GA -DAT 19

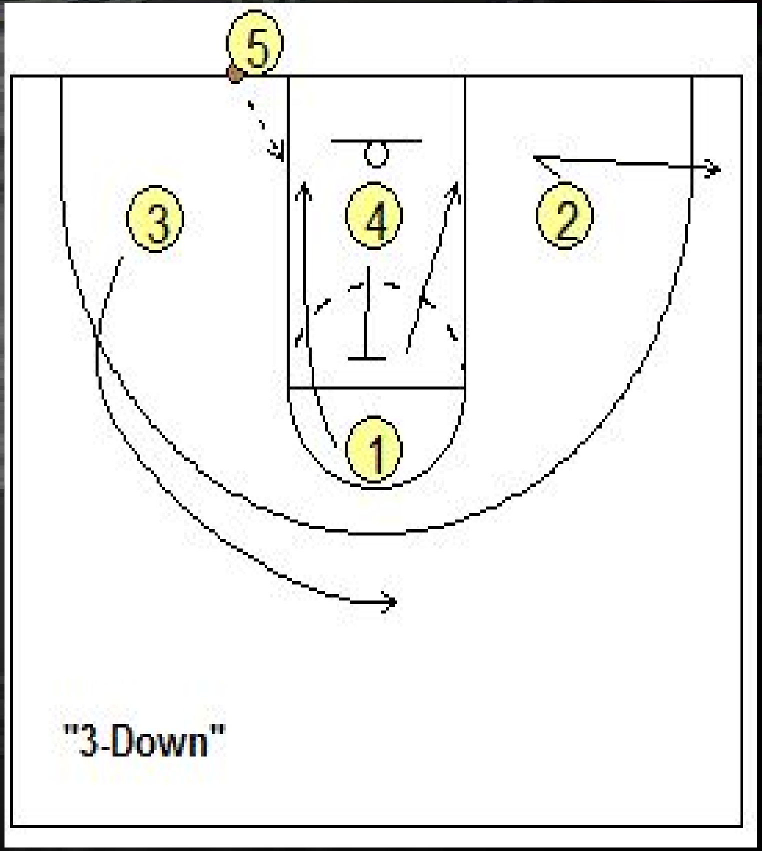
Predicting NBA Basketball Shots
By Dexter Aguia

National Basketball Assocation

- Basketball is a sport played by two teams of five players on a rectangular court. The objective is to shoot a ball through a hoop 18 inches (46 cm) in diameter and 10 feet (3.048 m) high mounted to a backboard at each end.
- 1 point penalty shot, 2 point field goal, 3 point field goal,

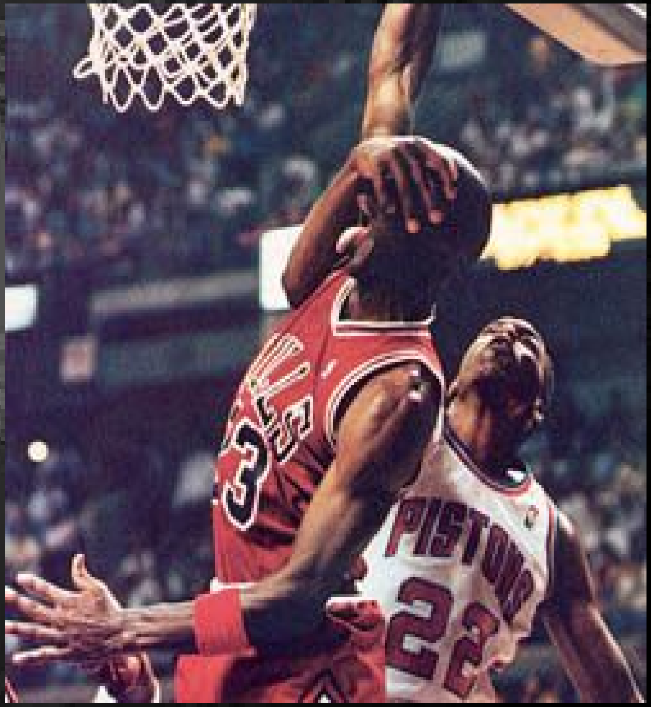
Problem to solve

- Predict probability of shooting accuracy based on player, time, shot distance data given matchup against a specific team
- Coaches create set plays
  ->Set play strategically planned and choreographed sequence of movements to get open shots and score points
- How to cater these plays based on the team you are playing against



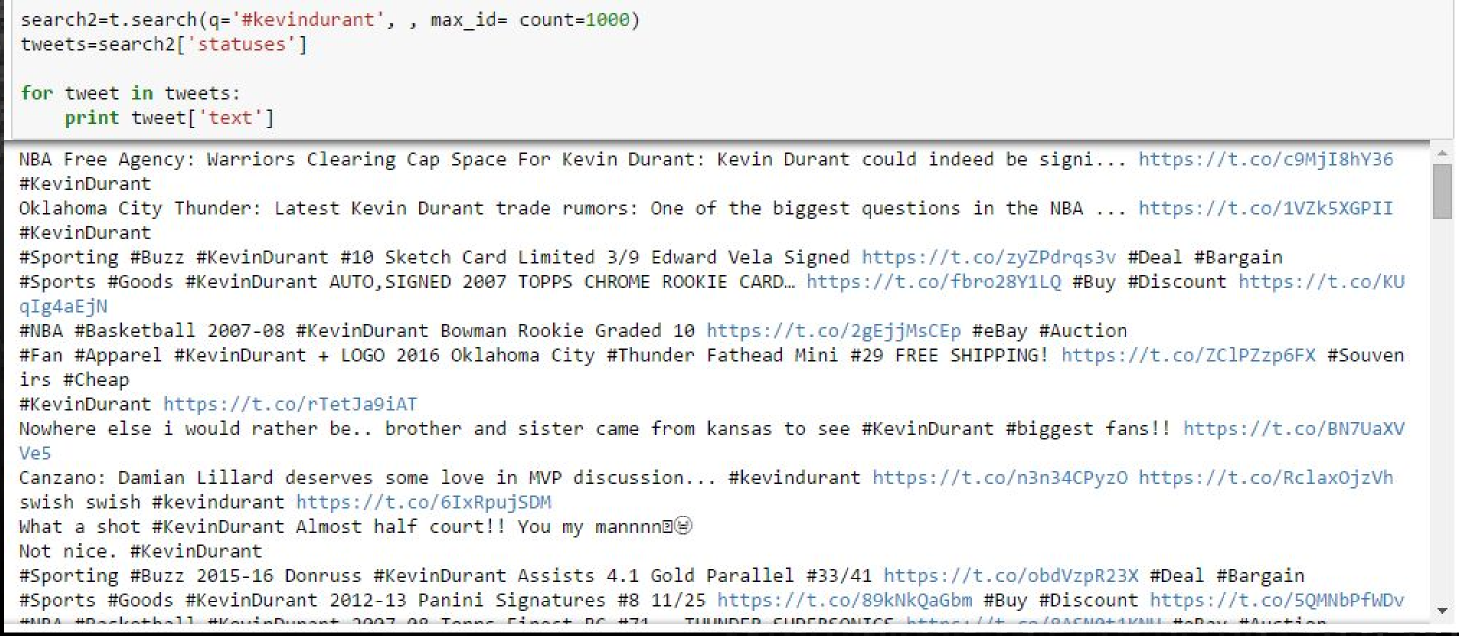
Inspiration

- The Golden State Warriors are ridiculous
- Super-Teams
- The Jordan Rule



Other Ideas

- Predict player performance based on tweets the within 24 hours prior before the game
- Predict Yelp Review rating based on an event data that is in the same location around the same time

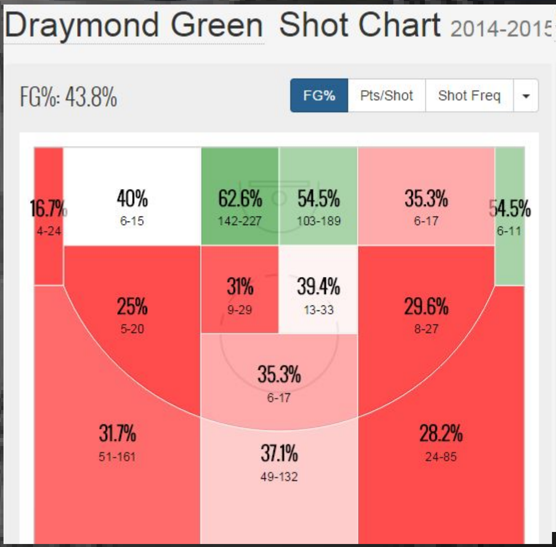


```
search2=t.search(q='#kevindurant', , max_id= count=1000)
tweets=search2['statuses']

for tweet in tweets:
    print tweet['text']
```

```
NBA Free Agency: Warriors Clearing Cap Space For Kevin Durant: Kevin Durant could indeed be signi... https://t.co/c9MjI8hY36
#KevinDurant
Oklahoma City Thunder: Latest Kevin Durant trade rumors: One of the biggest questions in the NBA ... https://t.co/1VZk5XGPII
#KevinDurant
#Sporting #Buzz #KevinDurant #10 Sketch Card Limited 3/9 Edward Vela Signed https://t.co/zyZPdrqs3v #Deal #Bargain
#Sports #Goods #KevinDurant AUTO,SIGNED 2007 TOPPS CHROME ROOKIE CARD… https://t.co/fbro28Y1LQ #Buy #Discount https://t.co/KU
qIg4aEjN
#NBA #Basketball 2007-08 #KevinDurant Bowman Rookie Graded 10 https://t.co/2gEjjMsCEp #eBay #Auction
#Fan #Apparel #KevinDurant + LOGO 2016 Oklahoma City #Thunder Fathead Mini #29 FREE SHIPPING! https://t.co/ZClPZzp6FX #Souven
irs #Cheap
#KevinDurant https://t.co/rTetJa9iAT
Nowhere else i would rather be.. brother and sister came from kansas to see #KevinDurant #biggest fans!! https://t.co/BN7UaXV
Ve5
Canzano: Damian Lillard deserves some love in MVP discussion... #kevindurant https://t.co/n3n34CPyzO https://t.co/RclaxOjzVh
swish swish #kevindurant https://t.co/6IxRpujSDM
What a shot #KevinDurant Almost half court!! You my mannnn😜
Not nice. #KevinDurant
#Sporting #Buzz 2015-16 Donruss #KevinDurant Assists 4.1 Gold Parallel #33/41 https://t.co/obdVzpR23X #Deal #Bargain
#Sports #Goods #KevinDurant 2012-13 Panini Signatures #8 11/25 https://t.co/89kNkQaGbm #Buy #Discount https://t.co/5QMNbPfWDv
#NBA #Basketball #KevinDurant 2007-08 Topps Finest RC #71 … THUNDER SUPERSONICS https://t.co/[illegible] #eBay #Auction
```

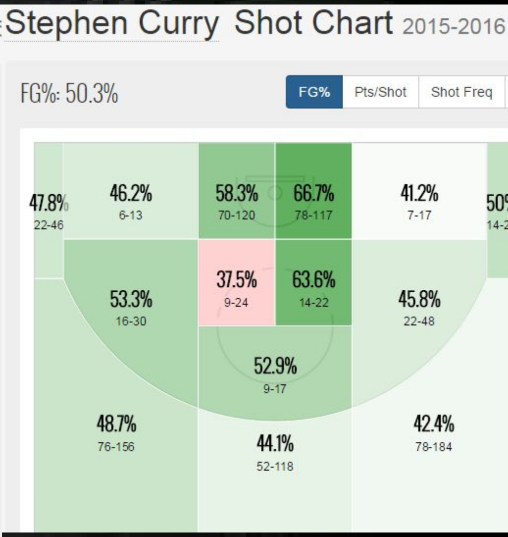
Background information

- Shot Charts
- Game Clock
- Player Positions



Draymond Green Shot Chart 2014-201

FG%: 43.8%

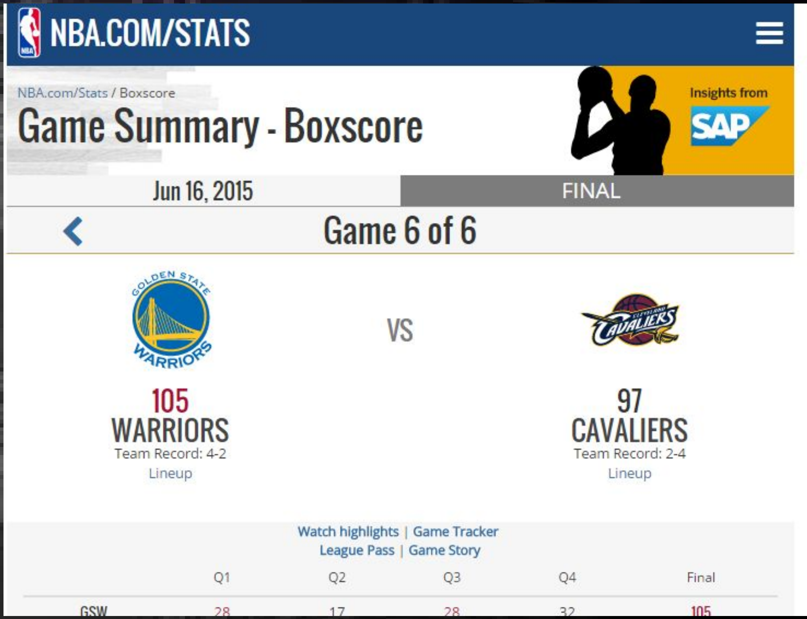


Stephen Curry Shot Chart 2015-2016

FG%: 50.3%

Data Retrieval

- stats.nba.com
- requests module
- Data in JSON
- Problems:
  - -> API documentation hard to find
- Data from Warrior vs Cavaliers championship series 2014-2015



NBA.COM/STATS

NBA.com/Stats / Boxscore

Game Summary - Boxscore

Insights from SAP

Jun 16, 2015 FINAL

Game 6 of 6

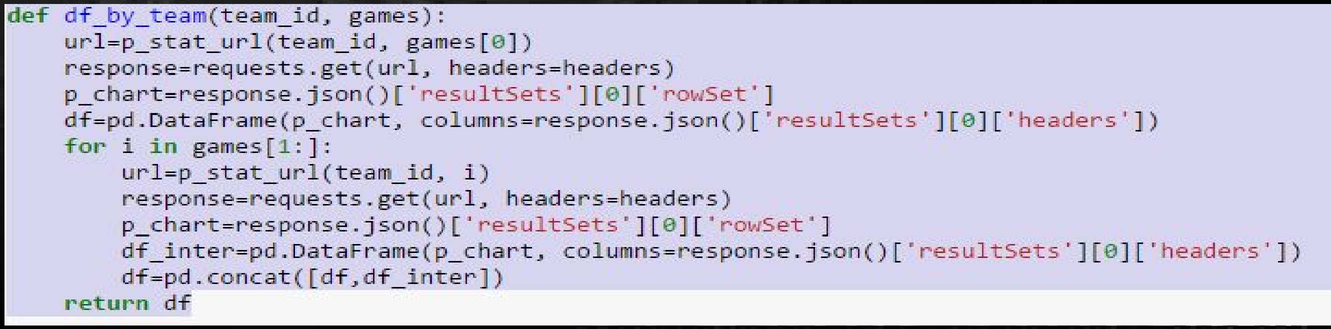
105 WARRIORS
Team Record: 4-2
Lineup

VS

97 CAVALIERS
Team Record: 2-4
Lineup

Watch highlights | Game Tracker
League Pass | Game Story

	Q1	Q2	Q3	Q4	Final
GSW	28	17	28	32	105

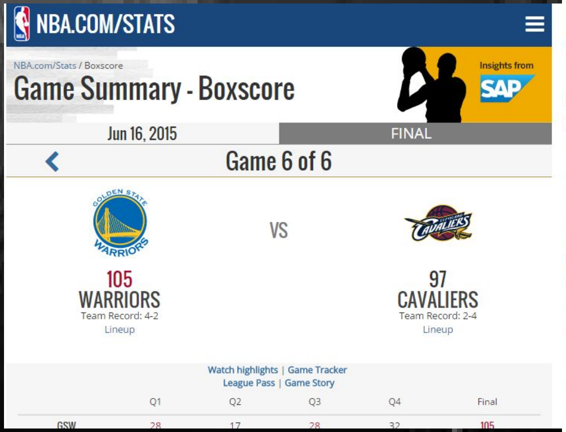


```
def df_by_team(team_id, games):
    url=p_stat_url(team_id, games[0])
    response=requests.get(url, headers=headers)
    p_chart=response.json()['resultSets'][0]['rowSet']
    df=pd.DataFrame(p_chart, columns=response.json()['resultSets'][0]['headers'])
    for i in games[1:]:
        url=p_stat_url(team_id, i)
        response=requests.get(url, headers=headers)
        p_chart=response.json()['resultSets'][0]['rowSet']
        df_inter=pd.DataFrame(p_chart, columns=response.json()['resultSets'][0]['headers'])
        df=pd.concat([df,df_inter])
    return df
```

Data(cont’d)

- Features:
  Game_id
  Game_event_id
  Player_id
  Team_id
  Period
  Minutes_remaining
  Event_type
  Action_type
  Shot_type
  Shot_zone_basic
  Shot_zone_area
  Shot_zone_range
  Shot_distance

- Features(cont’d)
  Loc_x
  Loc_y
  Shot_attempted_flag
  Shot_made_flag
  (dependent
  variable)



NBA.COM/STATS

NBA.com/Stats / Boxscore

Game Summary - Boxscore

Insights from SAP

Jun 16, 2015 FINAL

Game 6 of 6

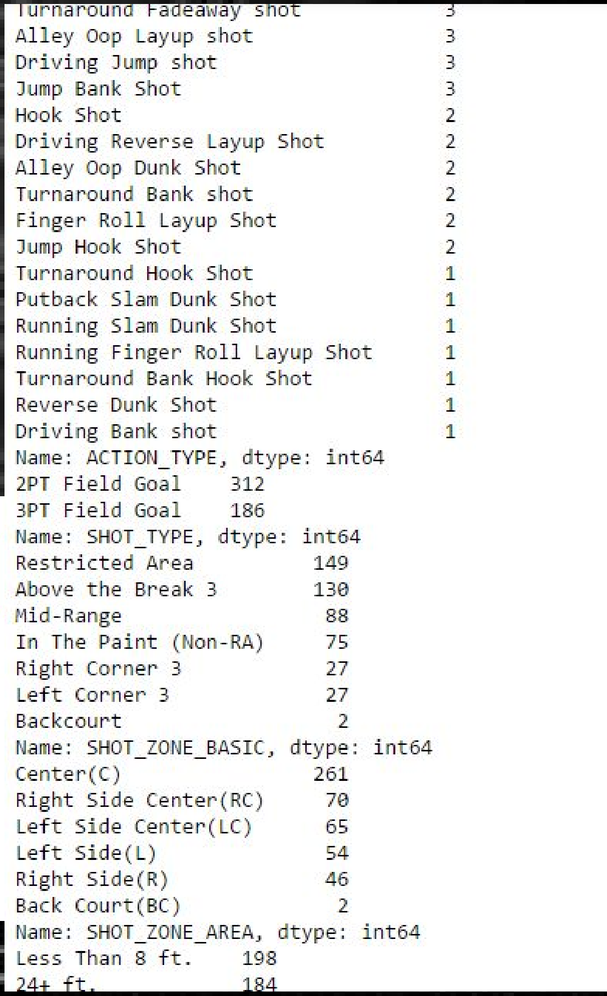
105 WARRIORS
Team Record: 4-2
Lineup

VS

97 CAVALIERS
Team Record: 2-4
Lineup

Watch highlights | Game Tracker
League Pass | Game Story

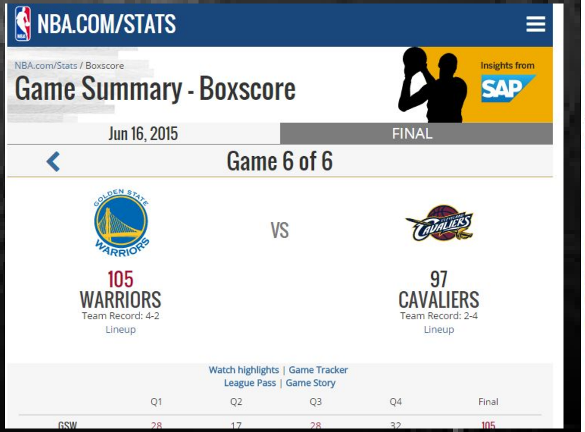
	Q1	Q2	Q3	Q4	Final
GSW	28	17	28	32	105



```
Turnaround Fadeaway shot                    3
Alley Oop Layup shot                        3
Driving Jump shot                           3
Jump Bank Shot                              3
Hook Shot                                   2
Driving Reverse Layup Shot                  2
Alley Oop Dunk Shot                         2
Turnaround Bank shot                        2
Finger Roll Layup Shot                      2
Jump Hook Shot                              2
Turnaround Hook Shot                        1
Putback Slam Dunk Shot                      1
Running Slam Dunk Shot                      1
Running Finger Roll Layup Shot              1
Turnaround Bank Hook Shot                   1
Reverse Dunk Shot                           1
Driving Bank shot                           1
Name: ACTION_TYPE, dtype: int64
2PT Field Goal     312
3PT Field Goal     186
Name: SHOT_TYPE, dtype: int64
Restricted Area           149
Above the Break 3         130
Mid-Range                  88
In The Paint (Non-RA)      75
Right Corner 3             27
Left Corner 3              27
Backcourt                   2
Name: SHOT_ZONE_BASIC, dtype: int64
Center(C)                261
Right Side Center(RC)     70
Left Side Center(LC)      65
Left Side(L)              54
Right Side(R)             46
Back Court(BC)             2
Name: SHOT_ZONE_AREA, dtype: int64
Less Than 8 ft.     198
24+ ft.             184
```

Data(cont’d)

- Feature Creation
  - Time_remaining
    - Combination of Minutes and seconds



NBA.COM/STATS

NBA.com/Stats / Boxscore

Game Summary - Boxscore

Jun 16, 2015 FINAL

Game 6 of 6

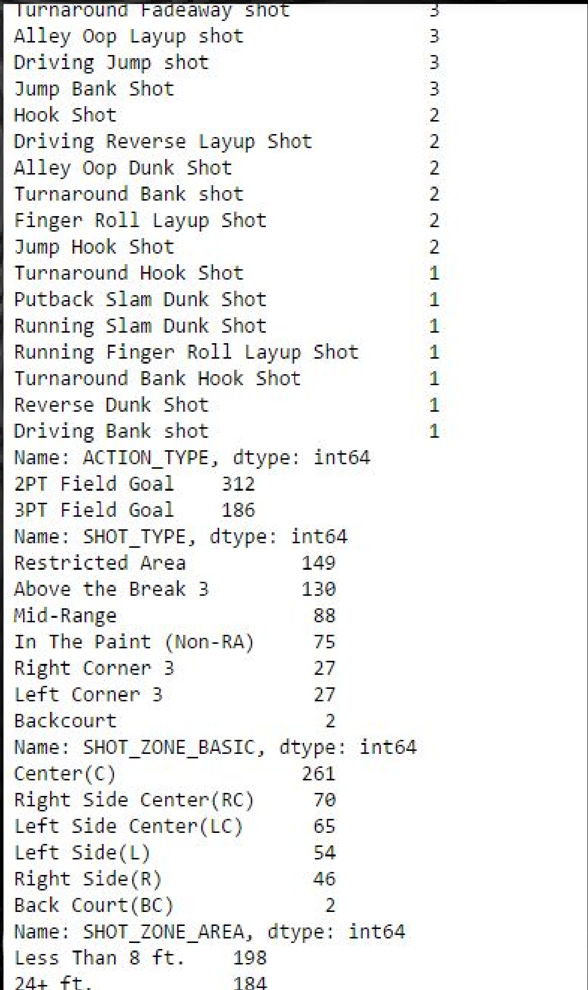
105
WARRIORS
Team Record: 4-2
Lineup

VS

97
CAVALIERS
Team Record: 2-4
Lineup

Watch highlights | Game Tracker
League Pass | Game Story

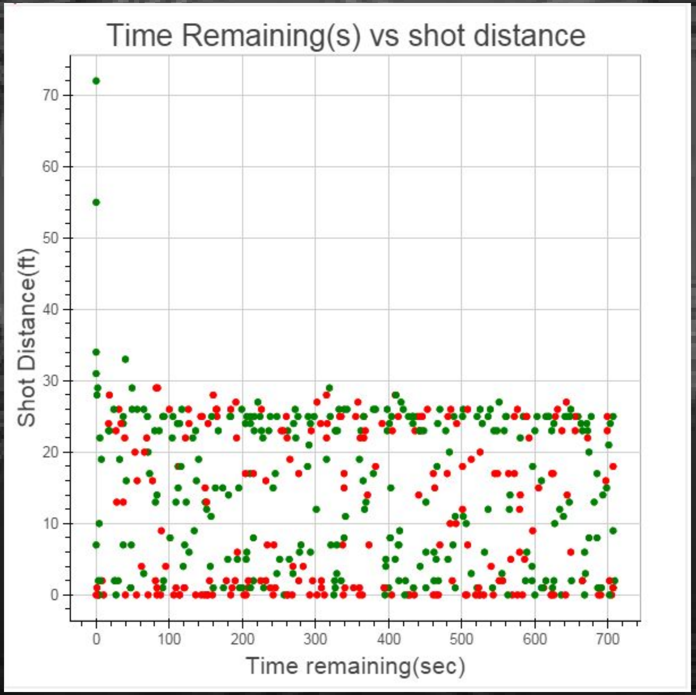
	Q1	Q2	Q3	Q4	Final
GSW	28	17	28	32	105

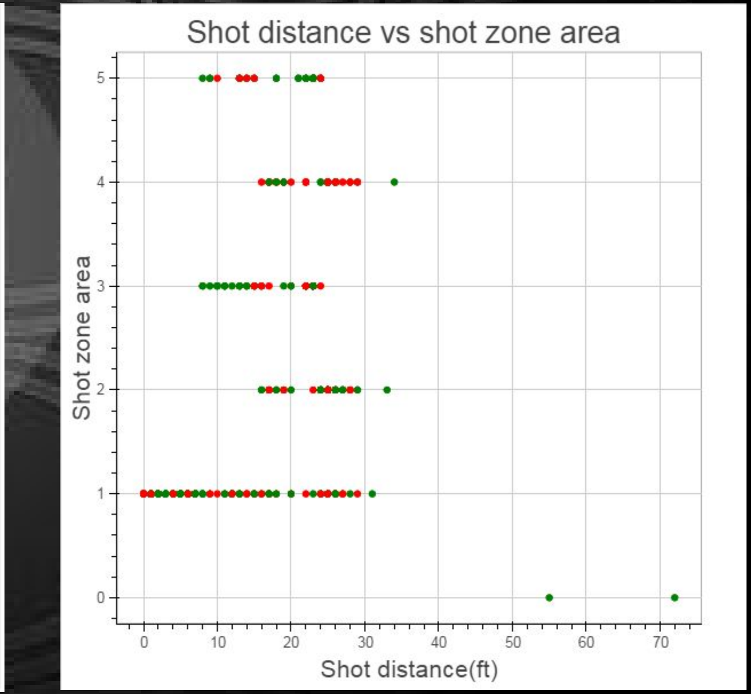


```
Turnaround Fadeaway shot              3
Alley Oop Layup shot                  3
Driving Jump shot                     3
Jump Bank Shot                        3
Hook Shot                             2
Driving Reverse Layup Shot            2
Alley Oop Dunk Shot                   2
Turnaround Bank shot                  2
Finger Roll Layup Shot                2
Jump Hook Shot                        2
Turnaround Hook Shot                  1
Putback Slam Dunk Shot                1
Running Slam Dunk Shot                1
Running Finger Roll Layup Shot        1
Turnaround Bank Hook Shot             1
Reverse Dunk Shot                     1
Driving Bank shot                     1
Name: ACTION_TYPE, dtype: int64
2PT Field Goal    312
3PT Field Goal    186
Name: SHOT_TYPE, dtype: int64
Restricted Area          149
Above the Break 3        130
Mid-Range                 88
In The Paint (Non-RA)     75
Right Corner 3            27
Left Corner 3             27
Backcourt                  2
Name: SHOT_ZONE_BASIC, dtype: int64
Center(C)                261
Right Side Center(RC)     70
Left Side Center(LC)      65
Left Side(L)              54
Right Side(R)             46
Back Court(BC)             2
Name: SHOT_ZONE_AREA, dtype: int64
Less Than 8 ft.    198
24+ ft.            184
```

Exploratory Graphs

- Time Remaining vs shot Distance
- Shot Distance vs shot Zone area





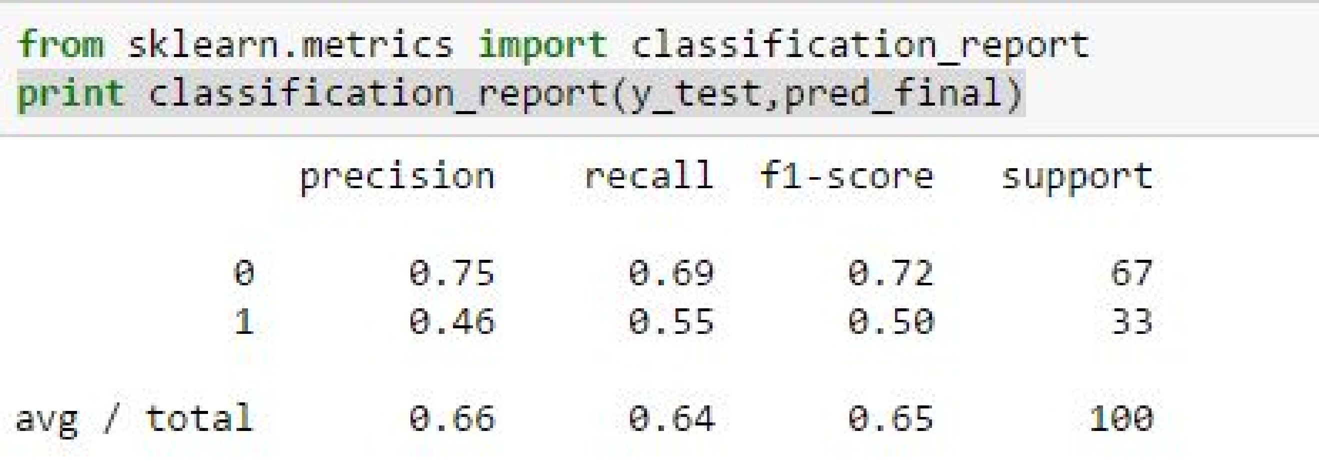
Legend:
Made
Missed

Classifiers

- Supervised Classification
- Classifiers tested
  - Logistic Regression
  - Random Forests
  - Support Vector Machines
  - Naive Bayes

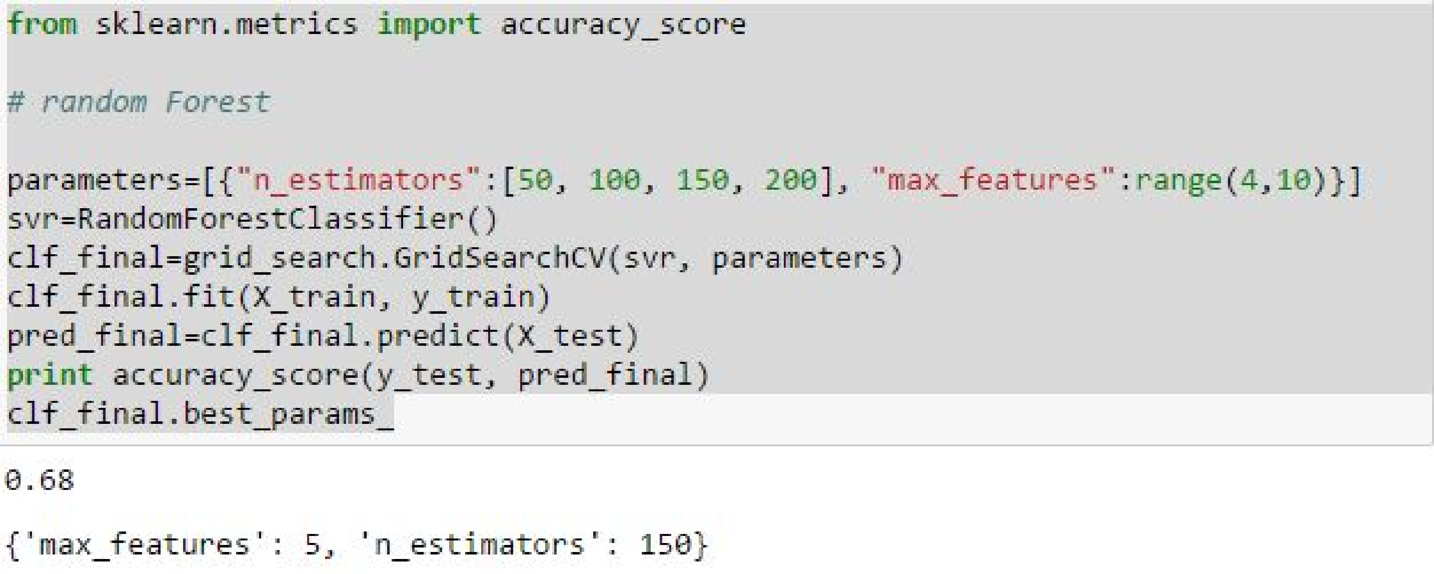
Random Forests

- Random Forests provided the best with an accuracy of 0.68, using max_features=5 and 150 estimators or forests
- Other Classifiers
  - Logistic Regression - 0.62
  - SVM - 0.63
  - Naive Bayes - 0.60



```
from sklearn.metrics import classification_report
print classification_report(y_test,pred_final)
```

```
             precision    recall  f1-score   support

          0       0.75      0.69      0.72        67
          1       0.46      0.55      0.50        33

avg / total       0.66      0.64      0.65       100
```

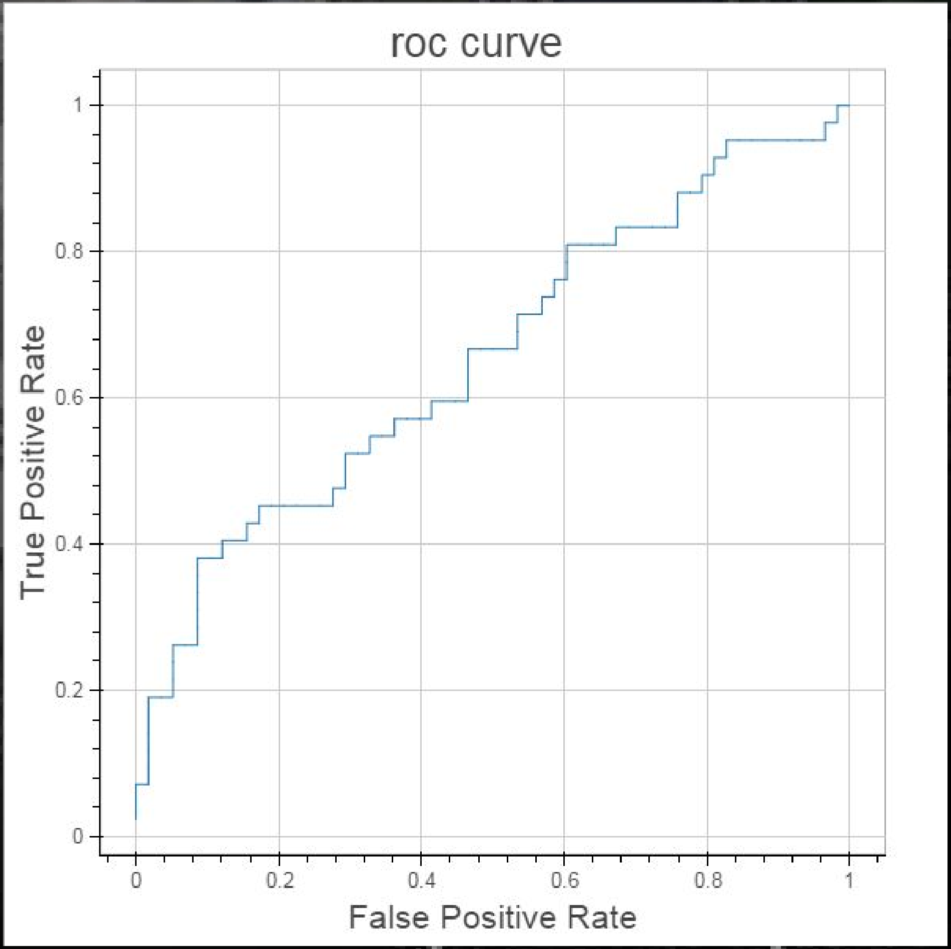


```
from sklearn.metrics import accuracy_score

# random Forest

parameters=[{"n_estimators":[50, 100, 150, 200], "max_features":range(4,10)}]
svr=RandomForestClassifier()
clf_final=grid_search.GridSearchCV(svr, parameters)
clf_final.fit(X_train, y_train)
pred_final=clf_final.predict(X_test)
print accuracy_score(y_test, pred_final)
clf_final.best_params_
```

```
0.68

{'max_features': 5, 'n_estimators': 150}
```

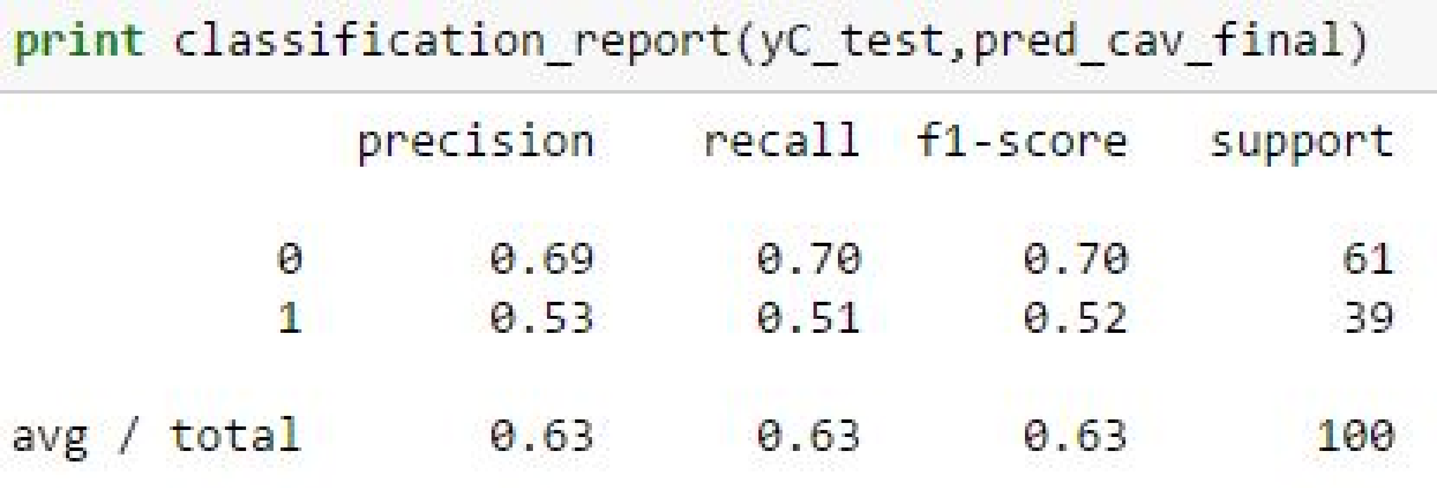
Random Forests

- Random Forests provided the best with an accuracy of 0.68, using max_features=5 and 150 estimators or forests
- Other Classifiers
  - Logistic Regression - 0.62
  - SVM - 0.63
  - Naive Bayes - 0.60

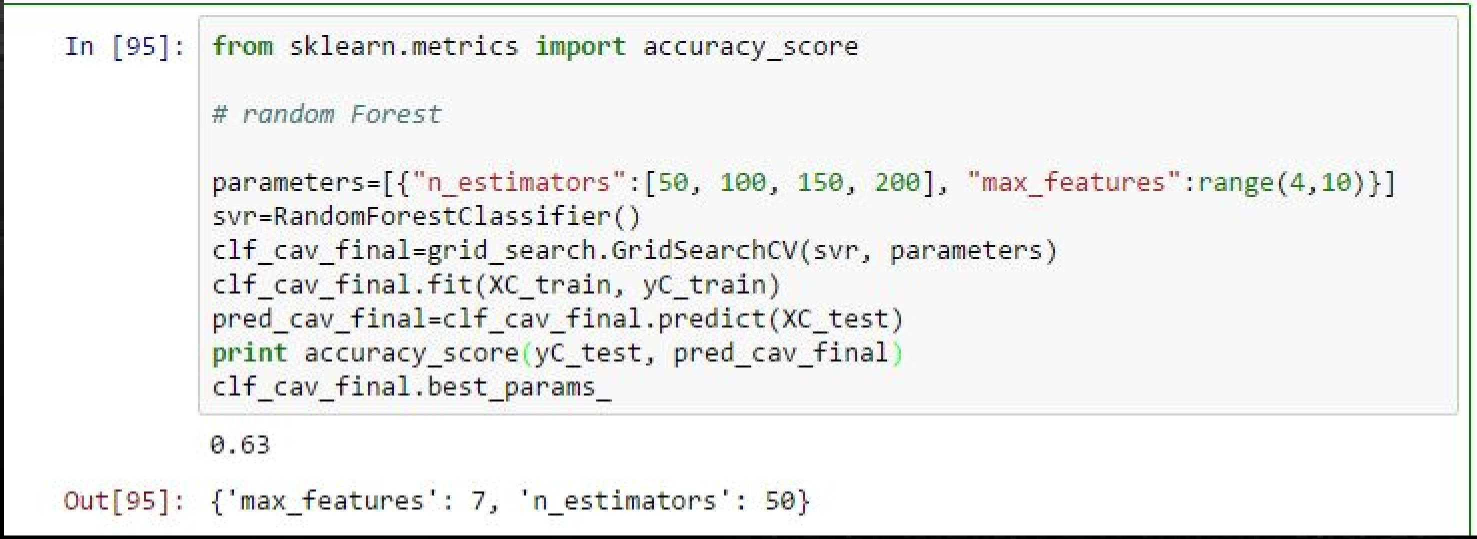


Cavaliers on same series

- Random Forests accuracy of 0.63



```
print classification_report(yC_test,pred_cav_final)

             precision    recall  f1-score   support

          0       0.69      0.70      0.70        61
          1       0.53      0.51      0.52        39

avg / total       0.63      0.63      0.63       100
```



```
In [95]: from sklearn.metrics import accuracy_score

         # random Forest

         parameters=[{"n_estimators":[50, 100, 150, 200], "max_features":range(4,10)}]
         svr=RandomForestClassifier()
         clf_cav_final=grid_search.GridSearchCV(svr, parameters)
         clf_cav_final.fit(XC_train, yC_train)
         pred_cav_final=clf_cav_final.predict(XC_test)
         print accuracy_score(yC_test, pred_cav_final)
         clf_cav_final.best_params_

         0.63

Out[95]: {'max_features': 7, 'n_estimators': 50}
```

Where to go from here?

- I plan on adding another variable -Previous shot made
    - Shooting Streak
- Webscraping game id’s and team id’s
- Pull data from Regular Season
- Individual Player level for prospective trades?
- Feature Weighting

Questions

?