IEOR 142 Final Project Linear Regression

May 12, 2021

[1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

[2]:
```
# compute out-of-sample R-squared using the test set
def OSR2(model, df_train, df_test, dependent_var):
    y_test = df_test[dependent_var]
    y_pred = model.predict(df_test)
    SSE = np.sum((y_test - y_pred)**2)
    SST = np.sum((y_test - np.mean(df_train[dependent_var]))**2)
    return 1 - SSE/SST
```

1 Model 2: Linear Regression

[3]:
```
dataset = pd.read_csv('update1422.csv')
```

Most Valuable Player (MVP)

[160]:
```
#mvp models
dataset = pd.read_csv('update142.csv')
train = dataset[dataset['Season'] <= 2017]
test = dataset[dataset['Season'] > 2017]

y_train = train['MVP']
y_test = test['MVP']
x_train = train.iloc[:,6:51]
x_test = test.iloc[:,6:51]
```

[5]:
```
x_train.head(5)
```

[5]:

	G	GS	MP	FGM	FGA	FGP	X3PM	X3PA	X3PP	X2PM	…	TOVP	USG	\
0	22	0	7.4	0.8	1.9	0.405	0.2	0.5	0.500	0.5	…	16.4	17.6	
1	56	2	15.4	2.6	6.3	0.410	1.1	3.1	0.345	1.5	…	13.2	22.7	
2	50	0	15.9	2.3	6.4	0.356	0.8	3.0	0.275	1.5	…	13.0	22.7	
3	44	1	12.9	1.3	4.0	0.339	0.6	2.0	0.295	0.8	…	14.5	17.7	
4	57	22	22.4	2.8	7.2	0.390	1.2	3.5	0.350	1.6	…	12.7	18.0	

```
   OWS  DWS   WS   WS48  OBPM  DBPM  BPM  VORP
0 -0.2  0.2  0.0 -0.001  -6.6   0.0 -6.6  -0.2
1  0.4  0.8  1.2  0.065  -0.1  -0.4 -0.5   0.3
2 -0.4  0.7  0.3  0.020  -1.9  -1.1 -2.9  -0.2
3  0.2  0.5  0.7  0.063  -0.6  -0.5 -1.0   0.1
4  1.0  1.2  2.2  0.084  -0.2  -0.4 -0.6   0.5

[5 rows x 45 columns]
```

[6]:
```
y_train.head(5)
```

[6]:
```
0    0.0
1    0.0
2    0.0
3    0.0
4    0.0
Name: MVP, dtype: float64
```

[7]:
```
x_train.columns
```

[7]:
```
Index(['G', 'GS', 'MP', 'FGM', 'FGA', 'FGP', 'X3PM', 'X3PA', 'X3PP', 'X2PM',
       'X2PA', 'X2PP', 'EFG', 'FTM', 'FTA', 'FTP', 'ORB', 'DRB', 'TRB', 'AST',
       'STL', 'BLK', 'TOV', 'PF', 'PTS', 'PER', 'TS', 'X3PAR', 'FTR', 'ORBP',
       'DRBP', 'TRBP', 'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
       'WS', 'WS48', 'OBPM', 'DBPM', 'BPM', 'VORP'],
      dtype='object')
```

[141]:
```
#taking out the 4 additional awards plus categorical stats
depedent_var = train.columns.difference(['Tm','Pos','Player','MVP',␣
↪'DPOY',"ROY",'SMOY','MIP'])
depedent_var
```

[141]:
```
Index(['AST', 'ASTP', 'Age', 'BLK', 'BLKP', 'BPM', 'DBPM', 'DRB', 'DRBP',
       'DWS', 'EFG', 'FGA', 'FGM', 'FGP', 'FTA', 'FTM', 'FTP', 'FTR', 'G',
       'GS', 'MP', 'OBPM', 'ORB', 'ORBP', 'OWS', 'PER', 'PF', 'PTS', 'STL',
       'STLP', 'Season', 'TOV', 'TOVP', 'TRB', 'TRBP', 'TS', 'USG',
       'Unnamed: 0', 'VORP', 'WS', 'WS48', 'X2PA', 'X2PM', 'X2PP', 'X3PA',
       'X3PAR', 'X3PM', 'X3PP'],
      dtype='object')
```

[142]:
```
#creating list for regression
depedent_var = "+".join(depedent_var)
```

[143]:
```
depedent_var
```

[143]:
```
'AST+ASTP+Age+BLK+BLKP+BPM+DBPM+DRB+DRBP+DWS+EFG+FGA+FGM+FGP+FTA+FTM+FTP+FTR+G+G
S+MP+OBPM+ORB+ORBP+OWS+PER+PF+PTS+STL+STLP+Season+TOV+TOVP+TRB+TRBP+TS+USG+Unnam
```

```
ed: 0+VORP+WS+WS48+X2PA+X2PM+X2PP+X3PA+X3PAR+X3PM+X3PP'
```

[144]:
```
train.head(5)
```

[144]:
```
   Unnamed: 0        Player  Season Pos  Age   Tm   G  GS    MP  FGM  …  \
0           1  A.J. Hammons    2017   C   24  DAL  22   0   7.4  0.8  …
1           2    A.J. Price    2010  PG   23  IND  56   2  15.4  2.6  …
2           3    A.J. Price    2011  PG   24  IND  50   0  15.9  2.3  …
3           4    A.J. Price    2012  PG   25  IND  44   1  12.9  1.3  …
4           5    A.J. Price    2013  PG   26  WAS  57  22  22.4  2.8  …

    WS48  OBPM  DBPM  BPM  VORP  DPOY  ROY  SMOY  MIP  MVP
0 -0.001  -6.6   0.0 -6.6  -0.2   0.0  0.0   0.0  0.0  0.0
1  0.065  -0.1  -0.4 -0.5   0.3   0.0  0.0   0.0  0.0  0.0
2  0.020  -1.9  -1.1 -2.9  -0.2   0.0  0.0   0.0  0.0  0.0
3  0.063  -0.6  -0.5 -1.0   0.1   0.0  0.0   0.0  0.0  0.0
4  0.084  -0.2  -0.4 -0.6   0.5   0.0  0.0   0.0  0.0  0.0

[5 rows x 56 columns]
```

[145]:
```
cols = [ 'Season', 'Age', 'G', 'GS', 'MP',
       'FGM', 'FGA', 'FGP', 'X3PM', 'X3PA', 'X3PP', 'X2PM', 'X2PA', 'X2PP',
       'EFG', 'FTM', 'FTA', 'FTP', 'ORB', 'DRB', 'TRB', 'AST', 'STL', 'BLK',
       'TOV', 'PF', 'PTS', 'PER', 'TS', 'X3PAR', 'FTR', 'ORBP', 'DRBP', 'TRBP',
       'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS', 'WS', 'WS48',
       'OBPM', 'DBPM', 'BPM', 'VORP']
```

[147]:
```
#calculate VIF
import statsmodels.api as sm

from statsmodels.stats.outliers_influence import variance_inflation_factor

# The dataframe passed to VIF must include the intercept term
def VIF(df, columns):
    values = sm.add_constant(df[columns]).values
    num_columns = len(columns)+1
    vif = [variance_inflation_factor(values, i) for i in range(num_columns)]
    return pd.Series(vif[1:], index=columns)
```

[148]:
```
VIF(train,cols)
```

[148]:
```
Season       1.194476
Age          1.171180
G            2.391408
GS           1.641163
MP          19.132866
FGM       1569.891795
```

```
FGA       3525.506741
FGP         36.893949
X3PM       269.847081
X3PA      1029.538971
X3PP         2.360528
X2PM       639.446195
X2PA      2108.013946
X2PP         7.219283
EFG         26.918430
FTM        141.599957
FTA         60.443550
FTP          2.412699
ORB        148.398531
DRB        452.846363
TRB        949.698229
AST          9.618181
STL          5.474090
BLK          6.334876
TOV         10.020104
PF           4.023531
PTS       2106.101220
PER         82.580696
TS          11.568447
X3PAR        8.638536
FTR          3.819038
ORBP       352.290197
DRBP       650.737814
TRBP      1610.725851
ASTP         8.143810
STLP         5.133920
BLKP         7.322589
TOVP         4.740989
USG         11.597479
OWS        685.531693
DWS        254.394251
WS        1379.407787
WS48        32.905514
OBPM      2664.256508
DBPM       729.156559
BPM       4008.612810
VORP        12.315918
dtype: float64
```

[149]:
```
#Due to the high VIF above we are going to remove FGA
depedent_var = train.columns.difference(['Tm','Pos','Player','MVP',␣
 ↪'DPOY',"ROY",'SMOY','MIP','FGA'])
depedent_var
```

```
depedent_var = "+".join(depedent_var)
```

[150]:
```
cols = [ 'Season', 'Age', 'G', 'GS', 'MP',
       'FGM', 'FGP', 'X3PM', 'X3PA', 'X3PP', 'X2PM', 'X2PA', 'X2PP',
       'EFG', 'FTM', 'FTA', 'FTP', 'ORB', 'DRB', 'TRB', 'AST', 'STL', 'BLK',
       'TOV', 'PF', 'PTS', 'PER', 'TS', 'X3PAR', 'FTR', 'ORBP', 'DRBP', 'TRBP',
       'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS', 'WS', 'WS48',
       'OBPM', 'DBPM', 'BPM', 'VORP']
```

[151]:
```
VIF(train,cols)
```

[151]:
```
Season       1.192306
Age          1.170550
G            2.390958
GS           1.641136
MP          19.116749
FGM       1569.539979
FGP         36.878469
X3PM       269.318653
X3PA        83.753328
X3PP         2.356868
X2PM       639.400745
X2PA       112.193462
X2PP         7.211029
EFG         26.865980
FTM        141.549555
FTA         60.438245
FTP          2.412662
ORB        148.089320
DRB        451.506745
TRB        947.420101
AST          9.616408
STL          5.473669
BLK          6.329678
TOV         10.014365
PF           4.022295
PTS       2105.344255
PER         82.536375
TS          11.539527
X3PAR        8.636201
FTR          3.818829
ORBP       352.189012
DRBP       650.453957
TRBP      1609.899949
ASTP         8.136427
STLP         5.124074
BLKP         7.321098
```

```
TOVP         4.731929
USG         11.596995
OWS        683.923741
DWS        253.582976
WS        1375.251650
WS48        32.884934
OBPM      2663.571876
DBPM       729.062872
BPM       4008.019099
VORP        12.314954
dtype: float64
```

[152]:
```
#removing pts due to high VIF
cols = [ 'Season', 'Age', 'G', 'GS', 'MP',
       'FGM', 'FGP', 'X3PM', 'X3PA', 'X3PP', 'X2PM', 'X2PA', 'X2PP',
       'EFG', 'FTM', 'FTA', 'FTP', 'ORB', 'DRB', 'TRB', 'AST', 'STL', 'BLK',
       'TOV', 'PF', 'PER', 'TS', 'X3PAR', 'FTR', 'ORBP', 'DRBP', 'TRBP',
       'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS', 'WS', 'WS48',
       'OBPM', 'DBPM', 'BPM', 'VORP']
```

[21]:
```
VIF(train,cols)
```

[21]:
```
Season       1.192514
Age          1.157173
G            2.157620
GS           3.608952
MP          23.900449
FGM       2283.109243
FGP         31.824353
X3PM       291.274036
X3PA        89.004363
X3PP         2.237480
X2PM      1920.781309
X2PA       176.130874
X2PP         7.776078
EFG         19.930258
FTM         72.487559
FTA         75.538858
FTP          2.264529
ORB        276.497219
DRB       1197.834521
TRB       2330.703775
AST         11.951138
STL          6.287637
BLK          5.729287
TOV         14.439666
PF           3.709223
```

```
PER        101.732875
TS           8.920161
X3PAR        8.271101
FTR          3.310331
ORBP       327.809995
DRBP       705.569352
TRBP      1673.008052
ASTP         9.241032
STLP         5.030717
BLKP         6.481220
TOVP         3.568362
USG         13.940051
OWS       1779.961837
DWS        540.702192
WS        3496.352067
WS48        35.859572
OBPM      3167.753461
DBPM       709.863068
BPM       4514.127858
VORP        24.534486
dtype: float64
```

[22]:
```
#removing bpm due to high VIF
cols = [ 'Season', 'Age', 'G', 'GS', 'MP',
       'FGM', 'FGP', 'X3PM', 'X3PA', 'X3PP', 'X2PM', 'X2PA', 'X2PP',
       'EFG', 'FTM', 'FTA', 'FTP', 'ORB', 'DRB', 'TRB', 'AST', 'STL', 'BLK',
       'TOV', 'PF', 'PER', 'TS', 'X3PAR', 'FTR', 'ORBP', 'DRBP', 'TRBP',
       'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS', 'WS', 'WS48',
       'OBPM', 'DBPM', 'VORP']
```

[23]:
```
VIF(train,cols)
```

[23]:
```
Season       1.191270
Age          1.156164
G            2.157619
GS           3.608781
MP          23.899740
FGM       2282.132921
FGP         31.819902
X3PM       291.245169
X3PA        88.987430
X3PP         2.237461
X2PM      1920.585170
X2PA       176.015777
X2PP         7.771931
EFG         19.928501
FTM         72.454841
```

```
FTA         75.471022
FTP          2.263889
ORB        276.380524
DRB       1197.362035
TRB       2329.768191
AST         11.951136
STL          6.284697
BLK          5.729277
TOV         14.439240
PF           3.709073
PER        101.630358
TS           8.911092
X3PAR        8.271061
FTR          3.304577
ORBP       327.529379
DRBP       705.241986
TRBP      1672.228619
ASTP         9.235713
STLP         5.030714
BLKP         6.474528
TOVP         3.562774
USG         13.922676
OWS       1779.322436
DWS        540.486595
WS        3494.802942
WS48        35.852381
OBPM        31.997043
DBPM         9.361493
VORP        24.522676
dtype: float64
```

[24]:
```
#removing ws due to high VIF
cols = [ 'Season', 'Age', 'G', 'GS', 'MP',
       'FGM', 'FGP', 'X3PM', 'X3PA', 'X3PP', 'X2PM', 'X2PA', 'X2PP',
       'EFG', 'FTM', 'FTA', 'FTP', 'ORB', 'DRB', 'TRB', 'AST', 'STL', 'BLK',
       'TOV', 'PF', 'PER', 'TS', 'X3PAR', 'FTR', 'ORBP', 'DRBP', 'TRBP',
       'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',  'WS48',
       'OBPM', 'DBPM', 'VORP']
```

[25]:
```
VIF(train,cols)
```

[25]:
```
Season       1.191115
Age          1.156049
G            2.153862
GS           3.607895
MP          23.898944
FGM       2281.864283
```

```
FGP         31.817768
X3PM       291.242649
X3PA        88.975874
X3PP         2.237320
X2PM      1920.166660
X2PA       175.970957
X2PP         7.771130
EFG         19.927337
FTM         72.454820
FTA         75.466325
FTP          2.263751
ORB        276.367278
DRB       1197.224624
TRB       2329.498754
AST         11.949438
STL          6.282586
BLK          5.729181
TOV         14.437050
PF           3.708898
PER        101.553367
TS           8.909877
X3PAR        8.269175
FTR          3.301326
ORBP       327.444951
DRBP       704.904861
TRBP      1671.389859
ASTP         9.234727
STLP         5.030692
BLKP         6.473371
TOVP         3.562668
USG         13.912545
OWS         20.766398
DWS          6.221203
WS48        35.848544
OBPM        31.942630
DBPM         9.354186
VORP        24.440627
dtype: float64
```

[26]:
```
#removing trb due to high VIF
cols = [ 'Season', 'Age', 'G', 'GS', 'MP',
       'FGM', 'FGP', 'X3PM', 'X3PA', 'X3PP', 'X2PM', 'X2PA', 'X2PP',
       'EFG', 'FTM', 'FTA', 'FTP', 'ORB', 'DRB', 'AST', 'STL', 'BLK',
       'TOV', 'PF', 'PER', 'TS', 'X3PAR', 'FTR', 'ORBP', 'DRBP', 'TRBP',
       'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',  'WS48',
       'OBPM', 'DBPM', 'VORP']
VIF(train,cols)
```

[26]:
```
Season       1.191098
Age          1.156027
G            2.151490
GS           3.605488
MP          23.887734
FGM       2281.758838
FGP         31.777961
X3PM       291.121029
X3PA        88.926900
X3PP         2.237278
X2PM      1920.092467
X2PA       175.970938
X2PP         7.767782
EFG         19.921656
FTM         72.437155
FTA         75.464105
FTP          2.263644
ORB         11.028235
DRB         13.013680
AST         11.939587
STL          6.282178
BLK          5.719172
TOV         14.420354
PF           3.707280
PER        101.553365
TS           8.909282
X3PAR        8.250469
FTR          3.299641
ORBP       327.009506
DRBP       704.484994
TRBP      1670.350156
ASTP         9.228112
STLP         5.026954
BLKP         6.460002
TOVP         3.553561
USG         13.911940
OWS         20.766174
DWS          6.221037
WS48        35.845835
OBPM        31.942169
DBPM         9.342066
VORP        24.440467
dtype: float64
```

[27]:
```
#removing fgm due to high VIF
cols = [ 'Season', 'Age', 'G', 'GS', 'MP',
       'FGP', 'X3PM', 'X3PA', 'X3PP', 'X2PM', 'X2PA', 'X2PP',
```

```
       'EFG', 'FTM', 'FTA', 'FTP', 'ORB', 'DRB', 'AST', 'STL', 'BLK',
       'TOV', 'PF', 'PER', 'TS', 'X3PAR', 'FTR', 'ORBP', 'DRBP', 'TRBP',
       'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',  'WS48',
       'OBPM', 'DBPM', 'VORP']
VIF(train,cols)
```

[27]:
```
Season       1.191091
Age          1.155668
G            2.151457
GS           3.605458
MP          23.874730
FGP         31.766296
X3PM        81.536482
X3PA        86.184756
X3PP         2.234934
X2PM       171.871076
X2PA       173.805932
X2PP         7.753346
EFG         19.831132
FTM         72.424124
FTA         75.463811
FTP          2.263583
ORB         11.027267
DRB         13.007214
AST         11.939387
STL          6.278886
BLK          5.717805
TOV         14.420343
PF           3.706706
PER        101.518087
TS           8.893118
X3PAR        8.250464
FTR          3.299591
ORBP       326.981006
DRBP       704.430155
TRBP      1670.288967
ASTP         9.207796
STLP         5.024333
BLKP         6.458940
TOVP         3.547873
USG         13.911815
OWS         20.750167
DWS          6.216409
WS48        35.843736
OBPM        31.936323
DBPM         9.341701
VORP        24.439215
```

```
dtype: float64
```

[28]:
```
#removing trbp due to high VIF
cols = [ 'Season', 'Age', 'G', 'GS', 'MP',
       'FGP', 'X3PM', 'X3PA', 'X3PP', 'X2PM', 'X2PA', 'X2PP',
       'EFG', 'FTM', 'FTA', 'FTP', 'ORB', 'DRB', 'AST', 'STL', 'BLK',
       'TOV', 'PF', 'PER', 'TS', 'X3PAR', 'FTR', 'ORBP', 'DRBP',
       'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',  'WS48',
       'OBPM', 'DBPM', 'VORP']
VIF(train,cols)
```

[28]:
```
Season      1.190743
Age         1.154935
G           2.150657
GS          3.604758
MP         23.874285
FGP        31.761892
X3PM       81.251721
X3PA       85.676643
X3PP        2.232876
X2PM      169.497259
X2PA      171.227676
X2PP        7.752946
EFG        19.831128
FTM        72.412926
FTA        75.434079
FTP         2.263343
ORB        10.843960
DRB        12.914462
AST        11.938813
STL         6.276723
BLK         5.688668
TOV        14.361535
PF          3.655416
PER        96.311802
TS          8.890759
X3PAR       8.241833
FTR         3.295449
ORBP       11.112562
DRBP        7.624789
ASTP        9.206936
STLP        5.023912
BLKP        6.458082
TOVP        3.540283
USG        12.927476
OWS        20.625135
DWS         6.190195
```

```
WS48       33.898806
OBPM       31.725791
DBPM        9.254250
VORP       24.296651
dtype: float64
```

[29]:
```
#removing X2PA due to high VIF
cols = [ 'Season', 'Age', 'G', 'GS', 'MP',
       'FGP', 'X3PM', 'X3PA', 'X3PP', 'X2PM', 'X2PP',
       'EFG', 'FTM', 'FTA', 'FTP', 'ORB', 'DRB', 'AST', 'STL', 'BLK',
       'TOV', 'PF', 'PER', 'TS', 'X3PAR', 'FTR', 'ORBP', 'DRBP',
       'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',  'WS48',
       'OBPM', 'DBPM', 'VORP']
VIF(train,cols)
```

[29]:
```
Season     1.190738
Age        1.154452
G          2.140194
GS         3.602211
MP        22.343352
FGP       31.691341
X3PM      79.032802
X3PA      83.297459
X3PP       2.232309
X2PM      15.664618
X2PP       6.991430
EFG       19.054164
FTM       67.416949
FTA       73.037851
FTP        2.253342
ORB       10.836093
DRB       12.670390
AST       11.906861
STL        6.250501
BLK        5.625577
TOV       14.345090
PF         3.629666
PER       94.908807
TS         8.780881
X3PAR      8.163553
FTR        3.286643
ORBP      11.101929
DRBP       7.534487
ASTP       9.162106
STLP       4.961063
BLKP       6.379067
TOVP       3.478914
```

```
USG       12.680292
OWS       18.516362
DWS        6.183209
WS48      33.747353
OBPM      31.184750
DBPM       9.171278
VORP      24.201876
dtype: float64
```

[30]:
```
#removing PER due to high VIF
cols = [ 'Season', 'Age', 'G', 'GS', 'MP',
       'FGP', 'X3PM', 'X3PA', 'X3PP', 'X2PM', 'X2PP',
       'EFG', 'FTM', 'FTA', 'FTP', 'ORB', 'DRB', 'AST', 'STL', 'BLK',
       'TOV', 'PF', 'TS', 'X3PAR', 'FTR', 'ORBP', 'DRBP',
       'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',  'WS48',
       'OBPM', 'DBPM', 'VORP']
VIF(train,cols)
```

[30]:
```
Season     1.182011
Age        1.150861
G          2.139219
GS         3.602163
MP        22.275823
FGP       31.019189
X3PM      79.032557
X3PA      83.237394
X3PP       2.225294
X2PM      15.462851
X2PP       6.926645
EFG       18.989398
FTM       66.608099
FTA       72.370881
FTP        2.252218
ORB       10.834954
DRB       12.608040
AST       11.905542
STL        6.229601
BLK        5.495950
TOV       14.343970
PF         3.520283
TS         7.707021
X3PAR      6.758435
FTR        3.261893
ORBP      10.888835
DRBP       5.823802
ASTP       8.224600
STLP       3.552044
```

```
BLKP       5.341370
TOVP       3.459957
USG        5.939072
OWS       17.674266
DWS        5.997991
WS48      23.222693
OBPM      15.253435
DBPM       9.116896
VORP      22.931276
dtype: float64
```

[31]:
```
#removing X3PA due to high VIF
cols = [ 'Season', 'Age', 'G', 'GS', 'MP',
       'FGP', 'X3PM','X3PP', 'X2PM', 'X2PP',
       'EFG', 'FTM', 'FTA', 'FTP', 'ORB', 'DRB', 'AST', 'STL', 'BLK',
       'TOV', 'PF', 'TS', 'X3PAR', 'FTR', 'ORBP', 'DRBP',
       'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',  'WS48',
       'OBPM', 'DBPM', 'VORP']
VIF(train,cols)
```

[31]:
```
Season     1.176277
Age        1.150852
G          2.136088
GS         3.580865
MP        21.741809
FGP       29.424692
X3PM       8.565349
X3PP       2.214324
X2PM      15.450172
X2PP       5.050942
EFG       18.543103
FTM       66.607056
FTA       72.192295
FTP        2.252135
ORB       10.735525
DRB       12.579199
AST       11.905477
STL        6.218975
BLK        5.493915
TOV       14.343678
PF         3.518936
TS         7.683229
X3PAR      6.627916
FTR        3.261631
ORBP      10.785946
DRBP       5.817433
ASTP       8.201403
```

```
STLP       3.551842
BLKP       5.325940
TOVP       3.437009
USG        5.933374
OWS       17.404690
DWS        5.992156
WS48      23.210584
OBPM      15.247378
DBPM       9.103365
VORP      22.917772
dtype: float64
```

[32]:
```
#removing FTM due to high VIF
cols = [ 'Season', 'Age', 'G', 'GS', 'MP',
       'FGP', 'X3PM','X3PP', 'X2PM', 'X2PP',
       'EFG', 'FTA', 'FTP', 'ORB', 'DRB', 'AST', 'STL', 'BLK',
       'TOV', 'PF', 'TS', 'X3PAR', 'FTR', 'ORBP', 'DRBP',
       'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',  'WS48',
       'OBPM', 'DBPM', 'VORP']
VIF(train,cols)
```

[32]:
```
Season     1.176277
Age        1.147968
G          2.136082
GS         3.576532
MP        21.741141
FGP       29.118624
X3PM       8.352450
X3PP       2.212089
X2PM      15.440217
X2PP       5.050904
EFG       17.900881
FTA        9.351436
FTP        1.634758
ORB       10.549852
DRB       12.568472
AST       11.902180
STL        6.199054
BLK        5.464350
TOV       14.343330
PF         3.517155
TS         7.675291
X3PAR      6.624863
FTR        3.142224
ORBP      10.784063
DRBP       5.814025
ASTP       8.182898
```

```
STLP       3.551730
BLKP       5.311123
TOVP       3.413876
USG        5.930774
OWS       17.225947
DWS        5.877662
WS48      23.114179
OBPM      15.241474
DBPM       9.103190
VORP      22.887696
dtype: float64
```

[33]:
```
#removing FGP
cols = [ 'Season', 'Age', 'G', 'GS', 'MP',
        'X3PM','X3PP', 'X2PM', 'X2PP',
       'EFG', 'FTA', 'FTP', 'ORB', 'DRB', 'AST', 'STL', 'BLK',
       'TOV', 'PF', 'TS', 'X3PAR', 'FTR', 'ORBP', 'DRBP',
       'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',  'WS48',
       'OBPM', 'DBPM', 'VORP']
VIF(train,cols)
```

[33]:
```
Season     1.176191
Age        1.147318
G          2.133058
GS         3.574171
MP        21.676414
X3PM       7.363673
X3PP       2.038686
X2PM      14.971425
X2PP       4.036321
EFG        5.266394
FTA        9.349609
FTP        1.625006
ORB       10.282187
DRB       12.553260
AST       11.845687
STL        6.198519
BLK        5.460107
TOV       14.335889
PF         3.498074
TS         7.675265
X3PAR      4.688249
FTR        3.139627
ORBP      10.589668
DRBP       5.795417
ASTP       8.148626
STLP       3.551492
```

```
BLKP       5.311048
TOVP       3.395314
USG        5.930508
OWS       17.103823
DWS        5.805003
WS48      23.000550
OBPM      15.227228
DBPM       9.088319
VORP      22.738849
dtype: float64
```

[34]:
```
#removing WS48 due to high VIF
cols = [ 'Season', 'Age', 'G', 'GS', 'MP',
        'X3PM','X3PP', 'X2PM', 'X2PP',
       'EFG', 'FTA', 'FTP', 'ORB', 'DRB', 'AST', 'STL', 'BLK',
       'TOV', 'PF', 'TS', 'X3PAR', 'FTR', 'ORBP', 'DRBP',
       'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
       'OBPM', 'DBPM', 'VORP']
VIF(train,cols)
```

[34]:
```
Season     1.176167
Age        1.145824
G          2.072139
GS         3.573711
MP        19.915127
X3PM       7.148818
X3PP       2.034893
X2PM      14.439520
X2PP       4.035960
EFG        5.261874
FTA        9.218844
FTP        1.612970
ORB       10.246025
DRB       12.552687
AST       11.574889
STL        6.189365
BLK        5.448306
TOV       14.089805
PF         3.479433
TS         6.158846
X3PAR      4.677784
FTR        3.107208
ORBP       9.017877
DRBP       5.787690
ASTP       8.069173
STLP       3.394789
BLKP       5.044777
```

```
TOVP       3.072474
USG        5.740124
OWS       14.866244
DWS        5.190462
OBPM       9.682280
DBPM       5.325734
VORP      20.186208
dtype: float64
```

[35]:
```
#removing VORP due to high VIF
cols = [ 'Season', 'Age', 'G', 'GS', 'MP',
        'X3PM','X3PP', 'X2PM', 'X2PP',
       'EFG', 'FTA', 'FTP', 'ORB', 'DRB', 'AST', 'STL', 'BLK',
       'TOV', 'PF', 'TS', 'X3PAR', 'FTR', 'ORBP', 'DRBP',
       'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
       'OBPM', 'DBPM', ]
VIF(train,cols)
```

[35]:
```
Season     1.164221
Age        1.144936
G          1.921076
GS         3.571630
MP        18.211867
X3PM       6.817424
X3PP       2.034562
X2PM      14.393876
X2PP       4.034296
EFG        5.257562
FTA        9.206455
FTP        1.601102
ORB       10.202223
DRB       12.086051
AST       11.419129
STL        5.758635
BLK        5.235173
TOV       14.082940
PF         3.261486
TS         5.873030
X3PAR      4.653031
FTR        3.106413
ORBP       9.017397
DRBP       5.772044
ASTP       8.064346
STLP       3.383403
BLKP       5.030034
TOVP       3.036467
USG        5.441081
```

```
OWS        4.595000
DWS        4.242302
OBPM       9.587956
DBPM       4.853477
dtype: float64
```

[36]:
```
#removing MP due to high VIF
cols = [ 'Season', 'Age', 'G', 'GS',
         'X3PM','X3PP', 'X2PM', 'X2PP',
        'EFG', 'FTA', 'FTP', 'ORB', 'DRB', 'AST', 'STL', 'BLK',
        'TOV', 'PF', 'TS', 'X3PAR', 'FTR', 'ORBP', 'DRBP',
        'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
        'OBPM', 'DBPM', ]
VIF(train,cols)
```

[36]:
```
Season     1.136723
Age        1.136248
G          1.889153
GS         3.298586
X3PM       4.999364
X3PP       2.025477
X2PM      11.813506
X2PP       4.034273
EFG        5.215808
FTA        9.197626
FTP        1.577641
ORB        9.793625
DRB       11.681708
AST       10.919832
STL        5.349282
BLK        5.235117
TOV       14.067784
PF         3.034813
TS         5.863261
X3PAR      4.485372
FTR        3.077002
ORBP       8.462237
DRBP       5.655832
ASTP       7.912304
STLP       3.358746
BLKP       5.026976
TOVP       3.030833
USG        4.892082
OWS        4.472444
DWS        4.173373
OBPM       9.323727
DBPM       4.748321
```

```
dtype: float64
```

[37]:
```
#removing TOV due to high VIF
cols = [ 'Season', 'Age', 'G', 'GS',
        'X3PM','X3PP', 'X2PM', 'X2PP',
       'EFG', 'FTA', 'FTP', 'ORB', 'DRB', 'AST', 'STL', 'BLK',
        'PF', 'TS', 'X3PAR', 'FTR', 'ORBP', 'DRBP',
       'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
       'OBPM', 'DBPM', ]
VIF(train,cols)
```

[37]:
```
Season     1.135775
Age        1.130004
G          1.888278
GS         3.295046
X3PM       4.912694
X3PP       2.023405
X2PM      11.452405
X2PP       4.034139
EFG        5.207698
FTA        8.060751
FTP        1.574891
ORB        9.732625
DRB       11.674759
AST        7.014743
STL        5.347932
BLK        5.234317
PF         2.891596
TS         5.849500
X3PAR      4.485192
FTR        3.058400
ORBP       8.401629
DRBP       5.654377
ASTP       7.752107
STLP       3.358745
BLKP       5.017649
TOVP       2.392870
USG        4.619550
OWS        3.950771
DWS        4.103320
OBPM       9.231896
DBPM       4.719999
dtype: float64
```

[38]:
```
#removing DRB due to high VIF
cols = [ 'Season', 'Age', 'G', 'GS',
        'X3PM','X3PP', 'X2PM', 'X2PP',
```

```
       'EFG', 'FTA', 'FTP', 'ORB', 'AST', 'STL', 'BLK',
        'PF', 'TS', 'X3PAR', 'FTR', 'ORBP', 'DRBP',
       'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
       'OBPM', 'DBPM', ]
VIF(train,cols)
```

[38]:
```
Season     1.131626
Age        1.105586
G          1.886255
GS         3.252811
X3PM       4.682543
X3PP       2.013417
X2PM      10.554792
X2PP       4.034062
EFG        5.206740
FTA        8.006049
FTP        1.573864
ORB        7.039333
AST        7.014608
STL        5.346409
BLK        4.961322
PF         2.889069
TS         5.846426
X3PAR      4.469559
FTR        3.042172
ORBP       7.391847
DRBP       3.028741
ASTP       7.719139
STLP       3.356726
BLKP       4.915779
TOVP       2.392109
USG        4.456246
OWS        3.893081
DWS        3.951947
OBPM       9.223139
DBPM       4.705253
dtype: float64
```

[39]:
```
#removing X2PM due to high VIF
cols = [ 'Season', 'Age', 'G', 'GS',
        'X3PM','X3PP', 'X2PP',
       'EFG', 'FTA', 'FTP', 'ORB', 'AST', 'STL', 'BLK',
        'PF', 'TS', 'X3PAR', 'FTR', 'ORBP', 'DRBP',
       'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
       'OBPM', 'DBPM', ]
VIF(train,cols)
```

[39]:
```
Season    1.130710
Age       1.104392
G         1.885560
GS        3.044699
X3PM      4.461442
X3PP      1.996280
X2PP      3.946231
EFG       5.202824
FTA       5.787522
FTP       1.559904
ORB       6.708103
AST       6.563561
STL       5.322994
BLK       4.943520
PF        2.874465
TS        5.811265
X3PAR     3.766829
FTR       2.504369
ORBP      7.205162
DRBP      3.017469
ASTP      7.228077
STLP      3.338498
BLKP      4.904424
TOVP      2.389749
USG       3.962929
OWS       3.889772
DWS       3.951828
OBPM      8.998154
DBPM      4.705084
dtype: float64
```

[40]:
```
#removing OBPM due to high VIF
cols = [ 'Season', 'Age', 'G', 'GS',
         'X3PM','X3PP', 'X2PP',
        'EFG', 'FTA', 'FTP', 'ORB', 'AST', 'STL', 'BLK',
         'PF', 'TS', 'X3PAR', 'FTR', 'ORBP', 'DRBP',
        'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
         'DBPM', ]
VIF(train,cols)
```

[40]:
```
Season    1.100899
Age       1.101521
G         1.882005
GS        3.044696
X3PM      4.246389
X3PP      1.993230
X2PP      3.945502
```

```
EFG       5.063072
FTA       5.782452
FTP       1.543005
ORB       6.517056
AST       6.546866
STL       5.194886
BLK       4.859312
PF        2.545097
TS        3.049602
X3PAR     3.680976
FTR       2.487958
ORBP      6.846407
DRBP      2.964372
ASTP      5.738207
STLP      3.276234
BLKP      4.802364
TOVP      1.835915
USG       3.648232
OWS       3.490515
DWS       3.850792
DBPM      4.390282
dtype: float64
```

[41]:
```
#removing ORBP due to high VIF
cols = [ 'Season', 'Age', 'G', 'GS',
         'X3PM','X3PP', 'X2PP',
        'EFG', 'FTA', 'FTP', 'ORB', 'AST', 'STL', 'BLK',
         'PF', 'TS', 'X3PAR', 'FTR', 'DRBP',
        'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
         'DBPM', ]
VIF(train,cols)
```

[41]:
```
Season    1.099357
Age       1.101409
G         1.879815
GS        2.915875
X3PM      4.153773
X3PP      1.992038
X2PP      3.945467
EFG       5.062506
FTA       5.580427
FTP       1.542795
ORB       4.394224
AST       6.346549
STL       5.190225
BLK       4.488437
PF        2.505498
```

```
TS        3.049306
X3PAR     3.169308
FTR       2.438070
DRBP      2.615965
ASTP      5.151984
STLP      3.221742
BLKP      4.156132
TOVP      1.730213
USG       3.633784
OWS       3.413857
DWS       3.839131
DBPM      4.159746
dtype: float64
```

[42]:
```
#removing AST due to high VIF
cols = [ 'Season', 'Age', 'G', 'GS',
        'X3PM','X3PP', 'X2PP',
       'EFG', 'FTA', 'FTP', 'ORB',  'STL', 'BLK',
        'PF', 'TS', 'X3PAR', 'FTR', 'DRBP',
       'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
        'DBPM', ]
VIF(train,cols)
```

[42]:
```
Season    1.099079
Age       1.098199
G         1.879814
GS        2.819469
X3PM      4.140265
X3PP      1.991777
X2PP      3.945325
EFG       5.060724
FTA       5.276088
FTP       1.540714
ORB       4.225238
STL       4.411035
BLK       4.429971
PF        2.500694
TS        3.046562
X3PAR     3.130705
FTR       2.422135
DRBP      2.585284
ASTP      2.704411
STLP      3.171858
BLKP      4.104606
TOVP      1.729317
USG       3.476758
OWS       3.409592
```

```
DWS       3.838223
DBPM      4.114434
dtype: float64
```

[43]:
```
#removing FTA due to high VIF
cols = [ 'Season', 'Age', 'G', 'GS',
         'X3PM','X3PP', 'X2PP',
        'EFG', 'FTP', 'ORB',  'STL', 'BLK',
         'PF', 'TS', 'X3PAR', 'FTR', 'DRBP',
        'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
         'DBPM', ]
VIF(train,cols)
```

[43]:
```
Season    1.098972
Age       1.098197
G         1.878774
GS        2.729311
X3PM      4.134291
X3PP      1.990011
X2PP      3.931779
EFG       5.027074
FTP       1.535914
ORB       4.149917
STL       3.724103
BLK       4.199374
PF        2.499454
TS        2.996593
X3PAR     3.126218
FTR       1.714497
DRBP      2.582271
ASTP      2.683918
STLP      2.821396
BLKP      3.816001
TOVP      1.722360
USG       2.531905
OWS       3.116697
DWS       3.631480
DBPM      3.894991
dtype: float64
```

[153]:
```
#removing EFG due to high VIF.
cols = [ 'Season', 'Age', 'G', 'GS',
         'X3PM','X3PP', 'X2PP',
         'FTP', 'ORB',  'STL', 'BLK',
         'PF', 'TS', 'X3PAR', 'FTR', 'DRBP',
        'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
         'DBPM', ]
```

```
VIF(train,cols)
```

[153]:
```
Season    1.094741
Age       1.093105
G         2.036491
GS        1.436180
X3PM      3.778693
X3PP      1.975416
X2PP      1.932047
FTP       1.430927
ORB       3.720472
STL       3.199244
BLK       4.613826
PF        2.526033
TS        3.138067
X3PAR     3.227589
FTR       1.696305
DRBP      2.379107
ASTP      2.538724
STLP      2.973322
BLKP      4.366174
TOVP      1.953010
USG       2.311579
OWS       2.740344
DWS       3.031416
DBPM      4.243942
dtype: float64
```

[156]:
```
#removed season b/c not relevant to predicting output
cols = [ 'Age', 'G', 'GS',
        'X3PM','X3PP', 'X2PP',
        'FTP', 'ORB',  'STL', 'BLK',
        'PF', 'TS', 'FTR', 'DRBP','X3PAR',
       'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
        'DBPM', ]
x_train = train[cols] #all of the variables
y_train = train['MVP']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    MVP   R-squared:                       0.022
```

Model:	OLS	Adj. R-squared:	0.010
Method:	Least Squares	F-statistic:	1.798
Date:	Wed, 12 May 2021	Prob (F-statistic):	0.0114
Time:	19:24:14	Log-Likelihood:	16233.
No. Observations:	1892	AIC:	-3.242e+04
Df Residuals:	1868	BIC:	-3.228e+04
Df Model:	23		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	1.36e-05	1.69e-05	0.805	0.421	-1.95e-05	4.67e-05
Age	2.025e-07	2.58e-07	0.785	0.433	-3.04e-07	7.09e-07
G	-9.361e-08	8.3e-08	-1.128	0.259	-2.56e-07	6.91e-08
GS	1.318e-07	1.38e-07	0.956	0.339	-1.39e-07	4.02e-07
X3PM	2.804e-06	3.76e-06	0.745	0.456	-4.57e-06	1.02e-05
X3PP	-4.379e-06	9.06e-06	-0.484	0.629	-2.21e-05	1.34e-05
X2PP	1.086e-06	2.13e-05	0.051	0.959	-4.07e-05	4.29e-05
FTP	2.23e-07	8.84e-06	0.025	0.980	-1.71e-05	1.76e-05
ORB	1.576e-06	3.44e-06	0.458	0.647	-5.17e-06	8.32e-06
STL	8.593e-06	6.75e-06	1.273	0.203	-4.65e-06	2.18e-05
BLK	-3.613e-06	7.16e-06	-0.504	0.614	-1.77e-05	1.04e-05
PF	-2.587e-06	2.78e-06	-0.930	0.352	-8.04e-06	2.87e-06
TS	-5.772e-05	2.72e-05	-2.123	0.034	-0.000	-4.39e-06
FTR	7.147e-06	8.64e-06	0.827	0.408	-9.79e-06	2.41e-05
DRBP	-1.36e-07	2.85e-07	-0.477	0.633	-6.95e-07	4.23e-07
X3PAR	2.589e-06	8.59e-06	0.301	0.763	-1.43e-05	1.94e-05
ASTP	-5.788e-08	2.09e-07	-0.277	0.782	-4.68e-07	3.52e-07
STLP	-3.083e-06	2.57e-06	-1.197	0.231	-8.13e-06	1.97e-06
BLKP	1.336e-07	1.41e-06	0.095	0.924	-2.63e-06	2.9e-06
TOVP	3.205e-07	2.85e-07	1.126	0.260	-2.38e-07	8.79e-07
USG	6.901e-07	3.66e-07	1.886	0.059	-2.77e-08	1.41e-06
OWS	5.188e-06	1.39e-06	3.744	0.000	2.47e-06	7.91e-06
DWS	-9.126e-07	2.39e-06	-0.382	0.703	-5.6e-06	3.78e-06
DBPM	3.445e-06	1.63e-06	2.113	0.035	2.48e-07	6.64e-06

Omnibus:	5681.464	Durbin-Watson:	2.010
Prob(Omnibus):	0.000	Jarque-Bera (JB):	257689789.971
Skew:	42.052	Prob(JB):	0.00
Kurtosis:	1809.023	Cond. No.	2.06e+03

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 2.06e+03. This might indicate that there are strong multicollinearity or other numerical problems.

[46]:
```
#removed X3PAR b/c not statistically significant
cols = [ 'Age', 'G', 'GS',
         'X3PM','X3PP', 'X2PP',
         'FTP', 'ORB',  'STL', 'BLK',
         'PF', 'TS', 'FTR', 'DRBP',
        'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
         'DBPM', ]
x_train = train[cols] #all of the variables
y_train = train['MVP']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

OLS Regression Results

Dep. Variable:	MVP	R-squared:	0.201
Model:	OLS	Adj. R-squared:	0.195
Method:	Least Squares	F-statistic:	36.67
Date:	Wed, 12 May 2021	Prob (F-statistic):	1.94e-138
Time:	19:19:01	Log-Likelihood:	4695.5
No. Observations:	3231	AIC:	-9345.
Df Residuals:	3208	BIC:	-9205.
Df Model:	22		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.0595	0.017	3.599	0.000	0.027	0.092
Age	-0.0004	0.000	-1.458	0.145	-0.001	0.000
G	-6.818e-05	7.73e-05	-0.882	0.378	-0.000	8.34e-05
GS	-0.0001	5.63e-05	-2.318	0.021	-0.000	-2.01e-05
X3PM	0.0115	0.002	4.836	0.000	0.007	0.016
X3PP	-0.0169	0.009	-1.972	0.049	-0.034	-9.88e-05
X2PP	0.0716	0.022	3.257	0.001	0.029	0.115
FTP	-0.0147	0.009	-1.552	0.121	-0.033	0.004
ORB	-0.0031	0.002	-1.278	0.201	-0.008	0.002
STL	0.0196	0.005	4.300	0.000	0.011	0.028
BLK	0.0135	0.004	3.081	0.002	0.005	0.022
PF	-0.0145	0.002	-6.249	0.000	-0.019	-0.010
TS	-0.1977	0.024	-8.218	0.000	-0.245	-0.151
FTR	0.0123	0.008	1.462	0.144	-0.004	0.029
DRBP	0.0002	0.000	0.885	0.376	-0.000	0.001
ASTP	-0.0005	0.000	-2.684	0.007	-0.001	-0.000
STLP	-0.0120	0.002	-4.966	0.000	-0.017	-0.007

```
BLKP          -0.0043      0.001     -3.757      0.000      -0.007      -0.002
TOVP           0.0016      0.000      5.934      0.000       0.001       0.002
USG            0.0023      0.000      7.459      0.000       0.002       0.003
OWS            0.0130      0.001     15.363      0.000       0.011       0.015
DWS           -0.0044      0.002     -2.620      0.009      -0.008      -0.001
DBPM           0.0112      0.001      7.580      0.000       0.008       0.014
==============================================================================
Omnibus:                     4869.946   Durbin-Watson:                   1.244
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          1899462.383
Skew:                           9.285   Prob(JB):                         0.00
Kurtosis:                     120.322   Cond. No.                     2.32e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 2.32e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[47]:
```
#removed 'G' games b/c not statistically significant
cols = [ 'Age', 'GS',
        'X3PM','X3PP', 'X2PP',
        'FTP', 'ORB',  'STL', 'BLK',
        'PF', 'TS', 'FTR', 'DRBP',
       'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
        'DBPM', ]
x_train = train[cols] #all of the variables
y_train = train['MVP']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    MVP   R-squared:                       0.201
Model:                            OLS   Adj. R-squared:                  0.196
Method:                 Least Squares   F-statistic:                     38.38
Date:                Wed, 12 May 2021   Prob (F-statistic):          4.51e-139
Time:                        19:19:02   Log-Likelihood:                 4695.1
No. Observations:                3231   AIC:                            -9346.
Df Residuals:                    3209   BIC:                            -9212.
Df Model:                          21
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
```

```
------------------------------------------------------------------------------
const          0.0574      0.016      3.509      0.000       0.025       0.090
Age           -0.0003      0.000     -1.398      0.162      -0.001       0.000
GS            -0.0001    5.5e-05     -2.555      0.011      -0.000   -3.27e-05
X3PM           0.0114      0.002      4.792      0.000       0.007       0.016
X3PP          -0.0173      0.009     -2.022      0.043      -0.034      -0.001
X2PP           0.0693      0.022      3.176      0.002       0.027       0.112
FTP           -0.0156      0.009     -1.656      0.098      -0.034       0.003
ORB           -0.0032      0.002     -1.294      0.196      -0.008       0.002
STL            0.0198      0.005      4.365      0.000       0.011       0.029
BLK            0.0139      0.004      3.165      0.002       0.005       0.022
PF            -0.0148      0.002     -6.481      0.000      -0.019      -0.010
TS            -0.1981      0.024     -8.239      0.000      -0.245      -0.151
FTR            0.0126      0.008      1.501      0.133      -0.004       0.029
DRBP           0.0002      0.000      0.969      0.333      -0.000       0.001
ASTP          -0.0005      0.000     -2.722      0.007      -0.001      -0.000
STLP          -0.0121      0.002     -4.993      0.000      -0.017      -0.007
BLKP          -0.0044      0.001     -3.809      0.000      -0.007      -0.002
TOVP           0.0016      0.000      5.999      0.000       0.001       0.002
USG            0.0023      0.000      7.682      0.000       0.002       0.003
OWS            0.0130      0.001     15.389      0.000       0.011       0.015
DWS           -0.0049      0.002     -3.134      0.002      -0.008      -0.002
DBPM           0.0114      0.001      7.763      0.000       0.008       0.014
==============================================================================
Omnibus:                     4869.595   Durbin-Watson:                   1.244
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          1898225.383
Skew:                           9.284   Prob(JB):                         0.00
Kurtosis:                     120.283   Cond. No.                     1.54e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.54e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[48]:
```
#removed 'DRBP' games b/c not statistically significant
cols = [ 'Age', 'GS',
         'X3PM','X3PP', 'X2PP',
         'FTP', 'ORB',  'STL', 'BLK',
         'PF', 'TS', 'FTR',
        'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
         'DBPM', ]
x_train = train[cols] #all of the variables
y_train = train['MVP']
#adding an intercept
x_train = sm.add_constant(x_train)
```

```
#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    MVP   R-squared:                       0.201
Model:                            OLS   Adj. R-squared:                  0.196
Method:                 Least Squares   F-statistic:                     40.25
Date:                Wed, 12 May 2021   Prob (F-statistic):          1.11e-139
Time:                        19:19:02   Log-Likelihood:                 4694.6
No. Observations:                3231   AIC:                            -9347.
Df Residuals:                    3210   BIC:                            -9220.
Df Model:                          20
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0600      0.016      3.711      0.000       0.028       0.092
Age           -0.0003      0.000     -1.347      0.178      -0.001       0.000
GS            -0.0001   5.49e-05     -2.622      0.009      -0.000   -3.63e-05
X3PM           0.0114      0.002      4.777      0.000       0.007       0.016
X3PP          -0.0172      0.009     -2.012      0.044      -0.034      -0.000
X2PP           0.0697      0.022      3.195      0.001       0.027       0.113
FTP           -0.0161      0.009     -1.707      0.088      -0.035       0.002
ORB           -0.0023      0.002     -1.018      0.309      -0.007       0.002
STL            0.0198      0.005      4.366      0.000       0.011       0.029
BLK            0.0130      0.004      3.032      0.002       0.005       0.021
PF            -0.0148      0.002     -6.482      0.000      -0.019      -0.010
TS            -0.2000      0.024     -8.343      0.000      -0.247      -0.153
FTR            0.0129      0.008      1.539      0.124      -0.004       0.029
ASTP          -0.0005      0.000     -3.114      0.002      -0.001      -0.000
STLP          -0.0125      0.002     -5.287      0.000      -0.017      -0.008
BLKP          -0.0041      0.001     -3.686      0.000      -0.006      -0.002
TOVP           0.0016      0.000      6.362      0.000       0.001       0.002
USG            0.0024      0.000      8.186      0.000       0.002       0.003
OWS            0.0130      0.001     15.402      0.000       0.011       0.015
DWS           -0.0047      0.002     -3.033      0.002      -0.008      -0.002
DBPM           0.0117      0.001      8.220      0.000       0.009       0.014
==============================================================================
Omnibus:                     4871.045   Durbin-Watson:                   1.244
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          1898740.151
Skew:                           9.290   Prob(JB):                         0.00
Kurtosis:                     120.298   Cond. No.                     1.48e+03
==============================================================================
```

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.48e+03. This might indicate that there are strong multicollinearity or other numerical problems.

[49]:
```
#removed 'ORB' games b/c not statistically significant
cols = [ 'Age', 'GS',
        'X3PM','X3PP', 'X2PP',
        'FTP',  'STL', 'BLK',
        'PF', 'TS', 'FTR',
       'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
        'DBPM', ]
x_train = train[cols] #all of the variables
y_train = train['MVP']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

OLS Regression Results

Dep. Variable:	MVP	R-squared:	0.200
Model:	OLS	Adj. R-squared:	0.196
Method:	Least Squares	F-statistic:	42.32
Date:	Wed, 12 May 2021	Prob (F-statistic):	2.79e-140
Time:	19:19:02	Log-Likelihood:	4694.1
No. Observations:	3231	AIC:	-9348.
Df Residuals:	3211	BIC:	-9227.
Df Model:	19		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.0592	0.016	3.668	0.000	0.028	0.091
Age	-0.0003	0.000	-1.336	0.182	-0.001	0.000
GS	-0.0002	5.44e-05	-2.787	0.005	-0.000	-4.5e-05
X3PM	0.0122	0.002	5.535	0.000	0.008	0.017
X3PP	-0.0159	0.008	-1.879	0.060	-0.032	0.001
X2PP	0.0682	0.022	3.132	0.002	0.026	0.111
FTP	-0.0153	0.009	-1.626	0.104	-0.034	0.003
STL	0.0197	0.005	4.337	0.000	0.011	0.029
BLK	0.0115	0.004	2.856	0.004	0.004	0.019
PF	-0.0156	0.002	-7.227	0.000	-0.020	-0.011
TS	-0.1999	0.024	-8.338	0.000	-0.247	-0.153
FTR	0.0129	0.008	1.543	0.123	-0.003	0.029

```
ASTP          -0.0005      0.000     -2.973      0.003      -0.001      -0.000
STLP          -0.0126      0.002     -5.311      0.000      -0.017      -0.008
BLKP          -0.0040      0.001     -3.612      0.000      -0.006      -0.002
TOVP           0.0016      0.000      6.362      0.000       0.001       0.002
USG            0.0024      0.000      8.164      0.000       0.002       0.003
OWS            0.0128      0.001     15.502      0.000       0.011       0.014
DWS           -0.0048      0.002     -3.126      0.002      -0.008      -0.002
DBPM           0.0119      0.001      8.436      0.000       0.009       0.015
==============================================================================
Omnibus:                     4869.255   Durbin-Watson:                   1.241
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          1892741.851
Skew:                           9.284   Prob(JB):                         0.00
Kurtosis:                     120.109   Cond. No.                     1.48e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.48e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[50]:
```
#removed season b/c not relevant to predicting output
cols = [  'GS',
        'X3PM','X3PP', 'X2PP',
        'FTP',  'STL', 'BLK',
        'PF', 'TS', 'FTR',
       'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
        'DBPM', ]
x_train = train[cols] #all of the variables
y_train = train['MVP']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    MVP   R-squared:                       0.200
Model:                            OLS   Adj. R-squared:                  0.195
Method:                 Least Squares   F-statistic:                     44.56
Date:                Wed, 12 May 2021   Prob (F-statistic):          9.84e-141
Time:                        19:19:02   Log-Likelihood:                 4693.2
No. Observations:                3231   AIC:                            -9348.
Df Residuals:                    3212   BIC:                            -9233.
Df Model:                          18
Covariance Type:            nonrobust
```

```
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0491      0.014      3.443      0.001       0.021       0.077
GS            -0.0002   5.44e-05     -2.806      0.005      -0.000    -4.6e-05
X3PM           0.0120      0.002      5.455      0.000       0.008       0.016
X3PP          -0.0156      0.008     -1.846      0.065      -0.032       0.001
X2PP           0.0683      0.022      3.134      0.002       0.026       0.111
FTP           -0.0157      0.009     -1.676      0.094      -0.034       0.003
STL            0.0201      0.005      4.443      0.000       0.011       0.029
BLK            0.0117      0.004      2.895      0.004       0.004       0.020
PF            -0.0156      0.002     -7.242      0.000      -0.020      -0.011
TS            -0.1986      0.024     -8.290      0.000      -0.246      -0.152
FTR            0.0134      0.008      1.597      0.110      -0.003       0.030
ASTP          -0.0005      0.000     -3.051      0.002      -0.001      -0.000
STLP          -0.0124      0.002     -5.236      0.000      -0.017      -0.008
BLKP          -0.0040      0.001     -3.557      0.000      -0.006      -0.002
TOVP           0.0016      0.000      6.345      0.000       0.001       0.002
USG            0.0024      0.000      8.306      0.000       0.002       0.003
OWS            0.0128      0.001     15.463      0.000       0.011       0.014
DWS           -0.0049      0.002     -3.174      0.002      -0.008      -0.002
DBPM           0.0117      0.001      8.338      0.000       0.009       0.014
==============================================================================
Omnibus:                     4873.501   Durbin-Watson:                   1.240
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          1901763.462
Skew:                           9.299   Prob(JB):                         0.00
Kurtosis:                     120.390   Cond. No.                     1.33e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.33e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[51]:
```
import statsmodels.formula.api as smf

MVPLinear = smf.ols(formula='MVP ~ GS + X3PM + X3PP + X2PP + STL+ BLK + PF + TS␣
 ↪+ ASTP + STLP + BLKP + TOVP + USG + OWS + DWS + DBPM',
                    data=train).fit()
print(MVPLinear.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    MVP   R-squared:                       0.198
Model:                            OLS   Adj. R-squared:                  0.194
Method:                 Least Squares   F-statistic:                     49.71
Date:                Wed, 12 May 2021   Prob (F-statistic):          3.33e-141
```

```
Time:                        19:19:02   Log-Likelihood:                 4690.3
No. Observations:                3231   AIC:                            -9347.
Df Residuals:                    3214   BIC:                            -9243.
Df Model:                          16
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      0.0394      0.013      2.946      0.003       0.013       0.066
GS            -0.0002   5.44e-05     -2.904      0.004      -0.000   -5.13e-05
X3PM           0.0109      0.002      5.057      0.000       0.007       0.015
X3PP          -0.0192      0.008     -2.311      0.021      -0.035      -0.003
X2PP           0.0706      0.022      3.284      0.001       0.028       0.113
STL            0.0206      0.005      4.551      0.000       0.012       0.029
BLK            0.0113      0.004      2.810      0.005       0.003       0.019
PF            -0.0156      0.002     -7.235      0.000      -0.020      -0.011
TS            -0.1992      0.023     -8.678      0.000      -0.244      -0.154
ASTP          -0.0006      0.000     -3.502      0.000      -0.001      -0.000
STLP          -0.0118      0.002     -5.014      0.000      -0.016      -0.007
BLKP          -0.0035      0.001     -3.191      0.001      -0.006      -0.001
TOVP           0.0018      0.000      7.153      0.000       0.001       0.002
USG            0.0024      0.000      8.322      0.000       0.002       0.003
OWS            0.0131      0.001     16.372      0.000       0.012       0.015
DWS           -0.0048      0.002     -3.131      0.002      -0.008      -0.002
DBPM           0.0114      0.001      8.184      0.000       0.009       0.014
==============================================================================
Omnibus:                     4877.428   Durbin-Watson:                   1.241
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          1907434.996
Skew:                           9.314   Prob(JB):                         0.00
Kurtosis:                     120.565   Cond. No.                     1.28e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.28e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[52]:
```python
# compute out-of-sample R-squared using the test set
def OSR2(model, df_train, df_test, dependent_var):
    y_test = df_test[dependent_var]
    y_pred = model.predict(df_test)
    SSE = np.sum((y_test - y_pred)**2)
    SST = np.sum((y_test - np.mean(df_train[dependent_var]))**2)
    return 1 - SSE/SST
```

[53]:
```python
OSR2(MVPLinear, train, test, 'MVP')
```

[53]:
```
0.22559973569316927
```

[54]:
```
dataset = pd.read_csv('mip_stats.csv')
train = dataset[dataset['Season'] <= 2017]
test = dataset[dataset['Season'] > 2017]
test_2010 = test = dataset[dataset['Season'] > 2010]
```

[55]:
```
dataset
```

[55]:

	Player	Season	G	GS	MP	FGM	FGA	FGP	X3PM	\
0	A.J. Price	2011.0	-6.0	-2.0	0.5	-0.3	0.1	-0.054	-0.3	
1	A.J. Price	2012.0	-6.0	1.0	-3.0	-1.0	-2.4	-0.017	-0.2	
2	A.J. Price	2013.0	13.0	21.0	9.5	1.5	3.2	0.051	0.6	
3	A.J. Price	2014.0	-29.0	-22.0	-18.9	-2.1	-5.6	0.023	-1.0	
4	A.J. Price	2015.0	-2.0	0.0	9.0	1.3	3.7	-0.041	0.4	
…	…	…	…	…	…	…	…	…	…	
3405	Zaza Pachulia	2016.0	3.0	24.0	2.7	-0.4	-1.0	0.012	0.0	
3406	Zaza Pachulia	2017.0	-6.0	1.0	-8.3	-0.6	-1.8	0.068	0.0	
3407	Zaza Pachulia	2018.0	-1.0	-13.0	-4.0	-0.1	-0.6	0.030	0.0	
3408	Zaza Pachulia	2019.0	-1.0	-54.0	-1.2	-0.9	-1.0	-0.124	0.0	
3409	Zydrunas Ilgauskas	2011.0	8.0	45.0	-5.0	-0.7	-2.4	0.065	-0.2	

	X3PA	…	USG	OWS	DWS	WS	WS48	OBPM	DBPM	BPM	VORP	MIP
0	-0.1	…	0.0	-0.8	-0.1	-0.9	-0.045	-1.8	-0.7	-2.4	-0.5	0.0
1	-1.0	…	-5.0	0.6	-0.2	0.4	0.043	1.3	0.6	1.9	0.3	0.0
2	1.5	…	0.3	0.8	0.7	1.5	0.021	0.4	0.1	0.4	0.4	0.0
3	-2.7	…	5.2	-1.1	-1.2	-2.2	-0.096	-2.2	-2.0	-4.2	-0.6	0.0
4	1.4	…	-5.0	0.9	0.5	1.4	0.057	1.5	1.3	2.8	0.1	0.0
…	…	…	…	…	…	…	…	…	…	…	…	…
3405	0.0	…	-10.1	-1.3	0.5	-1.0	-0.014	-2.1	0.8	-1.4	-0.7	0.0
3406	0.0	…	-0.9	-1.1	-0.2	-1.3	0.033	-0.8	1.5	0.8	-0.1	0.0
3407	0.0	…	1.7	-0.5	-1.1	-1.4	-0.017	0.4	-0.9	-0.6	-0.3	0.0
3408	0.1	…	-2.0	-0.7	-0.1	-1.0	-0.033	-2.1	0.9	-1.2	-0.3	0.0
3409	-0.4	…	-3.0	0.5	-0.1	0.4	0.034	0.0	1.1	1.1	0.4	0.0

[3410 rows x 48 columns]

[157]:
```
#use this to test award winner for each year and award
datasetmip = pd.read_csv('smoydata142.csv')
trainmip = datasetmip[datasetmip['Season'] <= 2017]
testmip = datasetmip[datasetmip['Season'] > 2017]
data2009 = datasetmip[datasetmip['Season'] == 2020.0]
testt = dataset[dataset['Season'] == 2020]
graphh = testt[testt['MIP']>0]
funn = graphh[['Player','MIP']].sort_values(by='MIP',ascending = False)
funn
```

[157]:

	Player	MIP
450	Brandon Ingram	0.652
334	Bam Adebayo	0.590
1971	Jayson Tatum	0.114
1123	Devonte' Graham	0.100
3813	Shai Gilgeous-Alexander	0.042
3344	Pascal Siakam	0.026
705	Christian Wood	0.022
4155	Trae Young	0.020
1420	Fred VanVleet	0.012
1968	Jaylen Brown	0.006
2915	Markelle Fultz	0.004
1221	Duncan Robinson	0.002
3889	Spencer Dinwiddie	0.002

Sixth Man of the Year (SMOY)

[158]:
```
datasett = pd.read_csv('smoydata142.csv')
train = datasett[datasett['Season'] <= 2017]
test = datasett[datasett['Season'] > 2017]

y_train = train['MVP']
y_test = test['MVP']
```

[159]:
```
#removed season b/c not relevant to predicting the output variable
cols = [ 'Age', 'G', 'GS',
        'X3PM','X3PP', 'X2PP',
        'FTP', 'ORB',  'STL', 'BLK',
        'PF', 'TS', 'FTR', 'DRBP','X3PAR',
       'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
        'DBPM', ]

dataset = pd.read_csv('update1422.csv')
x_train = train[cols] #all of the variables
y_train = train['SMOY']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

OLS Regression Results

Dep. Variable:	SMOY	R-squared:	0.149
Model:	OLS	Adj. R-squared:	0.139
Method:	Least Squares	F-statistic:	14.24
Date:	Wed, 12 May 2021	Prob (F-statistic):	8.97e-51

```
Time:                         19:28:00   Log-Likelihood:                 2684.8
No. Observations:                1892   AIC:                            -5322.
Df Residuals:                    1868   BIC:                            -5188.
Df Model:                          23
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0446      0.022      2.050      0.041       0.002       0.087
Age            0.0005      0.000      1.521      0.128      -0.000       0.001
G          -5.714e-05      0.000     -0.535      0.593      -0.000       0.000
GS             -0.0009      0.000     -5.050      0.000      -0.001      -0.001
X3PM           0.0332      0.005      6.857      0.000       0.024       0.043
X3PP           -0.0152      0.012     -1.302      0.193      -0.038       0.008
X2PP           0.0644      0.027      2.349      0.019       0.011       0.118
FTP            -0.0052      0.011     -0.458      0.647      -0.028       0.017
ORB            0.0023      0.004      0.521      0.603      -0.006       0.011
STL            0.0271      0.009      3.115      0.002       0.010       0.044
BLK            0.0182      0.009      1.976      0.048       0.000       0.036
PF             -0.0115      0.004     -3.205      0.001      -0.019      -0.004
TS             -0.1642      0.035     -4.691      0.000      -0.233      -0.096
FTR            0.0030      0.011      0.274      0.784      -0.019       0.025
DRBP           -0.0008      0.000     -2.137      0.033      -0.002   -6.43e-05
X3PAR          -0.0407      0.011     -3.676      0.000      -0.062      -0.019
ASTP           -0.0008      0.000     -3.042      0.002      -0.001      -0.000
STLP           -0.0115      0.003     -3.457      0.001      -0.018      -0.005
BLKP           -0.0045      0.002     -2.497      0.013      -0.008      -0.001
TOVP           0.0014      0.000      3.844      0.000       0.001       0.002
USG            0.0013      0.000      2.826      0.005       0.000       0.002
OWS            0.0136      0.002      7.608      0.000       0.010       0.017
DWS            0.0046      0.003      1.479      0.139      -0.001       0.011
DBPM           0.0050      0.002      2.368      0.018       0.001       0.009
==============================================================================
Omnibus:                     2876.639   Durbin-Watson:                   1.769
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           960973.714
Skew:                           9.353   Prob(JB):                         0.00
Kurtosis:                     111.812   Cond. No.                     2.06e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 2.06e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[61]:
```
#removed FTR b/c not statistically significant
```

```
cols = [ 'Age', 'G', 'GS',
        'X3PM','X3PP', 'X2PP',
        'FTP', 'ORB',  'STL', 'BLK',
        'PF', 'TS', 'DRBP','X3PAR',
       'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
        'DBPM', ]
x_train = train[cols] #all of the variables
y_train = train['SMOY']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

OLS Regression Results

Dep. Variable:	SMOY	R-squared:	0.149
Model:	OLS	Adj. R-squared:	0.139
Method:	Least Squares	F-statistic:	14.89
Date:	Wed, 12 May 2021	Prob (F-statistic):	2.37e-51
Time:	19:19:03	Log-Likelihood:	2684.7
No. Observations:	1892	AIC:	-5323.
Df Residuals:	1869	BIC:	-5196.
Df Model:	22		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.0450	0.022	2.071	0.038	0.002	0.088
Age	0.0005	0.000	1.511	0.131	-0.000	0.001
G	-5.78e-05	0.000	-0.541	0.588	-0.000	0.000
GS	-0.0009	0.000	-5.070	0.000	-0.001	-0.001
X3PM	0.0332	0.005	6.855	0.000	0.024	0.043
X3PP	-0.0154	0.012	-1.325	0.185	-0.038	0.007
X2PP	0.0638	0.027	2.335	0.020	0.010	0.117
FTP	-0.0054	0.011	-0.476	0.634	-0.028	0.017
ORB	0.0023	0.004	0.513	0.608	-0.006	0.011
STL	0.0271	0.009	3.116	0.002	0.010	0.044
BLK	0.0181	0.009	1.967	0.049	5.11e-05	0.036
PF	-0.0115	0.004	-3.199	0.001	-0.018	-0.004
TS	-0.1625	0.034	-4.723	0.000	-0.230	-0.095
DRBP	-0.0008	0.000	-2.141	0.032	-0.002	-6.59e-05
X3PAR	-0.0415	0.011	-3.908	0.000	-0.062	-0.021
ASTP	-0.0008	0.000	-3.125	0.002	-0.001	-0.000
STLP	-0.0114	0.003	-3.449	0.001	-0.018	-0.005
BLKP	-0.0045	0.002	-2.484	0.013	-0.008	-0.001

```
TOVP           0.0014      0.000      3.985      0.000       0.001       0.002
USG            0.0013      0.000      2.824      0.005       0.000       0.002
OWS            0.0137      0.002      7.755      0.000       0.010       0.017
DWS            0.0046      0.003      1.478      0.140      -0.001       0.011
DBPM           0.0049      0.002      2.353      0.019       0.001       0.009
==============================================================================
Omnibus:                     2876.791   Durbin-Watson:                   1.769
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           961369.871
Skew:                           9.354   Prob(JB):                         0.00
Kurtosis:                     111.835   Cond. No.                     2.03e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 2.03e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[62]:
```
#removed X3PP b/c not statistically significant
cols = [ 'Age', 'G', 'GS',
        'X3PM', 'X2PP',
        'FTP', 'ORB',  'STL', 'BLK',
        'PF', 'TS', 'DRBP','X3PAR',
       'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
        'DBPM', ]
x_train = train[cols] #all of the variables
y_train = train['SMOY']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   SMOY   R-squared:                       0.148
Model:                            OLS   Adj. R-squared:                  0.139
Method:                 Least Squares   F-statistic:                     15.51
Date:                Wed, 12 May 2021   Prob (F-statistic):           1.35e-51
Time:                        19:19:03   Log-Likelihood:                 2683.9
No. Observations:                1892   AIC:                            -5324.
Df Residuals:                    1870   BIC:                            -5202.
Df Model:                          21
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
```

```
const          0.0439      0.022      2.022      0.043       0.001       0.086
Age            0.0005      0.000      1.563      0.118      -0.000       0.001
G          -6.454e-05      0.000     -0.605      0.545      -0.000       0.000
GS             -0.0009      0.000     -5.103      0.000      -0.001      -0.001
X3PM           0.0319      0.005      6.725      0.000       0.023       0.041
X2PP           0.0678      0.027      2.493      0.013       0.014       0.121
FTP            -0.0067      0.011     -0.593      0.553      -0.029       0.015
ORB            0.0030      0.004      0.689      0.491      -0.006       0.012
STL            0.0267      0.009      3.073      0.002       0.010       0.044
BLK            0.0182      0.009      1.981      0.048       0.000       0.036
PF             -0.0111      0.004     -3.110      0.002      -0.018      -0.004
TS             -0.1684      0.034     -4.937      0.000      -0.235      -0.101
DRBP           -0.0008      0.000     -2.177      0.030      -0.002   -7.92e-05
X3PAR          -0.0435      0.011     -4.135      0.000      -0.064      -0.023
ASTP           -0.0008      0.000     -3.196      0.001      -0.001      -0.000
STLP           -0.0116      0.003     -3.542      0.000      -0.018      -0.005
BLKP           -0.0044      0.002     -2.458      0.014      -0.008      -0.001
TOVP           0.0015      0.000      4.077      0.000       0.001       0.002
USG            0.0013      0.000      2.831      0.005       0.000       0.002
OWS            0.0137      0.002      7.789      0.000       0.010       0.017
DWS            0.0044      0.003      1.424      0.155      -0.002       0.010
DBPM           0.0051      0.002      2.453      0.014       0.001       0.009
==============================================================================
Omnibus:                     2878.350   Durbin-Watson:                   1.767
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           963345.056
Skew:                           9.364   Prob(JB):                         0.00
Kurtosis:                     111.946   Cond. No.                     2.01e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 2.01e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[63]:
```
#removed FTP b/c not statistically significant
cols = [ 'Age', 'G', 'GS',
        'X3PM', 'X2PP',
         'ORB',  'STL', 'BLK',
        'PF', 'TS', 'DRBP','X3PAR',
       'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
        'DBPM', ]
x_train = train[cols] #all of the variables
y_train = train['SMOY']
#adding an intercept
x_train = sm.add_constant(x_train)
```

```
#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   SMOY   R-squared:                       0.148
Model:                            OLS   Adj. R-squared:                  0.139
Method:                 Least Squares   F-statistic:                     16.28
Date:                Wed, 12 May 2021   Prob (F-statistic):           3.89e-52
Time:                        19:19:03   Log-Likelihood:                 2683.7
No. Observations:                1892   AIC:                            -5325.
Df Residuals:                    1871   BIC:                            -5209.
Df Model:                          20
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0411      0.021      1.940      0.053      -0.000       0.083
Age            0.0005      0.000      1.546      0.122      -0.000       0.001
G          -6.933e-05      0.000     -0.652      0.515      -0.000       0.000
GS             -0.0009      0.000     -5.109      0.000      -0.001      -0.001
X3PM           0.0317      0.005      6.701      0.000       0.022       0.041
X2PP           0.0698      0.027      2.592      0.010       0.017       0.123
ORB            0.0032      0.004      0.739      0.460      -0.005       0.012
STL            0.0266      0.009      3.065      0.002       0.010       0.044
BLK            0.0184      0.009      1.994      0.046       0.000       0.036
PF             -0.0112      0.004     -3.141      0.002      -0.018      -0.004
TS             -0.1728      0.033     -5.193      0.000      -0.238      -0.108
DRBP           -0.0008      0.000     -2.157      0.031      -0.002   -7.17e-05
X3PAR          -0.0435      0.011     -4.138      0.000      -0.064      -0.023
ASTP           -0.0009      0.000     -3.240      0.001      -0.001      -0.000
STLP           -0.0116      0.003     -3.521      0.000      -0.018      -0.005
BLKP           -0.0044      0.002     -2.438      0.015      -0.008      -0.001
TOVP           0.0015      0.000      4.140      0.000       0.001       0.002
USG            0.0013      0.000      2.778      0.006       0.000       0.002
OWS            0.0137      0.002      7.810      0.000       0.010       0.017
DWS            0.0043      0.003      1.414      0.158      -0.002       0.010
DBPM           0.0051      0.002      2.458      0.014       0.001       0.009
==============================================================================
Omnibus:                     2878.889   Durbin-Watson:                   1.767
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           964330.897
Skew:                           9.367   Prob(JB):                         0.00
Kurtosis:                     112.002   Cond. No.                     1.98e+03
==============================================================================

Warnings:
```

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.98e+03. This might indicate that there are strong multicollinearity or other numerical problems.

[64]:
```
#removed DBPM b/c not statistically significant
cols = [ 'Age', 'G', 'GS',
        'X3PM', 'X2PP',
         'ORB',  'STL', 'BLK',
        'PF', 'TS', 'DRBP','X3PAR',
       'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',   ]
x_train = train[cols] #all of the variables
y_train = train['SMOY']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

OLS Regression Results

Dep. Variable:	SMOY	R-squared:	0.145
Model:	OLS	Adj. R-squared:	0.137
Method:	Least Squares	F-statistic:	16.77
Date:	Wed, 12 May 2021	Prob (F-statistic):	1.61e-51
Time:	19:19:03	Log-Likelihood:	2680.6
No. Observations:	1892	AIC:	-5321.
Df Residuals:	1872	BIC:	-5210.
Df Model:	19		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.0233	0.020	1.168	0.243	-0.016	0.062
Age	0.0006	0.000	1.666	0.096	-9.81e-05	0.001
G	-9.612e-05	0.000	-0.907	0.364	-0.000	0.000
GS	-0.0009	0.000	-5.337	0.000	-0.001	-0.001
X3PM	0.0305	0.005	6.474	0.000	0.021	0.040
X2PP	0.0622	0.027	2.321	0.020	0.010	0.115
ORB	0.0009	0.004	0.208	0.836	-0.007	0.009
STL	0.0246	0.009	2.846	0.004	0.008	0.042
BLK	0.0209	0.009	2.282	0.023	0.003	0.039
PF	-0.0107	0.004	-2.991	0.003	-0.018	-0.004
TS	-0.1350	0.030	-4.569	0.000	-0.193	-0.077
DRBP	-0.0006	0.000	-1.602	0.109	-0.001	0.000
X3PAR	-0.0423	0.011	-4.019	0.000	-0.063	-0.022
ASTP	-0.0007	0.000	-2.738	0.006	-0.001	-0.000

```
STLP          -0.0067      0.003     -2.552      0.011      -0.012      -0.002
BLKP          -0.0036      0.002     -2.031      0.042      -0.007      -0.000
TOVP           0.0013      0.000      3.795      0.000       0.001       0.002
USG            0.0007      0.000      1.696      0.090      -0.000       0.001
OWS            0.0129      0.002      7.456      0.000       0.009       0.016
DWS            0.0079      0.003      2.884      0.004       0.003       0.013
==============================================================================
Omnibus:                     2886.709   Durbin-Watson:                   1.770
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           978137.436
Skew:                           9.414   Prob(JB):                         0.00
Kurtosis:                     112.787   Cond. No.                     1.80e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.8e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[65]:
```
#removed USG b/c not statistically significant
cols = [ 'Age', 'G', 'GS',
        'X3PM', 'X2PP',
         'ORB',  'STL', 'BLK',
        'PF', 'TS', 'DRBP','X3PAR',
       'ASTP', 'STLP', 'BLKP', 'TOVP', 'OWS', 'DWS',   ]
x_train = train[cols] #all of the variables
y_train = train['SMOY']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   SMOY   R-squared:                       0.144
Model:                            OLS   Adj. R-squared:                  0.136
Method:                 Least Squares   F-statistic:                     17.53
Date:                Wed, 12 May 2021   Prob (F-statistic):           1.48e-51
Time:                        19:19:03   Log-Likelihood:                 2679.2
No. Observations:                1892   AIC:                            -5320.
Df Residuals:                    1873   BIC:                            -5215.
Df Model:                          18
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
```

```
const          0.0394      0.018      2.253      0.024       0.005       0.074
Age            0.0005      0.000      1.385      0.166      -0.000       0.001
G             -0.0001      0.000     -1.001      0.317      -0.000       0.000
GS            -0.0010      0.000     -5.437      0.000      -0.001      -0.001
X3PM           0.0330      0.004      7.386      0.000       0.024       0.042
X2PP           0.0624      0.027      2.326      0.020       0.010       0.115
ORB            0.0015      0.004      0.360      0.719      -0.007       0.010
STL            0.0226      0.009      2.631      0.009       0.006       0.039
BLK            0.0212      0.009      2.314      0.021       0.003       0.039
PF            -0.0104      0.004     -2.917      0.004      -0.017      -0.003
TS            -0.1374      0.030     -4.653      0.000      -0.195      -0.079
DRBP          -0.0005      0.000     -1.521      0.128      -0.001       0.000
X3PAR         -0.0475      0.010     -4.717      0.000      -0.067      -0.028
ASTP          -0.0005      0.000     -2.236      0.025      -0.001   -6.46e-05
STLP          -0.0061      0.003     -2.357      0.019      -0.011      -0.001
BLKP          -0.0038      0.002     -2.169      0.030      -0.007      -0.000
TOVP           0.0011      0.000      3.417      0.001       0.000       0.002
OWS            0.0127      0.002      7.384      0.000       0.009       0.016
DWS            0.0080      0.003      2.934      0.003       0.003       0.013
==============================================================================
Omnibus:                     2886.633   Durbin-Watson:                   1.767
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           976727.636
Skew:                           9.415   Prob(JB):                         0.00
Kurtosis:                     112.705   Cond. No.                     1.74e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.74e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[66]:
```
#removed TOVP b/c not statistically significant
cols = [ 'Age', 'G', 'GS',
        'X3PM', 'X2PP',
         'ORB',  'STL', 'BLK',
        'PF', 'TS', 'DRBP','X3PAR',
       'ASTP', 'STLP', 'BLKP', 'OWS', 'DWS',   ]
x_train = train[cols] #all of the variables
y_train = train['SMOY']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
```

```
==============================================================================
Dep. Variable:                   SMOY   R-squared:                       0.139
Model:                            OLS   Adj. R-squared:                  0.131
Method:                 Least Squares   F-statistic:                     17.77
Date:                Wed, 12 May 2021   Prob (F-statistic):           8.57e-50
Time:                        19:19:03   Log-Likelihood:                 2673.3
No. Observations:                1892   AIC:                            -5311.
Df Residuals:                    1874   BIC:                            -5211.
Df Model:                          17
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0430      0.018      2.455      0.014       0.009       0.077
Age            0.0005      0.000      1.608      0.108      -0.000       0.001
G             -0.0001      0.000     -1.215      0.225      -0.000    7.91e-05
GS            -0.0009      0.000     -5.231      0.000      -0.001      -0.001
X3PM           0.0316      0.004      7.076      0.000       0.023       0.040
X2PP           0.0619      0.027      2.302      0.021       0.009       0.115
ORB            0.0022      0.004      0.508      0.611      -0.006       0.011
STL            0.0205      0.009      2.388      0.017       0.004       0.037
BLK            0.0168      0.009      1.843      0.065      -0.001       0.035
PF            -0.0085      0.004     -2.412      0.016      -0.015      -0.002
TS            -0.1254      0.029     -4.263      0.000      -0.183      -0.068
DRBP          -0.0006      0.000     -1.570      0.117      -0.001       0.000
X3PAR         -0.0505      0.010     -5.018      0.000      -0.070      -0.031
ASTP          -0.0002      0.000     -0.765      0.444      -0.001       0.000
STLP          -0.0055      0.003     -2.100      0.036      -0.011      -0.000
BLKP          -0.0027      0.002     -1.570      0.117      -0.006       0.001
OWS            0.0110      0.002      6.657      0.000       0.008       0.014
DWS            0.0083      0.003      3.034      0.002       0.003       0.014
==============================================================================
Omnibus:                     2899.616   Durbin-Watson:                   1.754
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          1000081.384
Skew:                           9.493   Prob(JB):                         0.00
Kurtosis:                     114.020   Cond. No.                     1.70e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.7e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[67]:
```
#removed PF b/c not statistically significant
cols = [ 'Age', 'G', 'GS',
        'X3PM', 'X2PP',
```

```
         'ORB',  'STL', 'BLK',
        'TS', 'DRBP','X3PAR',
       'ASTP', 'STLP', 'BLKP', 'OWS', 'DWS',   ]
x_train = train[cols] #all of the variables
y_train = train['SMOY']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   SMOY   R-squared:                       0.136
Model:                            OLS   Adj. R-squared:                  0.129
Method:                 Least Squares   F-statistic:                     18.47
Date:                Wed, 12 May 2021   Prob (F-statistic):           3.09e-49
Time:                        19:19:03   Log-Likelihood:                2670.4
No. Observations:                1892   AIC:                            -5307.
Df Residuals:                    1875   BIC:                            -5212.
Df Model:                          16
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0412      0.018      2.348      0.019       0.007       0.076
Age            0.0005      0.000      1.383      0.167      -0.000       0.001
G             -0.0002      0.000     -1.453      0.146      -0.000    5.36e-05
GS            -0.0010      0.000     -5.493      0.000      -0.001      -0.001
X3PM           0.0299      0.004      6.767      0.000       0.021       0.039
X2PP           0.0644      0.027      2.392      0.017       0.012       0.117
ORB           -0.0011      0.004     -0.268      0.788      -0.009       0.007
STL            0.0144      0.008      1.750      0.080      -0.002       0.030
BLK            0.0121      0.009      1.361      0.174      -0.005       0.030
TS            -0.1336      0.029     -4.567      0.000      -0.191      -0.076
DRBP          -0.0006      0.000     -1.625      0.104      -0.001       0.000
X3PAR         -0.0474      0.010     -4.744      0.000      -0.067      -0.028
ASTP          -0.0001      0.000     -0.706      0.480      -0.001       0.000
STLP          -0.0046      0.003     -1.773      0.076      -0.010       0.000
BLKP          -0.0024      0.002     -1.355      0.176      -0.006       0.001
OWS            0.0115      0.002      7.015      0.000       0.008       0.015
DWS            0.0082      0.003      2.994      0.003       0.003       0.014
==============================================================================
Omnibus:                     2899.560   Durbin-Watson:                   1.749
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           994272.930
Skew:                           9.496   Prob(JB):                         0.00
```

```
Kurtosis:                       113.687   Cond. No.                         1.69e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.69e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[68]:
```
#removed ASTP b/c not statistically significant
cols = [ 'Age', 'G', 'GS',
        'X3PM', 'X2PP',
         'ORB',  'STL', 'BLK',
        'TS', 'DRBP','X3PAR',
        'STLP', 'BLKP', 'OWS', 'DWS',   ]
x_train = train[cols] #all of the variables
y_train = train['SMOY']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   SMOY   R-squared:                       0.136
Model:                            OLS   Adj. R-squared:                  0.129
Method:                 Least Squares   F-statistic:                     19.67
Date:                Wed, 12 May 2021   Prob (F-statistic):           8.66e-50
Time:                        19:19:03   Log-Likelihood:                 2670.1
No. Observations:                1892   AIC:                            -5308.
Df Residuals:                    1876   BIC:                            -5220.
Df Model:                          15
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0383      0.017      2.246      0.025       0.005       0.072
Age            0.0004      0.000      1.352      0.176      -0.000       0.001
G             -0.0002      0.000     -1.465      0.143      -0.000    5.24e-05
GS            -0.0010      0.000     -5.502      0.000      -0.001      -0.001
X3PM           0.0301      0.004      6.821      0.000       0.021       0.039
X2PP           0.0649      0.027      2.414      0.016       0.012       0.118
ORB           -0.0006      0.004     -0.142      0.887      -0.008       0.007
STL            0.0134      0.008      1.660      0.097      -0.002       0.029
BLK            0.0118      0.009      1.325      0.185      -0.006       0.029
TS            -0.1326      0.029     -4.539      0.000      -0.190      -0.075
```

```
DRBP          -0.0005      0.000     -1.522      0.128      -0.001       0.000
X3PAR         -0.0467      0.010     -4.700      0.000      -0.066      -0.027
STLP          -0.0047      0.003     -1.827      0.068      -0.010       0.000
BLKP          -0.0022      0.002     -1.252      0.211      -0.006       0.001
OWS            0.0114      0.002      6.980      0.000       0.008       0.015
DWS            0.0081      0.003      2.978      0.003       0.003       0.014
==============================================================================
Omnibus:                     2900.142   Durbin-Watson:                   1.750
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           994814.444
Skew:                           9.500   Prob(JB):                         0.00
Kurtosis:                     113.717   Cond. No.                     1.67e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.67e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[69]:
```
#removed BLK b/c not statistically significant
cols = [ 'Age', 'G', 'GS',
        'X3PM', 'X2PP',
         'ORB',  'STL',
        'TS', 'DRBP','X3PAR',
        'STLP', 'BLKP', 'OWS', 'DWS',   ]
x_train = train[cols] #all of the variables
y_train = train['SMOY']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   SMOY   R-squared:                       0.135
Model:                            OLS   Adj. R-squared:                  0.129
Method:                 Least Squares   F-statistic:                     20.94
Date:                Wed, 12 May 2021   Prob (F-statistic):           4.28e-50
Time:                        19:19:03   Log-Likelihood:                2669.2
No. Observations:                1892   AIC:                            -5308.
Df Residuals:                    1877   BIC:                            -5225.
Df Model:                          14
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
```

```
const          0.0369      0.017      2.166      0.030       0.003       0.070
Age            0.0004      0.000      1.332      0.183      -0.000       0.001
G             -0.0002      0.000     -1.429      0.153      -0.000    5.62e-05
GS            -0.0009      0.000     -5.383      0.000      -0.001      -0.001
X3PM           0.0298      0.004      6.760      0.000       0.021       0.038
X2PP           0.0673      0.027      2.509      0.012       0.015       0.120
ORB            0.0020      0.003      0.590      0.555      -0.005       0.009
STL            0.0146      0.008      1.818      0.069      -0.001       0.030
TS            -0.1340      0.029     -4.588      0.000      -0.191      -0.077
DRBP          -0.0006      0.000     -1.793      0.073      -0.001    5.79e-05
X3PAR         -0.0443      0.010     -4.533      0.000      -0.064      -0.025
STLP          -0.0049      0.003     -1.900      0.058      -0.010       0.000
BLKP          -0.0005      0.001     -0.418      0.676      -0.003       0.002
OWS            0.0113      0.002      6.912      0.000       0.008       0.014
DWS            0.0086      0.003      3.153      0.002       0.003       0.014
==============================================================================
Omnibus:                     2902.994   Durbin-Watson:                   1.749
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           999216.918
Skew:                           9.518   Prob(JB):                         0.00
Kurtosis:                     113.963   Cond. No.                     1.66e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.66e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[70]:
```
#removed BLKP b/c not statistically significant
cols = [ 'Age', 'G', 'GS',
        'X3PM', 'X2PP',
         'ORB',  'STL',
        'TS', 'DRBP','X3PAR',
        'STLP',  'OWS', 'DWS',   ]
x_train = train[cols] #all of the variables
y_train = train['SMOY']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   SMOY   R-squared:                       0.135
Model:                            OLS   Adj. R-squared:                  0.129
Method:                 Least Squares   F-statistic:                     22.55
```

```
Date:                Wed, 12 May 2021   Prob (F-statistic):           9.59e-51
Time:                        19:19:03   Log-Likelihood:                 2669.1
No. Observations:                1892   AIC:                            -5310.
Df Residuals:                    1878   BIC:                            -5233.
Df Model:                          13
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0375      0.017      2.211      0.027       0.004       0.071
Age            0.0004      0.000      1.328      0.184      -0.000       0.001
G             -0.0001      0.000     -1.393      0.164      -0.000    5.96e-05
GS            -0.0009      0.000     -5.380      0.000      -0.001      -0.001
X3PM           0.0297      0.004      6.754      0.000       0.021       0.038
X2PP           0.0662      0.027      2.481      0.013       0.014       0.119
ORB            0.0019      0.003      0.548      0.584      -0.005       0.009
STL            0.0149      0.008      1.861      0.063      -0.001       0.031
TS            -0.1354      0.029     -4.671      0.000      -0.192      -0.079
DRBP          -0.0007      0.000     -1.978      0.048      -0.001   -5.54e-06
X3PAR         -0.0434      0.010     -4.561      0.000      -0.062      -0.025
STLP          -0.0049      0.003     -1.899      0.058      -0.010       0.000
OWS            0.0114      0.002      7.053      0.000       0.008       0.015
DWS            0.0084      0.003      3.127      0.002       0.003       0.014
==============================================================================
Omnibus:                     2902.809   Durbin-Watson:                   1.749
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           998835.238
Skew:                           9.517   Prob(JB):                         0.00
Kurtosis:                     113.941   Cond. No.                     1.66e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.66e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[71]:
```
import statsmodels.formula.api as smf

SMOYLinear = smf.ols(formula='SMOY ~ Age + G+  GS+␣
 ↪X3PM+X2PP+ORB+STL+TS+DRBP+X3PAR+STLP+OWS+DWS',
                  data=train).fit()
print(SMOYLinear.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   SMOY   R-squared:                       0.135
Model:                            OLS   Adj. R-squared:                  0.129
Method:                 Least Squares   F-statistic:                     22.55
```

```
Date:                Wed, 12 May 2021   Prob (F-statistic):           9.59e-51
Time:                        19:19:03   Log-Likelihood:                 2669.1
No. Observations:                1892   AIC:                            -5310.
Df Residuals:                    1878   BIC:                            -5233.
Df Model:                          13
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      0.0375      0.017      2.211      0.027       0.004       0.071
Age            0.0004      0.000      1.328      0.184      -0.000       0.001
G             -0.0001      0.000     -1.393      0.164      -0.000    5.96e-05
GS            -0.0009      0.000     -5.380      0.000      -0.001      -0.001
X3PM           0.0297      0.004      6.754      0.000       0.021       0.038
X2PP           0.0662      0.027      2.481      0.013       0.014       0.119
ORB            0.0019      0.003      0.548      0.584      -0.005       0.009
STL            0.0149      0.008      1.861      0.063      -0.001       0.031
TS            -0.1354      0.029     -4.671      0.000      -0.192      -0.079
DRBP          -0.0007      0.000     -1.978      0.048      -0.001   -5.54e-06
X3PAR         -0.0434      0.010     -4.561      0.000      -0.062      -0.025
STLP          -0.0049      0.003     -1.899      0.058      -0.010       0.000
OWS            0.0114      0.002      7.053      0.000       0.008       0.015
DWS            0.0084      0.003      3.127      0.002       0.003       0.014
==============================================================================
Omnibus:                     2902.809   Durbin-Watson:                   1.749
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           998835.238
Skew:                           9.517   Prob(JB):                         0.00
Kurtosis:                     113.941   Cond. No.                     1.66e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.66e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[72]:
```
OSR2(SMOYLinear, train, test, 'SMOY')
```

[72]:
```
0.10762095053200527
```

Defensive Player of the Year

[163]:
```
#Loading in dataset
dataset = pd.read_csv('update142.csv')
train = dataset[dataset['Season'] <= 2017]
test = dataset[dataset['Season'] > 2017]

y_train = train['MVP']
```

```
y_test = test['MVP']
x_train = train.iloc[:,6:51]
x_test = test.iloc[:,6:51]
```

[164]:
```
#removed season b/c not relevant to prediciting our output variable
cols = [ 'Age', 'G', 'GS',
        'X3PM','X3PP', 'X2PP',
        'FTP', 'ORB',  'STL', 'BLK',
        'PF', 'TS', 'FTR', 'DRBP','X3PAR',
       'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
        'DBPM', ]
x_train = train[cols] #all of the variables
y_train = train['DPOY']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   DPOY   R-squared:                       0.133
Model:                            OLS   Adj. R-squared:                  0.126
Method:                 Least Squares   F-statistic:                     21.33
Date:                Wed, 12 May 2021   Prob (F-statistic):           1.90e-82
Time:                        19:31:06   Log-Likelihood:                 5579.1
No. Observations:                3231   AIC:                        -1.111e+04
Df Residuals:                    3207   BIC:                        -1.096e+04
Df Model:                          23
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0106      0.013      0.811      0.418      -0.015       0.036
Age           -0.0003      0.000     -1.341      0.180      -0.001       0.000
G             -0.0002    5.9e-05     -3.218      0.001      -0.000   -7.41e-05
GS         -6.374e-05   4.32e-05     -1.475      0.140      -0.000     2.1e-05
X3PM          -0.0015      0.002     -0.701      0.483      -0.006       0.003
X3PP           0.0043      0.007      0.649      0.517      -0.009       0.017
X2PP           0.0142      0.017      0.846      0.398      -0.019       0.047
FTP            0.0040      0.007      0.559      0.576      -0.010       0.018
ORB         5.672e-05      0.002      0.031      0.976      -0.004       0.004
STL            0.0126      0.003      3.617      0.000       0.006       0.019
BLK            0.0296      0.003      8.676      0.000       0.023       0.036
PF            -0.0081      0.002     -4.587      0.000      -0.012      -0.005
TS            -0.0286      0.018     -1.560      0.119      -0.065       0.007
```

```
FTR            0.0290      0.007      4.373      0.000       0.016       0.042
DRBP          -0.0003      0.000     -1.570      0.117      -0.001    7.77e-05
X3PAR          0.0177      0.006      2.792      0.005       0.005       0.030
ASTP       -3.608e-05      0.000     -0.272      0.786      -0.000       0.000
STLP          -0.0044      0.002     -2.397      0.017      -0.008      -0.001
BLKP          -0.0019      0.001     -2.075      0.038      -0.004      -0.000
TOVP           0.0004      0.000      1.885      0.060   -1.54e-05       0.001
USG           -0.0002      0.000     -0.960      0.337      -0.001       0.000
OWS            0.0001      0.001      0.206      0.837      -0.001       0.001
DWS            0.0099      0.001      7.850      0.000       0.007       0.012
DBPM           0.0010      0.001      0.865      0.387      -0.001       0.003
==============================================================================
Omnibus:                     5939.334   Durbin-Watson:                   1.107
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          7206173.730
Skew:                          13.579   Prob(JB):                         0.00
Kurtosis:                     232.761   Cond. No.                     2.32e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 2.32e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[202]:
```
#removed ORB b/c not statistically significant
cols = [ 'Age', 'G', 'GS',
        'X3PM','X3PP', 'X2PP',
        'FTP',   'STL', 'BLK',
        'PF', 'TS', 'FTR', 'DRBP','X3PAR',
       'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
        'DBPM', ]
x_train = train[cols] #all of the variables
y_train = train['DPOY']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   DPOY   R-squared:                       0.133
Model:                            OLS   Adj. R-squared:                  0.127
Method:                 Least Squares   F-statistic:                     22.31
Date:                Wed, 12 May 2021   Prob (F-statistic):           3.98e-83
Time:                        19:40:21   Log-Likelihood:                 5579.1
No. Observations:                3231   AIC:                         -1.111e+04
```

```
Df Residuals:                     3208   BIC:                         -1.097e+04
Df Model:                           22
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0106      0.013      0.811      0.418      -0.015       0.036
Age           -0.0003      0.000     -1.343      0.179      -0.001       0.000
G             -0.0002   5.89e-05     -3.219      0.001      -0.000   -7.41e-05
GS         -6.355e-05   4.28e-05     -1.486      0.137      -0.000    2.03e-05
X3PM          -0.0015      0.002     -0.741      0.459      -0.005       0.002
X3PP           0.0042      0.006      0.651      0.515      -0.008       0.017
X2PP           0.0142      0.017      0.849      0.396      -0.019       0.047
FTP            0.0040      0.007      0.559      0.577      -0.010       0.018
STL            0.0126      0.003      3.620      0.000       0.006       0.019
BLK            0.0296      0.003      9.378      0.000       0.023       0.036
PF            -0.0081      0.002     -4.801      0.000      -0.011      -0.005
TS            -0.0286      0.018     -1.560      0.119      -0.065       0.007
FTR            0.0290      0.007      4.373      0.000       0.016       0.042
DRBP          -0.0003      0.000     -1.663      0.096      -0.001    5.54e-05
X3PAR          0.0177      0.006      2.792      0.005       0.005       0.030
ASTP       -3.639e-05      0.000     -0.275      0.783      -0.000       0.000
STLP          -0.0044      0.002     -2.404      0.016      -0.008      -0.001
BLKP          -0.0019      0.001     -2.105      0.035      -0.004      -0.000
TOVP           0.0004      0.000      1.888      0.059   -1.46e-05       0.001
USG           -0.0002      0.000     -0.964      0.335      -0.001       0.000
OWS            0.0001      0.001      0.215      0.830      -0.001       0.001
DWS            0.0099      0.001      7.854      0.000       0.007       0.012
DBPM           0.0010      0.001      0.876      0.381      -0.001       0.003
==============================================================================
Omnibus:                     5939.342   Durbin-Watson:                   1.107
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          7206219.366
Skew:                          13.579   Prob(JB):                         0.00
Kurtosis:                     232.762   Cond. No.                     2.32e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 2.32e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[203]:
```
#removed ASTP b/c not statistically significant
cols = [ 'Age', 'G', 'GS',
        'X3PM','X3PP', 'X2PP',
        'FTP',   'STL', 'BLK',
        'PF', 'TS', 'FTR', 'DRBP','X3PAR',
```

```
        'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
        'DBPM', ]
x_train = train[cols] #all of the variables
y_train = train['DPOY']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   DPOY   R-squared:                       0.133
Model:                            OLS   Adj. R-squared:                  0.127
Method:                 Least Squares   F-statistic:                     23.37
Date:                Wed, 12 May 2021   Prob (F-statistic):           8.42e-84
Time:                        19:40:39   Log-Likelihood:                 5579.1
No. Observations:                3231   AIC:                        -1.111e+04
Df Residuals:                    3209   BIC:                        -1.098e+04
Df Model:                          21
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0105      0.013      0.803      0.422      -0.015       0.036
Age           -0.0003      0.000     -1.366      0.172      -0.001       0.000
G             -0.0002   5.89e-05     -3.233      0.001      -0.000   -7.49e-05
GS         -6.354e-05   4.28e-05     -1.486      0.137      -0.000    2.03e-05
X3PM          -0.0014      0.002     -0.713      0.476      -0.005       0.002
X3PP           0.0042      0.006      0.641      0.521      -0.009       0.017
X2PP           0.0144      0.017      0.861      0.389      -0.018       0.047
FTP            0.0040      0.007      0.556      0.578      -0.010       0.018
STL            0.0124      0.003      3.656      0.000       0.006       0.019
BLK            0.0297      0.003      9.395      0.000       0.023       0.036
PF            -0.0081      0.002     -4.794      0.000      -0.011      -0.005
TS            -0.0281      0.018     -1.539      0.124      -0.064       0.008
FTR            0.0292      0.007      4.461      0.000       0.016       0.042
DRBP          -0.0003      0.000     -1.661      0.097      -0.001     5.3e-05
X3PAR          0.0177      0.006      2.788      0.005       0.005       0.030
STLP          -0.0043      0.002     -2.390      0.017      -0.008      -0.001
BLKP          -0.0019      0.001     -2.087      0.037      -0.004      -0.000
TOVP           0.0004      0.000      2.064      0.039    1.76e-05       0.001
USG           -0.0003      0.000     -1.260      0.208      -0.001       0.000
OWS          8.45e-05      0.001      0.140      0.889      -0.001       0.001
DWS            0.0100      0.001      7.969      0.000       0.008       0.012
DBPM           0.0009      0.001      0.833      0.405      -0.001       0.003
```

```
==============================================================================
Omnibus:                     5939.300   Durbin-Watson:                   1.106
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          7205295.929
Skew:                          13.579   Prob(JB):                         0.00
Kurtosis:                     232.747   Cond. No.                     2.29e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 2.29e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[204]:
```
#removed OWS b/c not statistically significant
cols = [ 'Age', 'G', 'GS',
        'X3PM','X3PP', 'X2PP',
        'FTP',   'STL', 'BLK',
        'PF', 'TS', 'FTR', 'DRBP','X3PAR',
        'STLP', 'BLKP', 'TOVP', 'USG',  'DWS',
        'DBPM', ]
x_train = train[cols] #all of the variables
y_train = train['DPOY']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   DPOY   R-squared:                       0.133
Model:                            OLS   Adj. R-squared:                  0.127
Method:                 Least Squares   F-statistic:                     24.55
Date:                Wed, 12 May 2021   Prob (F-statistic):           1.69e-84
Time:                        19:40:51   Log-Likelihood:                 5579.1
No. Observations:                3231   AIC:                        -1.112e+04
Df Residuals:                    3210   BIC:                        -1.099e+04
Df Model:                          20
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0099      0.012      0.800      0.424      -0.014       0.034
Age           -0.0003      0.000     -1.360      0.174      -0.001       0.000
G             -0.0002   5.89e-05     -3.235      0.001      -0.000    -7.5e-05
GS         -6.252e-05   4.21e-05     -1.484      0.138      -0.000    2.01e-05
X3PM          -0.0014      0.002     -0.700      0.484      -0.005       0.002
```

```
X3PP           0.0041      0.006      0.638      0.524      -0.009       0.017
X2PP           0.0146      0.017      0.881      0.379      -0.018       0.047
FTP            0.0040      0.007      0.559      0.576      -0.010       0.018
STL            0.0124      0.003      3.664      0.000       0.006       0.019
BLK            0.0297      0.003      9.402      0.000       0.023       0.036
PF            -0.0081      0.002     -4.843      0.000      -0.011      -0.005
TS            -0.0273      0.017     -1.571      0.116      -0.061       0.007
FTR            0.0294      0.006      4.565      0.000       0.017       0.042
DRBP          -0.0003      0.000     -1.666      0.096      -0.001    5.21e-05
X3PAR          0.0176      0.006      2.785      0.005       0.005       0.030
STLP          -0.0043      0.002     -2.397      0.017      -0.008      -0.001
BLKP          -0.0019      0.001     -2.109      0.035      -0.004      -0.000
TOVP           0.0003      0.000      2.065      0.039    1.76e-05       0.001
USG           -0.0003      0.000     -1.256      0.209      -0.001       0.000
DWS            0.0100      0.001      8.410      0.000       0.008       0.012
DBPM           0.0009      0.001      0.829      0.407      -0.001       0.003
==============================================================================
Omnibus:                     5939.068   Durbin-Watson:                   1.107
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          7202919.454
Skew:                          13.578   Prob(JB):                         0.00
Kurtosis:                     232.709   Cond. No.                     2.26e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 2.26e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[205]:
```python
#removed FTP b/c not statistically significant
cols = [ 'Age', 'G', 'GS',
        'X3PM','X3PP', 'X2PP',
           'STL', 'BLK',
        'PF', 'TS', 'FTR', 'DRBP','X3PAR',
        'STLP', 'BLKP', 'TOVP', 'USG',  'DWS',
        'DBPM', ]
x_train = train[cols] #all of the variables
y_train = train['DPOY']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   DPOY   R-squared:                       0.133
```

```
Model:                            OLS   Adj. R-squared:                  0.127
Method:                 Least Squares   F-statistic:                     25.83
Date:                Wed, 12 May 2021   Prob (F-statistic):           3.81e-85
Time:                        19:41:15   Log-Likelihood:                 5578.9
No. Observations:                3231   AIC:                        -1.112e+04
Df Residuals:                    3211   BIC:                        -1.100e+04
Df Model:                          19
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0118      0.012      0.998      0.318      -0.011       0.035
Age           -0.0002      0.000     -1.334      0.182      -0.001       0.000
G             -0.0002   5.85e-05     -3.195      0.001      -0.000   -7.23e-05
GS         -6.278e-05   4.21e-05     -1.490      0.136      -0.000    1.98e-05
X3PM          -0.0013      0.002     -0.646      0.518      -0.005       0.003
X3PP           0.0045      0.006      0.701      0.483      -0.008       0.017
X2PP           0.0132      0.016      0.807      0.420      -0.019       0.045
STL            0.0124      0.003      3.679      0.000       0.006       0.019
BLK            0.0297      0.003      9.394      0.000       0.023       0.036
PF            -0.0081      0.002     -4.838      0.000      -0.011      -0.005
TS            -0.0251      0.017     -1.484      0.138      -0.058       0.008
FTR            0.0292      0.006      4.538      0.000       0.017       0.042
DRBP          -0.0003      0.000     -1.717      0.086      -0.001     4.3e-05
X3PAR          0.0175      0.006      2.766      0.006       0.005       0.030
STLP          -0.0044      0.002     -2.443      0.015      -0.008      -0.001
BLKP          -0.0019      0.001     -2.151      0.032      -0.004      -0.000
TOVP           0.0003      0.000      2.050      0.040    1.51e-05       0.001
USG           -0.0002      0.000     -1.164      0.244      -0.001       0.000
DWS            0.0100      0.001      8.394      0.000       0.008       0.012
DBPM           0.0009      0.001      0.853      0.394      -0.001       0.003
==============================================================================
Omnibus:                     5939.496   Durbin-Watson:                   1.106
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          7206095.273
Skew:                          13.580   Prob(JB):                         0.00
Kurtosis:                     232.760   Cond. No.                     2.21e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 2.21e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[206]:
```
#removed X3PM b/c not statistically significant
cols = [ 'Age', 'G', 'GS',
        'X3PP', 'X2PP',
```

```
           'STL', 'BLK',
        'PF', 'TS', 'FTR', 'DRBP','X3PAR',
        'STLP', 'BLKP', 'TOVP', 'USG',  'DWS',
        'DBPM', ]
x_train = train[cols] #all of the variables
y_train = train['DPOY']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   DPOY   R-squared:                       0.132
Model:                            OLS   Adj. R-squared:                  0.128
Method:                 Least Squares   F-statistic:                     27.24
Date:                Wed, 12 May 2021   Prob (F-statistic):           8.78e-86
Time:                        19:41:27   Log-Likelihood:                 5578.7
No. Observations:                3231   AIC:                        -1.112e+04
Df Residuals:                    3212   BIC:                        -1.100e+04
Df Model:                          18
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0136      0.012      1.176      0.240      -0.009       0.036
Age           -0.0003      0.000     -1.379      0.168      -0.001       0.000
G             -0.0002   5.83e-05     -3.263      0.001      -0.000    -7.6e-05
GS         -6.744e-05   4.15e-05     -1.626      0.104      -0.000    1.39e-05
X3PP           0.0035      0.006      0.554      0.579      -0.009       0.016
X2PP           0.0143      0.016      0.877      0.381      -0.018       0.046
STL            0.0118      0.003      3.637      0.000       0.005       0.018
BLK            0.0302      0.003      9.921      0.000       0.024       0.036
PF            -0.0082      0.002     -4.914      0.000      -0.011      -0.005
TS            -0.0268      0.017     -1.604      0.109      -0.060       0.006
FTR            0.0290      0.006      4.513      0.000       0.016       0.042
DRBP          -0.0003      0.000     -1.710      0.087      -0.001    4.41e-05
X3PAR          0.0153      0.005      2.857      0.004       0.005       0.026
STLP          -0.0043      0.002     -2.383      0.017      -0.008      -0.001
BLKP          -0.0020      0.001     -2.342      0.019      -0.004      -0.000
TOVP           0.0003      0.000      2.059      0.040    1.65e-05       0.001
USG           -0.0003      0.000     -1.344      0.179      -0.001       0.000
DWS            0.0100      0.001      8.412      0.000       0.008       0.012
DBPM           0.0009      0.001      0.894      0.371      -0.001       0.003
==============================================================================
```

```
Omnibus:                       5939.149   Durbin-Watson:                   1.106
Prob(Omnibus):                    0.000   Jarque-Bera (JB):          7202996.725
Skew:                            13.578   Prob(JB):                         0.00
Kurtosis:                       232.710   Cond. No.                     2.19e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 2.19e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[207]:
```
#removed X3PP b/c not statistically significant
cols = [ 'Age', 'G', 'GS',
         'X2PP',
           'STL', 'BLK',
        'PF', 'TS', 'FTR', 'DRBP','X3PAR',
        'STLP', 'BLKP', 'TOVP', 'USG',  'DWS',
        'DBPM', ]
x_train = train[cols] #all of the variables
y_train = train['DPOY']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   DPOY   R-squared:                       0.132
Model:                            OLS   Adj. R-squared:                  0.128
Method:                 Least Squares   F-statistic:                     28.83
Date:                Wed, 12 May 2021   Prob (F-statistic):           1.87e-86
Time:                        19:41:37   Log-Likelihood:                 5578.6
No. Observations:                3231   AIC:                        -1.112e+04
Df Residuals:                    3213   BIC:                        -1.101e+04
Df Model:                          17
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0139      0.012      1.207      0.228      -0.009       0.037
Age           -0.0003      0.000     -1.378      0.168      -0.001       0.000
G             -0.0002    5.8e-05     -3.225      0.001      -0.000   -7.34e-05
GS         -6.623e-05   4.14e-05     -1.599      0.110      -0.000     1.5e-05
X2PP           0.0129      0.016      0.802      0.422      -0.019       0.045
STL            0.0121      0.003      3.746      0.000       0.006       0.018
```

```
BLK            0.0300      0.003      9.916      0.000       0.024       0.036
PF            -0.0083      0.002     -4.989      0.000      -0.012      -0.005
TS            -0.0253      0.016     -1.534      0.125      -0.058       0.007
FTR            0.0286      0.006      4.480      0.000       0.016       0.041
DRBP          -0.0003      0.000     -1.736      0.083      -0.001    3.95e-05
X3PAR          0.0163      0.005      3.196      0.001       0.006       0.026
STLP          -0.0043      0.002     -2.398      0.017      -0.008      -0.001
BLKP          -0.0020      0.001     -2.334      0.020      -0.004      -0.000
TOVP           0.0003      0.000      2.051      0.040    1.53e-05       0.001
USG           -0.0002      0.000     -1.285      0.199      -0.001       0.000
DWS            0.0100      0.001      8.402      0.000       0.008       0.012
DBPM           0.0009      0.001      0.900      0.368      -0.001       0.003
==============================================================================
Omnibus:                     5938.747   Durbin-Watson:                   1.106
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          7197341.968
Skew:                          13.576   Prob(JB):                         0.00
Kurtosis:                     232.619   Cond. No.                     2.15e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 2.15e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[208]:
```
#removed X2PP b/c not statistically significant
cols = [ 'Age', 'G', 'GS',

           'STL', 'BLK',
        'PF', 'TS', 'FTR', 'DRBP','X3PAR',
        'STLP', 'BLKP', 'TOVP', 'USG',  'DWS',
        'DBPM', ]
x_train = train[cols] #all of the variables
y_train = train['DPOY']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   DPOY   R-squared:                       0.132
Model:                            OLS   Adj. R-squared:                  0.128
Method:                 Least Squares   F-statistic:                     30.60
Date:                Wed, 12 May 2021   Prob (F-statistic):           4.53e-87
Time:                        19:41:50   Log-Likelihood:                 5578.2
```

```
No. Observations:                3231   AIC:                         -1.112e+04
Df Residuals:                    3214   BIC:                         -1.102e+04
Df Model:                          16
Covariance Type:            nonrobust
================================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
const          0.0166      0.011      1.502      0.133      -0.005       0.038
Age           -0.0003      0.000     -1.383      0.167      -0.001       0.000
G             -0.0002   5.78e-05     -3.172      0.002      -0.000   -7.01e-05
GS         -6.634e-05   4.14e-05     -1.601      0.109      -0.000    1.49e-05
STL            0.0121      0.003      3.751      0.000       0.006       0.018
BLK            0.0305      0.003     10.280      0.000       0.025       0.036
PF            -0.0083      0.002     -5.013      0.000      -0.012      -0.005
TS            -0.0185      0.014     -1.307      0.191      -0.046       0.009
FTR            0.0284      0.006      4.450      0.000       0.016       0.041
DRBP          -0.0003      0.000     -1.676      0.094      -0.001    4.99e-05
X3PAR          0.0156      0.005      3.106      0.002       0.006       0.025
STLP          -0.0042      0.002     -2.366      0.018      -0.008      -0.001
BLKP          -0.0021      0.001     -2.360      0.018      -0.004      -0.000
TOVP           0.0003      0.000      2.044      0.041    1.41e-05       0.001
USG           -0.0003      0.000     -1.391      0.164      -0.001       0.000
DWS            0.0100      0.001      8.378      0.000       0.008       0.012
DBPM           0.0009      0.001      0.856      0.392      -0.001       0.003
================================================================================
Omnibus:                     5938.807   Durbin-Watson:                   1.107
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          7195953.930
Skew:                          13.577   Prob(JB):                         0.00
Kurtosis:                     232.596   Cond. No.                     1.73e+03
================================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.73e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[209]:
```
#removed DBPM b/c not statistically significant
cols = [ 'Age', 'G', 'GS',
           'STL', 'BLK',
        'PF', 'TS', 'FTR', 'DRBP','X3PAR',
        'STLP', 'BLKP', 'TOVP', 'USG',  'DWS',
        ]
x_train = train[cols] #all of the variables
y_train = train['DPOY']
#adding an intercept
x_train = sm.add_constant(x_train)
```

```
#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   DPOY   R-squared:                       0.132
Model:                            OLS   Adj. R-squared:                  0.128
Method:                 Least Squares   F-statistic:                     32.59
Date:                Wed, 12 May 2021   Prob (F-statistic):           1.11e-87
Time:                        19:42:03   Log-Likelihood:                 5577.9
No. Observations:                3231   AIC:                        -1.112e+04
Df Residuals:                    3215   BIC:                        -1.103e+04
Df Model:                          15
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0135      0.010      1.295      0.195      -0.007       0.034
Age           -0.0002      0.000     -1.296      0.195      -0.001       0.000
G             -0.0002   5.73e-05     -3.316      0.001      -0.000   -7.77e-05
GS         -6.745e-05   4.14e-05     -1.629      0.103      -0.000    1.37e-05
STL            0.0120      0.003      3.715      0.000       0.006       0.018
BLK            0.0303      0.003     10.245      0.000       0.025       0.036
PF            -0.0083      0.002     -5.039      0.000      -0.012      -0.005
TS            -0.0148      0.013     -1.100      0.271      -0.041       0.012
FTR            0.0276      0.006      4.373      0.000       0.015       0.040
DRBP          -0.0003      0.000     -1.609      0.108      -0.001    6.14e-05
X3PAR          0.0159      0.005      3.177      0.002       0.006       0.026
STLP          -0.0033      0.001     -2.294      0.022      -0.006      -0.000
BLKP          -0.0017      0.001     -2.204      0.028      -0.003      -0.000
TOVP           0.0003      0.000      2.062      0.039    1.71e-05       0.001
USG           -0.0003      0.000     -1.983      0.047      -0.001   -3.87e-06
DWS            0.0104      0.001      9.823      0.000       0.008       0.012
==============================================================================
Omnibus:                     5939.023   Durbin-Watson:                   1.108
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          7194613.321
Skew:                          13.578   Prob(JB):                         0.00
Kurtosis:                     232.574   Cond. No.                     1.62e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.62e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[210]:
```
#removed TS b/c not statistically significant
cols = [ 'Age', 'G', 'GS',
           'STL', 'BLK',
        'PF',  'FTR', 'DRBP','X3PAR',
        'STLP', 'BLKP', 'TOVP', 'USG',  'DWS',
        ]
x_train = train[cols] #all of the variables
y_train = train['DPOY']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   DPOY   R-squared:                       0.132
Model:                            OLS   Adj. R-squared:                  0.128
Method:                 Least Squares   F-statistic:                     34.83
Date:                Wed, 12 May 2021   Prob (F-statistic):           3.32e-88
Time:                        19:42:15   Log-Likelihood:                 5577.3
No. Observations:                3231   AIC:                        -1.112e+04
Df Residuals:                    3216   BIC:                        -1.103e+04
Df Model:                          14
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0072      0.009      0.827      0.408      -0.010       0.024
Age           -0.0002      0.000     -1.321      0.187      -0.001       0.000
G             -0.0002    5.7e-05     -3.444      0.001      -0.000   -8.46e-05
GS         -6.707e-05   4.14e-05     -1.620      0.105      -0.000    1.41e-05
STL            0.0120      0.003      3.717      0.000       0.006       0.018
BLK            0.0305      0.003     10.341      0.000       0.025       0.036
PF            -0.0085      0.002     -5.181      0.000      -0.012      -0.005
FTR            0.0255      0.006      4.239      0.000       0.014       0.037
DRBP          -0.0003      0.000     -1.600      0.110      -0.001    6.31e-05
X3PAR          0.0149      0.005      3.024      0.003       0.005       0.025
STLP          -0.0032      0.001     -2.218      0.027      -0.006      -0.000
BLKP          -0.0018      0.001     -2.338      0.019      -0.003      -0.000
TOVP           0.0004      0.000      2.206      0.027     4.1e-05       0.001
USG           -0.0003      0.000     -1.975      0.048      -0.001    -2.4e-06
DWS            0.0102      0.001      9.770      0.000       0.008       0.012
==============================================================================
Omnibus:                     5941.621   Durbin-Watson:                   1.108
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          7218470.187
```

```
Skew:                          13.590   Prob(JB):                         0.00
Kurtosis:                     232.957   Cond. No.                         979.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

[211]:
```
#removed Age b/c not statistically significant
cols = [  'G', 'GS',
            'STL', 'BLK',
         'PF',  'FTR', 'DRBP','X3PAR',
         'STLP', 'BLKP', 'TOVP', 'USG',  'DWS',
         ]
x_train = train[cols] #all of the variables
y_train = train['DPOY']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   DPOY   R-squared:                       0.131
Model:                            OLS   Adj. R-squared:                  0.128
Method:                 Least Squares   F-statistic:                     37.37
Date:                Wed, 12 May 2021   Prob (F-statistic):           1.25e-88
Time:                        19:42:29   Log-Likelihood:                 5576.4
No. Observations:                3231   AIC:                        -1.112e+04
Df Residuals:                    3217   BIC:                        -1.104e+04
Df Model:                          13
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0003      0.007      0.036      0.971      -0.013       0.014
G             -0.0002   5.69e-05     -3.354      0.001      -0.000   -7.92e-05
GS         -6.983e-05   4.14e-05     -1.689      0.091      -0.000    1.13e-05
STL            0.0122      0.003      3.784      0.000       0.006       0.018
BLK            0.0308      0.003     10.455      0.000       0.025       0.037
PF            -0.0086      0.002     -5.200      0.000      -0.012      -0.005
FTR            0.0259      0.006      4.315      0.000       0.014       0.038
DRBP          -0.0003      0.000     -1.633      0.103      -0.001    5.72e-05
X3PAR          0.0144      0.005      2.933      0.003       0.005       0.024
STLP          -0.0032      0.001     -2.194      0.028      -0.006      -0.000
BLKP          -0.0018      0.001     -2.341      0.019      -0.003      -0.000
```

```
TOVP           0.0004      0.000      2.160      0.031    3.34e-05       0.001
USG           -0.0003      0.000     -1.848      0.065      -0.001    1.92e-05
DWS            0.0101      0.001      9.680      0.000       0.008       0.012
==============================================================================
Omnibus:                     5944.767   Durbin-Watson:                   1.107
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          7240721.920
Skew:                          13.605   Prob(JB):                         0.00
Kurtosis:                     233.313   Cond. No.                         809.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

[212]:
```
#removed DRBP b/c not statistically significant
cols = [  'G', 'GS',
           'STL', 'BLK',
        'PF',  'FTR', 'X3PAR',
        'STLP', 'BLKP', 'TOVP', 'USG',  'DWS',
        ]
x_train = train[cols] #all of the variables
y_train = train['DPOY']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   DPOY   R-squared:                       0.130
Model:                            OLS   Adj. R-squared:                  0.127
Method:                 Least Squares   F-statistic:                     40.24
Date:                Wed, 12 May 2021   Prob (F-statistic):           7.06e-89
Time:                        19:42:44   Log-Likelihood:                 5575.0
No. Observations:                3231   AIC:                        -1.112e+04
Df Residuals:                    3218   BIC:                        -1.105e+04
Df Model:                          12
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.0042      0.006     -0.647      0.518      -0.017       0.008
G             -0.0002   5.63e-05     -3.150      0.002      -0.000    -6.7e-05
GS           -6.8e-05   4.14e-05     -1.645      0.100      -0.000    1.31e-05
STL            0.0128      0.003      3.985      0.000       0.006       0.019
BLK            0.0309      0.003     10.486      0.000       0.025       0.037
```

```
PF            -0.0090      0.002     -5.511      0.000      -0.012      -0.006
FTR            0.0257      0.006      4.286      0.000       0.014       0.038
X3PAR          0.0165      0.005      3.478      0.001       0.007       0.026
STLP          -0.0030      0.001     -2.055      0.040      -0.006      -0.000
BLKP          -0.0022      0.001     -2.838      0.005      -0.004      -0.001
TOVP           0.0004      0.000      2.123      0.034    2.72e-05       0.001
USG           -0.0003      0.000     -1.783      0.075      -0.001    3.02e-05
DWS            0.0095      0.001      9.659      0.000       0.008       0.011
==============================================================================
Omnibus:                     5944.119   Durbin-Watson:                   1.106
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          7231847.314
Skew:                          13.602   Prob(JB):                         0.00
Kurtosis:                     233.170   Cond. No.                         754.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

[213]:
```
#removed GS b/c not statistically significant
cols = [  'G',
           'STL', 'BLK',
        'PF',  'FTR', 'X3PAR',
        'STLP', 'BLKP', 'TOVP', 'USG',  'DWS',
        ]
x_train = train[cols] #all of the variables
y_train = train['DPOY']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   DPOY   R-squared:                       0.130
Model:                            OLS   Adj. R-squared:                  0.127
Method:                 Least Squares   F-statistic:                     43.63
Date:                Wed, 12 May 2021   Prob (F-statistic):           3.92e-89
Time:                        19:43:01   Log-Likelihood:                 5573.7
No. Observations:                3231   AIC:                        -1.112e+04
Df Residuals:                    3219   BIC:                        -1.105e+04
Df Model:                          11
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
```

```
const         -0.0033      0.006     -0.514      0.607      -0.016       0.009
G             -0.0002   5.47e-05     -3.644      0.000      -0.000   -9.21e-05
STL            0.0105      0.003      3.629      0.000       0.005       0.016
BLK            0.0299      0.003     10.366      0.000       0.024       0.036
PF            -0.0093      0.002     -5.754      0.000      -0.012      -0.006
FTR            0.0261      0.006      4.348      0.000       0.014       0.038
X3PAR          0.0169      0.005      3.579      0.000       0.008       0.026
STLP          -0.0023      0.001     -1.674      0.094      -0.005       0.000
BLKP          -0.0020      0.001     -2.617      0.009      -0.003      -0.000
TOVP           0.0004      0.000      2.133      0.033    2.88e-05       0.001
USG           -0.0003      0.000     -1.849      0.065      -0.001     1.9e-05
DWS            0.0093      0.001      9.519      0.000       0.007       0.011
==============================================================================
Omnibus:                     5947.654   Durbin-Watson:                   1.104
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          7260835.205
Skew:                          13.619   Prob(JB):                         0.00
Kurtosis:                     233.634   Cond. No.                         670.
==============================================================================
```

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[214]:
```
#removed STLP b/c not statistically significant
cols = [  'G',
           'STL', 'BLK',
        'PF',  'FTR', 'X3PAR',
         'BLKP', 'TOVP', 'USG',  'DWS',
        ]
x_train = train[cols] #all of the variables
y_train = train['DPOY']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   DPOY   R-squared:                       0.129
Model:                            OLS   Adj. R-squared:                  0.126
Method:                 Least Squares   F-statistic:                     47.68
Date:                Wed, 12 May 2021   Prob (F-statistic):           2.18e-89
Time:                        19:43:15   Log-Likelihood:                 5572.3
No. Observations:                3231   AIC:                        -1.112e+04
Df Residuals:                    3220   BIC:                        -1.106e+04
Df Model:                          10
```

```
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.0069      0.006     -1.141      0.254      -0.019       0.005
G             -0.0002   5.41e-05     -3.427      0.001      -0.000   -7.93e-05
STL            0.0079      0.002      3.235      0.001       0.003       0.013
BLK            0.0301      0.003     10.423      0.000       0.024       0.036
PF            -0.0088      0.002     -5.539      0.000      -0.012      -0.006
FTR            0.0259      0.006      4.314      0.000       0.014       0.038
X3PAR          0.0171      0.005      3.613      0.000       0.008       0.026
BLKP          -0.0019      0.001     -2.574      0.010      -0.003      -0.000
TOVP           0.0003      0.000      1.952      0.051   -1.43e-06       0.001
USG           -0.0003      0.000     -1.748      0.081      -0.001    3.62e-05
DWS            0.0092      0.001      9.462      0.000       0.007       0.011
==============================================================================
Omnibus:                     5948.044   Durbin-Watson:                   1.105
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          7252383.826
Skew:                          13.622   Prob(JB):                         0.00
Kurtosis:                     233.497   Cond. No.                         654.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

[216]:
```
#removed USG b/c not statistically significant
cols = [  'G',
           'STL', 'BLK',
        'PF',  'FTR', 'X3PAR',
         'BLKP', 'TOVP',   'DWS',
        ]
x_train = train[cols] #all of the variables
y_train = train['DPOY']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   DPOY   R-squared:                       0.128
Model:                            OLS   Adj. R-squared:                  0.126
Method:                 Least Squares   F-statistic:                     52.61
Date:                Wed, 12 May 2021   Prob (F-statistic):           1.31e-89
Time:                        19:43:39   Log-Likelihood:                 5570.7
```

```
No. Observations:                3231   AIC:                         -1.112e+04
Df Residuals:                    3221   BIC:                         -1.106e+04
Df Model:                           9
Covariance Type:            nonrobust
================================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
const         -0.0133      0.005     -2.753      0.006      -0.023      -0.004
G             -0.0002    5.4e-05     -3.330      0.001      -0.000   -7.39e-05
STL            0.0071      0.002      2.958      0.003       0.002       0.012
BLK            0.0298      0.003     10.341      0.000       0.024       0.035
PF            -0.0089      0.002     -5.570      0.000      -0.012      -0.006
FTR            0.0256      0.006      4.263      0.000       0.014       0.037
X3PAR          0.0182      0.005      3.882      0.000       0.009       0.027
BLKP          -0.0017      0.001     -2.281      0.023      -0.003      -0.000
TOVP           0.0004      0.000      2.386      0.017     6.9e-05       0.001
DWS            0.0091      0.001      9.331      0.000       0.007       0.011
================================================================================
Omnibus:                     5949.356   Durbin-Watson:                   1.105
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          7255172.866
Skew:                          13.629   Prob(JB):                         0.00
Kurtosis:                     233.540   Cond. No.                         598.
================================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

[217]:
```
import statsmodels.formula.api as smf

DPOYLinear = smf.ols(formula='DPOY ~ G + STL + BLK + PF + FTR + X3PAR+ BLKP +␣
 ↪TOVP + DWS',
                     data=train).fit()
print(DPOYLinear.summary())
```

```
                            OLS Regression Results
================================================================================
Dep. Variable:                   DPOY   R-squared:                       0.128
Model:                            OLS   Adj. R-squared:                  0.126
Method:                 Least Squares   F-statistic:                     52.61
Date:                Wed, 12 May 2021   Prob (F-statistic):           1.31e-89
Time:                        19:44:36   Log-Likelihood:                 5570.7
No. Observations:                3231   AIC:                         -1.112e+04
Df Residuals:                    3221   BIC:                         -1.106e+04
Df Model:                           9
Covariance Type:            nonrobust
================================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
```

```
------------------------------------------------------------------------------
Intercept     -0.0133      0.005     -2.753      0.006      -0.023      -0.004
G             -0.0002    5.4e-05     -3.330      0.001      -0.000   -7.39e-05
STL            0.0071      0.002      2.958      0.003       0.002       0.012
BLK            0.0298      0.003     10.341      0.000       0.024       0.035
PF            -0.0089      0.002     -5.570      0.000      -0.012      -0.006
FTR            0.0256      0.006      4.263      0.000       0.014       0.037
X3PAR          0.0182      0.005      3.882      0.000       0.009       0.027
BLKP          -0.0017      0.001     -2.281      0.023      -0.003      -0.000
TOVP           0.0004      0.000      2.386      0.017     6.9e-05       0.001
DWS            0.0091      0.001      9.331      0.000       0.007       0.011
==============================================================================
Omnibus:                     5949.356   Durbin-Watson:                   1.105
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          7255172.866
Skew:                          13.629   Prob(JB):                         0.00
Kurtosis:                     233.540   Cond. No.                         598.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

[218]:
```
OSR2(DPOYLinear, train, test, 'DPOY')
```

[218]:
```
0.11858601031363625
```

MIP

[184]:
```
#loading in the MIP data
datasetmip = pd.read_csv('mip_stats.csv')
trainmip = datasetmip[datasetmip['Season'] <= 2017]
testmip = datasetmip[datasetmip['Season'] > 2017]
testmip
```

[184]:

	Player	Season	G	GS	MP	FGM	FGA	FGP	X3PM	X3PA	\
11	Aaron Brooks	2018.0	-33.0	1.0	-7.9	-1.0	-2.4	0.003	-0.4	-1.0	
14	Aaron Gordon	2018.0	-22.0	-15.0	4.2	1.6	4.1	-0.020	1.0	2.6	
15	Aaron Gordon	2019.0	20.0	21.0	0.9	-0.5	-1.5	0.015	-0.4	-1.5	
16	Aaron Gordon	2020.0	-16.0	-16.0	-1.3	-0.6	-1.0	-0.012	-0.4	-0.6	
21	Aaron Holiday	2020.0	16.0	33.0	11.6	1.4	3.3	0.013	0.4	0.8	
…	…	…	…	…	…	…	…	…	…	…	
3390	Zach LaVine	2019.0	39.0	38.0	7.2	2.7	3.2	0.084	0.1	0.0	
3391	Zach LaVine	2020.0	-3.0	-2.0	0.3	0.6	2.0	-0.017	1.2	3.0	
3399	Zach Randolph	2018.0	-14.0	52.0	1.1	0.2	-0.3	0.024	0.6	1.2	
3407	Zaza Pachulia	2018.0	-1.0	-13.0	-4.0	-0.1	-0.6	0.030	0.0	0.0	
3408	Zaza Pachulia	2019.0	-1.0	-54.0	-1.2	-0.9	-1.0	-0.124	0.0	0.1	

…	USG	OWS	DWS	WS	WS48	OBPM	DBPM	BPM	VORP	MIP

```
11    …  0.7  0.3 -0.4 -0.2  0.017  -0.2  -0.1 -0.1   0.3  0.000
14    …  4.6 -1.1  0.3 -0.8 -0.005   1.2   0.1  1.3   0.5  0.004
15    … -2.9  0.9  1.3  2.2  0.021  -0.6   0.9  0.3   0.6  0.000
16    … -1.1 -0.4 -1.0 -1.4 -0.006  -0.4  -0.2 -0.6  -0.7  0.000
21    … -3.2  0.3  0.9  1.2 -0.002   0.1  -0.1  0.0   0.1  0.000
…   …  …  …  …  …     …   …   …  …   …    …
3390  …  1.0  1.8  0.6  2.5  0.037   3.2  -0.1  3.1   1.6  0.002
3391  …  1.2  0.1  1.1  1.2  0.029   0.8   0.8  1.6   0.9  0.000
3399  … -1.6 -0.5 -1.2 -1.7 -0.038  -0.5   0.1 -0.5  -0.3  0.000
3407  …  1.7 -0.5 -1.1 -1.4 -0.017   0.4  -0.9 -0.6  -0.3  0.000
3408  … -2.0 -0.7 -0.1 -1.0 -0.033  -2.1   0.9 -1.2  -0.3  0.000

[1020 rows x 48 columns]
```

[93]:
```
#removed age b/c not statistically significant
cols = [ 'G', 'GS',
        'X3PM','X3PP', 'X2PP',
        'FTP', 'ORB',  'STL', 'BLK',
        'PF', 'TS', 'FTR', 'DRBP','X3PAR',
       'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
        'DBPM', ]
x_train = trainmip[cols] #all of the variables
y_train = trainmip['MIP']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    MIP   R-squared:                       0.071
Model:                            OLS   Adj. R-squared:                  0.062
Method:                 Least Squares   F-statistic:                     8.225
Date:                Wed, 12 May 2021   Prob (F-statistic):           7.84e-26
Time:                        19:19:05   Log-Likelihood:                 4147.9
No. Observations:                2390   AIC:                            -8250.
Df Residuals:                    2367   BIC:                            -8117.
Df Model:                          22
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0055      0.001      6.036      0.000       0.004       0.007
G          -7.104e-05   5.54e-05     -1.282      0.200      -0.000    3.76e-05
GS          8.422e-05   4.49e-05      1.877      0.061   -3.76e-06       0.000
```

```
X3PM           0.0142      0.003      5.008      0.000       0.009       0.020
X3PP          -0.0025      0.006     -0.400      0.689      -0.015       0.010
X2PP           0.0146      0.017      0.869      0.385      -0.018       0.048
FTP            0.0052      0.008      0.633      0.527      -0.011       0.021
ORB            0.0141      0.003      5.310      0.000       0.009       0.019
STL            0.0018      0.005      0.387      0.699      -0.007       0.011
BLK            0.0160      0.004      3.587      0.000       0.007       0.025
PF            -0.0095      0.002     -3.824      0.000      -0.014      -0.005
TS            -0.0243      0.015     -1.599      0.110      -0.054       0.005
FTR        -1.122e-05      0.008     -0.001      0.999      -0.015       0.015
DRBP        -5.37e-05      0.000     -0.264      0.792      -0.000       0.000
X3PAR         -0.0114      0.006     -1.766      0.077      -0.024       0.001
ASTP       -7.935e-05      0.000     -0.545      0.586      -0.000       0.000
STLP          -0.0039      0.002     -2.289      0.022      -0.007      -0.001
BLKP          -0.0024      0.001     -2.932      0.003      -0.004      -0.001
TOVP           0.0004      0.000      2.130      0.033     3.2e-05       0.001
USG            0.0009      0.000      3.340      0.001       0.000       0.001
OWS            0.0024      0.001      3.729      0.000       0.001       0.004
DWS           -0.0020      0.001     -1.853      0.064      -0.004       0.000
DBPM           0.0020      0.001      1.886      0.059   -7.83e-05       0.004
==============================================================================
Omnibus:                     4360.865   Durbin-Watson:                   1.954
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          4780912.156
Skew:                          13.266   Prob(JB):                         0.00
Kurtosis:                     220.498   Cond. No.                         587.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

[94]:
```
#removed ftr not statistically significant
cols = [ 'G', 'GS',
        'X3PM','X3PP', 'X2PP',
        'FTP', 'ORB',  'STL', 'BLK',
        'PF', 'TS', 'DRBP','X3PAR',
       'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
        'DBPM', ]
x_train = trainmip[cols] #all of the variables
y_train = trainmip['MIP']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
```

Dep. Variable:	MIP	R-squared:	0.071
Model:	OLS	Adj. R-squared:	0.063
Method:	Least Squares	F-statistic:	8.620
Date:	Wed, 12 May 2021	Prob (F-statistic):	2.62e-26
Time:	19:19:06	Log-Likelihood:	4147.9
No. Observations:	2390	AIC:	-8252.
Df Residuals:	2368	BIC:	-8125.
Df Model:	21		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.0055	0.001	6.041	0.000	0.004	0.007
G	-7.103e-05	5.53e-05	-1.284	0.199	-0.000	3.75e-05
GS	8.422e-05	4.48e-05	1.880	0.060	-3.62e-06	0.000
X3PM	0.0142	0.003	5.009	0.000	0.009	0.020
X3PP	-0.0025	0.006	-0.401	0.689	-0.015	0.010
X2PP	0.0146	0.017	0.872	0.383	-0.018	0.048
FTP	0.0052	0.008	0.633	0.527	-0.011	0.021
ORB	0.0141	0.003	5.320	0.000	0.009	0.019
STL	0.0018	0.005	0.387	0.698	-0.007	0.011
BLK	0.0160	0.004	3.597	0.000	0.007	0.025
PF	-0.0095	0.002	-3.826	0.000	-0.014	-0.005
TS	-0.0243	0.015	-1.629	0.103	-0.054	0.005
DRBP	-5.37e-05	0.000	-0.264	0.792	-0.000	0.000
X3PAR	-0.0114	0.006	-1.890	0.059	-0.023	0.000
ASTP	-7.933e-05	0.000	-0.546	0.585	-0.000	0.000
STLP	-0.0039	0.002	-2.309	0.021	-0.007	-0.001
BLKP	-0.0024	0.001	-2.965	0.003	-0.004	-0.001
TOVP	0.0004	0.000	2.138	0.033	3.34e-05	0.001
USG	0.0009	0.000	3.341	0.001	0.000	0.001
OWS	0.0024	0.001	3.771	0.000	0.001	0.004
DWS	-0.0020	0.001	-1.856	0.064	-0.004	0.000
DBPM	0.0020	0.001	1.896	0.058	-6.72e-05	0.004

Omnibus:	4360.868	Durbin-Watson:	1.954
Prob(Omnibus):	0.000	Jarque-Bera (JB):	4780931.770
Skew:	13.266	Prob(JB):	0.00
Kurtosis:	220.498	Cond. No.	578.

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[95]:
```
#removed DRBP not statistically significant
cols = [ 'G', 'GS',
        'X3PM','X3PP', 'X2PP',
        'FTP', 'ORB',  'STL', 'BLK',
        'PF', 'TS','X3PAR',
       'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
        'DBPM', ]
x_train = trainmip[cols] #all of the variables
y_train = trainmip['MIP']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

OLS Regression Results

Dep. Variable:	MIP	R-squared:	0.071
Model:	OLS	Adj. R-squared:	0.063
Method:	Least Squares	F-statistic:	9.051
Date:	Wed, 12 May 2021	Prob (F-statistic):	8.83e-27
Time:	19:19:06	Log-Likelihood:	4147.9
No. Observations:	2390	AIC:	-8254.
Df Residuals:	2369	BIC:	-8132.
Df Model:	20		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.0054	0.001	6.038	0.000	0.004	0.007
G	-7.006e-05	5.52e-05	-1.269	0.205	-0.000	3.82e-05
GS	8.427e-05	4.48e-05	1.882	0.060	-3.56e-06	0.000
X3PM	0.0142	0.003	5.005	0.000	0.009	0.020
X3PP	-0.0024	0.006	-0.393	0.694	-0.015	0.010
X2PP	0.0146	0.017	0.869	0.385	-0.018	0.047
FTP	0.0052	0.008	0.635	0.525	-0.011	0.021
ORB	0.0140	0.003	5.316	0.000	0.009	0.019
STL	0.0018	0.005	0.382	0.702	-0.007	0.011
BLK	0.0161	0.004	3.630	0.000	0.007	0.025
PF	-0.0095	0.002	-3.821	0.000	-0.014	-0.005
TS	-0.0240	0.015	-1.614	0.107	-0.053	0.005
X3PAR	-0.0109	0.006	-1.895	0.058	-0.022	0.000
ASTP	-6.748e-05	0.000	-0.489	0.625	-0.000	0.000
STLP	-0.0038	0.002	-2.306	0.021	-0.007	-0.001
BLKP	-0.0024	0.001	-3.013	0.003	-0.004	-0.001
TOVP	0.0004	0.000	2.139	0.033	3.26e-05	0.001

```
USG            0.0009      0.000      3.352      0.001       0.000       0.001
OWS            0.0024      0.001      3.762      0.000       0.001       0.004
DWS           -0.0021      0.001     -1.897      0.058      -0.004    6.94e-05
DBPM           0.0019      0.001      1.895      0.058    -6.6e-05       0.004
==============================================================================
Omnibus:                     4361.123   Durbin-Watson:                   1.954
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          4782605.590
Skew:                          13.267   Prob(JB):                         0.00
Kurtosis:                     220.536   Cond. No.                         577.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

[96]:
```
#removed STL not statistically significant
cols = [ 'G', 'GS',
        'X3PM','X3PP', 'X2PP',
        'FTP', 'ORB', 'BLK',
        'PF', 'TS','X3PAR',
       'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
        'DBPM', ]
x_train = trainmip[cols] #all of the variables
y_train = trainmip['MIP']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    MIP   R-squared:                       0.071
Model:                            OLS   Adj. R-squared:                  0.063
Method:                 Least Squares   F-statistic:                     9.523
Date:                Wed, 12 May 2021   Prob (F-statistic):           3.00e-27
Time:                        19:19:06   Log-Likelihood:                4147.8
No. Observations:                2390   AIC:                            -8256.
Df Residuals:                    2370   BIC:                            -8140.
Df Model:                          19
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0054      0.001      6.041      0.000       0.004       0.007
G          -7.135e-05   5.51e-05     -1.295      0.195      -0.000    3.67e-05
GS          8.759e-05   4.39e-05      1.994      0.046    1.46e-06       0.000
```

```
X3PM           0.0144      0.003      5.199      0.000       0.009       0.020
X3PP          -0.0025      0.006     -0.403      0.687      -0.015       0.010
X2PP           0.0148      0.017      0.883      0.377      -0.018       0.048
FTP            0.0053      0.008      0.648      0.517      -0.011       0.021
ORB            0.0141      0.003      5.390      0.000       0.009       0.019
BLK            0.0162      0.004      3.644      0.000       0.007       0.025
PF            -0.0092      0.002     -3.849      0.000      -0.014      -0.005
TS            -0.0237      0.015     -1.597      0.110      -0.053       0.005
X3PAR         -0.0110      0.006     -1.916      0.056      -0.022       0.000
ASTP       -6.639e-05      0.000     -0.481      0.631      -0.000       0.000
STLP          -0.0036      0.002     -2.346      0.019      -0.007      -0.001
BLKP          -0.0024      0.001     -2.993      0.003      -0.004      -0.001
TOVP           0.0004      0.000      2.113      0.035    2.78e-05       0.001
USG            0.0009      0.000      3.340      0.001       0.000       0.001
OWS            0.0024      0.001      3.746      0.000       0.001       0.004
DWS           -0.0020      0.001     -1.863      0.063      -0.004       0.000
DBPM           0.0018      0.001      1.863      0.063   -9.65e-05       0.004
==============================================================================
Omnibus:                     4360.960   Durbin-Watson:                   1.954
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          4782892.258
Skew:                          13.266   Prob(JB):                         0.00
Kurtosis:                     220.543   Cond. No.                         577.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

[97]:
```
#removed X3PP not statistically significant
cols = [ 'G', 'GS',
        'X3PM', 'X2PP',
        'FTP', 'ORB', 'BLK',
        'PF', 'TS','X3PAR',
       'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
        'DBPM', ]
x_train = trainmip[cols] #all of the variables
y_train = trainmip['MIP']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    MIP   R-squared:                       0.071
Model:                            OLS   Adj. R-squared:                  0.064
```

```
Method:                 Least Squares   F-statistic:                     10.05
Date:                Wed, 12 May 2021   Prob (F-statistic):           9.98e-28
Time:                        19:19:06   Log-Likelihood:                 4147.8
No. Observations:                2390   AIC:                            -8258.
Df Residuals:                    2371   BIC:                            -8148.
Df Model:                          18
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0054      0.001      6.038      0.000       0.004       0.007
G          -7.186e-05   5.51e-05     -1.305      0.192      -0.000    3.61e-05
GS          8.789e-05   4.39e-05      2.002      0.045    1.79e-06       0.000
X3PM           0.0142      0.003      5.218      0.000       0.009       0.020
X2PP           0.0150      0.017      0.896      0.370      -0.018       0.048
FTP            0.0052      0.008      0.630      0.529      -0.011       0.021
ORB            0.0142      0.003      5.400      0.000       0.009       0.019
BLK            0.0161      0.004      3.635      0.000       0.007       0.025
PF            -0.0092      0.002     -3.841      0.000      -0.014      -0.005
TS            -0.0244      0.015     -1.656      0.098      -0.053       0.005
X3PAR         -0.0110      0.006     -1.925      0.054      -0.022       0.000
ASTP       -6.499e-05      0.000     -0.471      0.638      -0.000       0.000
STLP          -0.0036      0.002     -2.361      0.018      -0.007      -0.001
BLKP          -0.0024      0.001     -2.992      0.003      -0.004      -0.001
TOVP           0.0004      0.000      2.106      0.035    2.65e-05       0.001
USG            0.0009      0.000      3.349      0.001       0.000       0.001
OWS            0.0024      0.001      3.755      0.000       0.001       0.004
DWS           -0.0020      0.001     -1.863      0.063      -0.004       0.000
DBPM           0.0019      0.001      1.891      0.059   -6.89e-05       0.004
==============================================================================
Omnibus:                     4360.867   Durbin-Watson:                   1.954
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          4781927.099
Skew:                          13.266   Prob(JB):                         0.00
Kurtosis:                     220.521   Cond. No.                         576.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

[98]:
```
#removed ASTP not statistically significant
cols = [ 'G', 'GS',
         'X3PM', 'X2PP',
         'FTP', 'ORB', 'BLK',
         'PF', 'TS','X3PAR',
        'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
         'DBPM', ]
```

```
x_train = trainmip[cols] #all of the variables
y_train = trainmip['MIP']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    MIP   R-squared:                       0.071
Model:                            OLS   Adj. R-squared:                  0.064
Method:                 Least Squares   F-statistic:                     10.63
Date:                Wed, 12 May 2021   Prob (F-statistic):           3.32e-28
Time:                        19:19:06   Log-Likelihood:                 4147.6
No. Observations:                2390   AIC:                            -8259.
Df Residuals:                    2372   BIC:                            -8155.
Df Model:                          17
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0054      0.001      6.029      0.000       0.004       0.007
G          -7.142e-05   5.51e-05     -1.297      0.195      -0.000    3.65e-05
GS           8.68e-05   4.38e-05      1.980      0.048    8.27e-07       0.000
X3PM           0.0142      0.003      5.228      0.000       0.009       0.020
X2PP           0.0153      0.017      0.914      0.361      -0.018       0.048
FTP            0.0051      0.008      0.621      0.535      -0.011       0.021
ORB            0.0142      0.003      5.427      0.000       0.009       0.019
BLK            0.0160      0.004      3.612      0.000       0.007       0.025
PF            -0.0093      0.002     -3.877      0.000      -0.014      -0.005
TS            -0.0236      0.015     -1.612      0.107      -0.052       0.005
X3PAR         -0.0113      0.006     -1.988      0.047      -0.023      -0.000
STLP          -0.0036      0.002     -2.355      0.019      -0.007      -0.001
BLKP          -0.0023      0.001     -2.962      0.003      -0.004      -0.001
TOVP           0.0004      0.000      2.070      0.039    1.86e-05       0.001
USG            0.0008      0.000      3.558      0.000       0.000       0.001
OWS            0.0024      0.001      3.728      0.000       0.001       0.004
DWS           -0.0019      0.001     -1.813      0.070      -0.004       0.000
DBPM           0.0017      0.001      1.834      0.067      -0.000       0.004
==============================================================================
Omnibus:                     4361.786   Durbin-Watson:                   1.954
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          4788595.447
Skew:                          13.271   Prob(JB):                         0.00
Kurtosis:                     220.674   Cond. No.                         575.
==============================================================================
```

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[99]:
```
#removed FTP not statistically significant
cols = [ 'G', 'GS',
        'X3PM', 'X2PP',
         'ORB', 'BLK',
        'PF', 'TS','X3PAR',
       'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
        'DBPM', ]
x_train = trainmip[cols] #all of the variables
y_train = trainmip['MIP']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

OLS Regression Results

Dep. Variable:	MIP	R-squared:	0.071
Model:	OLS	Adj. R-squared:	0.064
Method:	Least Squares	F-statistic:	11.27
Date:	Wed, 12 May 2021	Prob (F-statistic):	1.15e-28
Time:	19:19:06	Log-Likelihood:	4147.5
No. Observations:	2390	AIC:	-8261.
Df Residuals:	2373	BIC:	-8163.
Df Model:	16		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.0054	0.001	6.030	0.000	0.004	0.007
G	-6.906e-05	5.49e-05	-1.258	0.209	-0.000	3.86e-05
GS	8.657e-05	4.38e-05	1.975	0.048	6.19e-07	0.000
X3PM	0.0143	0.003	5.261	0.000	0.009	0.020
X2PP	0.0147	0.017	0.879	0.380	-0.018	0.047
ORB	0.0143	0.003	5.461	0.000	0.009	0.019
BLK	0.0161	0.004	3.633	0.000	0.007	0.025
PF	-0.0092	0.002	-3.858	0.000	-0.014	-0.005
TS	-0.0224	0.015	-1.543	0.123	-0.051	0.006
X3PAR	-0.0114	0.006	-2.005	0.045	-0.023	-0.000
STLP	-0.0036	0.002	-2.357	0.019	-0.007	-0.001
BLKP	-0.0023	0.001	-2.981	0.003	-0.004	-0.001
TOVP	0.0004	0.000	2.080	0.038	2.04e-05	0.001

```
USG            0.0008      0.000      3.589      0.000       0.000       0.001
OWS            0.0024      0.001      3.738      0.000       0.001       0.004
DWS           -0.0019      0.001     -1.835      0.067      -0.004       0.000
DBPM           0.0017      0.001      1.854      0.064      -0.000       0.004
==============================================================================
Omnibus:                     4362.239   Durbin-Watson:                   1.954
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          4792217.802
Skew:                          13.274   Prob(JB):                         0.00
Kurtosis:                     220.756   Cond. No.                         571.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

[100]:
```
#removed X2PP not statistically significant
cols = [ 'G', 'GS',
        'X3PM',
         'ORB', 'BLK',
        'PF', 'TS','X3PAR',
       'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
        'DBPM', ]
x_train = trainmip[cols] #all of the variables
y_train = trainmip['MIP']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    MIP   R-squared:                       0.070
Model:                            OLS   Adj. R-squared:                  0.064
Method:                 Least Squares   F-statistic:                     11.97
Date:                Wed, 12 May 2021   Prob (F-statistic):           4.66e-29
Time:                        19:19:06   Log-Likelihood:                4147.1
No. Observations:                2390   AIC:                            -8262.
Df Residuals:                    2374   BIC:                            -8170.
Df Model:                          15
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0054      0.001      6.053      0.000       0.004       0.007
G          -6.433e-05   5.46e-05     -1.177      0.239      -0.000    4.28e-05
GS           8.67e-05   4.38e-05      1.978      0.048    7.52e-07       0.000
```

```
X3PM           0.0141      0.003      5.212      0.000       0.009       0.019
ORB            0.0143      0.003      5.474      0.000       0.009       0.019
BLK            0.0164      0.004      3.710      0.000       0.008       0.025
PF            -0.0092      0.002     -3.847      0.000      -0.014      -0.005
TS            -0.0184      0.014     -1.336      0.182      -0.045       0.009
X3PAR         -0.0113      0.006     -1.985      0.047      -0.022      -0.000
STLP          -0.0036      0.002     -2.347      0.019      -0.007      -0.001
BLKP          -0.0023      0.001     -2.999      0.003      -0.004      -0.001
TOVP           0.0004      0.000      2.079      0.038    2.03e-05       0.001
USG            0.0008      0.000      3.561      0.000       0.000       0.001
OWS            0.0024      0.001      3.828      0.000       0.001       0.004
DWS           -0.0020      0.001     -1.897      0.058      -0.004    6.79e-05
DBPM           0.0017      0.001      1.841      0.066      -0.000       0.004
==============================================================================
Omnibus:                     4362.208   Durbin-Watson:                   1.955
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          4790921.468
Skew:                          13.274   Prob(JB):                         0.00
Kurtosis:                     220.727   Cond. No.                         436.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

[101]:
```
#removed G not statistically significant
cols = [ 'GS',
        'X3PM',
         'ORB', 'BLK',
        'PF', 'TS','X3PAR',
       'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
        'DBPM', ]
x_train = trainmip[cols] #all of the variables
y_train = trainmip['MIP']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    MIP   R-squared:                       0.070
Model:                            OLS   Adj. R-squared:                  0.064
Method:                 Least Squares   F-statistic:                     12.73
Date:                Wed, 12 May 2021   Prob (F-statistic):           2.42e-29
Time:                        19:19:06   Log-Likelihood:                4146.4
No. Observations:                2390   AIC:                            -8263.
```

```
Df Residuals:                     2375   BIC:                            -8176.
Df Model:                           14
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0055      0.001      6.136      0.000       0.004       0.007
GS          7.397e-05   4.25e-05      1.741      0.082   -9.33e-06       0.000
X3PM           0.0140      0.003      5.168      0.000       0.009       0.019
ORB            0.0144      0.003      5.500      0.000       0.009       0.020
BLK            0.0163      0.004      3.685      0.000       0.008       0.025
PF            -0.0094      0.002     -3.955      0.000      -0.014      -0.005
TS            -0.0184      0.014     -1.333      0.183      -0.045       0.009
X3PAR         -0.0114      0.006     -2.008      0.045      -0.023      -0.000
STLP          -0.0036      0.002     -2.381      0.017      -0.007      -0.001
BLKP          -0.0023      0.001     -3.046      0.002      -0.004      -0.001
TOVP           0.0004      0.000      2.098      0.036    2.35e-05       0.001
USG            0.0008      0.000      3.720      0.000       0.000       0.001
OWS            0.0024      0.001      3.752      0.000       0.001       0.004
DWS           -0.0024      0.001     -2.328      0.020      -0.004      -0.000
DBPM           0.0019      0.001      1.971      0.049    9.87e-06       0.004
==============================================================================
Omnibus:                     4361.781   Durbin-Watson:                   1.952
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          4785748.987
Skew:                          13.272   Prob(JB):                         0.00
Kurtosis:                     220.608   Cond. No.                         399.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

[102]:
```python
#removed TS not statistically significant
cols = [ 'GS',
        'X3PM',
         'ORB', 'BLK',
        'PF', 'X3PAR',
       'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
        'DBPM', ]
x_train = trainmip[cols] #all of the variables
y_train = trainmip['MIP']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    MIP   R-squared:                       0.069
Model:                            OLS   Adj. R-squared:                  0.064
Method:                 Least Squares   F-statistic:                     13.56
Date:                Wed, 12 May 2021   Prob (F-statistic):           1.47e-29
Time:                        19:19:06   Log-Likelihood:                 4145.5
No. Observations:                2390   AIC:                            -8263.
Df Residuals:                    2376   BIC:                            -8182.
Df Model:                          13
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0055      0.001      6.160      0.000       0.004       0.007
GS           7.699e-05   4.24e-05      1.815      0.070   -6.21e-06       0.000
X3PM           0.0137      0.003      5.067      0.000       0.008       0.019
ORB            0.0144      0.003      5.490      0.000       0.009       0.019
BLK            0.0163      0.004      3.685      0.000       0.008       0.025
PF            -0.0096      0.002     -4.045      0.000      -0.014      -0.005
X3PAR         -0.0117      0.006     -2.047      0.041      -0.023      -0.000
STLP          -0.0032      0.001     -2.136      0.033      -0.006      -0.000
BLKP          -0.0023      0.001     -2.989      0.003      -0.004      -0.001
TOVP           0.0004      0.000      2.057      0.040    1.65e-05       0.001
USG            0.0008      0.000      3.746      0.000       0.000       0.001
OWS            0.0020      0.001      3.522      0.000       0.001       0.003
DWS           -0.0022      0.001     -2.210      0.027      -0.004      -0.000
DBPM           0.0014      0.001      1.626      0.104      -0.000       0.003
==============================================================================
Omnibus:                     4362.876   Durbin-Watson:                   1.951
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          4791400.123
Skew:                          13.279   Prob(JB):                         0.00
Kurtosis:                     220.737   Cond. No.                         170.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

[103]:
```
#removed DBPM not statistically significant
cols = [ 'GS',
        'X3PM',
         'ORB', 'BLK',
        'PF', 'X3PAR',
       'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
         ]
x_train = trainmip[cols] #all of the variables
```

```
y_train = trainmip['MIP']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    MIP   R-squared:                       0.068
Model:                            OLS   Adj. R-squared:                  0.063
Method:                 Least Squares   F-statistic:                     14.46
Date:                Wed, 12 May 2021   Prob (F-statistic):           1.30e-29
Time:                        19:19:06   Log-Likelihood:                 4144.1
No. Observations:                2390   AIC:                            -8262.
Df Residuals:                    2377   BIC:                            -8187.
Df Model:                          12
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0055      0.001      6.150      0.000       0.004       0.007
GS          6.811e-05   4.21e-05      1.618      0.106   -1.44e-05       0.000
X3PM           0.0135      0.003      4.997      0.000       0.008       0.019
ORB            0.0140      0.003      5.358      0.000       0.009       0.019
BLK            0.0163      0.004      3.693      0.000       0.008       0.025
PF            -0.0094      0.002     -3.956      0.000      -0.014      -0.005
X3PAR         -0.0104      0.006     -1.835      0.067      -0.021       0.001
STLP          -0.0019      0.001     -1.498      0.134      -0.004       0.001
BLKP          -0.0017      0.001     -2.514      0.012      -0.003      -0.000
TOVP           0.0003      0.000      1.896      0.058    -1.1e-05       0.001
USG            0.0007      0.000      3.375      0.001       0.000       0.001
OWS            0.0020      0.001      3.592      0.000       0.001       0.003
DWS           -0.0016      0.001     -1.732      0.083      -0.003       0.000
==============================================================================
Omnibus:                     4365.336   Durbin-Watson:                   1.949
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          4805743.639
Skew:                          13.294   Prob(JB):                         0.00
Kurtosis:                     221.063   Cond. No.                         169.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

[104]:
```
#removed STLP not statistically significant
cols = [ 'GS',
        'X3PM',
         'ORB', 'BLK',
        'PF', 'X3PAR',
       'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
         ]
x_train = trainmip[cols] #all of the variables
y_train = trainmip['MIP']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    MIP   R-squared:                       0.067
Model:                            OLS   Adj. R-squared:                  0.063
Method:                 Least Squares   F-statistic:                     15.57
Date:                Wed, 12 May 2021   Prob (F-statistic):           9.20e-30
Time:                        19:19:06   Log-Likelihood:                4143.0
No. Observations:                2390   AIC:                            -8262.
Df Residuals:                    2378   BIC:                            -8193.
Df Model:                          11
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0055      0.001      6.117      0.000       0.004       0.007
GS           7.028e-05   4.21e-05      1.670      0.095   -1.22e-05       0.000
X3PM           0.0135      0.003      5.002      0.000       0.008       0.019
ORB            0.0140      0.003      5.356      0.000       0.009       0.019
BLK            0.0159      0.004      3.610      0.000       0.007       0.025
PF            -0.0094      0.002     -3.944      0.000      -0.014      -0.005
X3PAR         -0.0103      0.006     -1.825      0.068      -0.021       0.001
BLKP          -0.0016      0.001     -2.365      0.018      -0.003      -0.000
TOVP           0.0003      0.000      1.848      0.065   -1.93e-05       0.001
USG            0.0007      0.000      3.336      0.001       0.000       0.001
OWS            0.0021      0.001      3.737      0.000       0.001       0.003
DWS           -0.0018      0.001     -1.990      0.047      -0.004   -2.69e-05
==============================================================================
Omnibus:                     4365.733   Durbin-Watson:                   1.951
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          4802558.752
Skew:                          13.298   Prob(JB):                         0.00
Kurtosis:                     220.989   Cond. No.                         169.
```

```
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

[105]:
```
#removed GS not statistically significant
cols = [
        'X3PM',
         'ORB', 'BLK',
        'PF', 'X3PAR',
       'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
         ]
x_train = trainmip[cols] #all of the variables
y_train = trainmip['MIP']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    MIP   R-squared:                       0.066
Model:                            OLS   Adj. R-squared:                  0.062
Method:                 Least Squares   F-statistic:                     16.83
Date:                Wed, 12 May 2021   Prob (F-statistic):           8.15e-30
Time:                        19:19:06   Log-Likelihood:                 4141.6
No. Observations:                2390   AIC:                            -8261.
Df Residuals:                    2379   BIC:                            -8198.
Df Model:                          10
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0055      0.001      6.087      0.000       0.004       0.007
X3PM           0.0145      0.003      5.480      0.000       0.009       0.020
ORB            0.0148      0.003      5.766      0.000       0.010       0.020
BLK            0.0162      0.004      3.669      0.000       0.008       0.025
PF            -0.0086      0.002     -3.689      0.000      -0.013      -0.004
X3PAR         -0.0105      0.006     -1.854      0.064      -0.022       0.001
BLKP          -0.0017      0.001     -2.476      0.013      -0.003      -0.000
TOVP           0.0003      0.000      1.883      0.060   -1.32e-05       0.001
USG            0.0007      0.000      3.210      0.001       0.000       0.001
OWS            0.0022      0.001      3.906      0.000       0.001       0.003
DWS           -0.0016      0.001     -1.785      0.074      -0.003       0.000
==============================================================================
```

```
Omnibus:                       4366.883   Durbin-Watson:                   1.950
Prob(Omnibus):                    0.000   Jarque-Bera (JB):          4806436.961
Skew:                            13.305   Prob(JB):                         0.00
Kurtosis:                       221.076   Cond. No.                         37.4
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

[106]:
```
#removed DWS not statistically significant
cols = [
        'X3PM',
         'ORB', 'BLK',
        'PF', 'X3PAR',
       'BLKP', 'TOVP', 'USG', 'OWS',
         ]
x_train = trainmip[cols] #all of the variables
y_train = trainmip['MIP']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    MIP   R-squared:                       0.065
Model:                            OLS   Adj. R-squared:                  0.061
Method:                 Least Squares   F-statistic:                     18.33
Date:                Wed, 12 May 2021   Prob (F-statistic):           8.38e-30
Time:                        19:19:07   Log-Likelihood:                 4140.0
No. Observations:                2390   AIC:                            -8260.
Df Residuals:                    2380   BIC:                            -8202.
Df Model:                           9
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0054      0.001      6.054      0.000       0.004       0.007
X3PM           0.0145      0.003      5.507      0.000       0.009       0.020
ORB            0.0150      0.003      5.876      0.000       0.010       0.020
BLK            0.0151      0.004      3.450      0.001       0.006       0.024
PF            -0.0091      0.002     -3.910      0.000      -0.014      -0.005
X3PAR         -0.0095      0.006     -1.691      0.091      -0.021       0.002
BLKP          -0.0017      0.001     -2.578      0.010      -0.003      -0.000
TOVP           0.0003      0.000      1.849      0.065   -1.91e-05       0.001
```

```
USG            0.0006      0.000      3.160      0.002       0.000       0.001
OWS            0.0017      0.000      3.477      0.001       0.001       0.003
==============================================================================
Omnibus:                     4371.799   Durbin-Watson:                   1.949
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          4837827.821
Skew:                          13.336   Prob(JB):                         0.00
Kurtosis:                     221.790   Cond. No.                         37.1
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

[107]:
```
#removed X3PAR not statistically significant
cols = [
        'X3PM',
         'ORB', 'BLK',
        'PF',
       'BLKP', 'TOVP', 'USG', 'OWS',
         ]
x_train = trainmip[cols] #all of the variables
y_train = trainmip['MIP']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    MIP   R-squared:                       0.064
Model:                            OLS   Adj. R-squared:                  0.061
Method:                 Least Squares   F-statistic:                     20.25
Date:                Wed, 12 May 2021   Prob (F-statistic):           7.05e-30
Time:                        19:19:07   Log-Likelihood:                 4138.6
No. Observations:                2390   AIC:                            -8259.
Df Residuals:                    2381   BIC:                            -8207.
Df Model:                           8
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0053      0.001      5.947      0.000       0.004       0.007
X3PM           0.0135      0.003      5.253      0.000       0.008       0.018
ORB            0.0151      0.003      5.916      0.000       0.010       0.020
BLK            0.0138      0.004      3.212      0.001       0.005       0.022
PF            -0.0086      0.002     -3.724      0.000      -0.013      -0.004
```

```
BLKP          -0.0012      0.001     -2.061      0.039      -0.002   -6.04e-05
TOVP           0.0004      0.000      2.190      0.029    3.84e-05       0.001
USG            0.0007      0.000      3.353      0.001       0.000       0.001
OWS            0.0018      0.000      3.736      0.000       0.001       0.003
==============================================================================
Omnibus:                     4374.607   Durbin-Watson:                   1.949
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          4852958.569
Skew:                          13.355   Prob(JB):                         0.00
Kurtosis:                     222.133   Cond. No.                         28.4
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

[108]:
```
import statsmodels.formula.api as smf

MIPLinear = smf.ols(formula='MIP ~ X3PM + ORB + BLK + PF + BLKP + TOVP + USG +␣
 ↪OWS',
                    data=trainmip).fit()
print(MIPLinear.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    MIP   R-squared:                       0.064
Model:                            OLS   Adj. R-squared:                  0.061
Method:                 Least Squares   F-statistic:                     20.25
Date:                Wed, 12 May 2021   Prob (F-statistic):           7.05e-30
Time:                        19:19:07   Log-Likelihood:                 4138.6
No. Observations:                2390   AIC:                            -8259.
Df Residuals:                    2381   BIC:                            -8207.
Df Model:                           8
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      0.0053      0.001      5.947      0.000       0.004       0.007
X3PM           0.0135      0.003      5.253      0.000       0.008       0.018
ORB            0.0151      0.003      5.916      0.000       0.010       0.020
BLK            0.0138      0.004      3.212      0.001       0.005       0.022
PF            -0.0086      0.002     -3.724      0.000      -0.013      -0.004
BLKP          -0.0012      0.001     -2.061      0.039      -0.002   -6.04e-05
TOVP           0.0004      0.000      2.190      0.029    3.84e-05       0.001
USG            0.0007      0.000      3.353      0.001       0.000       0.001
OWS            0.0018      0.000      3.736      0.000       0.001       0.003
==============================================================================
Omnibus:                     4374.607   Durbin-Watson:                   1.949
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          4852958.569
```

```
Skew:                          13.355   Prob(JB):                         0.00
Kurtosis:                     222.133   Cond. No.                         28.4
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

[109]:
```
OSR2(MIPLinear, trainmip, testmip, 'MIP')
```

[109]:
```
0.03547910086759887
```

Rookie of the Year

[185]:
```
#loading in ROY data
datasetroy = pd.read_csv('roydata142.csv')
trainroy = datasetroy[datasetroy['Season'] <= 2017]
testroy = datasetroy[datasetroy['Season'] > 2017]
testroy
```

[185]:

	Unnamed: 0	Player	Season	Pos	Age	Tm	G	GS	MP	FGM	…	USG	OWS	DWS	WS	WS48	OBPM	DBPM	BPM	VORP	ROY
4	5	Aaron Holiday	2019	PG	22	IND	50	0	12.9	2.1	…	21.9	0.1	0.8	0.9	0.065	-1.7	0.1	-1.6	0.1	0.00
5	6	Abdel Nader	2018	SF	24	BOS	48	1	10.9	1.0	…	17.1	-0.9	0.8	-0.1	-0.014	-5.5	0.2	-5.3	-0.4	0.00
6	7	Admiral Schofield	2020	PF	22	WAS	33	2	11.2	1.1	…	11.9	0.1	0.1	0.1	0.017	-4.3	-1.2	-5.5	-0.3	0.00
11	12	Alex Caruso	2018	PG	23	LAL	37	7	15.2	1.4	…	12.9	-0.1	0.6	0.5	0.041	-3.8	0.9	-2.9	-0.1	0.00
14	15	Allonzo Trier	2019	SG	23	NYK	64	3	22.8	3.6	…	21.5	0.4	0.5	0.9	0.030	-1.6	-1.5	-3.1	-0.4	0.00
..	…	…	…	..	…	..	..	…	…		…	…	…	…	…	…	…	…	…	…	…
555	556	Vincent Poirier	2020	C	26	BOS	22	0	5.9	0.8	…	15.4	0.1	0.2	0.3	0.113	-4.8	1.0	-3.8	-0.1	0.00
559	560	Wenyen Gabriel	2020	PF	22	TOT	30	1	7.8	0.7	…	13.4	0.0	0.2	0.2	0.048	-3.7	-0.1	-3.8	-0.1	0.00
560	561	Wesley Iwundu	2018	SF	23	ORL	62	12	16.5	1.5	…	11.6	0.0	0.7	0.7	0.032	-4.2	-0.1	-4.3	-0.6	0.00
570	571	Zach Collins	2018	C	20	POR	66	1	15.8	1.7	…	15.4	-0.7	1.3	0.6	0.030	-4.5	0.2	-4.3	-0.6	0.00
572	573	Zion Williamson	2020	PF	19	NOP	24	24	27.8	8.8	…	30.5	1.5	0.4	2.0	0.141	3.5	-1.4	2.1	0.7	0.28

[157 rows x 52 columns]

[111]:
```
#removed age b/c not relevant to predicting our output variable
cols = [ 'G', 'GS',
        'X3PM','X3PP', 'X2PP',
        'FTP', 'ORB',  'STL', 'BLK',
        'PF', 'TS', 'FTR', 'DRBP','X3PAR',
       'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
        'DBPM', ]
x_train = trainroy[cols] #all of the variables
y_train = trainroy['ROY']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    ROY   R-squared:                       0.434
Model:                            OLS   Adj. R-squared:                  0.403
Method:                 Least Squares   F-statistic:                     13.71
Date:                Wed, 12 May 2021   Prob (F-statistic):           1.96e-36
Time:                        19:19:07   Log-Likelihood:                321.54
No. Observations:                 416   AIC:                            -597.1
Df Residuals:                     393   BIC:                            -504.4
Df Model:                          22
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.0296      0.087     -0.342      0.733      -0.200       0.141
G             -0.0008      0.000     -1.840      0.066      -0.002    5.62e-05
GS             0.0020      0.000      4.580      0.000       0.001       0.003
X3PM           0.0407      0.026      1.580      0.115      -0.010       0.091
X3PP           0.0079      0.051      0.155      0.877      -0.092       0.107
X2PP           0.3159      0.126      2.505      0.013       0.068       0.564
FTP            0.1416      0.055      2.586      0.010       0.034       0.249
ORB           -0.0074      0.021     -0.360      0.719      -0.048       0.033
STL            0.1064      0.031      3.387      0.001       0.045       0.168
BLK            0.0862      0.037      2.342      0.020       0.014       0.159
PF            -0.0185      0.015     -1.224      0.222      -0.048       0.011
TS            -0.7847      0.161     -4.875      0.000      -1.101      -0.468
FTR            0.1299      0.044      2.936      0.004       0.043       0.217
DRBP           0.0029      0.002      1.691      0.092      -0.000       0.006
X3PAR         -0.0396      0.053     -0.752      0.453      -0.143       0.064
ASTP           0.0015      0.001      1.345      0.179      -0.001       0.004
STLP          -0.0383      0.013     -2.910      0.004      -0.064      -0.012
```

```
BLKP          -0.0128      0.008     -1.611      0.108      -0.029       0.003
TOVP           0.0014      0.002      0.892      0.373      -0.002       0.004
USG            0.0075      0.002      3.920      0.000       0.004       0.011
OWS            0.0310      0.007      4.570      0.000       0.018       0.044
DWS           -0.0088      0.014     -0.632      0.528      -0.036       0.019
DBPM           0.0194      0.010      1.949      0.052      -0.000       0.039
==============================================================================
Omnibus:                      278.849   Durbin-Watson:                   2.139
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             3489.838
Skew:                           2.719   Prob(JB):                         0.00
Kurtosis:                      16.106   Cond. No.                     2.19e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 2.19e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[112]:
```
#removed X3PP b/c not statistically significant
cols = [ 'G', 'GS',
        'X3PM', 'X2PP',
        'FTP', 'ORB',  'STL', 'BLK',
        'PF', 'TS', 'FTR', 'DRBP','X3PAR',
       'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
        'DBPM', ]
x_train = trainroy[cols] #all of the variables
y_train = trainroy['ROY']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    ROY   R-squared:                       0.434
Model:                            OLS   Adj. R-squared:                  0.404
Method:                 Least Squares   F-statistic:                     14.40
Date:                Wed, 12 May 2021   Prob (F-statistic):           5.13e-37
Time:                        19:19:07   Log-Likelihood:                321.53
No. Observations:                 416   AIC:                            -599.1
Df Residuals:                     394   BIC:                            -510.4
Df Model:                          21
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
```

```
------------------------------------------------------------------------------
const         -0.0296      0.087     -0.341      0.733      -0.200       0.141
G             -0.0008      0.000     -1.836      0.067      -0.002    5.78e-05
GS             0.0020      0.000      4.588      0.000       0.001       0.003
X3PM           0.0416      0.025      1.654      0.099      -0.008       0.091
X2PP           0.3120      0.123      2.527      0.012       0.069       0.555
FTP            0.1416      0.055      2.589      0.010       0.034       0.249
ORB           -0.0077      0.020     -0.378      0.706      -0.048       0.032
STL            0.1064      0.031      3.395      0.001       0.045       0.168
BLK            0.0864      0.037      2.350      0.019       0.014       0.159
PF            -0.0187      0.015     -1.239      0.216      -0.048       0.011
TS            -0.7807      0.159     -4.919      0.000      -1.093      -0.469
FTR            0.1304      0.044      2.959      0.003       0.044       0.217
DRBP           0.0030      0.002      1.723      0.086      -0.000       0.006
X3PAR         -0.0379      0.051     -0.736      0.462      -0.139       0.063
ASTP           0.0016      0.001      1.361      0.174      -0.001       0.004
STLP          -0.0381      0.013     -2.911      0.004      -0.064      -0.012
BLKP          -0.0129      0.008     -1.627      0.105      -0.029       0.003
TOVP           0.0014      0.002      0.891      0.374      -0.002       0.004
USG            0.0075      0.002      3.928      0.000       0.004       0.011
OWS            0.0310      0.007      4.575      0.000       0.018       0.044
DWS           -0.0088      0.014     -0.633      0.527      -0.036       0.019
DBPM           0.0194      0.010      1.949      0.052      -0.000       0.039
==============================================================================
Omnibus:                      279.009   Durbin-Watson:                   2.139
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             3498.433
Skew:                           2.720   Prob(JB):                         0.00
Kurtosis:                      16.124   Cond. No.                     2.15e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 2.15e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[113]:
```
#removed ORB b/c not statistically significant
cols = [ 'G', 'GS',
         'X3PM', 'X2PP',
         'FTP',  'STL', 'BLK',
         'PF', 'TS', 'FTR', 'DRBP','X3PAR',
        'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS', 'DWS',
         'DBPM', ]
x_train = trainroy[cols] #all of the variables
y_train = trainroy['ROY']
#adding an intercept
x_train = sm.add_constant(x_train)
```

```
#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    ROY   R-squared:                       0.434
Model:                            OLS   Adj. R-squared:                  0.405
Method:                 Least Squares   F-statistic:                     15.14
Date:                Wed, 12 May 2021   Prob (F-statistic):           1.39e-37
Time:                        19:19:07   Log-Likelihood:                321.45
No. Observations:                 416   AIC:                            -600.9
Df Residuals:                     395   BIC:                            -516.3
Df Model:                          20
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.0262      0.086     -0.304      0.761      -0.196       0.143
G             -0.0008      0.000     -1.839      0.067      -0.002    5.66e-05
GS             0.0020      0.000      4.577      0.000       0.001       0.003
X3PM           0.0449      0.024      1.909      0.057      -0.001       0.091
X2PP           0.3034      0.121      2.503      0.013       0.065       0.542
FTP            0.1447      0.054      2.677      0.008       0.038       0.251
STL            0.1061      0.031      3.389      0.001       0.045       0.168
BLK            0.0809      0.034      2.399      0.017       0.015       0.147
PF            -0.0210      0.014     -1.532      0.126      -0.048       0.006
TS            -0.7779      0.158     -4.912      0.000      -1.089      -0.467
FTR            0.1300      0.044      2.953      0.003       0.043       0.216
DRBP           0.0028      0.002      1.696      0.091      -0.000       0.006
X3PAR         -0.0384      0.051     -0.746      0.456      -0.139       0.063
ASTP           0.0016      0.001      1.452      0.147      -0.001       0.004
STLP          -0.0393      0.013     -3.078      0.002      -0.064      -0.014
BLKP          -0.0123      0.008     -1.585      0.114      -0.028       0.003
TOVP           0.0014      0.002      0.889      0.375      -0.002       0.004
USG            0.0075      0.002      3.933      0.000       0.004       0.011
OWS            0.0302      0.006      4.697      0.000       0.018       0.043
DWS           -0.0094      0.014     -0.676      0.499      -0.037       0.018
DBPM           0.0207      0.009      2.215      0.027       0.002       0.039
==============================================================================
Omnibus:                      278.459   Durbin-Watson:                   2.135
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             3474.206
Skew:                           2.715   Prob(JB):                         0.00
Kurtosis:                      16.075   Cond. No.                     2.14e+03
==============================================================================
```

```
Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 2.14e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[114]:
```
#removed DWS b/c not statistically significant
cols = [ 'G', 'GS',
        'X3PM', 'X2PP',
        'FTP',  'STL', 'BLK',
        'PF', 'TS', 'FTR', 'DRBP','X3PAR',
       'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS',
        'DBPM', ]
x_train = trainroy[cols] #all of the variables
y_train = trainroy['ROY']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    ROY   R-squared:                       0.433
Model:                            OLS   Adj. R-squared:                  0.406
Method:                 Least Squares   F-statistic:                     15.94
Date:                Wed, 12 May 2021   Prob (F-statistic):           4.23e-38
Time:                        19:19:07   Log-Likelihood:                321.21
No. Observations:                 416   AIC:                            -602.4
Df Residuals:                     396   BIC:                            -521.8
Df Model:                          19
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.0265      0.086     -0.308      0.759      -0.196       0.143
G             -0.0010      0.000     -2.381      0.018      -0.002      -0.000
GS             0.0019      0.000      4.551      0.000       0.001       0.003
X3PM           0.0471      0.023      2.024      0.044       0.001       0.093
X2PP           0.3036      0.121      2.506      0.013       0.065       0.542
FTP            0.1461      0.054      2.707      0.007       0.040       0.252
STL            0.1011      0.030      3.326      0.001       0.041       0.161
BLK            0.0794      0.034      2.362      0.019       0.013       0.146
PF            -0.0219      0.014     -1.604      0.109      -0.049       0.005
TS            -0.7626      0.157     -4.869      0.000      -1.071      -0.455
FTR            0.1285      0.044      2.926      0.004       0.042       0.215
DRBP           0.0025      0.002      1.588      0.113      -0.001       0.006
```

```
X3PAR         -0.0418      0.051     -0.818      0.414      -0.142       0.059
ASTP           0.0017      0.001      1.502      0.134      -0.001       0.004
STLP          -0.0368      0.012     -3.011      0.003      -0.061      -0.013
BLKP          -0.0116      0.008     -1.506      0.133      -0.027       0.004
TOVP           0.0013      0.002      0.856      0.393      -0.002       0.004
USG            0.0072      0.002      3.897      0.000       0.004       0.011
OWS            0.0290      0.006      4.696      0.000       0.017       0.041
DBPM           0.0179      0.008      2.138      0.033       0.001       0.034
==============================================================================
Omnibus:                      280.450   Durbin-Watson:                   2.130
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             3562.150
Skew:                           2.735   Prob(JB):                         0.00
Kurtosis:                      16.251   Cond. No.                     2.13e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 2.13e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[115]:
```
#removed X3PAR b/c not statistically significant
cols = [ 'G', 'GS',
        'X3PM', 'X2PP',
        'FTP',  'STL', 'BLK',
        'PF', 'TS', 'FTR', 'DRBP',
       'ASTP', 'STLP', 'BLKP', 'TOVP', 'USG', 'OWS',
        'DBPM', ]
x_train = trainroy[cols] #all of the variables
y_train = trainroy['ROY']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    ROY   R-squared:                       0.432
Model:                            OLS   Adj. R-squared:                  0.407
Method:                 Least Squares   F-statistic:                     16.80
Date:                Wed, 12 May 2021   Prob (F-statistic):           1.39e-38
Time:                        19:19:07   Log-Likelihood:                320.86
No. Observations:                 416   AIC:                            -603.7
Df Residuals:                     397   BIC:                            -527.1
Df Model:                          18
Covariance Type:            nonrobust
```

```
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.0600      0.076     -0.793      0.428      -0.209       0.089
G             -0.0009      0.000     -2.263      0.024      -0.002      -0.000
GS             0.0020      0.000      4.594      0.000       0.001       0.003
X3PM           0.0355      0.018      1.924      0.055      -0.001       0.072
X2PP           0.2956      0.121      2.449      0.015       0.058       0.533
FTP            0.1437      0.054      2.668      0.008       0.038       0.250
STL            0.1049      0.030      3.493      0.001       0.046       0.164
BLK            0.0731      0.033      2.235      0.026       0.009       0.137
PF            -0.0204      0.014     -1.510      0.132      -0.047       0.006
TS            -0.7456      0.155     -4.805      0.000      -1.051      -0.440
FTR            0.1354      0.043      3.143      0.002       0.051       0.220
DRBP           0.0028      0.002      1.780      0.076      -0.000       0.006
ASTP           0.0017      0.001      1.540      0.124      -0.000       0.004
STLP          -0.0359      0.012     -2.946      0.003      -0.060      -0.012
BLKP          -0.0091      0.007     -1.286      0.199      -0.023       0.005
TOVP           0.0014      0.002      0.910      0.363      -0.002       0.004
USG            0.0075      0.002      4.146      0.000       0.004       0.011
OWS            0.0290      0.006      4.710      0.000       0.017       0.041
DBPM           0.0160      0.008      1.992      0.047       0.000       0.032
==============================================================================
Omnibus:                      280.894   Durbin-Watson:                   2.129
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             3586.684
Skew:                           2.738   Prob(JB):                         0.00
Kurtosis:                      16.301   Cond. No.                     2.11e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 2.11e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[116]:
```
#removed TOVP b/c not statistically significant
cols = [ 'G', 'GS',
         'X3PM', 'X2PP',
         'FTP',  'STL', 'BLK',
         'PF', 'TS', 'FTR', 'DRBP',
        'ASTP', 'STLP', 'BLKP',  'USG', 'OWS',
         'DBPM', ]
x_train = trainroy[cols] #all of the variables
y_train = trainroy['ROY']
#adding an intercept
x_train = sm.add_constant(x_train)
```

```
#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    ROY   R-squared:                       0.431
Model:                            OLS   Adj. R-squared:                  0.407
Method:                 Least Squares   F-statistic:                     17.75
Date:                Wed, 12 May 2021   Prob (F-statistic):           4.79e-39
Time:                        19:19:07   Log-Likelihood:                320.43
No. Observations:                 416   AIC:                            -604.9
Df Residuals:                     398   BIC:                            -532.3
Df Model:                          17
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.0476      0.074     -0.639      0.523      -0.194       0.099
G             -0.0009      0.000     -2.265      0.024      -0.002      -0.000
GS             0.0020      0.000      4.632      0.000       0.001       0.003
X3PM           0.0344      0.018      1.868      0.062      -0.002       0.071
X2PP           0.2921      0.121      2.422      0.016       0.055       0.529
FTP            0.1397      0.054      2.604      0.010       0.034       0.245
STL            0.0999      0.030      3.384      0.001       0.042       0.158
BLK            0.0722      0.033      2.208      0.028       0.008       0.136
PF            -0.0184      0.013     -1.377      0.169      -0.045       0.008
TS            -0.7339      0.155     -4.746      0.000      -1.038      -0.430
FTR            0.1445      0.042      3.449      0.001       0.062       0.227
DRBP           0.0029      0.002      1.863      0.063      -0.000       0.006
ASTP           0.0023      0.001      2.402      0.017       0.000       0.004
STLP          -0.0344      0.012     -2.851      0.005      -0.058      -0.011
BLKP          -0.0087      0.007     -1.236      0.217      -0.023       0.005
USG            0.0071      0.002      4.046      0.000       0.004       0.011
OWS            0.0271      0.006      4.689      0.000       0.016       0.038
DBPM           0.0158      0.008      1.968      0.050    1.38e-05       0.032
==============================================================================
Omnibus:                      281.383   Durbin-Watson:                   2.119
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             3595.778
Skew:                           2.745   Prob(JB):                         0.00
Kurtosis:                      16.316   Cond. No.                     2.06e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 2.06e+03. This might indicate that there are
```

strong multicollinearity or other numerical problems.

[117]:
```
#removed BLKP b/c not statistically significant
cols = [ 'G', 'GS',
        'X3PM', 'X2PP',
        'FTP',  'STL', 'BLK',
        'PF', 'TS', 'FTR', 'DRBP',
       'ASTP', 'STLP', 'USG', 'OWS',
        'DBPM', ]
x_train = trainroy[cols] #all of the variables
y_train = trainroy['ROY']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

OLS Regression Results

Dep. Variable:	ROY	R-squared:	0.429
Model:	OLS	Adj. R-squared:	0.406
Method:	Least Squares	F-statistic:	18.74
Date:	Wed, 12 May 2021	Prob (F-statistic):	2.24e-39
Time:	19:19:07	Log-Likelihood:	319.63
No. Observations:	416	AIC:	-605.3
Df Residuals:	399	BIC:	-536.7
Df Model:	16		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-0.0676	0.073	-0.930	0.353	-0.210	0.075
G	-0.0008	0.000	-2.140	0.033	-0.002	-6.77e-05
GS	0.0020	0.000	4.670	0.000	0.001	0.003
X3PM	0.0323	0.018	1.758	0.079	-0.004	0.068
X2PP	0.2893	0.121	2.398	0.017	0.052	0.527
FTP	0.1515	0.053	2.867	0.004	0.048	0.255
STL	0.1050	0.029	3.588	0.000	0.047	0.162
BLK	0.0465	0.025	1.840	0.067	-0.003	0.096
PF	-0.0160	0.013	-1.213	0.226	-0.042	0.010
TS	-0.7391	0.155	-4.779	0.000	-1.043	-0.435
FTR	0.1339	0.041	3.262	0.001	0.053	0.215
DRBP	0.0025	0.002	1.667	0.096	-0.000	0.006
ASTP	0.0023	0.001	2.443	0.015	0.000	0.004
STLP	-0.0314	0.012	-2.654	0.008	-0.055	-0.008
USG	0.0072	0.002	4.125	0.000	0.004	0.011
OWS	0.0274	0.006	4.752	0.000	0.016	0.039

```
DBPM           0.0129      0.008      1.679      0.094      -0.002       0.028
==============================================================================
Omnibus:                      279.971   Durbin-Watson:                   2.108
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             3593.780
Skew:                           2.723   Prob(JB):                         0.00
Kurtosis:                      16.329   Cond. No.                     2.06e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 2.06e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[118]:
```
#removed PF b/c not statistically significant
cols = [ 'G', 'GS',
        'X3PM', 'X2PP',
        'FTP',  'STL', 'BLK',
         'TS', 'FTR', 'DRBP',
       'ASTP', 'STLP', 'USG', 'OWS',
        'DBPM', ]
x_train = trainroy[cols] #all of the variables
y_train = trainroy['ROY']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    ROY   R-squared:                       0.427
Model:                            OLS   Adj. R-squared:                  0.405
Method:                 Least Squares   F-statistic:                     19.87
Date:                Wed, 12 May 2021   Prob (F-statistic):           9.92e-40
Time:                        19:19:07   Log-Likelihood:                318.86
No. Observations:                 416   AIC:                            -605.7
Df Residuals:                     400   BIC:                            -541.2
Df Model:                          15
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.0658      0.073     -0.905      0.366      -0.209       0.077
G             -0.0009      0.000     -2.468      0.014      -0.002      -0.000
GS             0.0019      0.000      4.539      0.000       0.001       0.003
X3PM           0.0321      0.018      1.746      0.082      -0.004       0.068
```

```
X2PP           0.2827      0.121      2.344      0.020       0.046       0.520
FTP            0.1482      0.053      2.806      0.005       0.044       0.252
STL            0.0926      0.027      3.375      0.001       0.039       0.147
BLK            0.0350      0.023      1.493      0.136      -0.011       0.081
TS            -0.7499      0.154     -4.854      0.000      -1.054      -0.446
FTR            0.1327      0.041      3.233      0.001       0.052       0.213
DRBP           0.0024      0.002      1.601      0.110      -0.001       0.005
ASTP           0.0024      0.001      2.546      0.011       0.001       0.004
STLP          -0.0296      0.012     -2.524      0.012      -0.053      -0.007
USG            0.0071      0.002      4.079      0.000       0.004       0.011
OWS            0.0279      0.006      4.830      0.000       0.017       0.039
DBPM           0.0126      0.008      1.632      0.103      -0.003       0.028
==============================================================================
Omnibus:                      279.113   Durbin-Watson:                   2.115
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             3565.398
Skew:                           2.713   Prob(JB):                         0.00
Kurtosis:                      16.276   Cond. No.                     2.06e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 2.06e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[119]:
```
#removed BLK b/c not statistically significant
cols = [ 'G', 'GS',
        'X3PM', 'X2PP',
        'FTP',  'STL',
         'TS', 'FTR', 'DRBP',
       'ASTP', 'STLP', 'USG', 'OWS',
        'DBPM', ]
x_train = trainroy[cols] #all of the variables
y_train = trainroy['ROY']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    ROY   R-squared:                       0.424
Model:                            OLS   Adj. R-squared:                  0.404
Method:                 Least Squares   F-statistic:                     21.06
Date:                Wed, 12 May 2021   Prob (F-statistic):           6.12e-40
Time:                        19:19:07   Log-Likelihood:                317.71
```

```
No. Observations:                 416   AIC:                            -605.4
Df Residuals:                     401   BIC:                            -545.0
Df Model:                          14
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.0667      0.073     -0.916      0.360      -0.210       0.076
G             -0.0009      0.000     -2.457      0.014      -0.002      -0.000
GS             0.0021      0.000      5.086      0.000       0.001       0.003
X3PM           0.0254      0.018      1.422      0.156      -0.010       0.060
X2PP           0.3008      0.120      2.504      0.013       0.065       0.537
FTP            0.1445      0.053      2.735      0.007       0.041       0.248
STL            0.1020      0.027      3.814      0.000       0.049       0.155
TS            -0.7542      0.155     -4.875      0.000      -1.058      -0.450
FTR            0.1363      0.041      3.321      0.001       0.056       0.217
DRBP           0.0028      0.002      1.890      0.060      -0.000       0.006
ASTP           0.0022      0.001      2.313      0.021       0.000       0.004
STLP          -0.0361      0.011     -3.307      0.001      -0.058      -0.015
USG            0.0079      0.002      4.655      0.000       0.005       0.011
OWS            0.0273      0.006      4.740      0.000       0.016       0.039
DBPM           0.0172      0.007      2.442      0.015       0.003       0.031
==============================================================================
Omnibus:                      276.838   Durbin-Watson:                   2.116
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             3557.516
Skew:                           2.679   Prob(JB):                         0.00
Kurtosis:                      16.287   Cond. No.                     2.05e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 2.05e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[120]:
```python
#removed X3PM b/c not statistically significant
cols = [ 'G', 'GS',
         'X2PP',
        'FTP',  'STL',
         'TS', 'FTR', 'DRBP',
       'ASTP', 'STLP', 'USG', 'OWS',
        'DBPM', ]
x_train = trainroy[cols] #all of the variables
y_train = trainroy['ROY']
#adding an intercept
x_train = sm.add_constant(x_train)
```

```
#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    ROY   R-squared:                       0.421
Model:                            OLS   Adj. R-squared:                  0.402
Method:                 Least Squares   F-statistic:                     22.47
Date:                Wed, 12 May 2021   Prob (F-statistic):           3.32e-40
Time:                        19:19:07   Log-Likelihood:                316.66
No. Observations:                 416   AIC:                            -605.3
Df Residuals:                     402   BIC:                            -548.9
Df Model:                          13
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.0763      0.073     -1.051      0.294      -0.219       0.066
G             -0.0008      0.000     -2.237      0.026      -0.002      -0.000
GS             0.0021      0.000      5.212      0.000       0.001       0.003
X2PP           0.2638      0.117      2.246      0.025       0.033       0.495
FTP            0.1574      0.052      3.021      0.003       0.055       0.260
STL            0.1112      0.026      4.281      0.000       0.060       0.162
TS            -0.7045      0.151     -4.669      0.000      -1.001      -0.408
FTR            0.1247      0.040      3.096      0.002       0.046       0.204
DRBP           0.0023      0.001      1.586      0.114      -0.001       0.005
ASTP           0.0021      0.001      2.280      0.023       0.000       0.004
STLP          -0.0385      0.011     -3.566      0.000      -0.060      -0.017
USG            0.0083      0.002      4.956      0.000       0.005       0.012
OWS            0.0273      0.006      4.729      0.000       0.016       0.039
DBPM           0.0174      0.007      2.462      0.014       0.004       0.031
==============================================================================
Omnibus:                      279.881   Durbin-Watson:                   2.132
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             3659.776
Skew:                           2.713   Prob(JB):                         0.00
Kurtosis:                      16.479   Cond. No.                     1.99e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.99e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[121]:
```
#removed DRBP b/c not statistically significant
cols = [ 'G', 'GS',
```

```
         'X2PP',
        'FTP',  'STL',
         'TS', 'FTR',
       'ASTP', 'STLP', 'USG', 'OWS',
        'DBPM', ]
x_train = trainroy[cols] #all of the variables
y_train = trainroy['ROY']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    ROY   R-squared:                       0.417
Model:                            OLS   Adj. R-squared:                  0.400
Method:                 Least Squares   F-statistic:                     24.04
Date:                Wed, 12 May 2021   Prob (F-statistic):           2.20e-40
Time:                        19:19:07   Log-Likelihood:                315.36
No. Observations:                 416   AIC:                            -604.7
Df Residuals:                     403   BIC:                            -552.3
Df Model:                          12
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.0402      0.069     -0.582      0.561      -0.176       0.096
G             -0.0008      0.000     -2.129      0.034      -0.002    -6.1e-05
GS             0.0021      0.000      5.168      0.000       0.001       0.003
X2PP           0.2848      0.117      2.436      0.015       0.055       0.515
FTP            0.1504      0.052      2.890      0.004       0.048       0.253
STL            0.1149      0.026      4.431      0.000       0.064       0.166
TS            -0.7323      0.150     -4.877      0.000      -1.027      -0.437
FTR            0.1390      0.039      3.535      0.000       0.062       0.216
ASTP           0.0016      0.001      1.820      0.069      -0.000       0.003
STLP          -0.0438      0.010     -4.248      0.000      -0.064      -0.024
USG            0.0090      0.002      5.588      0.000       0.006       0.012
OWS            0.0284      0.006      4.951      0.000       0.017       0.040
DBPM           0.0211      0.007      3.161      0.002       0.008       0.034
==============================================================================
Omnibus:                      280.871   Durbin-Watson:                   2.122
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             3665.526
Skew:                           2.728   Prob(JB):                         0.00
Kurtosis:                      16.480   Cond. No.                     1.93e+03
==============================================================================
```

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.93e+03. This might indicate that there are strong multicollinearity or other numerical problems.

[122]:
```
#removed ASTP b/c not statistically significant
cols = [ 'G', 'GS',
         'X2PP',
        'FTP',  'STL',
         'TS', 'FTR',
        'STLP', 'USG', 'OWS',
        'DBPM', ]
x_train = trainroy[cols] #all of the variables
y_train = trainroy['ROY']
#adding an intercept
x_train = sm.add_constant(x_train)

#fit data to the model
model1= sm.OLS(y_train, x_train).fit()
print(model1.summary())
```

OLS Regression Results

Dep. Variable:	ROY	R-squared:	0.412
Model:	OLS	Adj. R-squared:	0.396
Method:	Least Squares	F-statistic:	25.78
Date:	Wed, 12 May 2021	Prob (F-statistic):	2.08e-40
Time:	19:19:08	Log-Likelihood:	313.66
No. Observations:	416	AIC:	-603.3
Df Residuals:	404	BIC:	-555.0
Df Model:	11		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-0.0361	0.069	-0.522	0.602	-0.172	0.100
G	-0.0008	0.000	-2.092	0.037	-0.002	-4.72e-05
GS	0.0020	0.000	4.939	0.000	0.001	0.003
X2PP	0.2985	0.117	2.551	0.011	0.068	0.529
FTP	0.1674	0.051	3.263	0.001	0.067	0.268
STL	0.1304	0.025	5.311	0.000	0.082	0.179
TS	-0.7867	0.148	-5.331	0.000	-1.077	-0.497
FTR	0.1336	0.039	3.398	0.001	0.056	0.211
STLP	-0.0422	0.010	-4.100	0.000	-0.062	-0.022
USG	0.0098	0.002	6.368	0.000	0.007	0.013
OWS	0.0289	0.006	5.021	0.000	0.018	0.040

```
DBPM           0.0200      0.007      2.996      0.003       0.007       0.033
==============================================================================
Omnibus:                      286.165   Durbin-Watson:                   2.125
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             3887.467
Skew:                           2.785   Prob(JB):                         0.00
Kurtosis:                      16.902   Cond. No.                     1.87e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.87e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[123]:
```
import statsmodels.formula.api as smf

ROYLinear = smf.ols(formula='ROY ~ G + GS + FTP + X2PP+ STL + TS + FTR + STLP +␣
 ↪USG + OWS + DBPM',
                    data=trainroy).fit()
print(ROYLinear.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    ROY   R-squared:                       0.412
Model:                            OLS   Adj. R-squared:                  0.396
Method:                 Least Squares   F-statistic:                     25.78
Date:                Wed, 12 May 2021   Prob (F-statistic):           2.08e-40
Time:                        19:19:08   Log-Likelihood:                313.66
No. Observations:                 416   AIC:                            -603.3
Df Residuals:                     404   BIC:                            -555.0
Df Model:                          11
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -0.0361      0.069     -0.522      0.602      -0.172       0.100
G             -0.0008      0.000     -2.092      0.037      -0.002   -4.72e-05
GS             0.0020      0.000      4.939      0.000       0.001       0.003
FTP            0.1674      0.051      3.263      0.001       0.067       0.268
X2PP           0.2985      0.117      2.551      0.011       0.068       0.529
STL            0.1304      0.025      5.311      0.000       0.082       0.179
TS            -0.7867      0.148     -5.331      0.000      -1.077      -0.497
FTR            0.1336      0.039      3.398      0.001       0.056       0.211
STLP          -0.0422      0.010     -4.100      0.000      -0.062      -0.022
USG            0.0098      0.002      6.368      0.000       0.007       0.013
OWS            0.0289      0.006      5.021      0.000       0.018       0.040
DBPM           0.0200      0.007      2.996      0.003       0.007       0.033
==============================================================================
```

```
Omnibus:                      286.165   Durbin-Watson:                   2.125
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             3887.467
Skew:                           2.785   Prob(JB):                         0.00
Kurtosis:                      16.902   Cond. No.                     1.87e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.87e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

[124]:
```
OSR2(ROYLinear, trainroy, testroy, 'ROY')
```

[124]:
```
0.42183685623993505
```

2021 Predictions

[131]:
```
#mvp 2021 prediction
nba2021 = pd.read_csv('nba2021.csv')
y_predicted = MVPLinear.predict(nba2021)
finallmvp = pd.concat([y_predicted,nba2021],axis=1)
finallmvp = finallmvp.rename(columns={0:'Predicted MVP'})
finallmvp = finallmvp.sort_values(by='Predicted MVP',ascending = False)
finallmvp = finallmvp.loc[finallmvp['GS'] > 30]
finallmvp['Player'].iloc[0]
```

[131]:
```
'Nikola Jokic'
```

[132]:
```
finallmvp.head(5)
```

[132]:

	Predicted MVP	Unnamed: 0	Player	Season	Pos	Age	Tm	\
259	0.150505	260	Nikola Jokic	2021	C	25	DEN	
294	0.134981	295	Damian Lillard	2021	PG	30	POR	
114	0.124799	115	Stephen Curry	2021	PG	32	GSW	
11	0.105019	12	Giannis Antetokounmpo	2021	PF	26	MIL	
82	0.096915	83	Jimmy Butler	2021	SF	31	MIA	

	G	GS	MP	…	OWS	DWS	WS	WS48	OBPM	DBPM	BPM	VORP	MVP	\
259	69	69	35.0	…	11.7	3.3	15.0	0.298	9.1	2.5	11.5	8.2	0	
294	64	64	35.7	…	9.0	0.8	9.8	0.205	7.3	-1.6	5.7	4.5	0	
114	62	62	34.1	…	6.5	2.5	8.9	0.203	8.0	0.0	8.1	5.4	0	
11	59	59	32.9	…	6.5	3.3	9.8	0.242	6.2	2.7	8.9	5.4	0	
82	51	51	33.6	…	6.4	2.6	9.1	0.253	5.0	2.4	7.4	4.1	0	

	DPOY
259	0
294	0

```
114     0
11      0
82      0

[5 rows x 54 columns]
```

[133]:
```
#dpoy 2021 prediction
nba2021 = pd.read_csv('nba2021.csv')
y_predicted = DPOYLinear.predict(nba2021)
finalldpoy = pd.concat([y_predicted,nba2021],axis=1)
finalldpoy = finalldpoy.rename(columns={0:'Predicted DPOY'})
finalldpoy = finalldpoy.sort_values(by='Predicted DPOY',ascending = False)
finalldpoy = finalldpoy.loc[finalldpoy['GS'] > 10]
finalldpoy['Player'].iloc[0]
```

[133]:
```
'Rudy Gobert'
```

[134]:
```
finalldpoy.head(5)
```

[134]:
```
     Predicted DPOY  Unnamed: 0         Player  Season Pos  Age   Tm   G  GS  \
177        0.003726         178    Rudy Gobert    2021   C   28  UTA  68  68
491        0.003559         492   Myles Turner    2021   C   24  IND  47  47
89         0.003025          90   Clint Capela    2021   C   26  ATL  61  61
82         0.002841          83   Jimmy Butler    2021  SF   31  MIA  51  51
116        0.002799         117  Anthony Davis    2021  PF   27  LAL  34  34

       MP  …  OWS  DWS    WS   WS48  OBPM  DBPM  BPM  VORP  MVP  DPOY
177  30.9  …  5.7  4.9  10.7  0.244   1.9   2.5  4.4   3.4    0     0
491  31.0  …  1.1  2.3   3.4  0.112  -1.3   1.7  0.4   0.9    0     0
89   30.2  …  4.7  3.1   7.8  0.204   2.7  -0.2  2.5   2.1    0     0
82   33.6  …  6.4  2.6   9.1  0.253   5.0   2.4  7.4   4.1    0     0
116  32.2  …  1.2  2.3   3.4  0.151   3.0   1.7  4.7   1.9    0     0

[5 rows x 54 columns]
```

[135]:
```
#smoy 2021 prediction
nba2021smoy = pd.read_csv('smoy2021.csv')
y_predicted = SMOYLinear.predict(nba2021smoy)
finallsmoy = pd.concat([y_predicted,nba2021smoy],axis=1)
finallsmoy = finallsmoy.rename(columns={0:'Predicted SMOY'})
finallsmoy = finallsmoy.sort_values(by='Predicted SMOY',ascending = False)
finallsmoyy = finallsmoy.loc[finallsmoy['G'] > 10]
finallsmoyy['Player'].iloc[0]
```

[135]:
```
'Jordan Clarkson'
```

[136]:
```
finallsmoyy.head(5)
```

[136]:
```
     Predicted SMOY  Unnamed: 0           Player  Season Pos  Age   Tm   G  \
54         0.097336          55   Jordan Clarkson    2021  SG   28  UTA  65
139        0.084820         140       Joe Ingles    2021  SF   33  UTA  64
115        0.069711         116  Tim Hardaway Jr.    2021  SG   28  DAL  67
26         0.063696          27     Chris Boucher    2021   C   28  TOR  60
218        0.061187         219       Patty Mills    2021  PG   32  SAS  65

     GS    MP  …   USG  OWS  DWS   WS   WS48  OBPM  DBPM  BPM  VORP  SMOY
54    1  26.6  …  29.6  1.3  2.3  3.6  0.101   1.6  -1.0  0.6   1.1     0
139  27  27.8  …  16.5  4.7  2.1  6.9  0.185   3.0   0.7  3.7   2.5     0
115  28  28.4  …  23.4  2.9  1.1  4.1  0.102   1.8  -1.8  0.0   1.0     0
26   14  24.2  …  20.5  4.0  2.1  6.1  0.203   2.8   0.1  2.9   1.8     0
218   0  24.8  …  18.3  1.4  0.9  2.3  0.068   0.2  -1.7 -1.5   0.2     0

[5 rows x 53 columns]
```

[137]:
```
#MIP 2021 prediction
nba2021mip = pd.read_csv('mip2021.csv')
y_predicted = MIPLinear.predict(nba2021mip)
finallmip = pd.concat([y_predicted,nba2021mip],axis=1)
finallmip = finallmip.rename(columns={0:'Predicted MIP'})
finallmip = finallmip.sort_values(by='Predicted MIP',ascending = False)
finallmip = finallmip.loc[finallmip['GS'] > 1]
finallmip['Player'].iloc[0]
```

[137]:
```
'Michael Porter Jr.'
```

[138]:
```
finallmip.head(10)
```

[138]:
```
     Predicted MIP              Player  Season     G    GS    MP  FGM  FGA  \
246       0.039612  Michael Porter Jr.  2021.0   4.0  44.0  15.4  3.9  6.5
230       0.035564       Malik Beasley  2021.0 -18.0  22.0  10.8  3.0  6.5
291       0.034211     Robert Williams  2021.0  23.0  12.0   5.5  1.4  2.0
247       0.033390       Mikal Bridges  2021.0  -4.0  37.0   5.0  1.6  2.8
225       0.032606       Luguentz Dort  2021.0  16.0  24.0   6.9  2.4  6.2
214       0.030863          Kyle Kuzma  2021.0   4.0  23.0   3.8  0.2  0.3
208       0.029583    Kevin Porter Jr.  2021.0 -24.0  20.0   8.9  2.3  5.7
45        0.029120       Chris Boucher  2021.0  -2.0  14.0  11.0  2.5  4.5
153       0.025760        Jakob Poeltl  2021.0   0.0  30.0   8.9  1.3  2.1
350       0.025653     Wayne Ellington  2021.0  10.0  30.0   6.5  1.5  2.5

        FGP  X3PM  …  TOVP  USG  OWS  DWS   WS   WS48  OBPM  DBPM  BPM  VORP
246   0.039   1.7  …  -2.5 -0.6  3.0  0.9  4.0  0.022   1.7  -0.1  1.6   1.6
230   0.015   1.6  …  -0.6  2.2  0.5 -0.3  0.2  0.004   2.4  -0.9  1.5   0.4
291  -0.006   0.0  …  -1.7  1.4  2.3  0.9  3.3  0.003   2.2  -1.2  1.0   1.3
247   0.025   0.9  …  -4.7  2.1  2.5  0.2  2.7  0.045   2.7  -0.8  2.0   1.2
225  -0.007   1.4  …   1.6  7.6 -0.8  0.2 -0.5 -0.039   1.4  -1.4  0.0  -0.3
```

```
214  0.002   0.7  …   1.2 -2.4  0.2  0.9  1.1  0.015   1.3   0.5  1.8   0.8
208 -0.017   0.8  …   1.5  4.9  0.0 -0.3 -0.3 -0.011   1.9  -1.8  0.0   0.1
45   0.042   0.9  …  -0.7  0.8  1.9  0.3  2.3 -0.021   1.8  -0.9  0.9   1.0
153 -0.001   0.0  …   0.3  0.2  0.7  0.9  1.5 -0.012  -1.0  -0.8 -1.8  -0.1
350  0.090   1.1  …  -0.6  1.7  1.1  0.3  1.5  0.054   3.2  -0.8  2.5   0.5

[10 rows x 48 columns]
```

[139]:
```
#ROY 2021 prediction
nba2021roy = pd.read_csv('roy2021.csv')
y_predicted = ROYLinear.predict(nba2021roy)
finallroy = pd.concat([y_predicted,nba2021roy],axis=1)
finallroy = finallroy.rename(columns={0:'Predicted ROY'})
finallroy = finallroy.sort_values(by='Predicted ROY',ascending = False)
finallroy = finallroy.loc[finallroy['GS'] > 1]
finallroy['Player'].iloc[0]
```

[139]:
```
'LaMelo Ball'
```

[140]:
```
finallroy.head(5)
```

[140]:
```
    Predicted ROY  Unnamed: 0            Player  Season Pos  Age   Tm   G  \
50       0.251742          51       LaMelo Ball    2021  PG   19  CHO  48
35       0.214883          36     Jae'Sean Tate    2021  SF   25  HOU  67
3        0.178309           4   Anthony Edwards    2021  SG   19  MIN  69
83       0.132846          84  Tyrese Haliburton    2021  PG   20  SAC  58
31       0.130995          32       Isaac Okoro    2021  SG   20  CLE  64

    GS    MP  …   USG  OWS  DWS   WS   WS48  OBPM  DBPM  BPM  VORP  ROY
50  28  28.8  …  26.0  1.0  1.8  2.8  0.096   1.1   0.6  1.7   1.3    0
35  55  29.4  …  16.2  2.1  1.5  3.7  0.090  -1.7   0.5 -1.2   0.4    0
3   52  31.9  …  27.0 -1.0  1.4  0.4  0.008  -1.2  -1.8 -3.0  -0.5    0
83  20  30.1  …  18.1  2.8  0.6  3.5  0.096   1.9  -0.9  1.1   1.4    0
31  64  32.3  …  13.8  0.0  0.9  1.0  0.023  -4.2  -0.7 -4.9  -1.5    0

[5 rows x 53 columns]
```

[ ]:
```

```