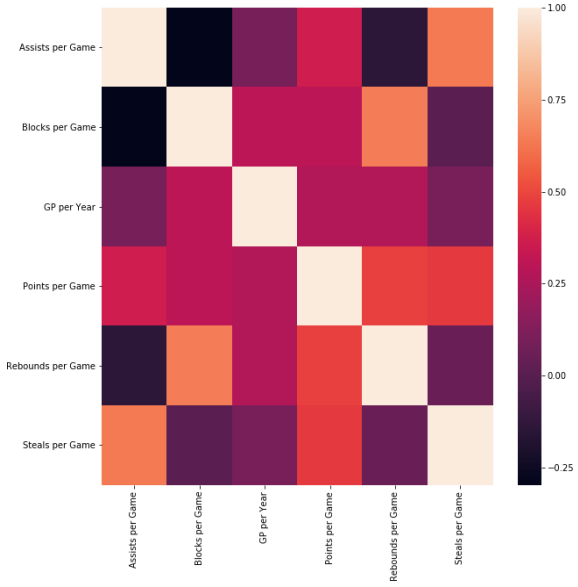
NBA HoF Predictions – Pure Statistics approach (KNN)

A. Data Cleanup(*data_clean.ipynb*)
   1. Pulled basketball data from Kaggle
      a. Hall of Fame
      b. Master – Player Information
      c. Players – Player Stats through 2011
   2. Merged tables
      a. Combined regular season and post season stats
   3. Added various columns for analysis
      a. Per game stats
      b. Length in career
      c. Years in retirement
B. Feature Selection(*sample_test.ipynb*)
   1. Ran combined dataset through grid and random forest to come up with correlating features:
      a. (Grid_Results.png & random_forest.png)



(random_forest.png)

```
GridSearchCV(cv='warn', error_score='raise-deprecating',
       estimator=SVC(C=1.0, cache_size=200, class_weight=None, coef0=0.0,
  decision_function_shape='ovr', degree=3, gamma='auto_deprecated',
  kernel='linear', max_iter=-1, probability=False, random_state=None,
  shrinking=True, tol=0.001, verbose=False),
       fit_params=None, iid='warn', n_jobs=None,
       param_grid={'C': [50, 100, 300, 500, 800], 'gamma': [1, 0.1, 0.01, 0.001, 0.0001]},
       pre_dispatch='2*n_jobs', refit=True, return_train_score='warn',
       scoring=None, verbose=3)
```

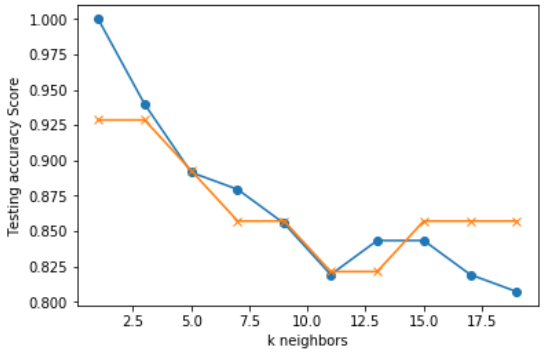
```
print(grid.best_params_)
print(grid.best_score_)
```

```
{'C': 50, 'gamma': 1}
0.891566265060241
```

(Grid_Results.png)

1. [(0.32915633026435404, 'Points per Game'),
2. (0.20278276060998965, 'GP per Year'),
3. (0.12624876122043238, 'D. Rebounds per Game'),
4. (0.10149189709341173, 'Blocks per Game'),
5. (0.08994479825188213, 'Steals per Game'),

6. (0.08065741416322285, 'Assists per Game'),
7. (0.06971803839670711, 'O. Rebounds per Game')]

C. Modeling our dataset *(KNN.ipynb)*
   1. Using the combined cleaned dataset, we filtered the data for only the features in the Random Forest
   2. Both Hall of Famers and Non were used for this model converting that column to Boolean.
   3. We then created our test/train split and scaled each feature.
   4. The KNN plot was then created as well as our result:
   5. 

(KNN.png)

```
knn = KNeighborsClassifier(n_neighbors=3)
knn.fit(X_train, y_train.ravel())
print('k=3 Test Acc: %.3f' % knn.score(X_test, y_test))
```

```
k=3 Test Acc: 0.857
```

(KNN_Results.png)

   6. Once, the model was trained we loaded & scaled a sample set of data from 2012 to predict our 2013 Hall of Fame candidates.
   7. The model picked 2 candidates and chose one correctly:
      i. Gary Dwayne Payton (2013 HOF)
      ii. Antoine Devon Walker

D. Making predictions on Out-of-Sample data
   1. We followed the same clean up procedure for our non-sample sets for years 2013-2015 hoping to predict candidates for 2014-2016.
      i. The following parameters were used to determine our candidate pool:
         1. Retirement: At least 4 but less than 10 (HoF Induction average is 8 yrs)
         2. Total games played greater than or equal to 400
         3. We also only selected players who started after 1967, since three pointers started then, and they went to an 80 game season.

2. We loaded and scaled each year's dataset and ran our predictions:
    i. 2014 (Alonzo Mourning)
        1. Antoine Devon Walker
        2. Jalen Rose
        3. James Arthur Jackson
        4. Gordan Giricek
    ii. 2015(Dikembe Mutombo)
        1. Allen Ezail Iverson
        2. Antoine Devon Walker
        3. Jalen Rose
        4. James Arthur Jackson
        5. Gordan Giricek
    iii. 2016(Allen Ezail Iverson & Shaquille Rashaun O'Neal)
        1. Allen Ezail Iverson
        2. Shaquille Rashaun O'Neal
        3. Antoine Devon Walker
        4. Jalen Rose
        5. James Arthur Jackson
        6. Al Thornton
        7. Gordan Giricek

The model could not predict for the first two samples, but did pick two winners for 2016.

Conclusion:

While KNN seems to be the model needed for this purpose, it does appear we need to consider several other features. If we compare our 2014 picks:

Antoine Devon Walker
- NBA champion (2006)
- 3× NBA All-Star (1998, 2002, 2003)
- NBA All-Rookie First Team (1997)
- NCAA champion (1996)
- First-team All-SEC (1996)

Alonzo Mourning
- NBA champion (2006)
- 7× NBA All-Star (1994–1997, 2000–2002)
- All-NBA First Team (1999)
- All-NBA Second Team (2000)
- 2× NBA Defensive Player of the Year (1999, 2000)
- 2× NBA All-Defensive First Team (1999, 2000)
- 2× NBA blocks leader (1999, 2000)
- NBA All-Rookie First Team (1993)
- J. Walter Kennedy Citizenship Award (2002)
- No. 33 retired by Miami Heat
- Consensus first-team All-American (1992)
- Consensus second-team All-American (1990)
- Third-team All-American – NABC (1991)
- Big East Player of the Year (1992)
- 2× First-team All-Big East (1990, 1992)
- 2× USA Basketball Male Athlete of the Year (1990, 2000)
- McDonald's All-American MVP (1988)
- Mr. Basketball USA (1988)
- Naismith Prep Player of the Year (1988)
- Virginia Mr. Basketball (1988)

Even though Walker had betters stats, it’s clear that Mourning was the best candidate for Hall of Fame.

2015 picks:

Allen Ezail Iverson	Dikembe Mutombo
-NBA Most Valuable Player (2001)	
- 11× NBA All-Star (2000–2010)	-8× NBA All-Star (1992, 1995–1998, 2000–2002)
-2× NBA All-Star Game MVP (2001, 2005)	- All-NBA Second Team (2001)
-3× All-NBA First Team (1999, 2001, 2005)	- 2× All-NBA Third Team (1998, 2002)
	- 4× NBA Defensive Player of the Year (1995, 1997, 1998, 2001)
-3× All-NBA Second Team (2000, 2002, 2003)	
-All-NBA Third Team (2006)	- 3× NBA All-Defensive First Team (1997, 1998, 2001)
	- 3× NBA All-Defensive Second Team (1995, 1999, 2002)
-NBA Rookie of the Year (1997)	
-NBA All-Rookie First Team (1997)	- NBA All-Rookie First Team (1992)
-4× NBA scoring champion (1999, 2001, 2002, 2005)	- 2× NBA rebounding leader (2000, 2001)
-3× NBA steals leader (2001–2003)	- 3× NBA blocks leader (1994–1996)
	- 2× J. Walter Kennedy Citizenship Award (2001, 2009)
-No. 3 retired by Philadelphia 76ers	
-Consensus first-team All-American (1996)	- No. 55 retired by Atlanta Hawks
-First-team All-Big East (1996)	- No. 55 retired by Denver Nuggets
	- Third-team All-American – AP, UPI (1991)
	- First-team All-Big East (1991)

Both players were similar in accomplishments and Mutombo was a very strong defensive player as Iverson was offensive. Without giving weight to individual achievements, we couldn't predict this outcome.

2016 Picks:

Our model correctly picked Allen Ezail Iverson and Shaquille Rashaun O'Neal as possible candidates out of 7 with a 22% accuracy. While that doesn't scream efficiency, it seems the model is on the right track.

Things we could do to improve our model:

1. Consider assigning a weight for special awards (MVP, Championships, All-Stars)
    a. This model takes these into account: https://www.basketball-reference.com/about/hof_prob.html
    b. Another approach: https://www.aboveandbiondi.com/predicting-future-basketball-halloffamers