

## CS2006 Python Practical 2

## Additional Requirement Hard 2

Source of the dataset is <http://www.nbamminer.com/player-basic-stats/> (<http://www.nbamminer.com/player-basic-stats/>). First thing to do is to refine the dataset.

```
In [3]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from ipywidgets import widgets
from ipywidgets import interact
from IPython.display import display
```

So first to make sure the data is initially read in correctly, the columns which should be numbers (and not containing any strings) are selected and forced into numeric columns. This means that if any of the columns had invalid data, in the format of string for example, then that cell gets changed to 'NaN'.

```
In [4]: low_memory = False
df=pd.read_csv("../data/NBA2016-17Stats.csv")
# set option to display all columns for every row
pd.set_option('display.max_columns', None)
df
```

Out[4]:

	Player	Season	Season Type	Team	Games	Min	Pts	OReb	Dreb	Reb	Ass
0	David Nwaba	2016-2017	REG	Lakers	18	19.10	5.61	0.83	2.33	3.17	0.0
1	Mike Tobey	2016-2017	REG	Hornets	2	12.59	1.00	1.00	0.50	1.50	0.0
2	Danuel House	2016-2017	REG	Wizards	1	0.83	0.00	0.00	1.00	1.00	0.0
3	Patricio Garino	2016-2017	REG	Magic	1	4.28	0.00	1.00	0.00	1.00	0.0
4	Marcus Georges-Hunt	2016-2017	REG	Heat, Magic	1	4.28	0.00	0.00	1.00	1.00	0.0
5	Nicolas Laprovittola	2016-2017	REG	Spurs	18	9.67	3.28	0.06	0.50	0.56	1.0
6	Derrick Jones, Jr.	2016-2017	REG	Suns	29	15.97	4.55	1.03	1.28	2.31	0.0
7	Okaro White	2016-2017	REG	Heat	31	13.55	3.03	0.84	1.77	2.61	0.0
8	Georgios Papagiannis	2016-2017	REG	Kings	18	14.57	4.94	1.00	2.67	3.67	0.0
9	Paul Zipser	2016-2017	REG	Bulls	40	18.81	5.13	0.35	2.33	2.68	0.0
10	Shawn Long	2016-2017	REG	76ers	14	14.17	9.57	2.00	2.50	4.50	0.0
11	Daniel Ochefu	2016-2017	REG	Wizards	18	3.24	0.89	0.17	0.89	1.06	0.0
12	Marshall Plumlee	2016-2017	REG	Knicks	17	6.76	1.18	0.76	1.06	1.82	0.0
13	Mindaugas Kuzminskas	2016-2017	REG	Knicks	64	14.65	6.00	0.67	1.17	1.84	1.0
14	Nicolas Brussino	2016-2017	REG	Mavericks	50	8.02	2.24	0.48	1.10	1.58	0.0
15	Bryn Forbes	2016-2017	REG	Spurs	33	7.26	1.94	0.06	0.48	0.55	0.0
16	Johnny O'Bryant III	2016-2017	REG	Nuggets, Hornets	11	7.33	3.45	0.82	0.91	1.73	0.0
17	James Michael McAdoo	2016-2017	REG	Warriors	49	8.54	2.71	0.67	1.08	1.76	0.0
18	Jarnell Stokes	2016-2017	REG	Nuggets	2	3.28	1.50	0.50	0.50	1.00	1.0
19	Andrew Wiggins	2016-2017	REG	Timberwolves	77	37.22	23.27	1.25	2.77	4.01	2.0
20	Jabari Parker	2016-2017	REG	Bucks	51	33.88	20.10	1.55	4.61	6.16	2.0
21	Joel Embiid	2016-2017	REG	76ers	31	25.36	20.23	1.97	5.84	7.81	2.0
22	Dante Exum	2016-2017	REG	Jazz	62	18.35	6.34	0.45	1.53	1.98	1.0
23	Josh Huestis	2016-2017	REG	Thunder	1	6.88	7.00	2.00	2.00	4.00	1.0
24	Dario Saric	2016-2017	REG	76ers	78	26.39	12.86	1.37	4.91	6.28	2.0
25	Clint Capela	2016-2017	REG	Rockets	63	23.77	12.35	2.75	5.37	8.11	0.0
26	Jusuf Nurkic	2016-2017	REG	Nuggets, Trail Blazers	65	21.40	10.18	2.42	4.75	7.17	1.0
27	Bruno	2016-2017	REG	Pacers	8	2.87	0.38	0.38	0.38	0.75	0.0

As shown below, the data types of the data set are numeric where expected.

```
In [5]: df.dtypes
```

```
Out[5]: Player      object
Season      object
Season Type  object
Team        object
Games       int64
Min         float64
Pts         float64
OREB        float64
DREB        float64
REB         float64
AST         float64
STL         float64
BLK         float64
TO          float64
PF          float64
FGM         float64
FGA         float64
FG%         float64
3PTM        float64
3PTA        float64
3PT%        float64
FTM         float64
FTA         float64
FT%         float64
dtype: object
```

To actually refine the data, any rows containing a cell which is 'NaN' gets dropped.

```
In [6]: refinedNBA = df.copy()
refinedNBA = refinedNBA.dropna()
```

More refining includes dropping any rows which are duplicates of another.

```
In [7]: refinedNBA = refinedNBA.drop_duplicates()
```

Change "Team" col value to equal only to the current team of player. Because if a player has changed his team this season both his teams are displayed e.g. "Hornets, Lakers". And we want to have only "Lakers" as the current team of the player.

```
In [8]: teams = refinedNBA['Team']

for i in range(0, len(teams)):
    if ", " in teams[i]:
        franchises = teams[i].split(", ", 5)
        numOfFranchises = len(franchises)
        teams[i] = franchises[numOfFranchises - 1]
```

/cs/home/ea50/.local/lib/python3.5/site-packages/ipykernel\_launcher.py:7:

SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy>

```
import sys
```

The final refining that takes place is to make sure that a team entry really exists in the NBA.

```
In [9]: refinedNBA.to_csv("refinedNBA.csv")
```

determine the total number of records in the dataset.

```
In [10]: len(refinedNBA)
```

```
Out[10]: 485
```

determine the type of each variable in the dataset.

```
In [11]: refinedNBA.dtypes
```

```
Out[11]: Player          object
Season                object
Season Type          object
Team                 object
Games                int64
Min                  float64
Pts                  float64
OReb                 float64
Dreb                 float64
Reb                  float64
Ast                  float64
Stl                  float64
Blk                  float64
TO                   float64
PF                   float64
FGM                  float64
FGA                  float64
FG%                  float64
3PTM                 float64
3PTA                 float64
3PT%                 float64
FTM                  float64
FTA                  float64
FT%                  float64
dtype: object
```

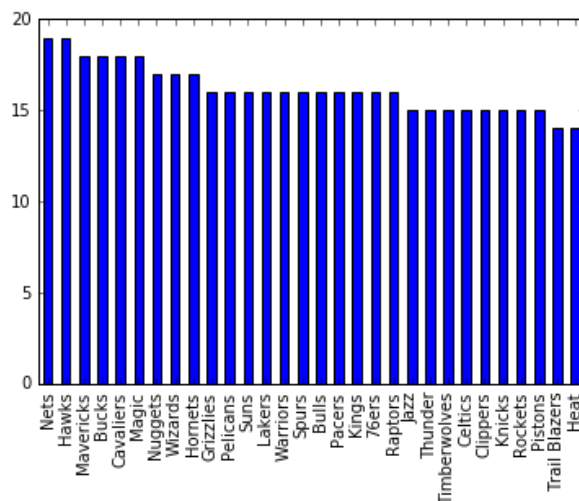
Finds number of players for each team.

```
In [12]: franchises = refinedNBA.groupby('Team')
franchises.size()
```

```
Out[12]: Team
76ers          16
Bucks          18
Bulls          16
Cavaliers      18
Celtics        15
Clippers       15
Grizzlies      16
Hawks          19
Heat           14
Hornets        17
Jazz           15
Kings          16
Knicks         15
Lakers         16
Magic          18
Mavericks      18
Nets           19
Nuggets        17
Pacers         16
Pelicans       16
Pistons        15
Raptors        16
Rockets        15
Spurs          16
Suns           16
Thunder        15
Timberwolves   15
Trail Blazers  14
Warriors       16
Wizards        17
dtype: int64
```

bar chart for the number of players for each team.

```
In [13]: refinedNBA['Team'].value_counts().plot(kind="bar")
plt.show()
```



Here you can see the ten most high scoring players in the NBA so far in the season.

```
In [14]: refinedNBA.sort_values('Pts', ascending = False).head(10)
```

```
Out[14]:
```

	Player	Season	Season Type	Team	Games	Min	Pts	OReb	Dreb	Reb	Ast	Stl
378	Russell Westbrook	2016-2017	REG	Thunder	78	34.74	31.79	1.68	9.03	10.71	10.45	1.7
353	James Harden	2016-2017	REG	Rockets	77	36.54	29.18	1.17	6.95	8.12	11.18	1.5
309	Isaiah Thomas	2016-2017	REG	Celtics	72	33.99	29.06	0.51	2.14	2.65	5.94	0.9
310	Anthony Davis	2016-2017	REG	Pelicans	74	36.27	28.05	2.36	9.69	12.05	2.14	1.3
349	DeMar DeRozan	2016-2017	REG	Raptors	72	35.46	27.04	0.96	4.31	5.26	4.00	1.0
250	DeMarcus Cousins	2016-2017	REG	Pelicans	72	34.24	26.97	2.08	8.96	11.04	4.63	1.3
315	Damian Lillard	2016-2017	REG	Trail Blazers	73	35.82	26.62	0.63	4.19	4.82	5.82	0.9
465	LeBron James	2016-2017	REG	Cavaliers	72	37.59	26.32	1.29	7.28	8.57	8.74	1.2
288	Kawhi Leonard	2016-2017	REG	Spurs	71	33.65	25.75	1.11	4.77	5.89	3.55	1.7
352	Stephen Curry	2016-2017	REG	Warriors	77	33.51	25.34	0.78	3.70	4.48	6.64	1.8

Here you can see the ten players who are getting most playtime per game so far this season.

In [15]: `refinedNBA.sort_values('Min', ascending = False).head(10)`

Out[15]:

	Player	Season	Season Type	Team	Games	Min	Pts	OReb	Dreb	Reb	Ass
396	Kyle Lowry	2016-2017	REG	Raptors	57	37.77	22.84	0.82	3.96	4.79	6.9
465	LeBron James	2016-2017	REG	Cavaliers	72	37.59	26.32	1.29	7.28	8.57	8.7
19	Andrew Wiggins	2016-2017	REG	Timberwolves	77	37.22	23.27	1.25	2.77	4.01	2.3
178	Zach LaVine	2016-2017	REG	Timberwolves	47	37.21	18.91	0.40	3.00	3.40	2.9
298	Jimmy Butler	2016-2017	REG	Bulls	72	37.06	23.92	1.74	4.39	6.13	5.5
51	Karl-Anthony Towns	2016-2017	REG	Timberwolves	77	36.91	24.82	3.64	8.32	11.96	2.6
353	James Harden	2016-2017	REG	Rockets	77	36.54	29.18	1.17	6.95	8.12	11.
246	John Wall	2016-2017	REG	Wizards	76	36.42	23.22	0.75	3.51	4.26	10.
310	Anthony Davis	2016-2017	REG	Pelicans	74	36.27	28.05	2.36	9.69	12.05	2.1
315	Damian Lillard	2016-2017	REG	Trail Blazers	73	35.82	26.62	0.63	4.19	4.82	5.8

Here you can see the ten payers who get most rebounds per game this season.

In [16]: `refinedNBA.sort_values('Reb', ascending = False).head(10)`

Out[16]:

	Player	Season	Season Type	Team	Games	Min	Pts	OReb	Dreb	Reb	Ass
266	Hassan Whiteside	2016-2017	REG	Heat	73	32.63	16.74	3.79	10.34	14.14	0.7
317	Andre Drummond	2016-2017	REG	Pistons	77	30.07	13.91	4.26	9.71	13.97	1.0
361	DeAndre Jordan	2016-2017	REG	Clippers	78	31.72	12.56	3.60	10.10	13.71	1.1
202	Rudy Gobert	2016-2017	REG	Jazz	77	34.07	14.00	3.94	8.94	12.87	1.2
454	Dwight Howard	2016-2017	REG	Hawks	71	29.99	13.44	4.08	8.79	12.87	1.3
310	Anthony Davis	2016-2017	REG	Pelicans	74	36.27	28.05	2.36	9.69	12.05	2.1
51	Karl-Anthony Towns	2016-2017	REG	Timberwolves	77	36.91	24.82	3.64	8.32	11.96	2.6
476	Tyson Chandler	2016-2017	REG	Suns	47	27.61	8.45	3.30	8.15	11.45	0.6
250	DeMarcus Cousins	2016-2017	REG	Pelicans	72	34.24	26.97	2.08	8.96	11.04	4.6
379	Kevin Love	2016-2017	REG	Cavaliers	57	31.08	19.14	2.49	8.53	11.02	1.9



Here you can see the top 10 players leading the league in Assists so far this season.

```
In [17]: refinedNBA.sort_values('Ast', ascending = False).head(10)
```

```
Out[17]:
```

	Player	Season	Season Type	Team	Games	Min	Pts	OReb	Dreb	Reb	Ast
353	James Harden	2016-2017	REG	Rockets	77	36.54	29.18	1.17	6.95	8.12	11.18
246	John Wall	2016-2017	REG	Wizards	76	36.42	23.22	0.75	3.51	4.26	10.70
378	Russell Westbrook	2016-2017	REG	Thunder	78	34.74	31.79	1.68	9.03	10.71	10.45
436	Chris Paul	2016-2017	REG	Clippers	58	31.50	18.09	0.71	4.26	4.97	9.28
351	Ricky Rubio	2016-2017	REG	Timberwolves	71	32.74	10.92	0.94	3.14	4.08	9.04
465	LeBron James	2016-2017	REG	Cavaliers	72	37.59	26.32	1.29	7.28	8.57	8.74
342	Jeff Teague	2016-2017	REG	Pacers	78	32.46	15.33	0.41	3.74	4.15	7.76
344	Jrue Holiday	2016-2017	REG	Pelicans	63	33.15	15.68	0.65	3.22	3.87	7.32
333	Draymond Green	2016-2017	REG	Warriors	74	32.77	10.24	1.31	6.62	7.93	7.09
396	Kyle Lowry	2016-2017	REG	Raptors	57	37.77	22.84	0.82	3.96	4.79	6.95

Here you can see the top ten players leading the league in steals this season.

In [18]: `refinedNBA.sort_values('Stl', ascending = False).head(10)`

Out[18]:

	Player	Season	Season Type	Team	Games	Min	Pts	OReb	Dreb	Reb	Ast
333	Draymond Green	2016-2017	REG	Warriors	74	32.77	10.24	1.31	6.62	7.93	7.09
246	John Wall	2016-2017	REG	Wizards	76	36.42	23.22	0.75	3.51	4.26	10.70
436	Chris Paul	2016-2017	REG	Clippers	58	31.50	18.09	0.71	4.26	4.97	9.28
298	Jimmy Butler	2016-2017	REG	Bulls	72	37.06	23.92	1.74	4.39	6.13	5.51
201	Robert Covington	2016-2017	REG	76ers	67	31.62	12.90	1.36	5.12	6.48	1.54
443	Trevor Ariza	2016-2017	REG	Rockets	77	34.59	11.79	0.66	5.09	5.75	2.13
352	Stephen Curry	2016-2017	REG	Warriors	77	33.51	25.34	0.78	3.70	4.48	6.64
288	Kawhi Leonard	2016-2017	REG	Spurs	71	33.65	25.75	1.11	4.77	5.89	3.55
378	Russell Westbrook	2016-2017	REG	Thunder	78	34.74	31.79	1.68	9.03	10.71	10.45
351	Ricky Rubio	2016-2017	REG	Timberwolves	71	32.74	10.92	0.94	3.14	4.08	9.04

Here you can see the top ten players leading the league in blocked shots this season.

In [19]: `refinedNBA.sort_values('Blk', ascending = False).head(10)`

Out[19]:

	Player	Season	Season Type	Team	Games	Min	Pts	OReb	Dreb	Reb	Ast
202	Rudy Gobert	2016-2017	REG	Jazz	77	34.07	14.00	3.94	8.94	12.87	1.25
21	Joel Embiid	2016-2017	REG	76ers	31	25.36	20.23	1.97	5.84	7.81	2.13
310	Anthony Davis	2016-2017	REG	Pelicans	74	36.27	28.05	2.36	9.69	12.05	2.14
266	Hassan Whiteside	2016-2017	REG	Heat	73	32.63	16.74	3.79	10.34	14.14	0.71
59	Myles Turner	2016-2017	REG	Pacers	77	31.48	14.34	1.68	5.44	7.12	1.34
29	Kristaps Porzingis	2016-2017	REG	Knicks	66	32.79	18.12	1.70	5.45	7.15	1.47
195	Giannis Antetokounmpo	2016-2017	REG	Bucks	77	35.50	23.08	1.83	6.88	8.71	5.35
376	Brook Lopez	2016-2017	REG	Nets	72	29.51	20.71	1.57	3.82	5.39	2.38
361	DeAndre Jordan	2016-2017	REG	Clippers	78	31.72	12.56	3.60	10.10	13.71	1.15
403	Kevin Durant	2016-2017	REG	Warriors	59	33.55	25.32	0.64	7.58	8.22	4.81

The ten players who have scored most three pointers per game so far this season.

In [20]: `refinedNBA.sort_values('3PTM', ascending = False).head(10)`

Out[20]:

	Player	Season	Season Type	Team	Games	Min	Pts	OReb	Dreb	Reb	Ast	Stl
<b>352</b>	Stephen Curry	2016-2017	REG	Warriors	77	33.51	25.34	0.78	3.70	4.48	6.64	1.82
<b>284</b>	Klay Thompson	2016-2017	REG	Warriors	76	34.15	22.50	0.63	3.11	3.74	2.05	0.83
<b>381</b>	Eric Gordon	2016-2017	REG	Rockets	72	31.13	16.33	0.39	2.25	2.64	2.57	0.67
<b>396</b>	Kyle Lowry	2016-2017	REG	Raptors	57	37.77	22.84	0.82	3.96	4.79	6.95	1.44
<b>309</b>	Isaiah Thomas	2016-2017	REG	Celtics	72	33.99	29.06	0.51	2.14	2.65	5.94	0.92
<b>353</b>	James Harden	2016-2017	REG	Rockets	77	36.54	29.18	1.17	6.95	8.12	11.18	1.51
<b>283</b>	Kemba Walker	2016-2017	REG	Hornets	78	34.66	23.17	0.56	3.44	4.00	5.46	1.09
<b>312</b>	Bradley Beal	2016-2017	REG	Wizards	74	34.82	23.04	0.68	2.39	3.07	3.47	1.09
<b>413</b>	Nick Young	2016-2017	REG	Lakers	60	25.94	13.18	0.42	1.87	2.28	0.97	0.65
<b>315</b>	Damian Lillard	2016-2017	REG	Trail Blazers	73	35.82	26.62	0.63	4.19	4.82	5.82	0.93

"double-double" - have a combination of more than 10 in at least two positive categories e.g. 15 points, 11 rebounds.  
 Perform query to find all players averaging double-double this season. Someone even averages tripple - double ! He is making history !! (You may be interested in looking his performance on youtube. I strongly suggest it !!!)

```
In [21]: avgDoubleDouble = refinedNBA[(refinedNBA['Pts'] >= 10) & (refinedNBA['Reb'] >= 10) |
                                             (refinedNBA['Pts'] >= 10) & (refinedNBA['Ast'] >= 10)]
avgDoubleDouble = avgDoubleDouble[['Player', 'Team', 'Pts', 'Reb', 'Ast']].copy()
avgDoubleDouble
#byRegionAndEconomicActivity.reset_index(name='Count')
```

Out[21]:

	Player	Team	Pts	Reb	Ast
51	Karl-Anthony Towns	Timberwolves	24.82	11.96	2.68
202	Rudy Gobert	Jazz	14.00	12.87	1.25
246	John Wall	Wizards	23.22	4.26	10.70
250	DeMarcus Cousins	Pelicans	26.97	11.04	4.63
266	Hassan Whiteside	Heat	16.74	14.14	0.71
289	Nikola Vucevic	Magic	14.62	10.35	2.76
310	Anthony Davis	Pelicans	28.05	12.05	2.14
317	Andre Drummond	Pistons	13.91	13.97	1.08
353	James Harden	Rockets	29.18	8.12	11.18
361	DeAndre Jordan	Clippers	12.56	13.71	1.15
378	Russell Westbrook	Thunder	31.79	10.71	10.45
379	Kevin Love	Cavaliers	19.14	11.02	1.91
423	Marcin Gortat	Wizards	10.67	10.47	1.46
454	Dwight Howard	Hawks	13.44	12.87	1.39

The second query to find players averaging more than 20 points per game.

```
In [22]: avgMostPoints = refinedNBA[(refinedNBA['Pts'] >= 20)]
avgMostPoints = avgMostPoints[['Player', 'Team', 'Pts']].copy()
avgMostPoints
```

Out[22]:

	Player	Team	Pts
19	Andrew Wiggins	Timberwolves	23.27
20	Jabari Parker	Bucks	20.10
21	Joel Embiid	76ers	20.23
51	Karl-Anthony Towns	Timberwolves	24.82
57	Devin Booker	Suns	21.95
195	Giannis Antetokounmpo	Bucks	23.08
214	CJ McCollum	Trail Blazers	23.21
246	John Wall	Wizards	23.22
250	DeMarcus Cousins	Pelicans	26.97
253	Gordon Hayward	Jazz	21.81
254	Paul George	Pacers	23.32
261	Eric Bledsoe	Suns	21.06
276	Kyrie Irving	Cavaliers	25.04
283	Kemba Walker	Hornets	23.17
284	Klay Thompson	Warriors	22.50
288	Kawhi Leonard	Spurs	25.75
298	Jimmy Butler	Bulls	23.92
309	Isaiah Thomas	Celtics	29.06
310	Anthony Davis	Pelicans	28.05
312	Bradley Beal	Wizards	23.04
315	Damian Lillard	Trail Blazers	26.62
349	DeMar DeRozan	Raptors	27.04
352	Stephen Curry	Warriors	25.34
353	James Harden	Rockets	29.18
356	Goran Dragic	Heat	20.35
359	Blake Griffin	Clippers	21.81
376	Brook Lopez	Nets	20.71
378	Russell Westbrook	Thunder	31.79
396	Kyle Lowry	Raptors	22.84
403	Kevin Durant	Warriors	25.32
405	Mike Conley	Grizzlies	20.52
464	Carmelo Anthony	Knicks	22.49
465	LeBron James	Cavaliers	26.32

Players averaging more than 10 assists per game this season.

```
In [23]: avgMostAssists = refinedNBA[(refinedNBA['Ast'] >= 10)]
avgMostAssists = avgMostAssists[['Player', 'Team', 'Ast']].copy()
avgMostAssists
```

```
Out[23]:
```

	Player	Team	Ast
246	John Wall	Wizards	10.70
353	James Harden	Rockets	11.18
378	Russell Westbrook	Thunder	10.45

Interactive player selector using widgets and different queries to filter data.

```
In [24]: def update(selectTeam = list(refinedNBA['Team'].unique()),
                minNumOfPtsPerGame = range(0,40),
                maxNumOfPtsPerGame = range(0,40),
                minNumOfAstPerGame = range(0,40),
                maxNumOfAstPerGame = range(0,42),
                minNumOfRebPerGame = range(0,40),
                maxNumOfRebPerGame = range(0,40)):
    selectedTeam = refinedNBA[(refinedNBA['Team'] == selectTeam) &
                              ((refinedNBA['Pts'] >= minNumOfPtsPerGame)
                               & (refinedNBA['Pts'] <= maxNumOfPtsPerGame)) &
                              ((refinedNBA['Ast'] >= minNumOfAstPerGame)
                               & (refinedNBA['Ast'] <= maxNumOfAstPerGame)) &
                              ((refinedNBA['Reb'] >= minNumOfRebPerGame)
                               & (refinedNBA['Reb'] <= maxNumOfRebPerGame))]
    display(selectedTeam)

interact(update);
```