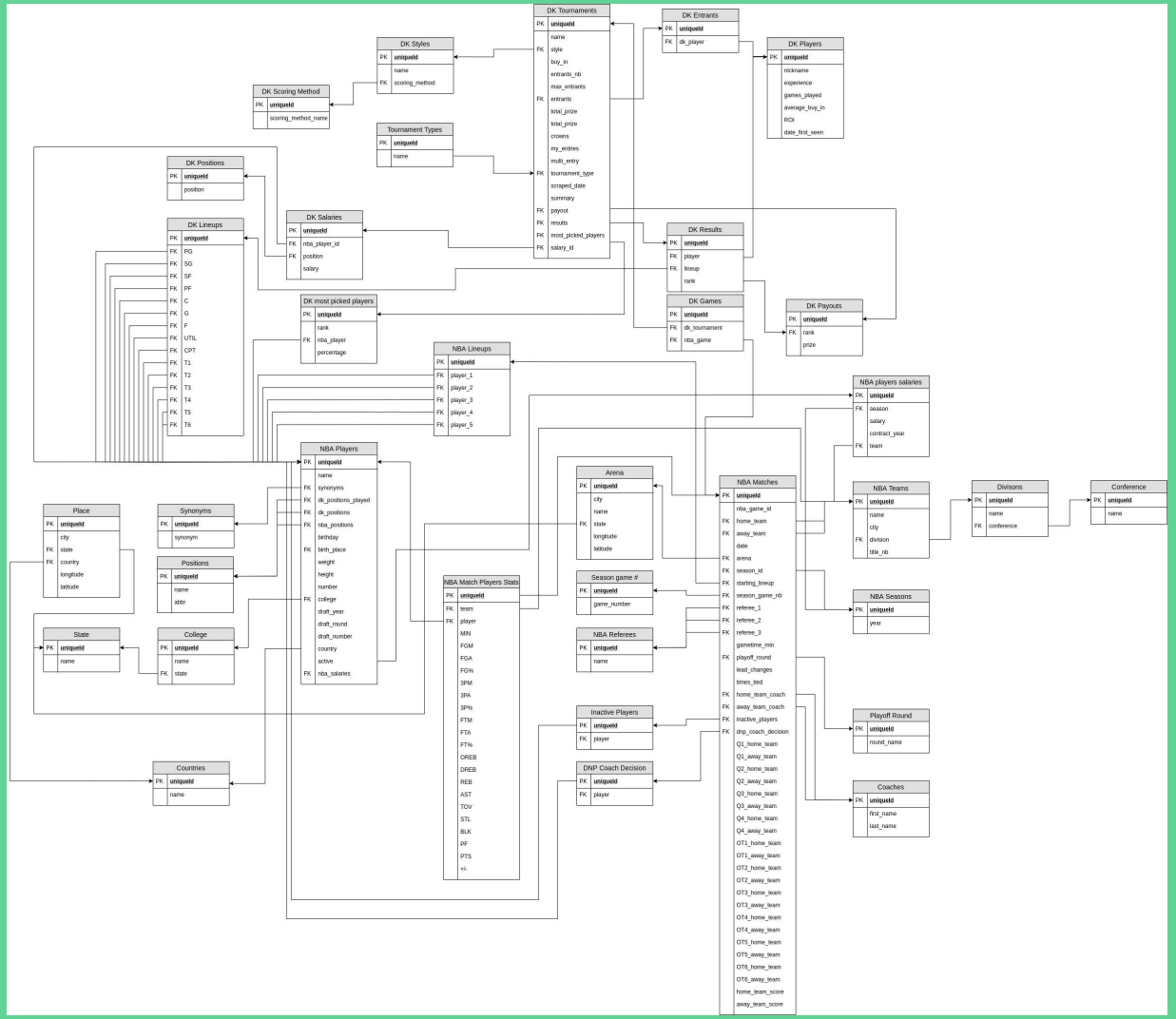
The bigger project

Building the best lineups for NBA fantasy leagues on DraftKings.co.uk

What data do I need ?

- NBA Data (players stats)
- Daily NBA players DFS price
- Prizepool and number of entrants
- Twitter Data (players out today?)

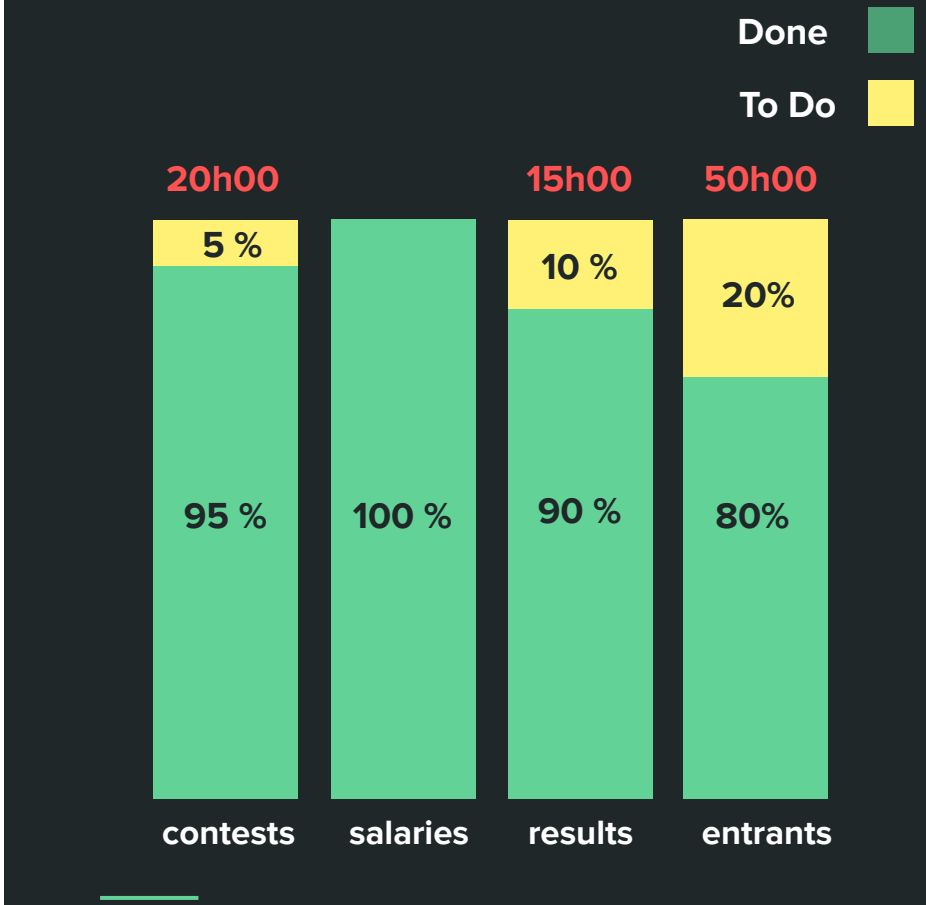
MySQL DataBase Schema



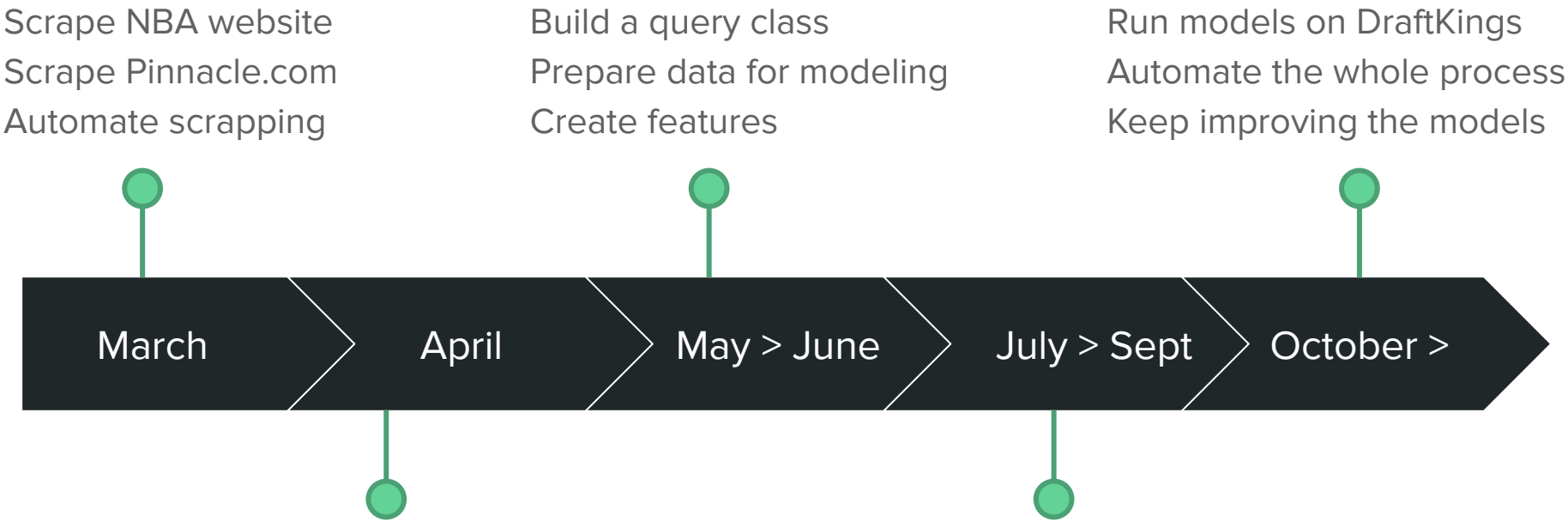
Collecting Data from DraftKings

What's left to be done in term of data scraping on DraftKings

The data collected is stored in csv files, generating 100Mb everyday.



What's next?


Scrape NBA website
Scrape Pinnacle.com
Automate scrapping
Build a query class
Prepare data for modeling
Create features
Run models on DraftKings
Automate the whole process
Keep improving the models
March
April
May > June
July > Sept
October >
Everything into a DB (MySQL)
Optimize the DB (indexing, chose optimal db engines etc...)
Create models
Back test models
Improve models

Conclusion

While this project is only the first of a long list of other projects fitting into a single huge project, it should definitely bring value when deployed during the next NBA season!

Questions & Answers