CollegeDataProj

March 15, 2019

0.0.1 This project was my first independent project. This consists of exploratory data analysis, merging data frames, webscraping, transforms, visualizing the data, and a machine learning technique (principal component analysis). The goal of this project was to identify players from the NCAA data sets that would end up as draft picks by identifying key features that would discern NBA talent from collegiate talent. In the Principal Component Analysis I found key features that identify players in a 3 dimensional graph and I plotted those data points red for distinction.

```
In [36]: import pandas as pd
         import numpy as np
         from sklearn.decomposition import PCA
```

```
In [37]: ncaa2015 = pd.read_excel("2015 Individual Player Stats.xlsx")
```

```
In [38]: ncaa2015.head()
```

```
Out[38]:         Team + Rank  Minutes Rank            Player Class Pos  \
         0  Abilene Christian1             1       Parker Wentz    JR   G
         1  Abilene Christian2             2       Austin Cooke    JR   F
         2  Abilene Christian3             3        Jalen Little   JR   G
         3  Abilene Christian4             4    Harrison Hawkins   SR   G
         4  Abilene Christian5             5   LaDarrien Williams  SR   G

                       School       Conf   G   MP  Totals FG   ...    Totals FTA  \
         0  Abilene Christian  Southland  27  780         97   ...            85
         1  Abilene Christian  Southland  27  768         92   ...            46
         2  Abilene Christian  Southland  27  682         25   ...            29
         3  Abilene Christian  Southland  27  655        124   ...            65
         4  Abilene Christian  Southland  27  651         97   ...            59

            Totals ORB  Totals DRB  Totals TRB  Totals AST  Totals STL  Totals BLK  \
         0           7          51          58          38          43           3
         1          36         128         164          23          14           2
         2           5          43          48          91          21           0
         3           9          23          32          43          24           3
         4          29          85         114          56          27           3

            Totals TOV  Totals PF  Totals PTS
```

```
0          40         58        322
1          33         73        268
2          44         54         85
3          64         71        353
4          60         61        272

[5 rows x 26 columns]
```

In [39]: ncaa2016 = pd.read_excel("2016 Individual Player Stats.xlsx")

In [40]: ncaa2016.head()

Out[40]:   Team + Rank Minutes Rank   Player Class Pos \
        0 Abilene Christian1   1   Parker Wentz   SR   G
        1 Abilene Christian2   2 Jaylen Franklin   FR   G
        2 Abilene Christian3   3   Jalen Little   SR   G
        3 Abilene Christian4   4   Jaren Lewis   FR   F
        4 Abilene Christian5   5   Jovan Crnic   JR   G

                     School   Conf   G   MP Totals FG   ...   Totals FTA \
        0 Abilene Christian Southland 31 918   87   ...   31
        1 Abilene Christian Southland 30 883   176   ...   206
        2 Abilene Christian Southland 26 737   48   ...   49
        3 Abilene Christian Southland 31 673   110   ...   73
        4 Abilene Christian Southland 31 640   71   ...   55

           Totals ORB Totals DRB Totals TRB Totals AST Totals STL Totals BLK \
        0   5   51   56   27   37   3
        1   23   92   115   82   52   4
        2   9   68   77   116   33   2
        3   36   144   180   40   34   18
        4   8   71   79   78   23   3

           Totals TOV Totals PF Totals PTS
        0   27   65   269
        1   85   84   486
        2   50   67   159
        3   34   64   292
        4   67   66   219

        [5 rows x 26 columns]

In [41]: ncaa2017 = pd.read_excel("2017 Individual Player Stats.xlsx")

In [42]: ncaa2017.head()

Out[42]:   Team + Rank Minutes Rank   Player Class Pos \
        0 Abilene Christian1   1   Jaren Lewis   SO   G
        1 Abilene Christian2   2 Jaylen Franklin   SO   G

2

```
      2   Abilene Christian3          3       Isaiah Tripp    JR    G
      3   Abilene Christian4          4        Drake Green    JR    G
      4   Abilene Christian5          5       Jalone Friday   FR    C


                      School        Conf    G    MP   Totals FG    ...     Totals FTA  \
      0   Abilene Christian   Southland   29   939         148    ...            125
      1   Abilene Christian   Southland   26   896         115    ...            108
      2   Abilene Christian   Southland   29   662          55    ...             47
      3   Abilene Christian   Southland   27   648          50    ...              9
      4   Abilene Christian   Southland   26   563         123    ...             91


          Totals ORB   Totals DRB   Totals TRB   Totals AST   Totals STL   Totals BLK  \
      0           39          151          190           61           53           12
      1           12           86           98          115           48            8
      2            8           34           42           46           16            1
      3            4           51           55           30           22            6
      4           41          101          142            9           12           34


          Totals TOV   Totals PF   Totals PTS
      0           75          72          397
      1           65          62          282
      2           31          69          175
      3           23          61          153
      4           41          95          357


      [5 rows x 26 columns]

In [43]: ncaa2015["year"] = "2015"

In [44]: ncaa2016["year"] = "2016"

In [45]: ncaa2017["year"] = "2017"

In [46]: ncaa2015.head()

Out[46]:            Team + Rank  Minutes Rank              Player Class Pos  \
      0   Abilene Christian1             1        Parker Wentz    JR    G
      1   Abilene Christian2             2         Austin Cooke   JR    F
      2   Abilene Christian3             3         Jalen Little   JR    G
      3   Abilene Christian4             4     Harrison Hawkins   SR    G
      4   Abilene Christian5             5   LaDarrien Williams   SR    G


                      School        Conf    G    MP   Totals FG   ...   Totals ORB  \
      0   Abilene Christian   Southland   27   780          97   ...            7
      1   Abilene Christian   Southland   27   768          92   ...           36
      2   Abilene Christian   Southland   27   682          25   ...            5
      3   Abilene Christian   Southland   27   655         124   ...            9
      4   Abilene Christian   Southland   27   651          97   ...           29
```

```
        Totals DRB  Totals TRB  Totals AST  Totals STL  Totals BLK  Totals TOV  \
      0          51          58          38          43           3          40
      1         128         164          23          14           2          33
      2          43          48          91          21           0          44
      3          23          32          43          24           3          64
      4          85         114          56          27           3          60

        Totals PF  Totals PTS  year
      0         58         322  2015
      1         73         268  2015
      2         54          85  2015
      3         71         353  2015
      4         61         272  2015

      [5 rows x 27 columns]
```

In [47]: ncaa = pd.concat([ncaa2015, ncaa2016, ncaa2017])

In [48]: ncaa.info()

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 13330 entries, 0 to 4466
Data columns (total 27 columns):
Team + Rank    13330 non-null object
Minutes Rank   13330 non-null int64
Player         13330 non-null object
Class          13295 non-null object
Pos            13329 non-null object
School         13330 non-null object
Conf           13330 non-null object
G              13330 non-null int64
MP             13330 non-null int64
Totals FG      13330 non-null int64
Totals FGA     13330 non-null int64
Totals 2P      13330 non-null int64
Totals 2PA     13330 non-null int64
Totals 3P      13330 non-null int64
Totals 3PA     13330 non-null int64
Totals FT      13330 non-null int64
Totals FTA     13330 non-null int64
Totals ORB     13330 non-null int64
Totals DRB     13330 non-null int64
Totals TRB     13330 non-null int64
Totals AST     13330 non-null int64
Totals STL     13330 non-null int64
Totals BLK     13330 non-null int64
Totals TOV     13330 non-null int64
Totals PF      13330 non-null int64
```

```
Totals PTS     13330 non-null int64
year           13330 non-null object
dtypes: int64(20), object(7)
memory usage: 2.8+ MB
```

In [49]: ncaa.describe()

Out[49]:   Minutes Rank   G   MP   Totals FG   Totals FGA \
        count 13330.000000 13330.000000 13330.000000 13330.000000 13330.000000
        mean   6.921830   24.262266   487.018530   60.442611   137.201800
        std   3.804556   8.705893   332.303125   52.895595   115.933211
        min   1.000000   1.000000   1.000000   1.000000   1.000000
        25%   4.000000   20.000000   181.000000   14.000000   36.000000
        50%   7.000000   27.000000   480.000000   48.000000   112.000000
        75%   10.000000   31.000000   769.000000   95.000000   215.000000
        max   19.000000   46.000000   1278.000000   313.000000   701.000000

                  Totals 2P   Totals 2PA   Totals 3P   Totals 3PA   Totals FT \
        count 13330.000000 13330.000000 13330.000000 13330.000000 13330.000000
        mean   43.524006   88.692873   16.918605   48.508927   34.915304
        std   42.244091   81.211921   21.690787   57.464477   35.877104
        min   0.000000   0.000000   0.000000   0.000000   0.000000
        25%   9.000000   20.000000   0.000000   2.000000   7.000000
        50%   31.000000   66.000000   7.000000   24.000000   23.000000
        75%   66.000000   137.000000   27.000000   79.000000   53.000000
        max   262.000000   569.000000   127.000000   340.000000   280.000000

                 Totals FTA   Totals ORB   Totals DRB   Totals TRB   Totals AST \
        count 13330.000000 13330.000000 13330.000000 13330.000000 13330.000000
        mean   50.042086   21.866692   56.074044   77.940735   31.974269
        std   47.893707   22.077375   46.264736   65.017973   35.841751
        min   0.000000   0.000000   0.000000   0.000000   0.000000
        25%   11.000000   5.000000   18.000000   25.000000   5.000000
        50%   36.000000   15.000000   48.000000   65.000000   20.000000
        75%   76.000000   32.000000   83.000000   114.000000   46.000000
        max   331.000000   173.000000   389.000000   474.000000   296.000000

                 Totals STL   Totals BLK   Totals TOV   Totals PF   Totals PTS
        count 13330.000000 13330.000000 13330.000000 13330.000000 13330.000000
        mean   15.032483   8.335109   30.139835   45.362641   172.719130
        std   13.827732   12.230708   24.072803   28.356689   152.672155
        min   0.000000   0.000000   0.000000   0.000000   2.000000
        25%   4.000000   1.000000   10.000000   21.000000   40.000000
        50%   12.000000   4.000000   26.000000   47.000000   135.000000
        75%   23.000000   11.000000   46.000000   67.000000   271.000000
        max   125.000000   130.000000   148.000000   130.000000   959.000000

In [50]: ncaa.to_excel("ncaa.xlsx", index = False)
```

```
In [51]: ncaa2015.columns.values

Out[51]: array(['Team + Rank', 'Minutes Rank', 'Player', 'Class', 'Pos', 'School',
               'Conf', 'G', 'MP', 'Totals FG', 'Totals FGA', 'Totals 2P',
               'Totals 2PA', 'Totals 3P', 'Totals 3PA', 'Totals FT', 'Totals FTA',
               'Totals ORB', 'Totals DRB', 'Totals TRB', 'Totals AST',
               'Totals STL', 'Totals BLK', 'Totals TOV', 'Totals PF', 'Totals PTS',
               'year'], dtype=object)

In [52]: ###cross fitting, overvalidation, linear regression, sci-kit learn

In [53]: #Make classification variables: year and whether they went into NBA so make an in and
        #Different regions: Classify rounds for draft picks. 1st, 2nd, not drafted at all. Wh
        #
        #Where are clusters forming? These are formed by a tuple of the factors you're putting
        #Principal Component Analysis
        #All these things go to what you do to separate and spot and define the dots.
        #Where are people clustering? Name these clusters.
        #With this regions set up, we identify these goals.

        #NEXT STEP: Look up how to set up an SVM. Put in data into it.
        #US or International prospects?
        #higher % of foreign players getting into the NBA draft?
        #convert categories to number values

In [54]: ncaa2015.head()

Out[54]:           Team + Rank  Minutes Rank            Player Class Pos  \
        0  Abilene Christian1             1      Parker Wentz    JR   G
        1  Abilene Christian2             2      Austin Cooke    JR   F
        2  Abilene Christian3             3      Jalen Little    JR   G
        3  Abilene Christian4             4   Harrison Hawkins   SR   G
        4  Abilene Christian5             5  LaDarrien Williams  SR   G


                    School       Conf   G   MP  Totals FG  ...  Totals ORB  \
        0  Abilene Christian  Southland  27  780         97  ...           7
        1  Abilene Christian  Southland  27  768         92  ...          36
        2  Abilene Christian  Southland  27  682         25  ...           5
        3  Abilene Christian  Southland  27  655        124  ...           9
        4  Abilene Christian  Southland  27  651         97  ...          29


           Totals DRB  Totals TRB  Totals AST  Totals STL  Totals BLK  Totals TOV  \
        0          51          58          38          43           3          40
        1         128         164          23          14           2          33
        2          43          48          91          21           0          44
        3          23          32          43          24           3          64
        4          85         114          56          27           3          60


           Totals PF  Totals PTS  year
```

```
         0      58        322  2015
         1      73        268  2015
         2      54         85  2015
         3      71        353  2015
         4      61        272  2015

         [5 rows x 27 columns]

In [55]: ncaa2015['Class'] = ncaa2015['Class'].astype('category')
         ncaa2015['Pos'] = ncaa2015['Pos'].astype('category')
         ncaa2015['Conf'] = ncaa2015['Conf'].astype('category')

In [56]: ncaa2015['Conf'].unique()

Out[56]: [Southland, MWC, MAC, SEC, SWAC, ..., Summit, Big Sky, A-Sun, Big Ten, Ind]
         Length: 33
         Categories (33, object): [Southland, MWC, MAC, SEC, ..., Big Sky, A-Sun, Big Ten, Ind]

In [57]: #! conda install -y -q html5lib
         import html5lib
         nba2015 = pd.read_html("http://www.nbadraft.net/nba_final_draft/2015")

In [58]: a = nba2015[0].iloc[:,:8].copy()
         b = nba2015[0].iloc[:,8:].copy()
         a['round'] = 1
         b['round'] = 2
         b.columns = a.columns

         c = a.append(b)
         #c.set_index('#',inplace=True)
         c

Out[58]:      #          Team              Player     H    W      P  \
         0    1      Minnesota   Karl-Anthony Towns   7-0  250      C
         1    2      LA Lakers     D'Angelo Russell   6-4  195  PG/SG
         2    3   Philadelphia        Jahlil Okafor  6-11  270      C
         3    4      *New York    Kristaps Porzingis   7-1  230     PF
         4    5        Orlando        Mario Hezonja   6-8  215  SG/SF
         5    6     Sacramento  Willie Cauley-Stein   7-0  242      C
         6    7         Denver      Emmanuel Mudiay   6-5  200     PG
         7    8        Detroit      Stanley Johnson   6-6  242     SF
         8    9      Charlotte       Frank Kaminsky   7-1  230      C
         9   10          Miami      Justise Winslow   6-6  222  SG/SF
         10  11        Indiana         Myles Turner  6-11  240      C
         11  12           Utah           Trey Lyles  6-10  240     PF
         12  13        Phoenix         Devin Booker   6-6  205     SG
         13  14   Oklahoma Cty       Cameron Payne   6-2  185     PG
         14  15    *Washington         Kelly Oubre   6-7  205     SF
         15  16         Boston         Terry Rozier   6-2  190     PG
```

16	17	Milwaukee	Rashad Vaughn	6-5	200	SG
17	18	*Houston	Sam Dekker	6-9	220	SF
18	19	*New York	Jerian Grant	6-4	200	PG
19	20	Toronto	Delon Wright	6-5	180	PG/SG
20	21	Dallas	Justin Anderson	6-6	230	SG/SF
21	22	Chicago	Bobby Portis	6-11	245	PF
22	23	*Brooklyn	Rondae Hollis-Jefferson	6-7	210	SF
23	24	*Minnesota	Tyus Jones	6-1	185	PG
24	25	Memphis	Jarell Martin	6-9	240	PF
25	26	San Antonio	Nikola Milutinov	7-0	225	PF/C
26	27	*LA Lakers	Larry Nance Jr.	6-9	230	PF
27	28	*Boston	RJ Hunter	6-6	185	SG
28	29	*Brooklyn	Chris McCullough	6-9	200	PF
29	30	Golden St.	Kevon Looney	6-9	222	PF
0	31	*Cleveland	Cedi Osman	6-8	215	SF
1	32	*Houston	Montrezl Harrell	6-8	255	PF
2	33	*Boston	Jordan Mickey	6-8	240	PF
3	34	*LA Lakers	Anthony Brown	6-7	210	SG/SF
4	35	*New York	Guillermo Hernangomez	6-11	250	PF/C
5	36	*Cleveland	Rakeem Christmas	6-9	245	PF/C
6	37	*Philadelphia	Richaun Holmes	6-9	245	PF
7	38	Detroit	Darrun Hilliard	6-7	220	SG
8	39	*Brooklyn	Juan Vaulet	6-7	200	SG/SF
9	40	Miami	Josh Richardson	6-6	200	SG
10	41	*Portland	Pat Connaughton	6-5	215	SG
11	42	Utah	Olivier Hanlan	6-4	185	PG/SG
12	43	Indiana	Joseph Young	6-2	185	PG/SG
13	44	*Memphis	Andrew Harrison	6-6	215	PG/SG
14	45	Boston	Marcus Thornton WM	6-3	180	PG/SG
15	46	*Toronto	Norman Powell	6-4	215	SG
16	47	*Philadelphia	Arturas Gudaitis	6-10	255	PF
17	48	Oklahoma Cty	Dakari Johnson	7-0	260	C
18	49	Washington	Aaron White	6-9	220	SF/PF
19	50	*Atlanta	Marcus Eriksson	6-7	190	SG
20	51	*Orlando	Tyler Harvey	6-4	180	SG
21	52	Dallas	Satnam Singh Bhamara	7-2	290	C
22	53	*Cleveland	Sir'Dominic Pointer	6-6	190	SF
23	54	*Portland	Daniel Diez	6-8	220	SF/PF
24	55	San Antonio	Cady Lalanne	6-9	240	C
25	56	*LA Clippers	Branden Dawson	6-7	230	SF/PF
26	57	*Denver	Nikola Radicevic	6-5	200	PG/SG
27	58	*Philadelphia	JP Tokoto	6-6	195	SG/SF
28	59	Atlanta	Dimitrios Agravanis	6-10	235	PF
29	60	*Philadelphia	Luka Mitrovic	6-8	200	SF/PF

	School	C	round
0	Kentucky	Fr.	1
1	Ohio State	Fr.	1

2	Duke	Fr.	1
3	Latvia	Intl.	1
4	Croatia	Intl.	1
5	Kentucky	Jr.	1
6	Congo	Intl.	1
7	Arizona	Fr.	1
8	Wisconsin	Sr.	1
9	Duke	Fr	1
10	Texas	Fr.	1
11	Kentucky	Fr.	1
12	Kentucky	Fr.	1
13	Murray St.	So.	1
14	Kansas	Fr.	1
15	Louisville	So.	1
16	UNLV	Fr.	1
17	Wisconsin	Jr.	1
18	Notre Dame	Sr.	1
19	Utah	Sr.	1
20	Virginia	Jr.	1
21	Arkansas	So.	1
22	Arizona	So.	1
23	Duke	Fr.	1
24	LSU	So.	1
25	Serbia	Intl.	1
26	Wyoming	Sr.	1
27	Georgia St.	Jr.	1
28	Syracuse	Fr.	1
29	UCLA	Fr.	1
0	Turkey	Intl.	2
1	Louisville	Jr.	2
2	LSU	So.	2
3	Stanford	Sr.	2
4	Spain	Intl.	2
5	Syracuse	Sr.	2
6	Bowling Green	Sr.	2
7	Villanova	Sr.	2
8	Argentina	Intl.	2
9	Tennessee	Sr.	2
10	Notre Dame	Sr.	2
11	Boston College	Jr.	2
12	Oregon	Sr.	2
13	Kentucky	So.	2
14	William & Mary	Sr.	2
15	UCLA	Sr.	2
16	Lithuania	Intl.	2
17	Kentucky	So.	2
18	Iowa	Sr.	2
19	Sweden	Intl.	2

```
     20   Eastern Washington    Jr.       2
     21              India     Intl.       2
     22          St. Johns      Sr.       2
     23              Spain     Intl.       2
     24       Massachusetts     Sr.       2
     25        Michigan St.     Sr.       2
     26             Serbia     Intl.       2
     27     North Carolina      Jr.       2
     28             Greece     Intl.       2
     29             Serbia     Intl.       2
```

In [59]: `ncaa2015.head()`

Out[59]:
```
           Team + Rank  Minutes Rank            Player Class Pos  \
0  Abilene Christian1              1      Parker Wentz    JR   G
1  Abilene Christian2              2       Austin Cooke    JR   F
2  Abilene Christian3              3       Jalen Little    JR   G
3  Abilene Christian4              4   Harrison Hawkins    SR   G
4  Abilene Christian5              5  LaDarrien Williams    SR   G

              School       Conf   G   MP  Totals FG  ...  Totals ORB  \
0  Abilene Christian  Southland  27  780         97  ...           7
1  Abilene Christian  Southland  27  768         92  ...          36
2  Abilene Christian  Southland  27  682         25  ...           5
3  Abilene Christian  Southland  27  655        124  ...           9
4  Abilene Christian  Southland  27  651         97  ...          29

   Totals DRB  Totals TRB  Totals AST  Totals STL  Totals BLK  Totals TOV  \
0          51          58          38          43           3          40
1         128         164          23          14           2          33
2          43          48          91          21           0          44
3          23          32          43          24           3          64
4          85         114          56          27           3          60

   Totals PF  Totals PTS  year
0         58         322  2015
1         73         268  2015
2         54          85  2015
3         71         353  2015
4         61         272  2015

[5 rows x 27 columns]
```

In [60]: `c.head()`

Out[60]:
```
   #          Team             Player     H    W     P       School     C  \
0  1     Minnesota  Karl-Anthony Towns   7-0  250     C     Kentucky   Fr.
1  2     LA Lakers    D'Angelo Russell   6-4  195  PG/SG  Ohio State   Fr.
2  3  Philadelphia       Jahlil Okafor  6-11  270     C         Duke   Fr.
```

```
          3   4       *New York   Kristaps Porzingis   7-1  230     PF      Latvia  Intl.
          4   5         Orlando        Mario Hezonja   6-8  215  SG/SF     Croatia  Intl.


                round
          0        1
          1        1
          2        1
          3        1
          4        1


In [61]: ncaadraft2015 = pd.merge(ncaa2015, c, on=['Player', 'School'], how='outer', indicator=

In [62]: ncaadraft2015[ncaadraft2015['Found']=='both']

Out[62]:                 Team + Rank  Minutes Rank                 Player Class  Pos  \
          149                Arizona2           2.0          Stanley Johnson    FR    F
          150                Arizona3           3.0  Rondae Hollis-Jefferson    SO    F
          175               Arkansas2           2.0             Bobby Portis    SO    F
          348          Boston College1           1.0            Olivier Hanlan    JR    G
          1002                 Duke2           2.0               Tyus Jones    FR    G
          1003                 Duke3           3.0            Jahlil Okafor    FR    C
          1004                 Duke4           4.0           Justise Winslow    FR    F
          1091    Eastern Washington1           1.0             Tyler Harvey    JR    G
          1675                  Iowa1           1.0               Aaron White    SR    F
          1758                Kansas4           4.0               Kelly Oubre    FR    G
          1808              Kentucky2           2.0      Willie Cauley-Stein    JR    F
          1809              Kentucky3           3.0          Andrew Harrison    SO    G
          1810              Kentucky4           4.0              Devin Booker    FR    G
          1812              Kentucky6           6.0       Karl-Anthony Towns    FR    F
          1813              Kentucky7           7.0               Trey Lyles    FR    F
          1814              Kentucky8           8.0            Dakari Johnson    SO    C
          1990            Louisville1           1.0             Terry Rozier    SO    G
          1991            Louisville2           2.0          Montrezl Harrell    JR    F
          2141         Massachusetts3           3.0             Cady Lalanne    SR    C
          2855             Notre Dame1           1.0              Jerian Grant    SR    G
          2856             Notre Dame2           2.0          Pat Connaughton    SR  G-F
          2890             Ohio State1           1.0          D'Angelo Russell    FR    G
          2953                Oregon1           1.0             Joseph Young    SR    G
          3697              Stanford2           2.0            Anthony Brown    SR  G-F
          3748              Syracuse2           2.0          Rakeem Christmas    SR    F
          3752              Syracuse6           6.0          Chris McCullough    FR    F
          3771             Tennessee1           1.0          Josh Richardson    SR    G
          3821                 Texas3           3.0              Myles Turner    FR    F
          4026                  UCLA3           3.0             Norman Powell    SR    G
          4027                  UCLA4           4.0              Kevon Looney    FR    F
          4035                  Utah1           1.0              Delon Wright    SR    G
          4108             Villanova2           2.0           Darrun Hilliard    SR    G
          4121              Virginia4           4.0           Justin Anderson    JR    G
```

```
4345           Wisconsin3           3.0          Frank Kaminsky   SR   C
4346           Wisconsin4           4.0            Sam Dekker     JR   F


                    School      Conf      G        MP   Totals FG   ...      \
149                Arizona     Pac-12   28.0     793.0      130.0   ...
150                Arizona     Pac-12   28.0     785.0      102.0   ...
175               Arkansas        SEC   28.0     818.0      203.0   ...
348         Boston College        ACC   27.0    1020.0      192.0   ...
1002                  Duke        ACC   28.0     918.0       92.0   ...
1003                  Duke        ACC   27.0     840.0      206.0   ...
1004                  Duke        ACC   28.0     806.0      116.0   ...
1091     Eastern Washington   Big Sky   25.0     904.0      174.0   ...
1675                  Iowa    Big Ten   28.0     847.0      127.0   ...
1758                Kansas     Big 12   28.0     552.0       83.0   ...
1808              Kentucky        SEC   28.0     706.0      102.0   ...
1809              Kentucky        SEC   28.0     701.0       64.0   ...
1810              Kentucky        SEC   27.0     584.0      105.0   ...
1812              Kentucky        SEC   28.0     569.0       94.0   ...
1813              Kentucky        SEC   25.0     536.0       75.0   ...
1814              Kentucky        SEC   28.0     496.0       66.0   ...
1990            Louisville        ACC   28.0     943.0      172.0   ...
1991            Louisville        ACC   27.0     924.0      161.0   ...
2141         Massachusetts       A-10   28.0     803.0      120.0   ...
2855            Notre Dame        ACC   29.0    1055.0      165.0   ...
2856            Notre Dame        ACC   29.0    1010.0      131.0   ...
2890            Ohio State    Big Ten   28.0     928.0      180.0   ...
2953                Oregon     Pac-12   29.0    1049.0      206.0   ...
3697              Stanford     Pac-12   27.0     961.0      130.0   ...
3748              Syracuse        ACC   28.0     956.0      191.0   ...
3752              Syracuse        ACC   16.0     450.0       54.0   ...
3771             Tennessee        SEC   27.0     973.0      152.0   ...
3821                 Texas     Big 12   28.0     634.0      104.0   ...
4026                  UCLA     Pac-12   29.0     981.0      166.0   ...
4027                  UCLA     Pac-12   29.0     921.0      127.0   ...
4035                  Utah     Pac-12   27.0     873.0      132.0   ...
4108             Villanova   Big East   27.0     788.0      132.0   ...
4121              Virginia        ACC   22.0     647.0       97.0   ...
4345             Wisconsin    Big Ten   27.0     870.0      173.0   ...
4346             Wisconsin    Big Ten   28.0     823.0      139.0   ...


      Totals PTS   year      #        Team    H       W       P    C  round  \
149        398.0   2015    8.0      Detroit   6-6   242.0     SF  Fr.    1.0
150        312.0   2015   23.0    *Brooklyn   6-7   210.0     SF  So.    1.0
175        493.0   2015   22.0      Chicago  6-11   245.0     PF  So.    1.0
348        524.0   2015   42.0         Utah   6-4   185.0  PG/SG  Jr.    2.0
1002       323.0   2015   24.0   *Minnesota   6-1   185.0     PG  Fr.    1.0
1003       496.0   2015    3.0  Philadelphia 6-11   270.0      C  Fr.    1.0
1004       335.0   2015   10.0        Miami   6-6   222.0  SG/SF  Fr.    1.0
```

```
1091     567.0  2015  51.0    *Orlando       6-4   180.0    SG    Jr.   2.0
1675     425.0  2015  49.0    Washington     6-9   220.0  SF/PF   Sr.   2.0
1758     240.0  2015  15.0    *Washington    6-7   205.0    SF    Fr.   1.0
1808     256.0  2015   6.0    Sacramento     7-0   242.0     C    Jr.   1.0
1809     235.0  2015  44.0    *Memphis       6-6   215.0  PG/SG   So.   2.0
1810     301.0  2015  13.0    Phoenix        6-6   205.0    SG    Fr.   1.0
1812     261.0  2015   1.0    Minnesota      7-0   250.0     C    Fr.   1.0
1813     198.0  2015  12.0    Utah          6-10   240.0    PF    Fr.   1.0
1814     205.0  2015  48.0    Oklahoma Cty   7-0   260.0     C    So.   2.0
1990     496.0  2015  16.0    Boston         6-2   190.0    PG    So.   1.0
1991     414.0  2015  32.0    *Houston       6-8   255.0    PF    Jr.   2.0
2141     327.0  2015  55.0    San Antonio    6-9   240.0     C    Sr.   2.0
2855     490.0  2015  19.0    *New York      6-4   200.0    PG    Sr.   1.0
2856     380.0  2015  41.0    *Portland      6-5   215.0    SG    Sr.   2.0
2890     523.0  2015   2.0    LA Lakers      6-4   195.0  PG/SG   Fr.   1.0
2953     578.0  2015  43.0    Indiana        6-2   185.0  PG/SG   Sr.   2.0
3697     415.0  2015  34.0    *LA Lakers     6-7   210.0  SG/SF   Sr.   2.0
3748     506.0  2015  36.0    *Cleveland     6-9   245.0   PF/C   Sr.   2.0
3752     149.0  2015  29.0    *Brooklyn      6-9   200.0    PF    Fr.   1.0
3771     428.0  2015  40.0    Miami          6-6   200.0    SG    Sr.   2.0
3821     313.0  2015  11.0    Indiana       6-11   240.0     C    Fr.   1.0
4026     466.0  2015  46.0    *Toronto       6-4   215.0    SG    Sr.   2.0
4027     356.0  2015  30.0    Golden St.     6-9   222.0    PF    Fr.   1.0
4035     384.0  2015  20.0    Toronto        6-5   180.0  PG/SG   Sr.   1.0
4108     398.0  2015  38.0    Detroit        6-7   220.0    SG    Sr.   2.0
4121     294.0  2015  21.0    Dallas         6-6   230.0  SG/SF   Jr.   1.0
4345     477.0  2015   9.0    Charlotte      7-1   230.0     C    Sr.   1.0
4346     367.0  2015  18.0    *Houston       6-9   220.0    SF    Jr.   1.0

       Found
149      both
150      both
175      both
348      both
1002     both
1003     both
1004     both
1091     both
1675     both
1758     both
1808     both
1809     both
1810     both
1812     both
1813     both
1814     both
1990     both
1991     both
```

```
           2141    both
           2855    both
           2856    both
           2890    both
           2953    both
           3697    both
           3748    both
           3752    both
           3771    both
           3821    both
           4026    both
           4027    both
           4035    both
           4108    both
           4121    both
           4345    both
           4346    both

           [35 rows x 35 columns]

In [63]: ncaadraft2015[(ncaadraft2015['Found']=='right_only') & (ncaadraft2015.C !='Intl.')]

Out[63]:       Team + Rank  Minutes Rank             Player Class  Pos  \
           4435         NaN          NaN       Cameron Payne   NaN  NaN
           4436         NaN          NaN       Rashad Vaughn   NaN  NaN
           4437         NaN          NaN        Jarell Martin   NaN  NaN
           4439         NaN          NaN      Larry Nance Jr.  NaN  NaN
           4440         NaN          NaN           RJ Hunter   NaN  NaN
           4442         NaN          NaN        Jordan Mickey  NaN  NaN
           4444         NaN          NaN       Richaun Holmes  NaN  NaN
           4446         NaN          NaN  Marcus Thornton WM   NaN  NaN
           4450         NaN          NaN  Sir'Dominic Pointer  NaN  NaN
           4452         NaN          NaN       Branden Dawson  NaN  NaN
           4454         NaN          NaN            JP Tokoto  NaN  NaN

                           School Conf    G   MP  Totals FG   ...   Totals PTS  year  \
           4435        Murray St.  NaN  NaN  NaN       NaN   ...          NaN   NaN
           4436              UNLV  NaN  NaN  NaN       NaN   ...          NaN   NaN
           4437               LSU  NaN  NaN  NaN       NaN   ...          NaN   NaN
           4439           Wyoming  NaN  NaN  NaN       NaN   ...          NaN   NaN
           4440       Georgia St.  NaN  NaN  NaN       NaN   ...          NaN   NaN
           4442               LSU  NaN  NaN  NaN       NaN   ...          NaN   NaN
           4444     Bowling Green  NaN  NaN  NaN       NaN   ...          NaN   NaN
           4446    William & Mary  NaN  NaN  NaN       NaN   ...          NaN   NaN
           4450         St. Johns  NaN  NaN  NaN       NaN   ...          NaN   NaN
           4452      Michigan St.  NaN  NaN  NaN       NaN   ...          NaN   NaN
           4454    North Carolina  NaN  NaN  NaN       NaN   ...          NaN   NaN
```

```
              #             Team    H      W      P     C  round       Found
      4435  14.0   Oklahoma Cty  6-2  185.0    PG   So.    1.0  right_only
      4436  17.0      Milwaukee  6-5  200.0    SG   Fr.    1.0  right_only
      4437  25.0        Memphis  6-9  240.0    PF   So.    1.0  right_only
      4439  27.0     *LA Lakers  6-9  230.0    PF   Sr.    1.0  right_only
      4440  28.0        *Boston  6-6  185.0    SG   Jr.    1.0  right_only
      4442  33.0        *Boston  6-8  240.0    PF   So.    2.0  right_only
      4444  37.0   *Philadelphia 6-9  245.0    PF   Sr.    2.0  right_only
      4446  45.0         Boston  6-3  180.0 PG/SG   Sr.    2.0  right_only
      4450  53.0     *Cleveland  6-6  190.0    SF   Sr.    2.0  right_only
      4452  56.0   *LA Clippers  6-7  230.0 SF/PF   Sr.    2.0  right_only
      4454  58.0  *Philadelphia  6-6  195.0 SG/SF   Jr.    2.0  right_only

      [11 rows x 35 columns]

In [64]: ncaa2015[ncaa2015.Player.str.endswith('Pointer')]

Out[64]:         Team + Rank  Minutes Rank               Player Class  Pos  \
      3687  St. John's (NY)3               3  Sir'Dominic Pointer    SR  G-F


                    School      Conf   G   MP  Totals FG   ...    Totals ORB  \
      3687  St. John's (NY)  Big East  28  962         148  ...            56


            Totals DRB  Totals TRB  Totals AST  Totals STL  Totals BLK  Totals TOV  \
      3687         156         212          81          57          72          54


            Totals PF  Totals PTS  year
      3687         86         382  2015

      [1 rows x 27 columns]

In [65]: #c.to_excel('nba2015.xlsx')

In [66]: d = pd.read_excel('nba2015.xlsx')
         ncaadraft2015 = pd.merge(ncaa2015, d, on=['Player', 'School'], how='outer', indicator=

In [67]: ncaadraft2015[(ncaadraft2015['Found']=='right_only') & (ncaadraft2015.C !='Intl.')]

Out[67]:       Team + Rank  Minutes Rank               Player Class  Pos  \
      4435          NaN           NaN       Cameron Payne   NaN  NaN
      4436          NaN           NaN       Rashad Vaughn   NaN  NaN
      4437          NaN           NaN        Jarell Martin   NaN  NaN
      4439          NaN           NaN     Larry Nance Jr.   NaN  NaN
      4440          NaN           NaN            RJ Hunter   NaN  NaN
      4442          NaN           NaN        Jordan Mickey   NaN  NaN
      4444          NaN           NaN       Richaun Holmes   NaN  NaN
      4446          NaN           NaN  Marcus Thornton WM   NaN  NaN
      4450          NaN           NaN  Sir'Dominic Pointer   NaN  NaN
```

```
              School  Conf    G   MP  Totals FG    ...          Totals PTS  year  \
4435      Murray St.   NaN  NaN  NaN        NaN    ...                 NaN   NaN
4436            UNLV   NaN  NaN  NaN        NaN    ...                 NaN   NaN
4437             LSU   NaN  NaN  NaN        NaN    ...                 NaN   NaN
4439         Wyoming   NaN  NaN  NaN        NaN    ...                 NaN   NaN
4440     Georgia St.   NaN  NaN  NaN        NaN    ...                 NaN   NaN
4442             LSU   NaN  NaN  NaN        NaN    ...                 NaN   NaN
4444   Bowling Green   NaN  NaN  NaN        NaN    ...                 NaN   NaN
4446  William & Mary   NaN  NaN  NaN        NaN    ...                 NaN   NaN
4450       St. Johns   NaN  NaN  NaN        NaN    ...                 NaN   NaN

         #           Team    H      W      P    C  round        Found
4435  14.0   Oklahoma Cty  6-2  185.0     PG  So.    1.0   right_only
4436  17.0      Milwaukee  6-5  200.0     SG  Fr.    1.0   right_only
4437  25.0        Memphis  6-9  240.0     PF  So.    1.0   right_only
4439  27.0      *LA Lakers  6-9  230.0     PF  Sr.    1.0   right_only
4440  28.0        *Boston  6-6  185.0     SG  Jr.    1.0   right_only
4442  33.0        *Boston  6-8  240.0     PF  So.    2.0   right_only
4444  37.0  *Philadelphia  6-9  245.0     PF  Sr.    2.0   right_only
4446  45.0         Boston  6-3  180.0  PG/SG  Sr.    2.0   right_only
4450  53.0      *Cleveland  6-6  190.0     SF  Sr.    2.0   right_only

[9 rows x 35 columns]
```

```
In [68]: from sklearn.decomposition import PCA
         NBA_PCA = PCA(n_components=12, copy=True, whiten=False, svd_solver='auto', tol=0.0, it

In [70]: [(c, type(ncaadraft2015[c][0])) for c in ncaadraft2015.columns]

Out[70]: [('Team + Rank', str),
          ('Minutes Rank', numpy.float64),
          ('Player', str),
          ('Class', str),
          ('Pos', str),
          ('School', str),
          ('Conf', str),
          ('G', numpy.float64),
          ('MP', numpy.float64),
          ('Totals FG', numpy.float64),
          ('Totals FGA', numpy.float64),
          ('Totals 2P', numpy.float64),
          ('Totals 2PA', numpy.float64),
          ('Totals 3P', numpy.float64),
          ('Totals 3PA', numpy.float64),
          ('Totals FT', numpy.float64),
          ('Totals FTA', numpy.float64),
          ('Totals ORB', numpy.float64),
          ('Totals DRB', numpy.float64),
```

```
              ('Totals TRB', numpy.float64),
              ('Totals AST', numpy.float64),
              ('Totals STL', numpy.float64),
              ('Totals BLK', numpy.float64),
              ('Totals TOV', numpy.float64),
              ('Totals PF', numpy.float64),
              ('Totals PTS', numpy.float64),
              ('year', str),
              ('#', numpy.float64),
              ('Team', float),
              ('H', float),
              ('W', numpy.float64),
              ('P', float),
              ('C', float),
              ('round', numpy.float64),
              ('Found', str)]

In [71]: y = ncaadraft2015["round"]

In [72]: X = ncaadraft2015[["G",
                            "Totals FT", "MP", "Totals FG", "Totals FGA", "Totals 2PA", "Totals

In [73]: X.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 4455 entries, 0 to 4454
Data columns (total 12 columns):
G             4432 non-null float64
Totals FT     4432 non-null float64
MP            4432 non-null float64
Totals FG     4432 non-null float64
Totals FGA    4432 non-null float64
Totals 2PA    4432 non-null float64
Totals 3P     4432 non-null float64
Totals 3PA    4432 non-null float64
Totals PF     4432 non-null float64
Totals FT     4432 non-null float64
Totals TOV    4432 non-null float64
Totals BLK    4432 non-null float64
dtypes: float64(12)
memory usage: 612.5 KB


In [74]: NBA_PCA.fit(X.fillna(0))

Out[74]: PCA(copy=True, iterated_power='auto', n_components=12, random_state=None,
           svd_solver='auto', tol=0.0, whiten=False)

In [75]: NBA_PCA.score(X.fillna(0), y.fillna(0))
```

```
Out[75]:  -15.432924581292502

In [76]:  NBA_PCA.score_samples(X.fillna(0))

Out[76]:  array([-18.41291705, -17.31809269, -16.24035275, ..., -16.13573713,
                 -16.13573713, -16.13573713])
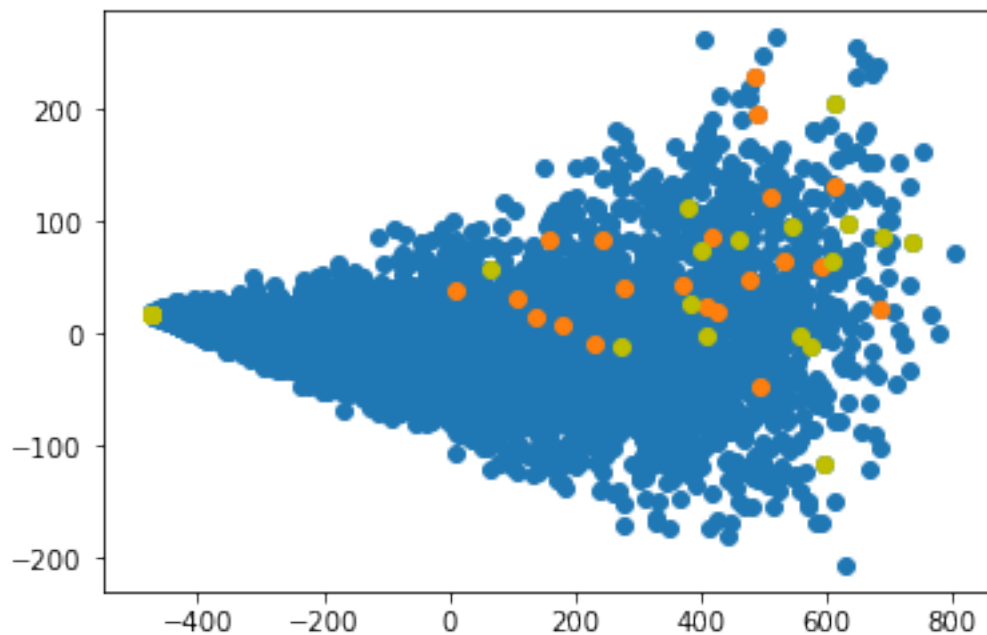
In [77]:  PCA2 = PCA(n_components = 2)

In [78]:  PCA2.fit(X.fillna(0))

Out[78]:  PCA(copy=True, iterated_power='auto', n_components=2, random_state=None,
              svd_solver='auto', tol=0.0, whiten=False)

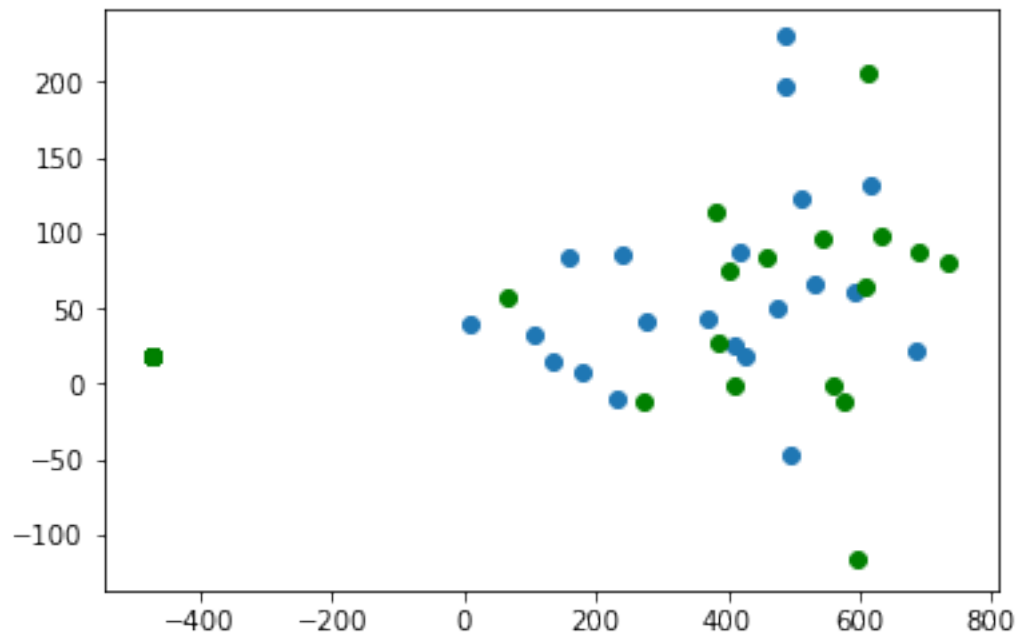In [79]:  X2 = PCA2.fit(X.fillna(0)).transform(X.fillna(0))

In [80]:  import matplotlib.pyplot as plt
          %matplotlib inline
          plt.scatter(X2[:,0],X2[:,1])
          plt.scatter(X2[y==1,0],X2[y==1,1])
          plt.scatter(X2[y==2,0],X2[y==2,1], color = 'y')

Out[80]:  <matplotlib.collections.PathCollection at 0x1ddeb9f4c18>
```



```
In [81]:  plt.scatter(X2[y==1,0],X2[y==1,1])
          plt.scatter(X2[y==2,0],X2[y==2,1], color = 'g')
```

```
Out[81]: <matplotlib.collections.PathCollection at 0x1dde8b32518>
```



```
In [82]: PCA2.components_.shape

Out[82]: (2, 12)

In [83]: PCA2.components_

Out[83]: array([[ 0.01810339,  0.07900715,  0.91436761,  0.12770712,  0.2859286 ,
                 0.18548448,  0.03653402,  0.10044412,  0.06095739,  0.07900715,
                 0.05785737,  0.01280984],
               [-0.02141211,  0.20188979, -0.29997777,  0.25576466,  0.38526974,
                 0.70050338, -0.12157717, -0.31523364,  0.02586588,  0.20188979,
                 0.06358627,  0.06794895]])

In [84]: X.columns

Out[84]: Index(['G', 'Totals FT', 'MP', 'Totals FG', 'Totals FGA', 'Totals 2PA',
               'Totals 3P', 'Totals 3PA', 'Totals PF', 'Totals FT', 'Totals TOV',
               'Totals BLK'],
              dtype='object')

In [85]: PCA3 = PCA(n_components = 3)

In [86]: X3 = PCA3.fit(X.fillna(0)).transform(X.fillna(0))
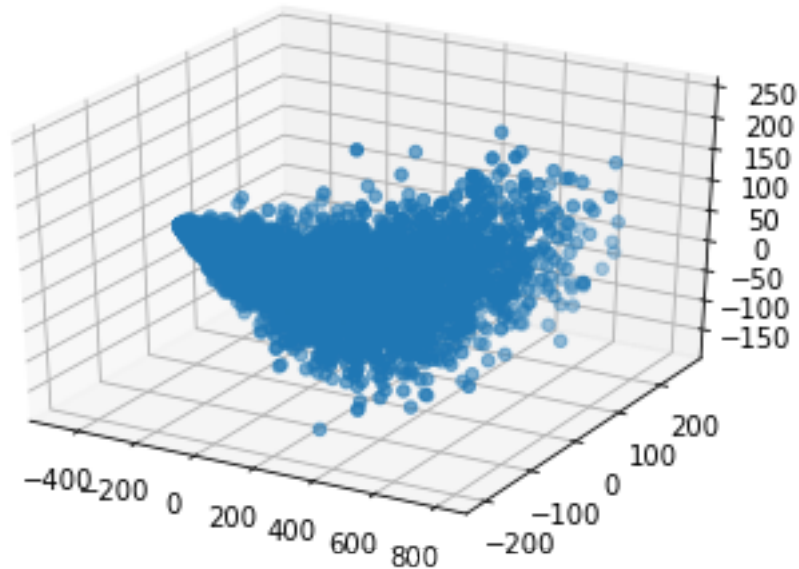
In [87]: X3
```

19

```
Out[87]: array([[ 365.90099956,  -56.48611856,   56.79248452],
               [ 339.73939064,  -67.99349321,   20.51001516],
               [ 200.32472025, -109.90129996,  -87.41819176],
               ...,
               [-474.51576903,   18.00355954,   20.23131023],
               [-474.51576903,   18.00355954,   20.23131023],
               [-474.51576903,   18.00355954,   20.23131023]])
```

```python
In [88]: from mpl_toolkits.mplot3d import Axes3D
```

```python
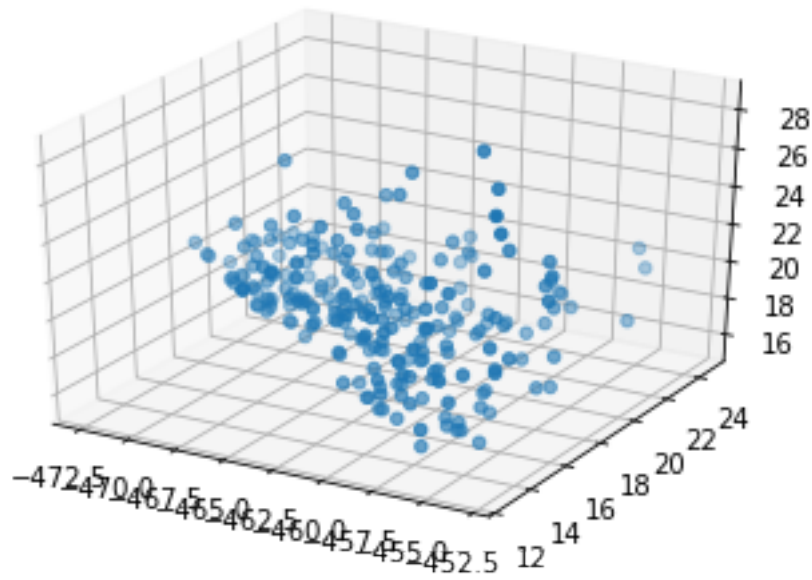In [89]: fig = plt.figure()
         Ax = fig.add_subplot(111, projection = "3d")
         Ax.scatter(X3[:,0], X3[:,1], X3[:,2])
```

```
Out[89]: <mpl_toolkits.mplot3d.art3d.Path3DCollection at 0x1dde8ddfb70>
```



```python
In [90]: fig = plt.figure()
         Ax = fig.add_subplot(111, projection = "3d")
         Ax.scatter(X3[X["MP"] < 20,0], X3[X["MP"] < 20,1], X3[X["MP"] < 20,2])
```

```
Out[90]: <mpl_toolkits.mplot3d.art3d.Path3DCollection at 0x1dde96fcb00>
```

20

```
In [91]: FR = ncaadraft2015.loc[ncaadraft2015["round"]==1,["G", "Totals FT", "MP", "Totals FG"

In [92]: FR

Out[92]:          G  Totals FT      MP  Totals FG  Totals FGA  Totals 2PA  Totals 3P  \
         149   28.0      112.0   793.0      130.0       277.0       203.0       26.0
         150   28.0      103.0   785.0      102.0       192.0       171.0        5.0
         175   28.0       77.0   818.0      203.0       360.0       339.0       10.0
         1002  28.0      104.0   918.0       92.0       218.0       128.0       35.0
         1003  27.0       84.0   840.0      206.0       310.0       310.0        0.0
         1004  28.0       71.0   806.0      116.0       253.0       169.0       32.0
         1758  28.0       44.0   552.0       83.0       187.0       110.0       30.0
         1808  28.0       52.0   706.0      102.0       168.0       168.0        0.0
         1810  27.0       43.0   584.0      105.0       214.0       108.0       48.0
         1812  28.0       71.0   569.0       94.0       170.0       162.0        2.0
         1813  25.0       44.0   536.0       75.0       151.0       123.0        4.0
         1990  28.0      109.0   943.0      172.0       403.0       276.0       43.0
         2855  29.0      117.0  1055.0      165.0       333.0       205.0       43.0
         2890  28.0       87.0   928.0      180.0       391.0       213.0       76.0
         3752  16.0       40.0   450.0       54.0       113.0       112.0        1.0
         3821  28.0       89.0   634.0      104.0       223.0       170.0       16.0
         4027  29.0       84.0   921.0      127.0       272.0       231.0       18.0
         4035  27.0      102.0   873.0      132.0       251.0       195.0       18.0
         4121  22.0       54.0   647.0       97.0       202.0       107.0       46.0
         4345  27.0      104.0   870.0      173.0       318.0       250.0       27.0
         4346  28.0       56.0   823.0      139.0       267.0       169.0       33.0
         4432   NaN        NaN     NaN        NaN         NaN         NaN        NaN
```

4433	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4434	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4435	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4436	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4437	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4438	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4439	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4440	NaN	NaN	NaN	NaN	NaN	NaN	NaN

	Totals 3PA	Totals PF	Totals FT	Totals TOV	Totals BLK
149	74.0	73.0	112.0	67.0	9.0
150	21.0	69.0	103.0	49.0	26.0
175	21.0	59.0	77.0	38.0	44.0
1002	90.0	33.0	104.0	55.0	1.0
1003	0.0	60.0	84.0	73.0	39.0
1004	84.0	79.0	71.0	48.0	20.0
1758	77.0	48.0	44.0	28.0	7.0
1808	0.0	55.0	52.0	36.0	43.0
1810	106.0	38.0	43.0	30.0	1.0
1812	8.0	79.0	71.0	38.0	66.0
1813	28.0	42.0	44.0	28.0	10.0
1990	127.0	50.0	109.0	58.0	5.0
2855	128.0	52.0	117.0	58.0	17.0
2890	178.0	54.0	87.0	77.0	5.0
3752	1.0	41.0	40.0	35.0	33.0
3821	53.0	70.0	89.0	38.0	74.0
4027	41.0	87.0	84.0	43.0	31.0
4035	56.0	32.0	102.0	50.0	23.0
4121	95.0	25.0	54.0	26.0	12.0
4345	68.0	44.0	104.0	39.0	43.0
4346	98.0	31.0	56.0	30.0	16.0
4432	NaN	NaN	NaN	NaN	NaN
4433	NaN	NaN	NaN	NaN	NaN
4434	NaN	NaN	NaN	NaN	NaN
4435	NaN	NaN	NaN	NaN	NaN
4436	NaN	NaN	NaN	NaN	NaN
4437	NaN	NaN	NaN	NaN	NaN
4438	NaN	NaN	NaN	NaN	NaN
4439	NaN	NaN	NaN	NaN	NaN
4440	NaN	NaN	NaN	NaN	NaN

```
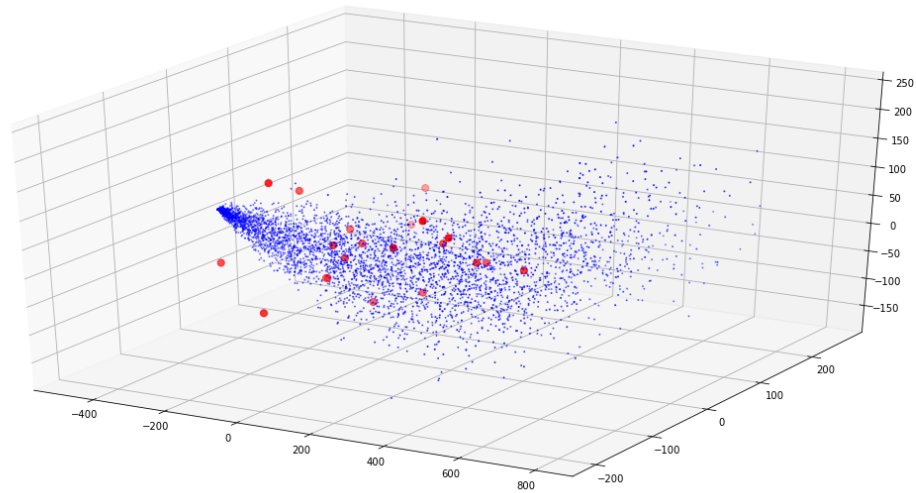In [93]: FR2 = PCA2.fit(FR.dropna()).transform(FR.dropna())

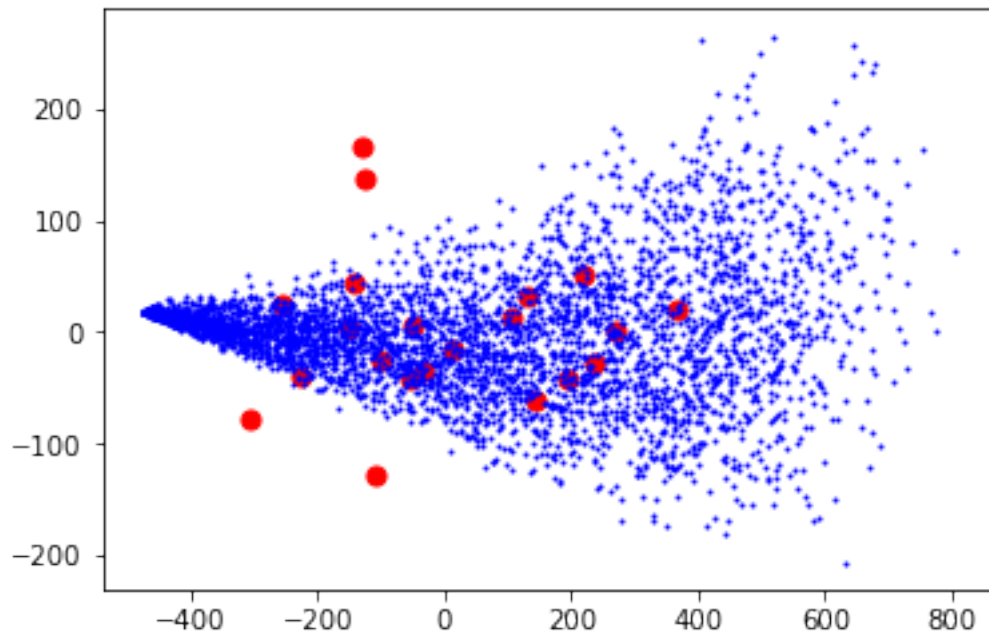In [94]: FR3 = PCA3.fit(FR.dropna()).transform(FR.dropna())

In [95]: fig = plt.figure(figsize=(20,10))
         Ax = fig.add_subplot(111, projection = "3d")
         Ax.scatter(FR3[:,0], FR3[:,1], FR3[:,2], color='r', s=50)
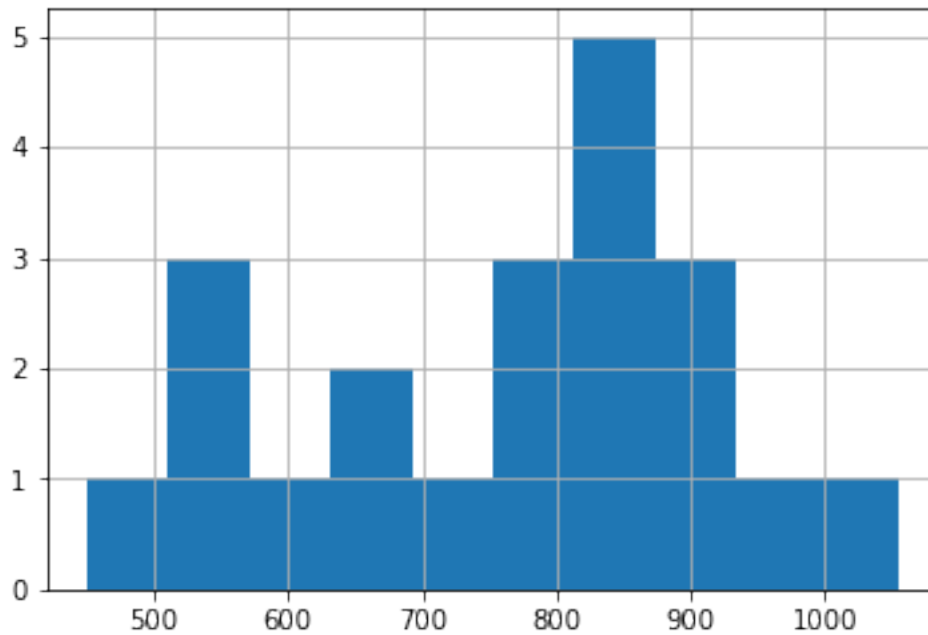         Ax.scatter(X3[:,0], X3[:,1], X3[:,2], color='b', s=1)
```

In [96]: plt.scatter(FR2[:,0], FR2[:,1], color='r', s=50)
         plt.scatter(X2[:,0], X2[:,1], color='b', s=1)

```
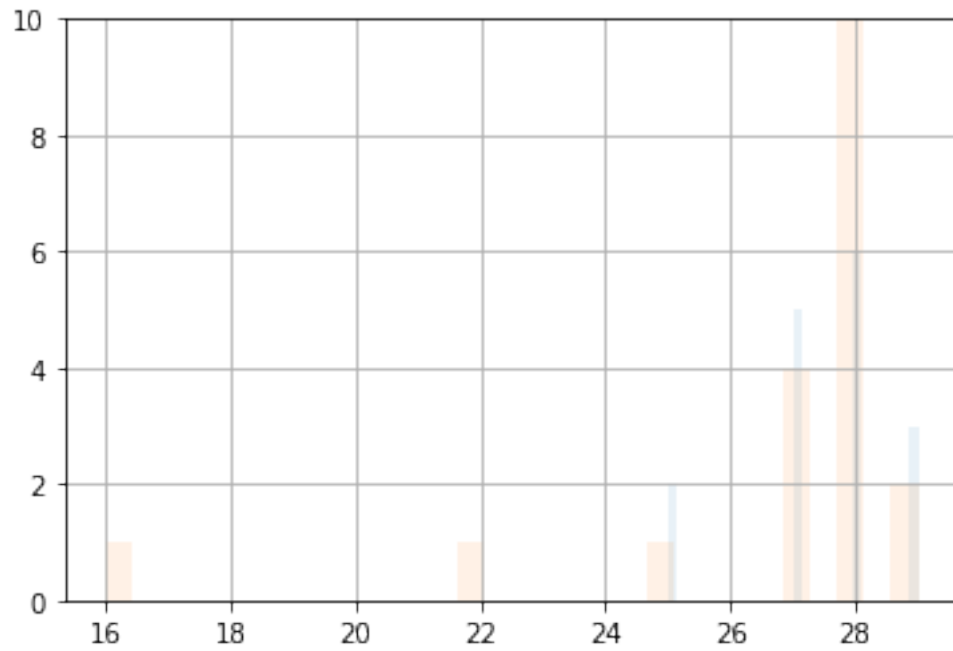In [97]: X.loc[y==1, "MP"].hist()
```

```
Out[97]: <matplotlib.axes._subplots.AxesSubplot at 0x1ddeb2980f0>
```



```
In [98]: X.loc[y==2, "G"].hist(alpha=0.1, bins=30)
         X.loc[y==1, "G"].hist(alpha=0.1, bins=30)
         #X["G"].hist(alpha=0.05, bins = 30)
         #plt.xlim(300,1000)
         plt.ylim(0,10)
```

```
Out[98]: (0, 10)
```

```
In [99]: for c in X.columns:
             X.loc[y==2, c].hist(alpha=0.1, bins=30)
             X.loc[y==1, c].hist(alpha=0.1, bins=30)
             X.loc[y==np.nan, c].hist(alpha=0.05, bins = 30)
             #plt.xlim(300,1000)
             #plt.ylim(0,10)
             plt.title(c)
             plt.show()
```

G

Totals FT

Totals FT

Totals FT

Totals FT

Totals FT

Totals FT

MP

Totals FG

Totals FGA

Totals 2PA

Totals 3P

Totals 3PA

Totals PF

Totals FT

Totals FT

Totals FT

Totals FT

Totals FT

Totals FT

Totals TOV

Totals BLK