# final

November 18, 2018

In [2]:
```
# dataset - Sloan Digital Sky Survey RD14
# Number of rows - 600
#Number of columns - 18
# number of categorical features - 7
#number of numerical features - 9

%matplotlib inline
import numpy as np
import pandas as pd
import pylab as pl
import matplotlib.pyplot as plt
from sklearn import preprocessing
import scipy.stats
```

In [3]:
```
df = pd.read_csv('sky1.csv')
```

In [4]:
```
df
```

Out[4]:
```
           objid          ra       dec         u         g         r  \
0   1.240000e+18  183.531326  0.089693  19.47406  17.04240  15.94699
1   1.240000e+18  183.598371  0.135285  18.66280  17.21449  16.67637
2   1.240000e+18  183.680207  0.126185  19.38298  18.19169  17.47428
3   1.240000e+18  183.870529  0.049911  17.76536  16.60272  16.16116
4   1.240000e+18  183.883288  0.102557  17.55025  16.26342  16.43869
5   1.240000e+18  183.847174  0.173694  19.43133  18.46779  18.16451
6   1.240000e+18  183.864379  0.019201  19.38322  17.88995  17.10537
7   1.240000e+18  183.900081  0.187473  18.97993  17.84496  17.38022
8   1.240000e+18  183.924588  0.097246  17.90616  16.97172  16.67541
9   1.240000e+18  183.973498  0.081626  18.67249  17.71375  17.49362
10  1.240000e+18  183.979195  0.135998  19.29772  17.80227  17.18266
11  1.240000e+18  184.085331  0.112110  18.83307  17.51785  16.94273
12  1.240000e+18  184.102098  0.191511  19.56250  18.19113  17.65759
13  1.240000e+18  184.160510  0.075645  19.57990  17.72815  16.98740
14  1.240000e+18  184.189574  0.099482  19.25667  17.54869  16.63578
15  1.240000e+18  184.350647  0.207230  18.73832  18.60962  18.39696
16  1.240000e+18  184.221797  0.046955  19.05958  18.09512  17.92766
17  1.240000e+18  184.245664  0.198257  19.22143  19.30248  19.13823
18  1.240000e+18  184.388870  0.068287  19.04397  17.51106  16.87335
```

| | | | | | | |
|---|---|---|---|---|---|---|
| 19 | 1.240000e+18 | 184.380919 | 0.174323 | 17.81661 | 16.86976 | 16.53884 |
| 20 | 1.240000e+18 | 184.466853 | 0.111965 | 19.39320 | 18.48274 | 18.16551 |
| 21 | 1.240000e+18 | 184.569411 | 0.137091 | 17.51339 | 16.41793 | 16.06695 |
| 22 | 1.240000e+18 | 184.654170 | 0.122673 | 19.07731 | 18.64518 | 18.49678 |
| 23 | 1.240000e+18 | 184.658795 | 0.159936 | 18.63032 | 17.23437 | 16.72749 |
| 24 | 1.240000e+18 | 184.516834 | 0.145550 | 19.21684 | 18.29956 | 18.19589 |
| 25 | 1.240000e+18 | 184.618563 | 0.080576 | 19.41270 | 17.47109 | 16.49870 |
| 26 | 1.240000e+18 | 184.627408 | 0.013234 | 18.48543 | 17.33900 | 16.90179 |
| 27 | 1.240000e+18 | 184.676762 | 0.059442 | 17.72835 | 16.19783 | 15.66073 |
| 28 | 1.240000e+18 | 184.711970 | 0.066572 | 17.70824 | 16.61069 | 16.19772 |
| 29 | 1.240000e+18 | 184.770204 | 0.198010 | 18.18974 | 17.23738 | 16.89113 |
| .. | ... | ... | ... | ... | ... | ... |
| 569 | 1.240000e+18 | 244.284910 | -0.455477 | 18.63084 | 16.69326 | 15.67451 |
| 570 | 1.240000e+18 | 244.491831 | -0.470566 | 18.43597 | 17.47617 | 17.10257 |
| 571 | 1.240000e+18 | 244.496791 | -0.498398 | 17.98780 | 16.51811 | 15.74451 |
| 572 | 1.240000e+18 | 244.916397 | -0.523076 | 18.45084 | 17.39167 | 16.97855 |
| 573 | 1.240000e+18 | 244.925582 | -0.455836 | 19.34707 | 17.76923 | 16.92714 |
| 574 | 1.240000e+18 | 244.935545 | -0.479247 | 18.53797 | 17.16947 | 16.50210 |
| 575 | 1.240000e+18 | 245.128493 | -0.499415 | 19.18215 | 18.79711 | 18.81150 |
| 576 | 1.240000e+18 | 245.367353 | -0.457074 | 18.12316 | 16.31978 | 15.37991 |
| 577 | 1.240000e+18 | 245.376694 | -0.457008 | 19.46922 | 17.96987 | 17.31625 |
| 578 | 1.240000e+18 | 245.978479 | -0.483455 | 19.38072 | 17.38953 | 16.26663 |
| 579 | 1.240000e+18 | 246.034378 | -0.473284 | 19.51987 | 17.56103 | 16.44300 |
| 580 | 1.240000e+18 | 246.268157 | -0.426990 | 19.46025 | 18.28482 | 17.78494 |
| 581 | 1.240000e+18 | 246.695626 | -0.510111 | 18.01060 | 16.81185 | 16.32209 |
| 582 | 1.240000e+18 | 246.782081 | -0.492432 | 17.60994 | 15.90911 | 15.02090 |
| 583 | 1.240000e+18 | 247.021139 | -0.448677 | 19.23512 | 19.08207 | 18.78046 |
| 584 | 1.240000e+18 | 248.269016 | -0.488920 | 19.27635 | 17.99304 | 17.47953 |
| 585 | 1.240000e+18 | 248.894520 | -0.452329 | 18.73576 | 16.83826 | 15.94847 |
| 586 | 1.240000e+18 | 249.496551 | -0.498468 | 18.99369 | 17.31122 | 16.48666 |
| 587 | 1.240000e+18 | 189.429821 | -0.131042 | 17.83883 | 16.62422 | 16.84056 |
| 588 | 1.240000e+18 | 189.448544 | -0.052745 | 18.97562 | 17.84829 | 17.49412 |
| 589 | 1.240000e+18 | 189.453801 | -0.097313 | 17.66081 | 16.62739 | 16.28842 |
| 590 | 1.240000e+18 | 189.468747 | -0.036000 | 17.28231 | 16.20900 | 15.76480 |
| 591 | 1.240000e+18 | 189.522249 | -0.027031 | 16.32175 | 14.72385 | 13.88480 |
| 592 | 1.240000e+18 | 189.497150 | -0.110585 | 19.25978 | 18.10524 | 17.42960 |
| 593 | 1.240000e+18 | 225.529253 | -0.029788 | 19.21136 | 18.73800 | 18.76889 |
| 594 | 1.240000e+18 | 225.494356 | -0.114317 | 19.43974 | 18.09289 | 17.40579 |
| 595 | 1.240000e+18 | 145.288576 | 0.346512 | 19.22352 | 18.00232 | 17.37943 |
| 596 | 1.240000e+18 | 145.292821 | 0.376326 | 19.31302 | 18.13527 | 17.67289 |
| 597 | 1.240000e+18 | 145.593807 | 0.278308 | 19.10790 | 18.92201 | 18.59314 |
| 598 | 1.240000e+18 | 145.637275 | 0.373311 | 19.20753 | 17.87736 | 17.30583 |

| | i | z | run | rerun | camcol | field | specobjid | class | \ |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 15.50342 | 15.22531 | 752.0 | 301.0 | 4 | 267 | 3.720000e+18 | STAR | |
| 1 | NaN | 16.39150 | 752.0 | 301.0 | 4 | 267 | 3.640000e+17 | STAR | |
| 2 | 17.08732 | 16.80125 | 752.0 | 301.0 | 4 | 268 | 3.230000e+17 | GALAXY | |
| 3 | 15.98233 | 15.90438 | 752.0 | 301.0 | 4 | 269 | 3.720000e+18 | STAR | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 4 | NaN | 16.61326 | 752.0 | NaN | 4 | 269 | 3.720000e+18 | NaN |
| 5 | 18.01475 | 18.04155 | 752.0 | 301.0 | 4 | 269 | 3.650000e+17 | STAR |
| 6 | 16.66393 | 16.36955 | 752.0 | 301.0 | 4 | 269 | 3.230000e+17 | GALAXY |
| 7 | 17.20673 | 17.07071 | NaN | 301.0 | 4 | 269 | 3.720000e+18 | STAR |
| 8 | NaN | 16.47596 | 752.0 | 301.0 | 4 | 270 | 3.640000e+17 | STAR |
| 9 | 17.28284 | 17.22644 | 752.0 | 301.0 | 4 | 270 | 3.240000e+17 | GALAXY |
| 10 | 16.92335 | 16.79928 | 752.0 | 301.0 | 4 | 270 | 3.720000e+18 | STAR |
| 11 | 16.71418 | 16.60521 | 752.0 | 301.0 | 4 | 271 | 3.720000e+18 | STAR |
| 12 | 17.47573 | 17.39203 | 752.0 | 301.0 | 4 | 271 | 3.720000e+18 | STAR |
| 13 | 16.68076 | 16.50426 | 752.0 | 301.0 | 4 | 271 | 3.720000e+18 | STAR |
| 14 | 16.14922 | 15.76639 | 752.0 | 301.0 | 4 | 271 | 3.240000e+17 | GALAXY |
| 15 | 18.31174 | 17.97663 | 752.0 | 301.0 | 4 | 272 | 3.230000e+17 | QSO |
| 16 | NaN | 17.90772 | 752.0 | 301.0 | 4 | 272 | 3.650000e+17 | STAR |
| 17 | 19.11351 | 19.23454 | 752.0 | 301.0 | 4 | 272 | 3.230000e+17 | QSO |
| 18 | 16.61114 | 16.48303 | 752.0 | 301.0 | 4 | 273 | 3.640000e+17 | STAR |
| 19 | 16.19576 | 16.08668 | 752.0 | 301.0 | 4 | 273 | 3.230000e+17 | GALAXY |
| 20 | 18.05122 | 18.04328 | 752.0 | 301.0 | 4 | 273 | 3.650000e+17 | STAR |
| 21 | 15.93751 | 15.89478 | 752.0 | 301.0 | 4 | 274 | 3.240000e+17 | STAR |
| 22 | 18.52677 | 18.45765 | 752.0 | 301.0 | 4 | 274 | 3.240000e+17 | QSO |
| 23 | 16.51772 | 16.44039 | 752.0 | 301.0 | 4 | 274 | 2.880000e+18 | STAR |
| 24 | 18.14758 | 18.26658 | 752.0 | 301.0 | 4 | 274 | 3.720000e+18 | STAR |
| 25 | 16.06502 | 15.66848 | 752.0 | 301.0 | 4 | 274 | 3.240000e+17 | GALAXY |
| 26 | 16.66712 | 16.48684 | 752.0 | 301.0 | 4 | 274 | 3.230000e+17 | NaN |
| 27 | 15.51586 | 15.43615 | 752.0 | 301.0 | 4 | 275 | 2.880000e+18 | STAR |
| 28 | NaN | 16.00229 | 752.0 | 301.0 | 4 | 275 | 3.720000e+18 | STAR |
| 29 | 16.75727 | 16.71052 | 752.0 | 301.0 | 4 | 275 | 2.880000e+18 | STAR |
| .. | ... | ... | ... | ... | ... | ... | ... | ... |
| 569 | 15.20557 | 14.82077 | 745.0 | 301.0 | 2 | 571 | 3.900000e+17 | GALAXY |
| 570 | 16.90971 | 16.82976 | 745.0 | 301.0 | 2 | 573 | 3.900000e+17 | GALAXY |
| 571 | 15.30433 | 14.98556 | 745.0 | 301.0 | 2 | 573 | 3.900000e+17 | GALAXY |
| 572 | 16.82599 | 16.76456 | 745.0 | 301.0 | 2 | 576 | 3.900000e+17 | STAR |
| 573 | 16.48486 | 16.14608 | 745.0 | 301.0 | 2 | 576 | 3.900000e+17 | GALAXY |
| 574 | 16.18933 | 15.95892 | 745.0 | 301.0 | 2 | 576 | 3.900000e+17 | GALAXY |
| 575 | 18.61492 | 18.51464 | 745.0 | 301.0 | 2 | 577 | 4.100000e+17 | QSO |
| 576 | 14.91012 | 14.53670 | 745.0 | 301.0 | 2 | 579 | 4.100000e+17 | GALAXY |
| 577 | 16.92729 | 16.71774 | 745.0 | 301.0 | 2 | 579 | 3.900000e+17 | GALAXY |
| 578 | 15.75659 | 15.36963 | 745.0 | 301.0 | 2 | 583 | 4.100000e+17 | NaN |
| 579 | 15.95302 | 15.57304 | 745.0 | 301.0 | 2 | 583 | 4.100000e+17 | GALAXY |
| 580 | 17.50377 | 17.31497 | 745.0 | 301.0 | 2 | 585 | 4.100000e+17 | GALAXY |
| 581 | 16.03653 | 15.92627 | 745.0 | 301.0 | 2 | 588 | 4.100000e+17 | GALAXY |
| 582 | 14.54955 | 14.19971 | 745.0 | 301.0 | 2 | 588 | 4.100000e+17 | GALAXY |
| 583 | 18.81124 | 18.99055 | 745.0 | 301.0 | 2 | 590 | 4.100000e+17 | QSO |
| 584 | 17.29088 | 18.28560 | 745.0 | 301.0 | 2 | 598 | 3.920000e+17 | STAR |
| 585 | 15.53803 | 15.15638 | 745.0 | 301.0 | 2 | 602 | 3.920000e+17 | GALAXY |
| 586 | 16.05332 | 15.74858 | 745.0 | 301.0 | 2 | 606 | 3.920000e+17 | GALAXY |
| 587 | 17.03584 | 17.15817 | 745.0 | 301.0 | 3 | 205 | 3.260000e+18 | STAR |
| 588 | 17.37400 | 17.35097 | 745.0 | 301.0 | 3 | 205 | 3.280000e+17 | STAR |
| 589 | 16.17215 | 16.14751 | 745.0 | 301.0 | 3 | 205 | 3.260000e+18 | STAR |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 590 | 15.51375 | 15.47968 | 745.0 | 301.0 | 3 | 205 | 3.260000e+18 | STAR |
| 591 | 13.46923 | 13.15061 | 745.0 | 301.0 | 3 | 205 | 3.270000e+17 | GALAXY |
| 592 | 17.08964 | 16.85153 | 745.0 | 301.0 | 3 | 205 | 3.280000e+17 | GALAXY |
| 593 | 18.96195 | 19.18182 | 745.0 | 301.0 | 3 | 446 | 4.520000e+18 | STAR |
| 594 | 17.00810 | 16.82082 | 745.0 | 301.0 | 3 | 446 | 3.490000e+17 | GALAXY |
| 595 | 17.01110 | 16.76524 | 756.0 | 301.0 | 4 | 197 | 3.000000e+17 | GALAXY |
| 596 | 17.49996 | 17.44187 | 756.0 | 301.0 | 4 | 197 | 3.000000e+17 | STAR |
| 597 | 18.47573 | 18.40461 | 756.0 | 301.0 | 4 | 199 | 3.000000e+17 | QSO |
| 598 | 17.08623 | 16.96132 | 756.0 | 301.0 | 4 | 199 | 3.000000e+17 | STAR |

| | redshift | plate | mjd | fiberid |
|---|---|---|---|---|
| 0 | -8.960000e-06 | 3306 | 54922 | 491 |
| 1 | -5.490000e-05 | 323 | 51615 | 541 |
| 2 | 1.231112e-01 | 287 | 52023 | 513 |
| 3 | -1.106160e-04 | 3306 | 54922 | 510 |
| 4 | 5.903570e-04 | 3306 | 54922 | 512 |
| 5 | 3.146030e-04 | 324 | 51666 | 594 |
| 6 | 1.002423e-01 | 287 | 52023 | 559 |
| 7 | 3.148480e-04 | 3306 | 54922 | 515 |
| 8 | 8.910000e-05 | 323 | 51615 | 595 |
| 9 | 4.050813e-02 | 288 | 52000 | 400 |
| 10 | -3.460000e-05 | 3306 | 54922 | 506 |
| 11 | 6.227740e-04 | 3306 | 54922 | 547 |
| 12 | 5.470000e-05 | 3306 | 54922 | 544 |
| 13 | 8.300000e-06 | 3306 | 54922 | 546 |
| 14 | 7.208736e-02 | 288 | 52000 | 389 |
| 15 | 2.719369e-01 | 287 | 52023 | 587 |
| 16 | 2.099740e-04 | 324 | 51666 | 35 |
| 17 | 1.178098e+00 | 287 | 52023 | 583 |
| 18 | 2.400980e-04 | 323 | 51615 | 628 |
| 19 | 7.277206e-02 | 287 | 52023 | 632 |
| 20 | -2.702180e-04 | 324 | 51666 | 623 |
| 21 | -3.550000e-05 | 288 | 52000 | 430 |
| 22 | 9.251733e-01 | 288 | 52000 | 421 |
| 23 | 4.310000e-06 | 2558 | 54140 | 356 |
| 24 | 8.800000e-05 | 3306 | 54922 | 595 |
| 25 | 1.167829e-01 | 288 | 52000 | 437 |
| 26 | 2.129653e-02 | 287 | 52023 | 637 |
| 27 | 1.753650e-04 | 2558 | 54140 | 358 |
| 28 | 1.723800e-04 | 3306 | 54922 | 632 |
| 29 | 5.198080e-04 | 2558 | 54140 | 349 |
| .. | ... | ... | ... | ... |
| 569 | 9.280799e-02 | 346 | 51693 | 162 |
| 570 | 1.660742e-02 | 346 | 51693 | 148 |
| 571 | 5.047690e-02 | 346 | 51693 | 146 |
| 572 | -3.407500e-04 | 346 | 51693 | 81 |
| 573 | 7.900908e-02 | 346 | 51693 | 106 |
| 574 | 5.936759e-02 | 346 | 51693 | 108 |

```
575  5.382954e-01    364  52000       307
576  5.909181e-02    364  52000       302
577  6.916405e-02    346  51693        67
578  9.362785e-02    364  52000       221
579  1.424379e-01    364  52000       190
580  4.653732e-02    364  52000       161
581  4.449500e-02    364  52000       146
582  4.619161e-02    364  52000       107
583  1.149604e+00    364  52000        74
584 -7.320000e-05    348  51671       236
585  7.018682e-02    348  51671       163
586  5.894065e-02    348  51671       106
587  6.484750e-04   2895  54567       116
588  2.573870e-04    291  51928       310
589  5.910000e-07   2895  54567       117
590  5.791130e-04   2895  54567       119
591  1.251297e-02    290  51941       516
592  1.383053e-01    291  51928       318
593 -1.976670e-04   4016  55632       396
594  4.019753e-02    310  51990       224
595  6.460119e-02    266  51630       359
596  2.133720e-04    266  51630       346
597  1.404700e+00    266  51630       396
598 -3.480000e-05    266  51630       390

[599 rows x 18 columns]
```

In [5]:
```python
#To count the number of NAN values in each column
#NAN value in class column(categorical feature)
len(df) - df['class'].count()
```

Out[5]: 20

In [6]:
```python
#NAN values in i column or feature (numerical feature)
len(df) - df['i'].count()
```

Out[6]: 11

In [7]:
```python
# to find the number of NAN values of every feature
len(df) - df.count()
```

Out[7]:
```
objid         0
ra            0
dec           0
u             0
g             0
r             0
i            11
z             0
```

```
run           9
rerun         9
camcol        0
field         0
specobjid     0
class        20
redshift      0
plate         0
mjd           0
fiberid       0
dtype: int64
```

In [8]:
```python
# 1) Data Cleaning
```

In [9]:
```python
# All the NAN's for categorical columns  replaced with its previous row value
```

In [10]:
```python
df['class'].ffill(inplace=True)
```

In [11]:
```python
df['run'].ffill(inplace=True)
```

In [12]:
```python
df['rerun'].ffill(inplace=True)
```

In [13]:
```python
# All the NAN's for numeric columns  replaced with average of the column
```

In [14]:
```python
df['i'].fillna(df['i'].mean(),inplace=True)
```

In [15]:
```python
# Number of NAN values after filling with respective values
```

In [16]:
```python
len(df) - df.count()
```

Out[16]:
```
objid        0
ra           0
dec          0
u            0
g            0
r            0
i            0
z            0
run          0
rerun        0
camcol       0
field        0
specobjid    0
class        0
redshift     0
plate        0
mjd          0
fiberid      0
dtype: int64
```

In [17]: # now our data is complete with no NAN values

In [18]: # By studying the data we observe that objid and rerun values are the same for every ro

In [19]:
```python
del df['objid']
del df['rerun']
```

In [20]: # our dataset after Data Cleaning process

In [21]:
```python
df.to_csv('cleanedsky.csv', encoding='utf-8',index=False)
df
```

Out[21]:

| | ra | dec | u | g | r | i | z \ |
|---|---|---|---|---|---|---|---|
| 0 | 183.531326 | 0.089693 | 19.47406 | 17.04240 | 15.94699 | 15.503420 | 15.22531 |
| 1 | 183.598371 | 0.135285 | 18.66280 | 17.21449 | 16.67637 | 16.544191 | 16.39150 |
| 2 | 183.680207 | 0.126185 | 19.38298 | 18.19169 | 17.47428 | 17.087320 | 16.80125 |
| 3 | 183.870529 | 0.049911 | 17.76536 | 16.60272 | 16.16116 | 15.982330 | 15.90438 |
| 4 | 183.883288 | 0.102557 | 17.55025 | 16.26342 | 16.43869 | 16.544191 | 16.61326 |
| 5 | 183.847174 | 0.173694 | 19.43133 | 18.46779 | 18.16451 | 18.014750 | 18.04155 |
| 6 | 183.864379 | 0.019201 | 19.38322 | 17.88995 | 17.10537 | 16.663930 | 16.36955 |
| 7 | 183.900081 | 0.187473 | 18.97993 | 17.84496 | 17.38022 | 17.206730 | 17.07071 |
| 8 | 183.924588 | 0.097246 | 17.90616 | 16.97172 | 16.67541 | 16.544191 | 16.47596 |
| 9 | 183.973498 | 0.081626 | 18.67249 | 17.71375 | 17.49362 | 17.282840 | 17.22644 |
| 10 | 183.979195 | 0.135998 | 19.29772 | 17.80227 | 17.18266 | 16.923350 | 16.79928 |
| 11 | 184.085331 | 0.112110 | 18.83307 | 17.51785 | 16.94273 | 16.714180 | 16.60521 |
| 12 | 184.102098 | 0.191511 | 19.56250 | 18.19113 | 17.65759 | 17.475730 | 17.39203 |
| 13 | 184.160510 | 0.075645 | 19.57990 | 17.72815 | 16.98740 | 16.680760 | 16.50426 |
| 14 | 184.189574 | 0.099482 | 19.25667 | 17.54869 | 16.63578 | 16.149220 | 15.76639 |
| 15 | 184.350647 | 0.207230 | 18.73832 | 18.60962 | 18.39696 | 18.311740 | 17.97663 |
| 16 | 184.221797 | 0.046955 | 19.05958 | 18.09512 | 17.92766 | 16.544191 | 17.90772 |
| 17 | 184.245664 | 0.198257 | 19.22143 | 19.30248 | 19.13823 | 19.113510 | 19.23454 |
| 18 | 184.388870 | 0.068287 | 19.04397 | 17.51106 | 16.87335 | 16.611140 | 16.48303 |
| 19 | 184.380919 | 0.174323 | 17.81661 | 16.86976 | 16.53884 | 16.195760 | 16.08668 |
| 20 | 184.466853 | 0.111965 | 19.39320 | 18.48274 | 18.16551 | 18.051220 | 18.04328 |
| 21 | 184.569411 | 0.137091 | 17.51339 | 16.41793 | 16.06695 | 15.937510 | 15.89478 |
| 22 | 184.654170 | 0.122673 | 19.07731 | 18.64518 | 18.49678 | 18.526770 | 18.45765 |
| 23 | 184.658795 | 0.159936 | 18.63032 | 17.23437 | 16.72749 | 16.517720 | 16.44039 |
| 24 | 184.516834 | 0.145550 | 19.21684 | 18.29956 | 18.19589 | 18.147580 | 18.26658 |
| 25 | 184.618563 | 0.080576 | 19.41270 | 17.47109 | 16.49870 | 16.065020 | 15.66848 |
| 26 | 184.627408 | 0.013234 | 18.48543 | 17.33900 | 16.90179 | 16.667120 | 16.48684 |
| 27 | 184.676762 | 0.059442 | 17.72835 | 16.19783 | 15.66073 | 15.515860 | 15.43615 |
| 28 | 184.711970 | 0.066572 | 17.70824 | 16.61069 | 16.19772 | 16.544191 | 16.00229 |
| 29 | 184.770204 | 0.198010 | 18.18974 | 17.23738 | 16.89113 | 16.757270 | 16.71052 |
| .. | ... | ... | ... | ... | ... | ... | ... |
| 569 | 244.284910 | -0.455477 | 18.63084 | 16.69326 | 15.67451 | 15.205570 | 14.82077 |
| 570 | 244.491831 | -0.470566 | 18.43597 | 17.47617 | 17.10257 | 16.909710 | 16.82976 |
| 571 | 244.496791 | -0.498398 | 17.98780 | 16.51811 | 15.74451 | 15.304330 | 14.98556 |
| 572 | 244.916397 | -0.523076 | 18.45084 | 17.39167 | 16.97855 | 16.825990 | 16.76456 |
| 573 | 244.925582 | -0.455836 | 19.34707 | 17.76923 | 16.92714 | 16.484860 | 16.14608 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 574 | 244.935545 | -0.479247 | 18.53797 | 17.16947 | 16.50210 | 16.189330 | 15.95892 |
| 575 | 245.128493 | -0.499415 | 19.18215 | 18.79711 | 18.81150 | 18.614920 | 18.51464 |
| 576 | 245.367353 | -0.457074 | 18.12316 | 16.31978 | 15.37991 | 14.910120 | 14.53670 |
| 577 | 245.376694 | -0.457008 | 19.46922 | 17.96987 | 17.31625 | 16.927290 | 16.71774 |
| 578 | 245.978479 | -0.483455 | 19.38072 | 17.38953 | 16.26663 | 15.756590 | 15.36963 |
| 579 | 246.034378 | -0.473284 | 19.51987 | 17.56103 | 16.44300 | 15.953020 | 15.57304 |
| 580 | 246.268157 | -0.426990 | 19.46025 | 18.28482 | 17.78494 | 17.503770 | 17.31497 |
| 581 | 246.695626 | -0.510111 | 18.01060 | 16.81185 | 16.32209 | 16.036530 | 15.92627 |
| 582 | 246.782081 | -0.492432 | 17.60994 | 15.90911 | 15.02090 | 14.549550 | 14.19971 |
| 583 | 247.021139 | -0.448677 | 19.23512 | 19.08207 | 18.78046 | 18.811240 | 18.99055 |
| 584 | 248.269016 | -0.488920 | 19.27635 | 17.99304 | 17.47953 | 17.290880 | 18.28560 |
| 585 | 248.894520 | -0.452329 | 18.73576 | 16.83826 | 15.94847 | 15.538030 | 15.15638 |
| 586 | 249.496551 | -0.498468 | 18.99369 | 17.31122 | 16.48666 | 16.053320 | 15.74858 |
| 587 | 189.429821 | -0.131042 | 17.83883 | 16.62422 | 16.84056 | 17.035840 | 17.15817 |
| 588 | 189.448544 | -0.052745 | 18.97562 | 17.84829 | 17.49412 | 17.374000 | 17.35097 |
| 589 | 189.453801 | -0.097313 | 17.66081 | 16.62739 | 16.28842 | 16.172150 | 16.14751 |
| 590 | 189.468747 | -0.036000 | 17.28231 | 16.20900 | 15.76480 | 15.513750 | 15.47968 |
| 591 | 189.522249 | -0.027031 | 16.32175 | 14.72385 | 13.88480 | 13.469230 | 13.15061 |
| 592 | 189.497150 | -0.110585 | 19.25978 | 18.10524 | 17.42960 | 17.089640 | 16.85153 |
| 593 | 225.529253 | -0.029788 | 19.21136 | 18.73800 | 18.76889 | 18.961950 | 19.18182 |
| 594 | 225.494356 | -0.114317 | 19.43974 | 18.09289 | 17.40579 | 17.008100 | 16.82082 |
| 595 | 145.288576 | 0.346512 | 19.22352 | 18.00232 | 17.37943 | 17.011100 | 16.76524 |
| 596 | 145.292821 | 0.376326 | 19.31302 | 18.13527 | 17.67289 | 17.499960 | 17.44187 |
| 597 | 145.593807 | 0.278308 | 19.10790 | 18.92201 | 18.59314 | 18.475730 | 18.40461 |
| 598 | 145.637275 | 0.373311 | 19.20753 | 17.87736 | 17.30583 | 17.086230 | 16.96132 |

| | run | camcol | field | specobjid | class | redshift | plate | mjd \ |
|---|---|---|---|---|---|---|---|---|
| 0 | 752.0 | 4 | 267 | 3.720000e+18 | STAR | -8.960000e-06 | 3306 | 54922 |
| 1 | 752.0 | 4 | 267 | 3.640000e+17 | STAR | -5.490000e-05 | 323 | 51615 |
| 2 | 752.0 | 4 | 268 | 3.230000e+17 | GALAXY | 1.231112e-01 | 287 | 52023 |
| 3 | 752.0 | 4 | 269 | 3.720000e+18 | STAR | -1.106160e-04 | 3306 | 54922 |
| 4 | 752.0 | 4 | 269 | 3.720000e+18 | STAR | 5.903570e-04 | 3306 | 54922 |
| 5 | 752.0 | 4 | 269 | 3.650000e+17 | STAR | 3.146030e-04 | 324 | 51666 |
| 6 | 752.0 | 4 | 269 | 3.230000e+17 | GALAXY | 1.002423e-01 | 287 | 52023 |
| 7 | 752.0 | 4 | 269 | 3.720000e+18 | STAR | 3.148480e-04 | 3306 | 54922 |
| 8 | 752.0 | 4 | 270 | 3.640000e+17 | STAR | 8.910000e-05 | 323 | 51615 |
| 9 | 752.0 | 4 | 270 | 3.240000e+17 | GALAXY | 4.050813e-02 | 288 | 52000 |
| 10 | 752.0 | 4 | 270 | 3.720000e+18 | STAR | -3.460000e-05 | 3306 | 54922 |
| 11 | 752.0 | 4 | 271 | 3.720000e+18 | STAR | 6.227740e-04 | 3306 | 54922 |
| 12 | 752.0 | 4 | 271 | 3.720000e+18 | STAR | 5.470000e-05 | 3306 | 54922 |
| 13 | 752.0 | 4 | 271 | 3.720000e+18 | STAR | 8.300000e-06 | 3306 | 54922 |
| 14 | 752.0 | 4 | 271 | 3.240000e+17 | GALAXY | 7.208736e-02 | 288 | 52000 |
| 15 | 752.0 | 4 | 272 | 3.230000e+17 | QSO | 2.719369e-01 | 287 | 52023 |
| 16 | 752.0 | 4 | 272 | 3.650000e+17 | STAR | 2.099740e-04 | 324 | 51666 |
| 17 | 752.0 | 4 | 272 | 3.230000e+17 | QSO | 1.178098e+00 | 287 | 52023 |
| 18 | 752.0 | 4 | 273 | 3.640000e+17 | STAR | 2.400980e-04 | 323 | 51615 |
| 19 | 752.0 | 4 | 273 | 3.230000e+17 | GALAXY | 7.277206e-02 | 287 | 52023 |
| 20 | 752.0 | 4 | 273 | 3.650000e+17 | STAR | -2.702180e-04 | 324 | 51666 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 21 | 752.0 | 4 | 274 | 3.240000e+17 | STAR | -3.550000e-05 | 288 | 52000 |
| 22 | 752.0 | 4 | 274 | 3.240000e+17 | QSO | 9.251733e-01 | 288 | 52000 |
| 23 | 752.0 | 4 | 274 | 2.880000e+18 | STAR | 4.310000e-06 | 2558 | 54140 |
| 24 | 752.0 | 4 | 274 | 3.720000e+18 | STAR | 8.800000e-05 | 3306 | 54922 |
| 25 | 752.0 | 4 | 274 | 3.240000e+17 | GALAXY | 1.167829e-01 | 288 | 52000 |
| 26 | 752.0 | 4 | 274 | 3.230000e+17 | GALAXY | 2.129653e-02 | 287 | 52023 |
| 27 | 752.0 | 4 | 275 | 2.880000e+18 | STAR | 1.753650e-04 | 2558 | 54140 |
| 28 | 752.0 | 4 | 275 | 3.720000e+18 | STAR | 1.723800e-04 | 3306 | 54922 |
| 29 | 752.0 | 4 | 275 | 2.880000e+18 | STAR | 5.198080e-04 | 2558 | 54140 |
| .. | ... | ... | ... | ... | ... | ... | ... | ... |
| 569 | 745.0 | 2 | 571 | 3.900000e+17 | GALAXY | 9.280799e-02 | 346 | 51693 |
| 570 | 745.0 | 2 | 573 | 3.900000e+17 | GALAXY | 1.660742e-02 | 346 | 51693 |
| 571 | 745.0 | 2 | 573 | 3.900000e+17 | GALAXY | 5.047690e-02 | 346 | 51693 |
| 572 | 745.0 | 2 | 576 | 3.900000e+17 | STAR | -3.407500e-04 | 346 | 51693 |
| 573 | 745.0 | 2 | 576 | 3.900000e+17 | GALAXY | 7.900908e-02 | 346 | 51693 |
| 574 | 745.0 | 2 | 576 | 3.900000e+17 | GALAXY | 5.936759e-02 | 346 | 51693 |
| 575 | 745.0 | 2 | 577 | 4.100000e+17 | QSO | 5.382954e-01 | 364 | 52000 |
| 576 | 745.0 | 2 | 579 | 4.100000e+17 | GALAXY | 5.909181e-02 | 364 | 52000 |
| 577 | 745.0 | 2 | 579 | 3.900000e+17 | GALAXY | 6.916405e-02 | 346 | 51693 |
| 578 | 745.0 | 2 | 583 | 4.100000e+17 | GALAXY | 9.362785e-02 | 364 | 52000 |
| 579 | 745.0 | 2 | 583 | 4.100000e+17 | GALAXY | 1.424379e-01 | 364 | 52000 |
| 580 | 745.0 | 2 | 585 | 4.100000e+17 | GALAXY | 4.653732e-02 | 364 | 52000 |
| 581 | 745.0 | 2 | 588 | 4.100000e+17 | GALAXY | 4.449500e-02 | 364 | 52000 |
| 582 | 745.0 | 2 | 588 | 4.100000e+17 | GALAXY | 4.619161e-02 | 364 | 52000 |
| 583 | 745.0 | 2 | 590 | 4.100000e+17 | QSO | 1.149604e+00 | 364 | 52000 |
| 584 | 745.0 | 2 | 598 | 3.920000e+17 | STAR | -7.320000e-05 | 348 | 51671 |
| 585 | 745.0 | 2 | 602 | 3.920000e+17 | GALAXY | 7.018682e-02 | 348 | 51671 |
| 586 | 745.0 | 2 | 606 | 3.920000e+17 | GALAXY | 5.894065e-02 | 348 | 51671 |
| 587 | 745.0 | 3 | 205 | 3.260000e+18 | STAR | 6.484750e-04 | 2895 | 54567 |
| 588 | 745.0 | 3 | 205 | 3.280000e+17 | STAR | 2.573870e-04 | 291 | 51928 |
| 589 | 745.0 | 3 | 205 | 3.260000e+18 | STAR | 5.910000e-07 | 2895 | 54567 |
| 590 | 745.0 | 3 | 205 | 3.260000e+18 | STAR | 5.791130e-04 | 2895 | 54567 |
| 591 | 745.0 | 3 | 205 | 3.270000e+17 | GALAXY | 1.251297e-02 | 290 | 51941 |
| 592 | 745.0 | 3 | 205 | 3.280000e+17 | GALAXY | 1.383053e-01 | 291 | 51928 |
| 593 | 745.0 | 3 | 446 | 4.520000e+18 | STAR | -1.976670e-04 | 4016 | 55632 |
| 594 | 745.0 | 3 | 446 | 3.490000e+17 | GALAXY | 4.019753e-02 | 310 | 51990 |
| 595 | 756.0 | 4 | 197 | 3.000000e+17 | GALAXY | 6.460119e-02 | 266 | 51630 |
| 596 | 756.0 | 4 | 197 | 3.000000e+17 | STAR | 2.133720e-04 | 266 | 51630 |
| 597 | 756.0 | 4 | 199 | 3.000000e+17 | QSO | 1.404700e+00 | 266 | 51630 |
| 598 | 756.0 | 4 | 199 | 3.000000e+17 | STAR | -3.480000e-05 | 266 | 51630 |

| | fiberid |
|---|---|
| 0 | 491 |
| 1 | 541 |
| 2 | 513 |
| 3 | 510 |
| 4 | 512 |
| 5 | 594 |

| | |
|---|---|
| 6 | 559 |
| 7 | 515 |
| 8 | 595 |
| 9 | 400 |
| 10 | 506 |
| 11 | 547 |
| 12 | 544 |
| 13 | 546 |
| 14 | 389 |
| 15 | 587 |
| 16 | 35 |
| 17 | 583 |
| 18 | 628 |
| 19 | 632 |
| 20 | 623 |
| 21 | 430 |
| 22 | 421 |
| 23 | 356 |
| 24 | 595 |
| 25 | 437 |
| 26 | 637 |
| 27 | 358 |
| 28 | 632 |
| 29 | 349 |
| .. | ... |
| 569 | 162 |
| 570 | 148 |
| 571 | 146 |
| 572 | 81 |
| 573 | 106 |
| 574 | 108 |
| 575 | 307 |
| 576 | 302 |
| 577 | 67 |
| 578 | 221 |
| 579 | 190 |
| 580 | 161 |
| 581 | 146 |
| 582 | 107 |
| 583 | 74 |
| 584 | 236 |
| 585 | 163 |
| 586 | 106 |
| 587 | 116 |
| 588 | 310 |
| 589 | 117 |
| 590 | 119 |
| 591 | 516 |

```
592      318
593      396
594      224
595      359
596      346
597      396
598      390

[599 rows x 16 columns]
```

In [22]:
```
# 2)  Normalization and Standardization
```

In [23]:
```
# Why is normalization important?
''' The normal distribution is important because of the Central limit theorem.  In simp

Another corollary is that the normal distribution makes math easy - things like calcula
```

Out[23]:
```
" The normal distribution is important because of the Central limit theorem.  In simple
```

In [24]:
```
# How does it affect dataset?
'''Normalization is performed on data to remove amplitude variation and only focus on t

''' normalization brings any dataset to a comparable range. It could be to squash down
```

Out[24]:
```
' normalization brings any dataset to a comparable range. It could be to squash down th
```

In [25]:
```
#  Different graphs used to check whether the data is normal
# Histograms
# boxplots
# normal probability plot or qq plot or quantile-quantile plot
```

In [26]:
```
# Normalizing all the numeric columns, to make mean 0 and variance 1
```

In [27]:
```
# Normalizing only numeric features
cols_to_norm = ['ra','dec','u','g','r','i','z','redshift','mjd']
df[cols_to_norm] = df[cols_to_norm].apply(lambda x: (x - x.min()) / (x.max() - x.min())
```

In [28]:
```
df
```

Out[28]:
```
         ra       dec         u         g         r         i         z  \
0  0.526143  0.501836  0.973617  0.548873  0.340158  0.186836  0.290166
1  0.526624  0.518848  0.791571  0.579327  0.460469  0.282428  0.453268
2  0.527210  0.515452  0.953179  0.752260  0.592084  0.332313  0.510575
3  0.528573  0.486993  0.590185  0.471064  0.375485  0.230822  0.385140
4  0.528664  0.506636  0.541915  0.411019  0.421264  0.282428  0.484283
5  0.528406  0.533179  0.964029  0.801121  0.705937  0.417496  0.684043
6  0.528529  0.475534  0.953233  0.698862  0.531232  0.293426  0.450198
7  0.528785  0.538321  0.862735  0.690900  0.576569  0.343281  0.548262
8  0.528960  0.504654  0.621781  0.536365  0.460311  0.282428  0.465081
```

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 9 | 0.529310 | 0.498826 | 0.793745 | 0.667680 | 0.595274 | 0.350271 | 0.570042 |
| 10 | 0.529351 | 0.519114 | 0.934047 | 0.683345 | 0.543981 | 0.317253 | 0.510300 |
| 11 | 0.530111 | 0.510201 | 0.829779 | 0.633012 | 0.504405 | 0.298041 | 0.483157 |
| 12 | 0.530231 | 0.539827 | 0.993463 | 0.752161 | 0.622321 | 0.367988 | 0.593201 |
| 13 | 0.530650 | 0.496595 | 0.997368 | 0.670228 | 0.511773 | 0.294972 | 0.469039 |
| 14 | 0.530858 | 0.505489 | 0.924835 | 0.638470 | 0.453774 | 0.246151 | 0.365841 |
| 15 | 0.532012 | 0.545692 | 0.808517 | 0.826220 | 0.744280 | 0.444773 | 0.674963 |
| 16 | 0.531089 | 0.485890 | 0.880608 | 0.735170 | 0.666869 | 0.282428 | 0.665325 |
| 17 | 0.531260 | 0.542344 | 0.916927 | 0.948833 | 0.866552 | 0.518414 | 0.850893 |
| 18 | 0.532285 | 0.493849 | 0.877105 | 0.631811 | 0.492961 | 0.288577 | 0.466069 |
| 19 | 0.532228 | 0.533414 | 0.601686 | 0.518321 | 0.437783 | 0.250425 | 0.410636 |
| 20 | 0.532844 | 0.510147 | 0.955472 | 0.803766 | 0.706102 | 0.420845 | 0.684285 |
| 21 | 0.533578 | 0.519522 | 0.533643 | 0.438362 | 0.359945 | 0.226706 | 0.383797 |
| 22 | 0.534185 | 0.514142 | 0.884587 | 0.832513 | 0.760745 | 0.464523 | 0.742238 |
| 23 | 0.534219 | 0.528045 | 0.784282 | 0.582845 | 0.468901 | 0.279997 | 0.460106 |
| 24 | 0.533202 | 0.522678 | 0.915897 | 0.771349 | 0.711113 | 0.429696 | 0.715515 |
| 25 | 0.533930 | 0.498435 | 0.959848 | 0.624737 | 0.431162 | 0.238417 | 0.352147 |
| 26 | 0.533994 | 0.473308 | 0.751769 | 0.601362 | 0.497652 | 0.293719 | 0.466602 |
| 27 | 0.534347 | 0.490549 | 0.581880 | 0.399412 | 0.292939 | 0.187978 | 0.319654 |
| 28 | 0.534599 | 0.493209 | 0.577367 | 0.472474 | 0.381516 | 0.282428 | 0.398834 |
| 29 | 0.535016 | 0.542252 | 0.685416 | 0.583378 | 0.495894 | 0.301999 | 0.497886 |
| .. | ... | ... | ... | ... | ... | ... | ... |
| 569 | 0.961272 | 0.298420 | 0.784399 | 0.487087 | 0.295212 | 0.159479 | 0.233587 |
| 570 | 0.962754 | 0.292790 | 0.740670 | 0.625636 | 0.530771 | 0.316000 | 0.514563 |
| 571 | 0.962789 | 0.282405 | 0.640101 | 0.456091 | 0.306759 | 0.168550 | 0.256635 |
| 572 | 0.965795 | 0.273197 | 0.744007 | 0.610682 | 0.510313 | 0.308311 | 0.505444 |
| 573 | 0.965861 | 0.298286 | 0.945121 | 0.677498 | 0.501833 | 0.276979 | 0.418944 |
| 574 | 0.965932 | 0.289551 | 0.763559 | 0.571360 | 0.431723 | 0.249835 | 0.392768 |
| 575 | 0.967314 | 0.282025 | 0.908113 | 0.859399 | 0.812658 | 0.472620 | 0.750209 |
| 576 | 0.969025 | 0.297824 | 0.670475 | 0.420993 | 0.246618 | 0.132342 | 0.193857 |
| 577 | 0.969091 | 0.297849 | 0.972531 | 0.713005 | 0.566017 | 0.317615 | 0.498896 |
| 578 | 0.973402 | 0.287981 | 0.952672 | 0.610304 | 0.392882 | 0.210089 | 0.310350 |
| 579 | 0.973802 | 0.291776 | 0.983897 | 0.640654 | 0.421975 | 0.228130 | 0.338799 |
| 580 | 0.975476 | 0.309049 | 0.970518 | 0.768741 | 0.643327 | 0.370563 | 0.582424 |
| 581 | 0.978538 | 0.278035 | 0.645217 | 0.508073 | 0.402031 | 0.235801 | 0.388201 |
| 582 | 0.979157 | 0.284631 | 0.555309 | 0.348318 | 0.187399 | 0.099225 | 0.146726 |
| 583 | 0.980869 | 0.300957 | 0.919999 | 0.909828 | 0.807538 | 0.490651 | 0.816769 |
| 584 | 0.989807 | 0.285942 | 0.929251 | 0.717105 | 0.592950 | 0.351010 | 0.718175 |
| 585 | 0.994287 | 0.299594 | 0.807943 | 0.512747 | 0.340402 | 0.190015 | 0.280525 |
| 586 | 0.998599 | 0.282379 | 0.865822 | 0.596445 | 0.429176 | 0.237343 | 0.363350 |
| 587 | 0.568390 | 0.419475 | 0.606672 | 0.474869 | 0.487552 | 0.327585 | 0.560494 |
| 588 | 0.568524 | 0.448689 | 0.861767 | 0.691489 | 0.595357 | 0.358644 | 0.587459 |
| 589 | 0.568561 | 0.432060 | 0.566724 | 0.475430 | 0.396477 | 0.248257 | 0.419144 |
| 590 | 0.568668 | 0.454937 | 0.481789 | 0.401388 | 0.310106 | 0.187784 | 0.325742 |
| 591 | 0.569052 | 0.458284 | 0.266239 | 0.138565 | 0.000000 | 0.000000 | 0.000000 |
| 592 | 0.568872 | 0.427107 | 0.925533 | 0.736961 | 0.584714 | 0.332526 | 0.517608 |
| 593 | 0.826940 | 0.457255 | 0.914667 | 0.848939 | 0.805630 | 0.504494 | 0.843520 |
| 594 | 0.826690 | 0.425715 | 0.965916 | 0.734775 | 0.580787 | 0.325037 | 0.513312 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 595 | 0.252242 | 0.597662 | 0.917396 | 0.718748 | 0.576439 | 0.325313 | 0.505539 |
| 596 | 0.252272 | 0.608786 | 0.937480 | 0.742275 | 0.624845 | 0.370213 | 0.600172 |
| 597 | 0.254428 | 0.572213 | 0.891451 | 0.881503 | 0.776640 | 0.459835 | 0.734820 |
| 598 | 0.254739 | 0.607661 | 0.913808 | 0.696634 | 0.564298 | 0.332213 | 0.532963 |

| | run | camcol | field | specobjid | class | redshift | plate | mjd \ |
|---|---|---|---|---|---|---|---|---|
| 0 | 752.0 | 4 | 267 | 3.720000e+18 | STAR | 0.000305 | 3306 | 0.645394 |
| 1 | 752.0 | 4 | 267 | 3.640000e+17 | STAR | 0.000288 | 323 | 0.000000 |
| 2 | 752.0 | 4 | 268 | 3.230000e+17 | GALAXY | 0.043760 | 287 | 0.079625 |
| 3 | 752.0 | 4 | 269 | 3.720000e+18 | STAR | 0.000269 | 3306 | 0.645394 |
| 4 | 752.0 | 4 | 269 | 3.720000e+18 | STAR | 0.000516 | 3306 | 0.645394 |
| 5 | 752.0 | 4 | 269 | 3.650000e+17 | STAR | 0.000419 | 324 | 0.009953 |
| 6 | 752.0 | 4 | 269 | 3.230000e+17 | GALAXY | 0.035688 | 287 | 0.079625 |
| 7 | 752.0 | 4 | 269 | 3.720000e+18 | STAR | 0.000419 | 3306 | 0.645394 |
| 8 | 752.0 | 4 | 270 | 3.640000e+17 | STAR | 0.000339 | 323 | 0.000000 |
| 9 | 752.0 | 4 | 270 | 3.240000e+17 | GALAXY | 0.014605 | 288 | 0.075137 |
| 10 | 752.0 | 4 | 270 | 3.720000e+18 | STAR | 0.000296 | 3306 | 0.645394 |
| 11 | 752.0 | 4 | 271 | 3.720000e+18 | STAR | 0.000528 | 3306 | 0.645394 |
| 12 | 752.0 | 4 | 271 | 3.720000e+18 | STAR | 0.000327 | 3306 | 0.645394 |
| 13 | 752.0 | 4 | 271 | 3.720000e+18 | STAR | 0.000311 | 3306 | 0.645394 |
| 14 | 752.0 | 4 | 271 | 3.240000e+17 | GALAXY | 0.025751 | 288 | 0.075137 |
| 15 | 752.0 | 4 | 272 | 3.230000e+17 | QSO | 0.096288 | 287 | 0.079625 |
| 16 | 752.0 | 4 | 272 | 3.650000e+17 | STAR | 0.000382 | 324 | 0.009953 |
| 17 | 752.0 | 4 | 272 | 3.230000e+17 | QSO | 0.416117 | 287 | 0.079625 |
| 18 | 752.0 | 4 | 273 | 3.640000e+17 | STAR | 0.000393 | 323 | 0.000000 |
| 19 | 752.0 | 4 | 273 | 3.230000e+17 | GALAXY | 0.025993 | 287 | 0.079625 |
| 20 | 752.0 | 4 | 273 | 3.650000e+17 | STAR | 0.000212 | 324 | 0.009953 |
| 21 | 752.0 | 4 | 274 | 3.240000e+17 | STAR | 0.000295 | 288 | 0.075137 |
| 22 | 752.0 | 4 | 274 | 3.240000e+17 | QSO | 0.326847 | 288 | 0.075137 |
| 23 | 752.0 | 4 | 274 | 2.880000e+18 | STAR | 0.000309 | 2558 | 0.492779 |
| 24 | 752.0 | 4 | 274 | 3.720000e+18 | STAR | 0.000339 | 3306 | 0.645394 |
| 25 | 752.0 | 4 | 274 | 3.240000e+17 | GALAXY | 0.041526 | 288 | 0.075137 |
| 26 | 752.0 | 4 | 274 | 3.230000e+17 | GALAXY | 0.007824 | 287 | 0.079625 |
| 27 | 752.0 | 4 | 275 | 2.880000e+18 | STAR | 0.000370 | 2558 | 0.492779 |
| 28 | 752.0 | 4 | 275 | 3.720000e+18 | STAR | 0.000369 | 3306 | 0.645394 |
| 29 | 752.0 | 4 | 275 | 2.880000e+18 | STAR | 0.000491 | 2558 | 0.492779 |
| .. | ... | ... | ... | ... | ... | ... | ... | ... |
| 569 | 745.0 | 2 | 571 | 3.900000e+17 | GALAXY | 0.033064 | 346 | 0.015222 |
| 570 | 745.0 | 2 | 573 | 3.900000e+17 | GALAXY | 0.006169 | 346 | 0.015222 |
| 571 | 745.0 | 2 | 573 | 3.900000e+17 | GALAXY | 0.018124 | 346 | 0.015222 |
| 572 | 745.0 | 2 | 576 | 3.900000e+17 | STAR | 0.000188 | 346 | 0.015222 |
| 573 | 745.0 | 2 | 576 | 3.900000e+17 | GALAXY | 0.028194 | 346 | 0.015222 |
| 574 | 745.0 | 2 | 576 | 3.900000e+17 | GALAXY | 0.021262 | 346 | 0.015222 |
| 575 | 745.0 | 2 | 577 | 4.100000e+17 | QSO | 0.190299 | 364 | 0.075137 |
| 576 | 745.0 | 2 | 579 | 4.100000e+17 | GALAXY | 0.021164 | 364 | 0.075137 |
| 577 | 745.0 | 2 | 579 | 3.900000e+17 | GALAXY | 0.024719 | 346 | 0.015222 |
| 578 | 745.0 | 2 | 583 | 4.100000e+17 | GALAXY | 0.033354 | 364 | 0.075137 |
| 579 | 745.0 | 2 | 583 | 4.100000e+17 | GALAXY | 0.050581 | 364 | 0.075137 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 580 | 745.0 | 2 | 585 | 4.100000e+17 | GALAXY | 0.016733 | 364 | 0.075137 |
| 581 | 745.0 | 2 | 588 | 4.100000e+17 | GALAXY | 0.016012 | 364 | 0.075137 |
| 582 | 745.0 | 2 | 588 | 4.100000e+17 | GALAXY | 0.016611 | 364 | 0.075137 |
| 583 | 745.0 | 2 | 590 | 4.100000e+17 | QSO | 0.406060 | 364 | 0.075137 |
| 584 | 745.0 | 2 | 598 | 3.920000e+17 | STAR | 0.000282 | 348 | 0.010929 |
| 585 | 745.0 | 2 | 602 | 3.920000e+17 | GALAXY | 0.025080 | 348 | 0.010929 |
| 586 | 745.0 | 2 | 606 | 3.920000e+17 | GALAXY | 0.021111 | 348 | 0.010929 |
| 587 | 745.0 | 3 | 205 | 3.260000e+18 | STAR | 0.000537 | 2895 | 0.576112 |
| 588 | 745.0 | 3 | 205 | 3.280000e+17 | STAR | 0.000399 | 291 | 0.061085 |
| 589 | 745.0 | 3 | 205 | 3.260000e+18 | STAR | 0.000308 | 2895 | 0.576112 |
| 590 | 745.0 | 3 | 205 | 3.260000e+18 | STAR | 0.000512 | 2895 | 0.576112 |
| 591 | 745.0 | 3 | 205 | 3.270000e+17 | GALAXY | 0.004724 | 290 | 0.063622 |
| 592 | 745.0 | 3 | 205 | 3.280000e+17 | GALAXY | 0.049123 | 291 | 0.061085 |
| 593 | 745.0 | 3 | 446 | 4.520000e+18 | STAR | 0.000238 | 4016 | 0.783958 |
| 594 | 745.0 | 3 | 446 | 3.490000e+17 | GALAXY | 0.014496 | 310 | 0.073185 |
| 595 | 756.0 | 4 | 197 | 3.000000e+17 | GALAXY | 0.023109 | 266 | 0.002927 |
| 596 | 756.0 | 4 | 197 | 3.000000e+17 | STAR | 0.000383 | 266 | 0.002927 |
| 597 | 756.0 | 4 | 199 | 3.000000e+17 | QSO | 0.496096 | 266 | 0.002927 |
| 598 | 756.0 | 4 | 199 | 3.000000e+17 | STAR | 0.000296 | 266 | 0.002927 |

| | fiberid |
|---|---|
| 0 | 491 |
| 1 | 541 |
| 2 | 513 |
| 3 | 510 |
| 4 | 512 |
| 5 | 594 |
| 6 | 559 |
| 7 | 515 |
| 8 | 595 |
| 9 | 400 |
| 10 | 506 |
| 11 | 547 |
| 12 | 544 |
| 13 | 546 |
| 14 | 389 |
| 15 | 587 |
| 16 | 35 |
| 17 | 583 |
| 18 | 628 |
| 19 | 632 |
| 20 | 623 |
| 21 | 430 |
| 22 | 421 |
| 23 | 356 |
| 24 | 595 |
| 25 | 437 |
| 26 | 637 |

```
27      358
28      632
29      349
..      ...
569     162
570     148
571     146
572      81
573     106
574     108
575     307
576     302
577      67
578     221
579     190
580     161
581     146
582     107
583      74
584     236
585     163
586     106
587     116
588     310
589     117
590     119
591     516
592     318
593     396
594     224
595     359
596     346
597     396
598     390

[599 rows x 16 columns]
```

```
In [29]: scipy.stats.probplot(df['i'], sparams=(), dist='norm', fit=True, plot=None)
```

```
Out[29]: ((array([-3.0467878 , -2.76949799, -2.61391463, -2.50376377, -2.41756205,
                  -2.3462824 , -2.28524272, -2.23168992, -2.18386283, -2.14056426,
                  -2.10094199, -2.0643671 , -2.03036168, -1.9985538 , -1.96864807,
                  -1.94040588, -1.91363157, -1.88816269, -1.86386277, -1.84061606,
                  -1.81832347, -1.79689952, -1.77626993, -1.75636971, -1.73714168,
                  -1.71853525, -1.70050543, -1.68301204, -1.66601902, -1.6494939 ,
                  -1.63340733, -1.61773268, -1.6024457 , -1.58752429, -1.57294819,
                  -1.55869883, -1.5447591 , -1.53111323, -1.51774667, -1.5046459 ,
                  -1.4917984 , -1.47919252, -1.4668174 , -1.45466293, -1.44271964,
```

```
-1.43097867, -1.41943171, -1.40807096, -1.3968891 , -1.38587922,
-1.37503481, -1.36434974, -1.35381821, -1.34343473, -1.33319411,
-1.32309142, -1.313122  , -1.30328141, -1.29356543, -1.28397005,
-1.27449145, -1.26512599, -1.25587021, -1.24672078, -1.23767455,
-1.22872848, -1.21987968, -1.21112538, -1.20246293, -1.19388978,
-1.1854035 , -1.17700173, -1.16868224, -1.16044286, -1.15228152,
-1.14419621, -1.13618502, -1.12824609, -1.12037764, -1.11257795,
-1.10484537, -1.09717829, -1.08957517, -1.08203451, -1.07455489,
-1.0671349 , -1.05977321, -1.05246851, -1.04521953, -1.03802507,
-1.03088394, -1.023795  , -1.01675714, -1.00976928, -1.00283038,
-0.99593944, -0.98909547, -0.98229751, -0.97554465, -0.96883598,
-0.96217064, -0.95554777, -0.94896655, -0.94242618, -0.93592587,
-0.92946488, -0.92304245, -0.91665787, -0.91031044, -0.90399948,
-0.89772431, -0.8914843 , -0.88527881, -0.87910723, -0.87296895,
-0.86686338, -0.86078997, -0.85474813, -0.84873734, -0.84275706,
-0.83680677, -0.83088596, -0.82499413, -0.81913081, -0.81329551,
-0.80748777, -0.80170715, -0.79595319, -0.79022546, -0.78452354,
-0.77884701, -0.77319547, -0.76756852, -0.76196577, -0.75638683,
-0.75083134, -0.74529893, -0.73978923, -0.7343019 , -0.72883659,
-0.72339297, -0.7179707 , -0.71256945, -0.70718892, -0.70182878,
-0.69648873, -0.69116847, -0.6858677 , -0.68058613, -0.67532347,
-0.67007946, -0.66485381, -0.65964625, -0.65445652, -0.64928435,
-0.6441295 , -0.63899171, -0.63387073, -0.62876631, -0.62367823,
-0.61860625, -0.61355012, -0.60850964, -0.60348456, -0.59847468,
-0.59347978, -0.58849964, -0.58353405, -0.57858281, -0.57364572,
-0.56872256, -0.56381315, -0.5589173 , -0.5540348 , -0.54916548,
-0.54430914, -0.53946561, -0.53463469, -0.52981623, -0.52501003,
-0.52021593, -0.51543376, -0.51066334, -0.50590452, -0.50115713,
-0.49642101, -0.49169599, -0.48698193, -0.48227866, -0.47758604,
-0.47290392, -0.46823213, -0.46357055, -0.45891901, -0.45427739,
-0.44964553, -0.44502329, -0.44041055, -0.43580716, -0.43121298,
-0.42662789, -0.42205175, -0.41748443, -0.4129258 , -0.40837574,
-0.40383412, -0.39930081, -0.39477569, -0.39025865, -0.38574955,
-0.38124828, -0.37675472, -0.37226876, -0.36779027, -0.36331915,
-0.35885528, -0.35439855, -0.34994885, -0.34550607, -0.3410701 ,
-0.33664082, -0.33221815, -0.32780196, -0.32339215, -0.31898863,
-0.31459128, -0.31020001, -0.30581471, -0.30143528, -0.29706163,
-0.29269365, -0.28833125, -0.28397433, -0.27962279, -0.27527654,
-0.27093549, -0.26659953, -0.26226859, -0.25794255, -0.25362134,
-0.24930486, -0.24499302, -0.24068572, -0.23638289, -0.23208444,
-0.22779026, -0.22350028, -0.21921441, -0.21493257, -0.21065466,
-0.2063806 , -0.20211031, -0.1978437 , -0.19358069, -0.1893212 ,
-0.18506514, -0.18081242, -0.17656298, -0.17231672, -0.16807356,
-0.16383343, -0.15959625, -0.15536192, -0.15113038, -0.14690155,
-0.14267534, -0.13845167, -0.13423048, -0.13001168, -0.12579518,
-0.12158093, -0.11736883, -0.11315881, -0.1089508 , -0.10474472,
-0.10054048, -0.09633803, -0.09213727, -0.08793815, -0.08374057,
-0.07954446, -0.07534976, -0.07115638, -0.06696425, -0.0627733 ,
```

```
-0.05858345, -0.05439463, -0.05020676, -0.04601978, -0.0418336 ,
-0.03764815, -0.03346336, -0.02927916, -0.02509547, -0.02091222,
-0.01672934, -0.01254675, -0.00836438, -0.00418215,  0.        ,
 0.00418215,  0.00836438,  0.01254675,  0.01672934,  0.02091222,
 0.02509547,  0.02927916,  0.03346336,  0.03764815,  0.0418336 ,
 0.04601978,  0.05020676,  0.05439463,  0.05858345,  0.0627733 ,
 0.06696425,  0.07115638,  0.07534976,  0.07954446,  0.08374057,
 0.08793815,  0.09213727,  0.09633803,  0.10054048,  0.10474472,
 0.1089508 ,  0.11315881,  0.11736883,  0.12158093,  0.12579518,
 0.13001168,  0.13423048,  0.13845167,  0.14267534,  0.14690155,
 0.15113038,  0.15536192,  0.15959625,  0.16383343,  0.16807356,
 0.17231672,  0.17656298,  0.18081242,  0.18506514,  0.1893212 ,
 0.19358069,  0.1978437 ,  0.20211031,  0.2063806 ,  0.21065466,
 0.21493257,  0.21921441,  0.22350028,  0.22779026,  0.23208444,
 0.23638289,  0.24068572,  0.24499302,  0.24930486,  0.25362134,
 0.25794255,  0.26226859,  0.26659953,  0.27093549,  0.27527654,
 0.27962279,  0.28397433,  0.28833125,  0.29269365,  0.29706163,
 0.30143528,  0.30581471,  0.31020001,  0.31459128,  0.31898863,
 0.32339215,  0.32780196,  0.33221815,  0.33664082,  0.3410701 ,
 0.34550607,  0.34994885,  0.35439855,  0.35885528,  0.36331915,
 0.36779027,  0.37226876,  0.37675472,  0.38124828,  0.38574955,
 0.39025865,  0.39477569,  0.39930081,  0.40383412,  0.40837574,
 0.4129258 ,  0.41748443,  0.42205175,  0.42662789,  0.43121298,
 0.43580716,  0.44041055,  0.44502329,  0.44964553,  0.45427739,
 0.45891901,  0.46357055,  0.46823213,  0.47290392,  0.47758604,
 0.48227866,  0.48698193,  0.49169599,  0.49642101,  0.50115713,
 0.50590452,  0.51066334,  0.51543376,  0.52021593,  0.52501003,
 0.52981623,  0.53463469,  0.53946561,  0.54430914,  0.54916548,
 0.5540348 ,  0.5589173 ,  0.56381315,  0.56872256,  0.57364572,
 0.57858281,  0.58353405,  0.58849964,  0.59347978,  0.59847468,
 0.60348456,  0.60850964,  0.61355012,  0.61860625,  0.62367823,
 0.62876631,  0.63387073,  0.63899171,  0.6441295 ,  0.64928435,
 0.65445652,  0.65964625,  0.66485381,  0.67007946,  0.67532347,
 0.68058613,  0.6858677 ,  0.69116847,  0.69648873,  0.70182878,
 0.70718892,  0.71256945,  0.7179707 ,  0.72339297,  0.72883659,
 0.7343019 ,  0.73978923,  0.74529893,  0.75083134,  0.75638683,
 0.76196577,  0.76756852,  0.77319547,  0.77884701,  0.78452354,
 0.79022546,  0.79595319,  0.80170715,  0.80748777,  0.81329551,
 0.81913081,  0.82499413,  0.83088596,  0.83680677,  0.84275706,
 0.84873734,  0.85474813,  0.86078997,  0.86686338,  0.87296895,
 0.87910723,  0.88527881,  0.8914843 ,  0.89772431,  0.90399948,
 0.91031044,  0.91665787,  0.92304245,  0.92946488,  0.93592587,
 0.94242618,  0.94896655,  0.95554777,  0.96217064,  0.96883598,
 0.97554465,  0.98229751,  0.98909547,  0.99593944,  1.00283038,
 1.00976928,  1.01675714,  1.023795  ,  1.03088394,  1.03802507,
 1.04521953,  1.05246851,  1.05977321,  1.0671349 ,  1.07455489,
 1.08203451,  1.08957517,  1.09717829,  1.10484537,  1.11257795,
 1.12037764,  1.12824609,  1.13618502,  1.14419621,  1.15228152,
```

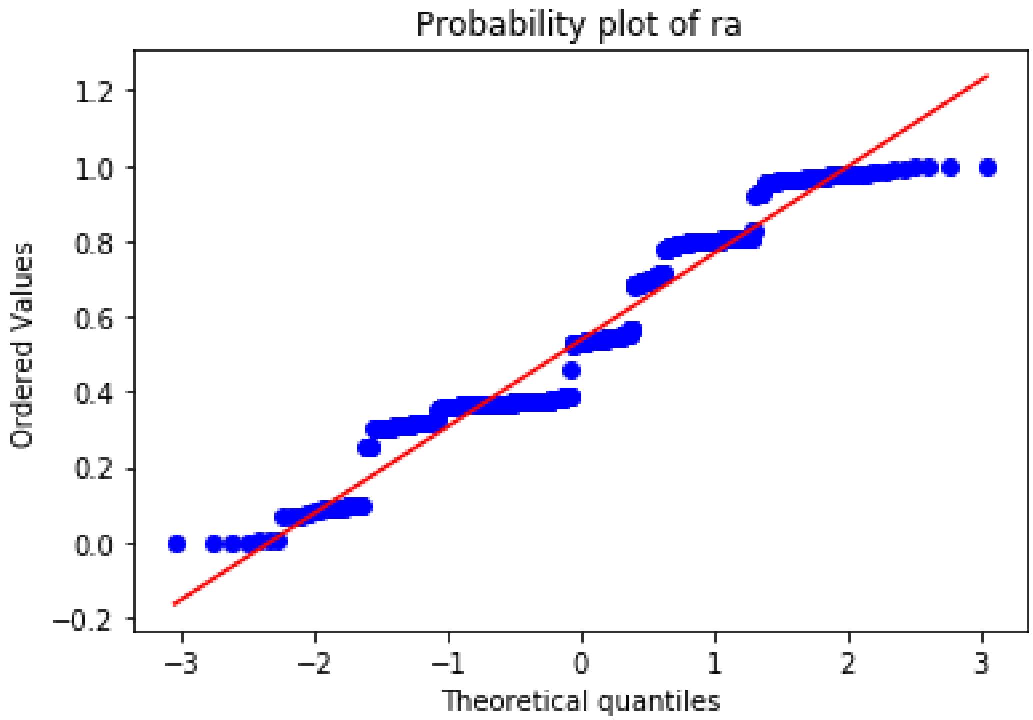
```
       1.16044286,  1.16868224,  1.17700173,  1.1854035 ,  1.19388978,
       1.20246293,  1.21112538,  1.21987968,  1.22872848,  1.23767455,
       1.24672078,  1.25587021,  1.26512599,  1.27449145,  1.28397005,
       1.29356543,  1.30328141,  1.313122  ,  1.32309142,  1.33319411,
       1.34343473,  1.35381821,  1.36434974,  1.37503481,  1.38587922,
       1.3968891 ,  1.40807096,  1.41943171,  1.43097867,  1.44271964,
       1.45466293,  1.4668174 ,  1.47919252,  1.4917984 ,  1.5046459 ,
       1.51774667,  1.53111323,  1.5447591 ,  1.55869883,  1.57294819,
       1.58752429,  1.6024457 ,  1.61773268,  1.63340733,  1.6494939 ,
       1.66601902,  1.68301204,  1.70050543,  1.71853525,  1.73714168,
       1.75636971,  1.77626993,  1.79689952,  1.81832347,  1.84061606,
       1.86386277,  1.88816269,  1.91363157,  1.94040588,  1.96864807,
       1.9985538 ,  2.03036168,  2.0643671 ,  2.10094199,  2.14056426,
       2.18386283,  2.23168992,  2.28524272,  2.3462824 ,  2.41756205,
       2.50376377,  2.61391463,  2.76949799,  3.0467878 ]),
array([0.        , 0.02599841, 0.0483725 , 0.04890063, 0.05535844,
       0.05831594, 0.05982224, 0.06825661, 0.07346621, 0.0758616 ,
       0.07802737, 0.08555521, 0.08562042, 0.08562042, 0.08615222,
       0.09536729, 0.09544628, 0.09720792, 0.09862881, 0.0986784 ,
       0.0992249 , 0.10218882, 0.10219158, 0.10249468, 0.10264071,
       0.1036042 , 0.10603908, 0.10738832, 0.10791001, 0.10833068,
       0.1083766 , 0.10866408, 0.10988107, 0.11107692, 0.11272192,
       0.11457632, 0.11678709, 0.11973173, 0.12130784, 0.12326603,
       0.12359668, 0.12609678, 0.12642926, 0.12841869, 0.12958699,
       0.12985151, 0.13125402, 0.13234242, 0.1327888 , 0.13359155,
       0.13629922, 0.13765397, 0.1382666 , 0.13837222, 0.1394349 ,
       0.14063627, 0.1442321 , 0.14434048, 0.14492188, 0.14634644,
       0.14660912, 0.14796387, 0.14904768, 0.14933424, 0.15003412,
       0.15167636, 0.15177739, 0.15225775, 0.15243502, 0.15531904,
       0.1553769 , 0.15591146, 0.15794496, 0.15875047, 0.15947882,
       0.16198259, 0.16238121, 0.16317753, 0.16365422, 0.16441196,
       0.1645929 , 0.16515776, 0.16664202, 0.16684317, 0.1685497 ,
       0.16864614, 0.16957931, 0.16983557, 0.17081374, 0.17116644,
       0.1720693 , 0.17278847, 0.17393197, 0.17408628, 0.17516365,
       0.17594435, 0.17602518, 0.17762425, 0.17822677, 0.18070758,
       0.18175372, 0.18382397, 0.18482235, 0.18494175, 0.18499594,
       0.18502901, 0.18511902, 0.18599525, 0.18674381, 0.18683565,
       0.18752819, 0.18762463, 0.18778444, 0.18787445, 0.18797824,
       0.1900145 , 0.19328704, 0.19466751, 0.19589827, 0.19636669,
       0.19691869, 0.19737058, 0.19859767, 0.19881719, 0.19894853,
       0.19982384, 0.20055311, 0.20057056, 0.20092233, 0.20102337,
       0.20210166, 0.20262152, 0.20274367, 0.20332783, 0.20393953,
       0.20398637, 0.20457695, 0.20463849, 0.20468533, 0.20491036,
       0.20571587, 0.20618888, 0.20685294, 0.20780265, 0.20785775,
       0.20965062, 0.21007496, 0.21008873, 0.21017599, 0.21143063,
       0.21210571, 0.2124878 , 0.21325381, 0.21344393, 0.21493186,
       0.21535712, 0.21657502, 0.21728959, 0.2186223 , 0.21872058,
       0.21888591, 0.21897592, 0.21903103, 0.21914308, 0.22045834,
```

0.22108015, 0.22134375, 0.22197934, 0.22281882, 0.22421123,
0.22433247, 0.22468333, 0.22546496, 0.22566427, 0.22583051,
0.22583694, 0.22670582, 0.22694738, 0.2270677 , 0.22736896,
0.22813038, 0.22850603, 0.22895976, 0.22916642, 0.23009867,
0.23040912, 0.23072415, 0.23082243, 0.23118339, 0.23237098,
0.23288625, 0.2337983 , 0.23407292, 0.23407843, 0.23459921,
0.23483801, 0.23489404, 0.23580058, 0.2373427 , 0.23767427,
0.23841732, 0.23890135, 0.2395994 , 0.23977942, 0.24004027,
0.24013671, 0.24091649, 0.24168893, 0.24244851, 0.24393828,
0.24402186, 0.24441405, 0.2447842 , 0.24529487, 0.2459029 ,
0.24601588, 0.24615089, 0.24657798, 0.24707947, 0.24753504,
0.24800438, 0.24803928, 0.24807418, 0.24825696, 0.2492728 ,
0.2498349 , 0.2502574 , 0.25042548, 0.25043467, 0.25106474,
0.25217794, 0.25410674, 0.25442361, 0.25449066, 0.25555977,
0.25575265, 0.2567593 , 0.25738662, 0.25740866, 0.25744265,
0.25763553, 0.25822886, 0.2590509 , 0.25908488, 0.25920796,
0.25926766, 0.25972323, 0.25979487, 0.26034687, 0.26067201,
0.26080336, 0.26144537, 0.26163825, 0.26219668, 0.2626596 ,
0.26272389, 0.26291401, 0.2630426 , 0.26425775, 0.26435511,
0.26447083, 0.26548483, 0.26551514, 0.26560332, 0.26718952,
0.2675064 , 0.26752385, 0.26802901, 0.26803177, 0.26825312,
0.2683165 , 0.26911649, 0.26918997, 0.27074403, 0.27098559,
0.27112336, 0.27130338, 0.27139615, 0.2716074 , 0.27248914,
0.27258007, 0.27304665, 0.2734021 , 0.27417179, 0.27450244,
0.27456581, 0.27575983, 0.2758278 , 0.2761309 , 0.27626224,
0.27672056, 0.27674995, 0.27697865, 0.27711734, 0.27848771,
0.27960641, 0.27999677, 0.2803761 , 0.28040182, 0.28110629,
0.28136346, 0.28137173, 0.28184658, 0.28198986, 0.2824281 ,
0.2824281 , 0.2824281 , 0.2824281 , 0.2824281 , 0.2824281 ,
0.2824281 , 0.2824281 , 0.2824281 , 0.2824281 , 0.2824281 ,
0.28270719, 0.28290375, 0.28336482, 0.28368629, 0.28399398,
0.28409226, 0.28411246, 0.2846415 , 0.28509707, 0.28524127,
0.28634803, 0.28675676, 0.28692392, 0.28803068, 0.28848992,
0.28857718, 0.28911541, 0.28959485, 0.28992367, 0.29026534,
0.29046924, 0.29163203, 0.29230344, 0.29283248, 0.29287381,
0.29342582, 0.29351675, 0.29371881, 0.29405314, 0.29457575,
0.29472087, 0.29480629, 0.29497161, 0.29497437, 0.29678744,
0.29690868, 0.29708136, 0.29804116, 0.29835712, 0.29876492,
0.29891923, 0.29946664, 0.29990292, 0.30073138, 0.30199888,
0.30404984, 0.3040976 , 0.30430518, 0.30433641, 0.30455868,
0.30495913, 0.30784591, 0.30831065, 0.30854395, 0.30928057,
0.31140408, 0.31144633, 0.3117476 , 0.31241074, 0.31446629,
0.314661  , 0.31600014, 0.31627293, 0.31670462, 0.31712711,
0.31725295, 0.31761483, 0.31830552, 0.31900264, 0.32004787,
0.32061549, 0.32214016, 0.32219894, 0.32228436, 0.32241203,
0.32267196, 0.32503704, 0.32531258, 0.32570844, 0.32595276,
0.32632842, 0.326412  , 0.32655712, 0.32758489, 0.32970933,
0.33038717, 0.33052402, 0.33094468, 0.33129554, 0.33152608,

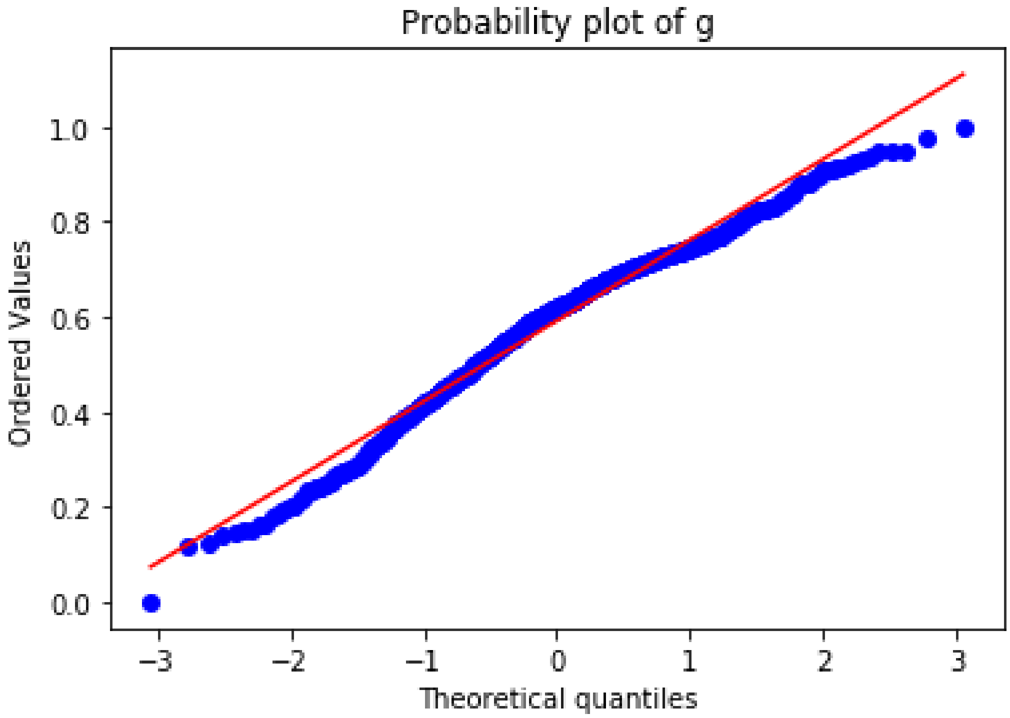
```
            0.33155914, 0.33156282, 0.3322131 , 0.33226545, 0.33229209,
            0.33231321, 0.3325263 , 0.33284868, 0.33296992, 0.33331435,
            0.33404362, 0.33506221, 0.33507232, 0.33511916, 0.33547461,
            0.33573086, 0.33586037, 0.33600916, 0.33687253, 0.33693131,
            0.33791776, 0.33836597, 0.33946264, 0.34024426, 0.34034988,
            0.34154207, 0.34189201, 0.34205917, 0.34215377, 0.34216388,
            0.34288396, 0.34304745, 0.34328074, 0.34388051, 0.3439246 ,
            0.34402104, 0.34402563, 0.34424239, 0.3443223 , 0.34526741,
            0.34542722, 0.34584054, 0.34616384, 0.34778404, 0.34966508,
            0.35011238, 0.35027127, 0.3503273 , 0.35054773, 0.35100973,
            0.35324714, 0.35425011, 0.35465975, 0.35485263, 0.35490499,
            0.35551761, 0.35558925, 0.35604114, 0.35854032, 0.35864411,
            0.35888291, 0.35974169, 0.36083284, 0.3608411 , 0.36163099,
            0.36184867, 0.36214259, 0.36251824, 0.36292513, 0.36308219,
            0.36351571, 0.36365532, 0.36420916, 0.36422937, 0.36590742,
            0.36624818, 0.36644749, 0.36670558, 0.36682866, 0.36724197,
            0.36759926, 0.36798777, 0.36911199, 0.37004332, 0.37014344,
            0.37021324, 0.37056318, 0.37418933, 0.37526395, 0.37613374,
            0.37621457, 0.37646991, 0.37649195, 0.37724694, 0.37732501,
            0.37779711, 0.37823063, 0.37970386, 0.38055989, 0.38076103,
            0.38175115, 0.38287353, 0.38406571, 0.38417501, 0.38433758,
            0.3846269 , 0.38795546, 0.38885281, 0.38971343, 0.38988151,
            0.38996601, 0.39033799, 0.39152834, 0.3936142 , 0.39434806,
            0.39559811, 0.39617216, 0.3962438 , 0.39692531, 0.39776388,
            0.40071494, 0.40113744, 0.40266028, 0.40336291, 0.40341435,
            0.40623866, 0.40801775, 0.40818216, 0.41053071, 0.41487969,
            0.41612239, 0.41738071, 0.41749552, 0.41836899, 0.41990744,
            0.4208452 , 0.42175082, 0.42229823, 0.42314507, 0.42534666,
            0.42728005, 0.42845937, 0.42969564, 0.4311845 , 0.43205613,
            0.43402259, 0.43712796, 0.43774609, 0.44327257, 0.44424707,
            0.44477336, 0.44579838, 0.44586819, 0.44847942, 0.45034025,
            0.4505515 , 0.45091613, 0.45154254, 0.45194942, 0.45203576,
            0.45367616, 0.45468832, 0.45819966, 0.45889862, 0.45983546,
            0.46259089, 0.46400076, 0.46452337, 0.47032723, 0.47129346,
            0.47261974, 0.47386152, 0.47545784, 0.47620731, 0.47829685,
            0.47964701, 0.4797159 , 0.48599644, 0.48815119, 0.49065128,
            0.49438948, 0.49748842, 0.49752149, 0.49784112, 0.4992519 ,
            0.50449365, 0.50468837, 0.5099246 , 0.51119026, 0.51841408,
            0.57123018, 0.57637549, 0.62184469, 1.        ])),
    (0.10540330888896787, 0.282428095613796, 0.9865696042348661))
```

In [30]:
```
res = scipy.stats.probplot(df['ra'], plot=plt)
plt.title("Probability plot of ra")
```

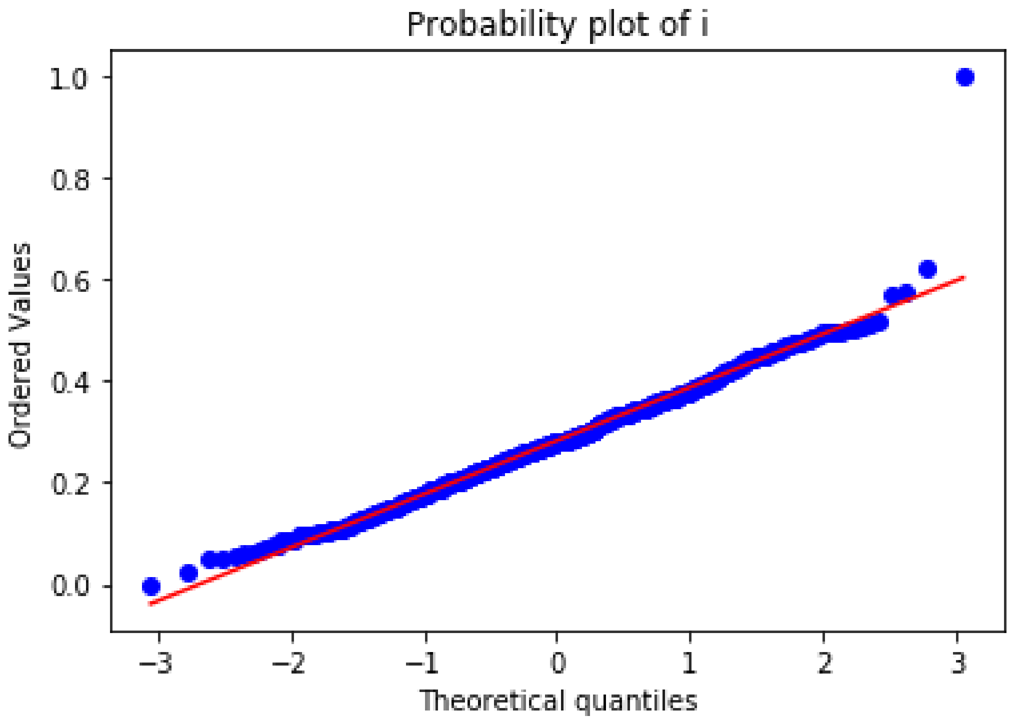
Out[30]:
```
Text(0.5,1,'Probability plot of ra')
```



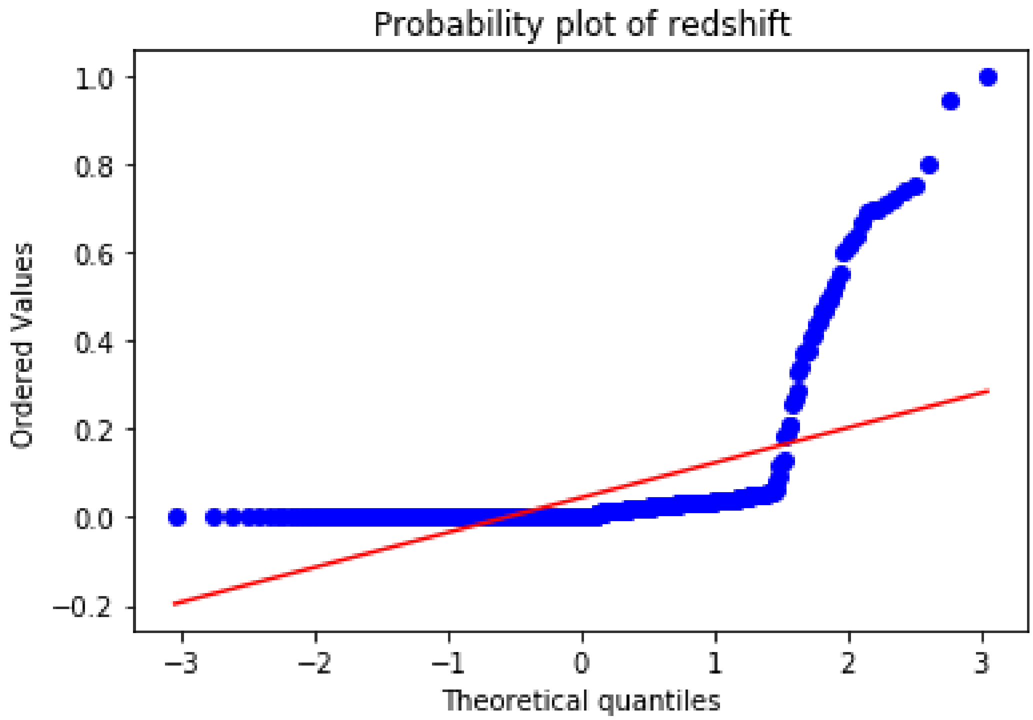

```
In [31]: res = scipy.stats.probplot(df['g'], plot=plt)
         plt.title("Probability plot of g")

Out[31]: Text(0.5,1,'Probability plot of g')
```



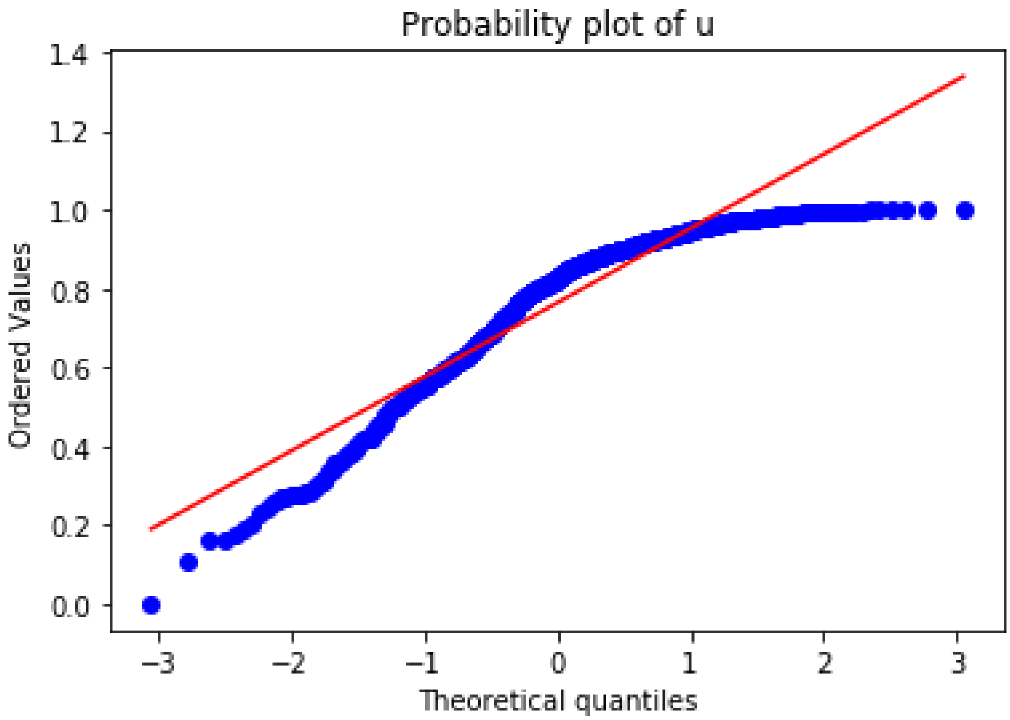

```
In [32]: res = scipy.stats.probplot(df['i'], plot=plt)
         plt.title("Probability plot of i")

Out[32]: Text(0.5,1,'Probability plot of i')
```



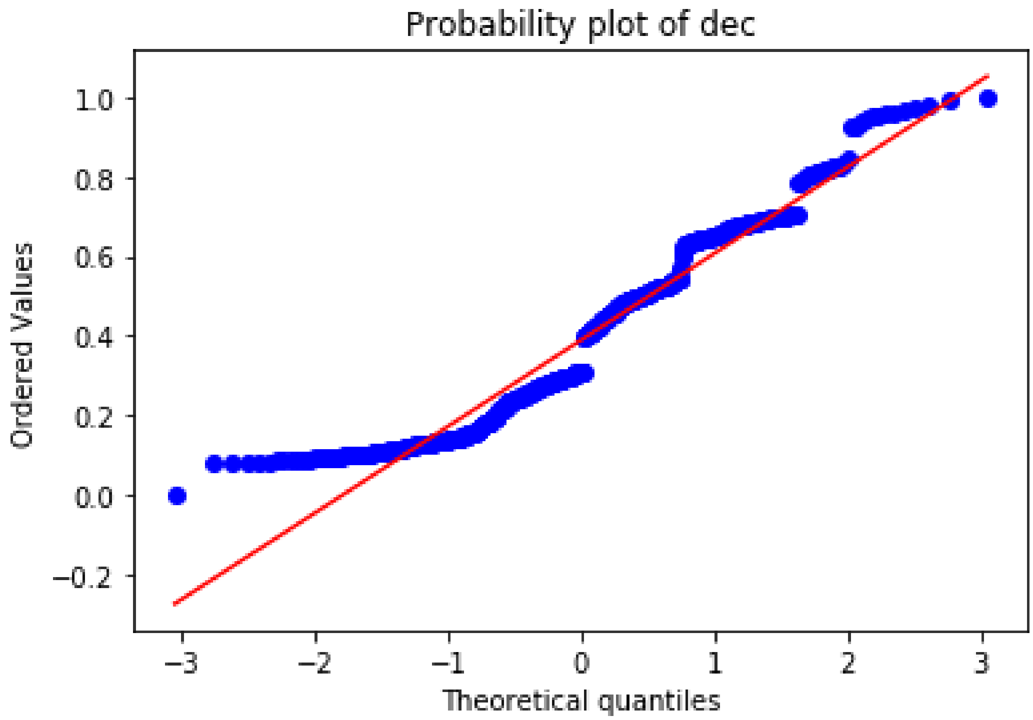

```
In [33]: res = scipy.stats.probplot(df['redshift'], plot=plt)
         plt.title("Probability plot of redshift")

Out[33]: Text(0.5,1,'Probability plot of redshift')
```



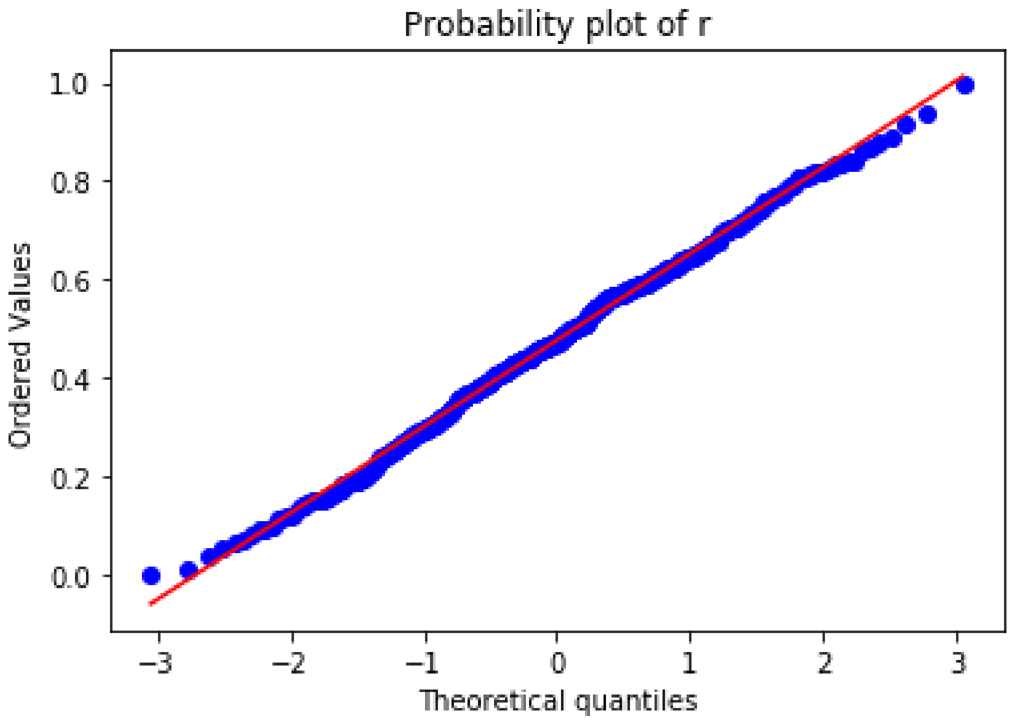

```
In [34]: res = scipy.stats.probplot(df['u'], plot=plt)
         plt.title("Probability plot of u")

Out[34]: Text(0.5,1,'Probability plot of u')
```



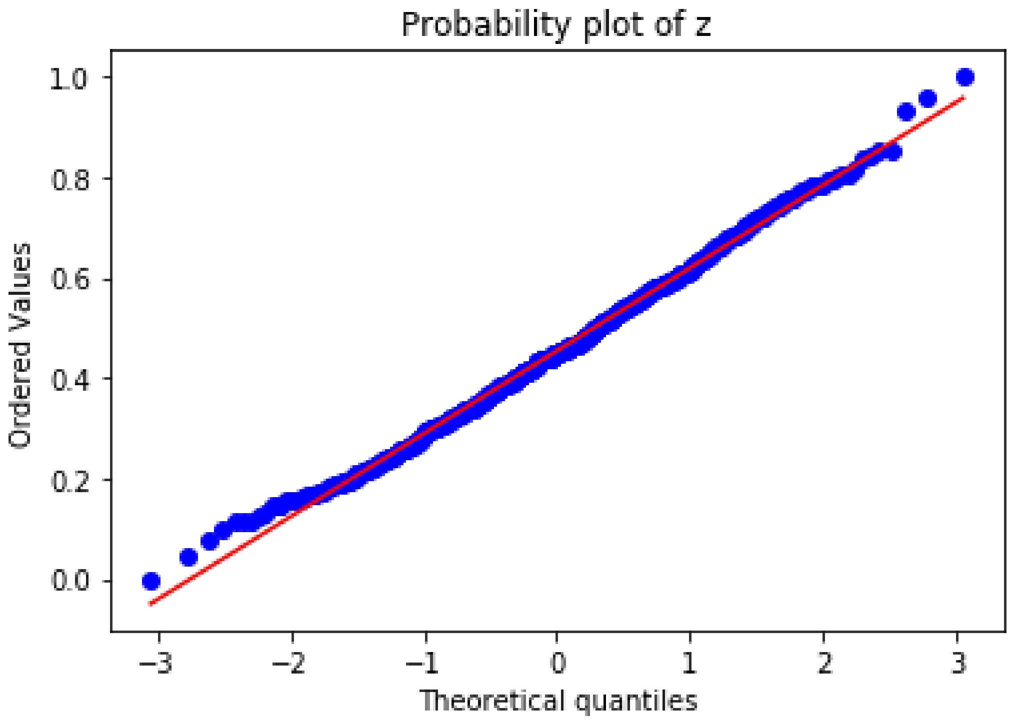

```
In [35]: res = scipy.stats.probplot(df['dec'], plot=plt)
         plt.title("Probability plot of dec")

Out[35]: Text(0.5,1,'Probability plot of dec')
```



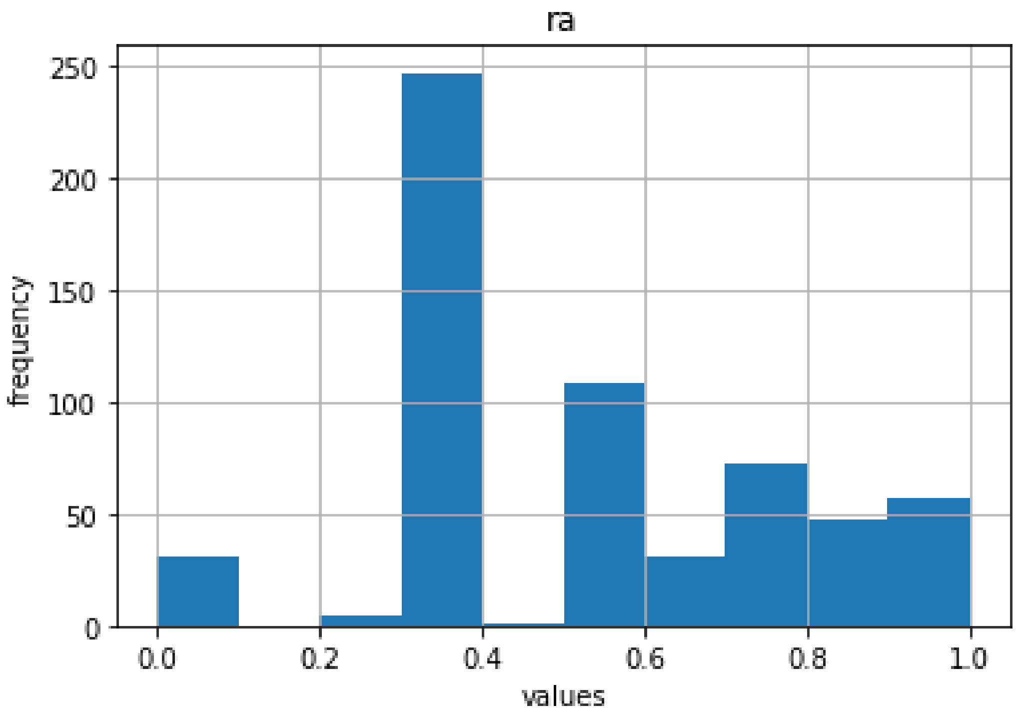

```
In [36]: res = scipy.stats.probplot(df['r'], plot=plt)
         plt.title("Probability plot of r")

Out[36]: Text(0.5,1,'Probability plot of r')
```



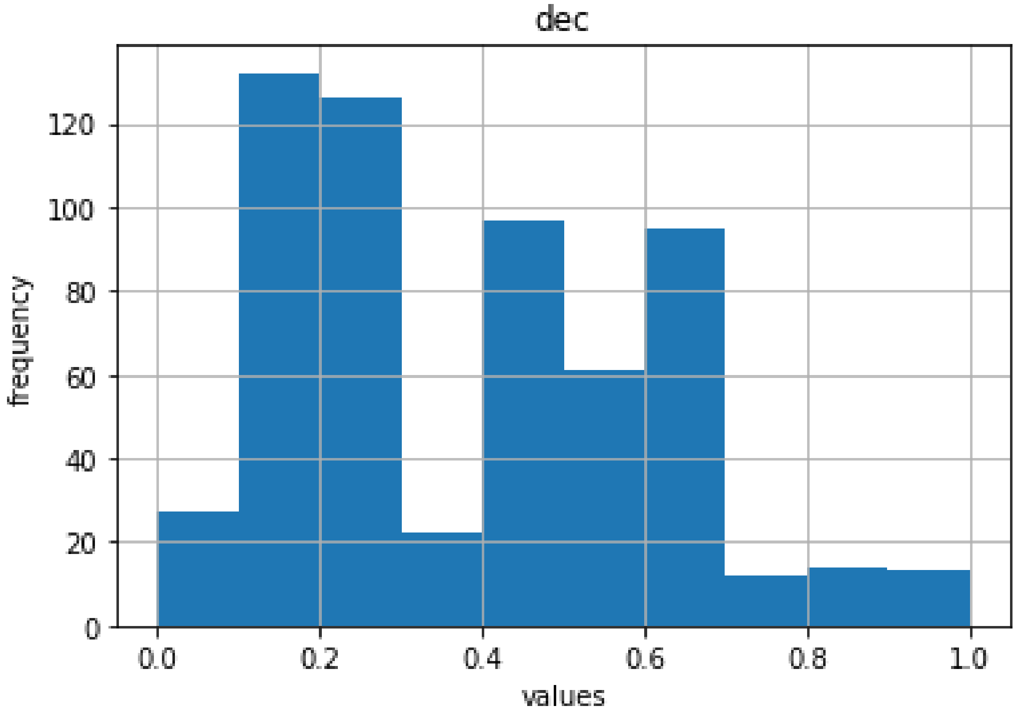

```
In [37]: res = scipy.stats.probplot(df['z'], plot=plt)
         plt.title("Probability plot of z")

Out[37]: Text(0.5,1,'Probability plot of z')
```



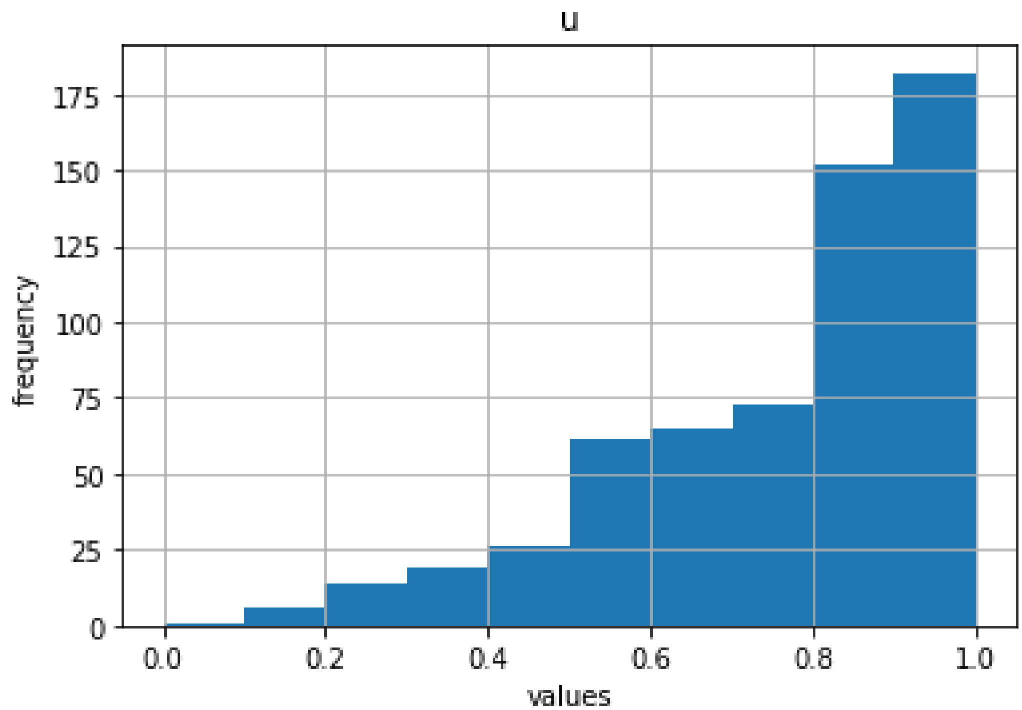

```
In [38]: df.hist(column='ra')
         pl.ylabel("frequency")
         pl.xlabel("values")

Out[38]: Text(0.5,0,'values')
```



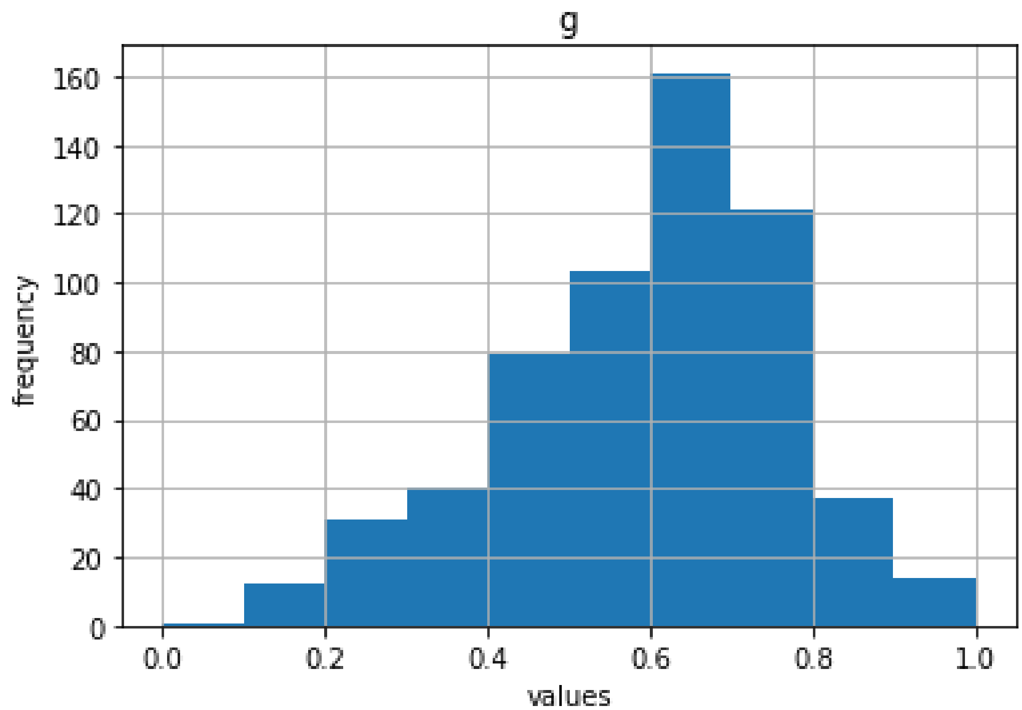

```
In [39]: df.hist(column='dec')
         pl.ylabel("frequency")
         pl.xlabel("values")

Out[39]: Text(0.5,0,'values')
```



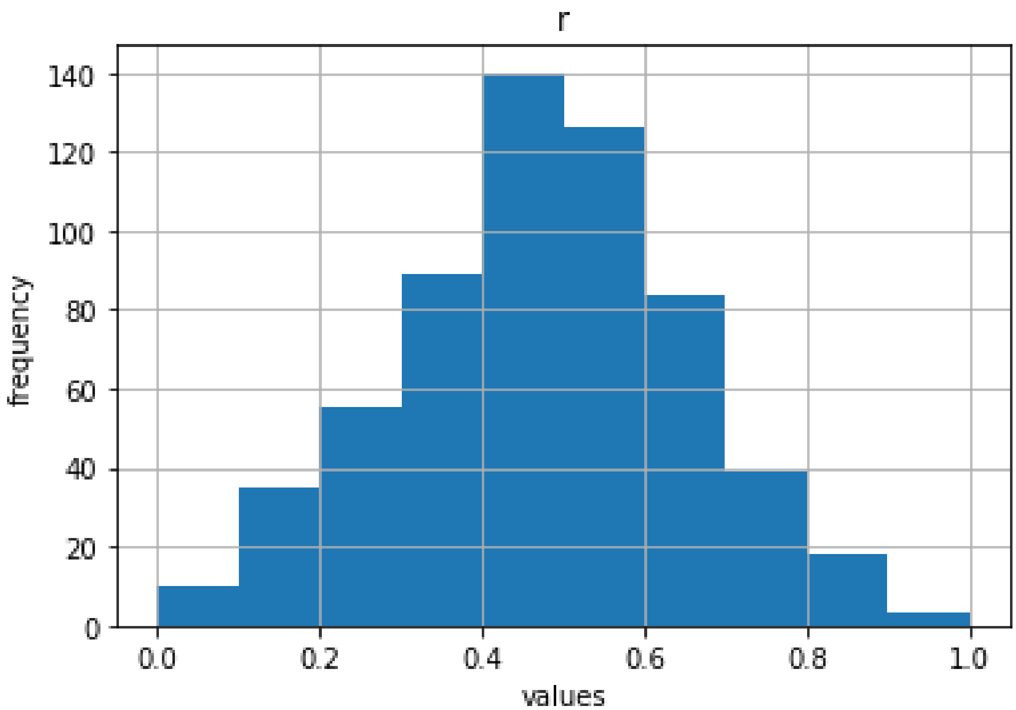

```
In [40]: df.hist(column='u')
         pl.ylabel("frequency")
         pl.xlabel("values")

Out[40]: Text(0.5,0,'values')
```



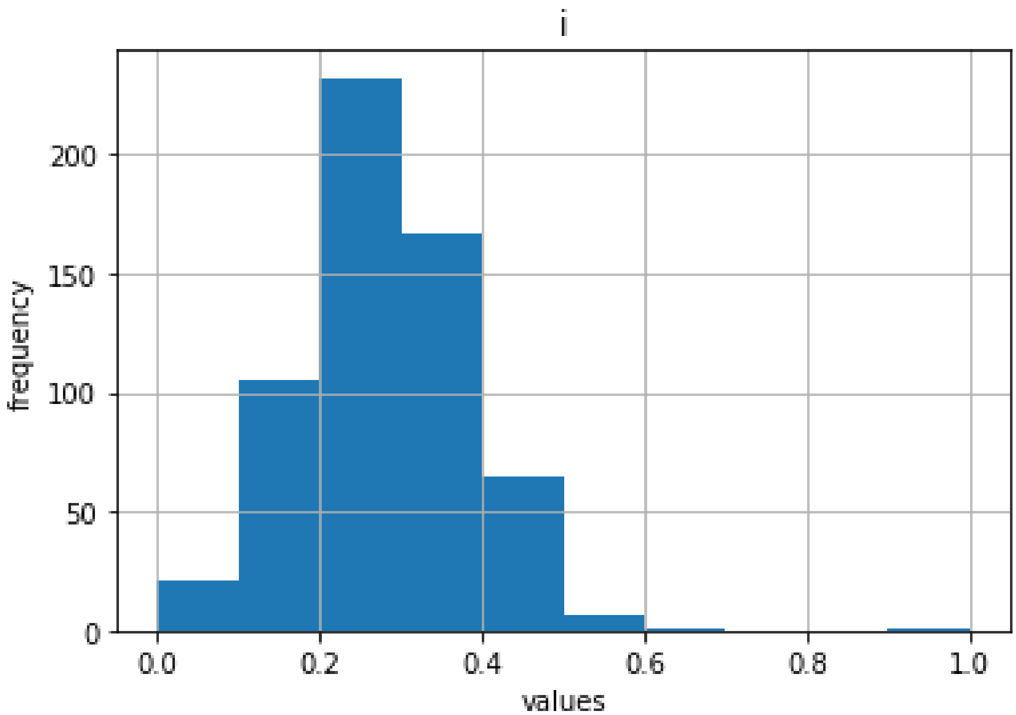

```
In [41]: df.hist(column='g')
         pl.ylabel("frequency")
         pl.xlabel("values")

Out[41]: Text(0.5,0,'values')
```



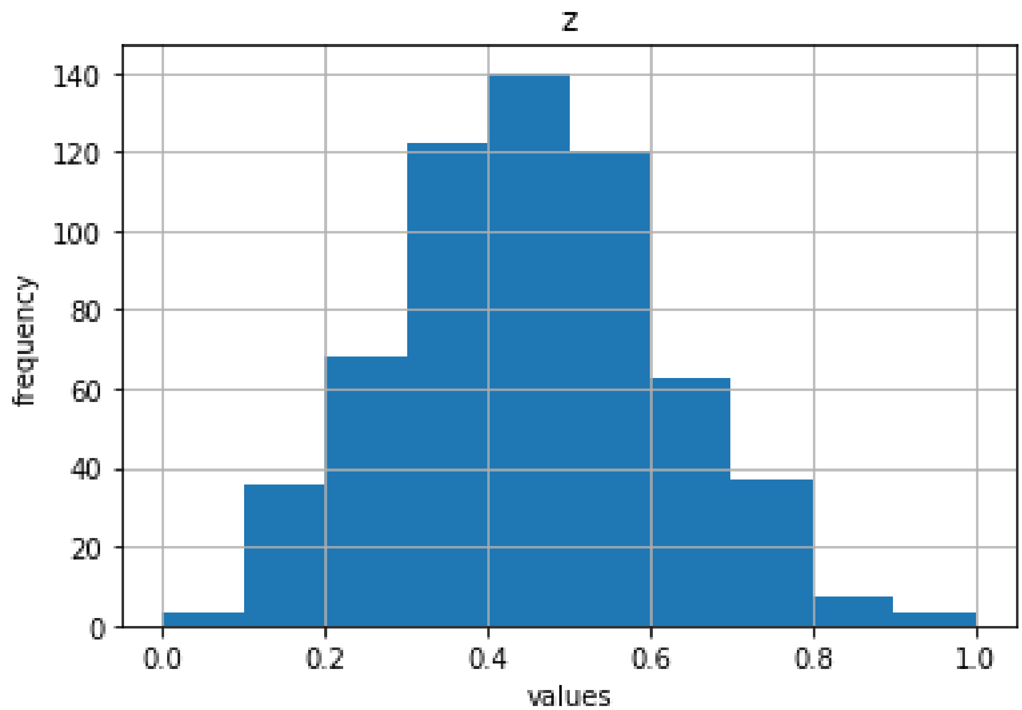

```
In [42]: df.hist(column='r')
         pl.ylabel("frequency")
         pl.xlabel("values")

Out[42]: Text(0.5,0,'values')
```



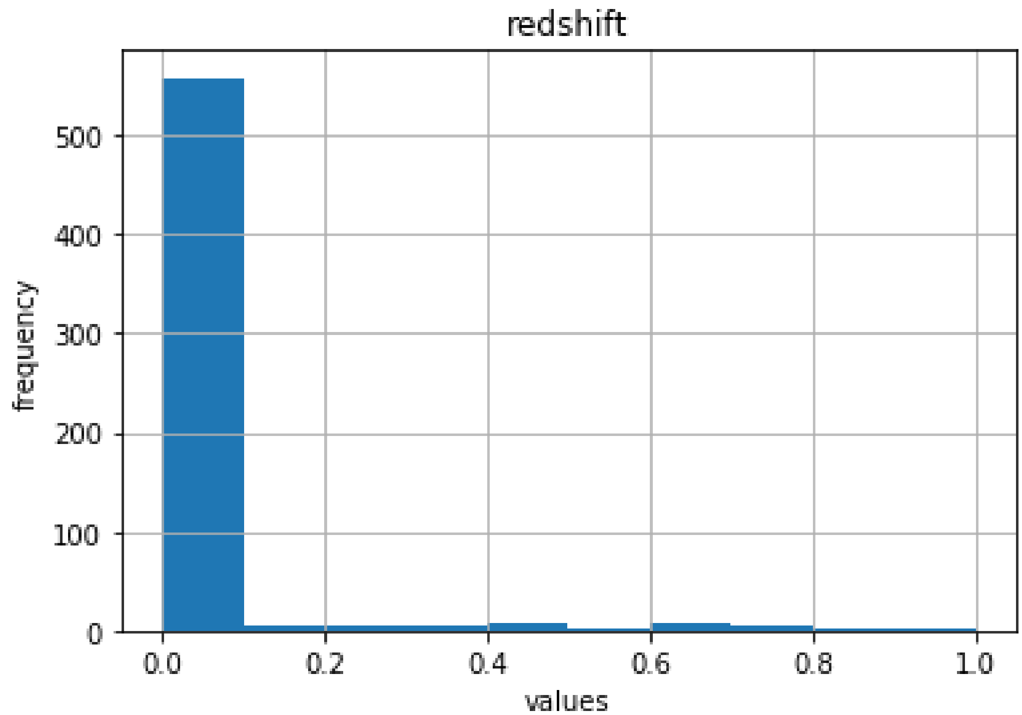

```
In [43]: df.hist(column='i')
         pl.ylabel("frequency")
         pl.xlabel("values")

Out[43]: Text(0.5,0,'values')
```




```
In [44]: df.hist(column='z')
         pl.ylabel("frequency")
         pl.xlabel("values")

Out[44]: Text(0.5,0,'values')
```




```
In [45]: df.hist(column='redshift')
         pl.ylabel("frequency")
         pl.xlabel("values")

Out[45]: Text(0.5,0,'values')
```




In [46]: *# 3)  Graph visualization*

In [37]: *#loading cleaned dataset*

```
df1 = pd.read_csv('cleanedsky.csv')
```

In [48]: df1

Out[48]:

| | ra | dec | u | g | r | i | z | \ |
|---|---|---|---|---|---|---|---|---|
| 0 | 183.531326 | 0.089693 | 19.47406 | 17.04240 | 15.94699 | 15.503420 | 15.22531 | |
| 1 | 183.598370 | 0.135285 | 18.66280 | 17.21449 | 16.67637 | 16.544191 | 16.39150 | |
| 2 | 183.680207 | 0.126185 | 19.38298 | 18.19169 | 17.47428 | 17.087320 | 16.80125 | |
| 3 | 183.870529 | 0.049911 | 17.76536 | 16.60272 | 16.16116 | 15.982330 | 15.90438 | |
| 4 | 183.883288 | 0.102557 | 17.55025 | 16.26342 | 16.43869 | 16.544191 | 16.61326 | |
| 5 | 183.847174 | 0.173694 | 19.43133 | 18.46779 | 18.16451 | 18.014750 | 18.04155 | |
| 6 | 183.864379 | 0.019201 | 19.38322 | 17.88995 | 17.10537 | 16.663930 | 16.36955 | |
| 7 | 183.900081 | 0.187473 | 18.97993 | 17.84496 | 17.38022 | 17.206730 | 17.07071 | |
| 8 | 183.924588 | 0.097246 | 17.90616 | 16.97172 | 16.67541 | 16.544191 | 16.47596 | |
| 9 | 183.973498 | 0.081626 | 18.67249 | 17.71375 | 17.49362 | 17.282840 | 17.22644 | |
| 10 | 183.979195 | 0.135998 | 19.29772 | 17.80227 | 17.18266 | 16.923350 | 16.79928 | |
| 11 | 184.085331 | 0.112110 | 18.83307 | 17.51785 | 16.94273 | 16.714180 | 16.60521 | |
| 12 | 184.102098 | 0.191511 | 19.56250 | 18.19113 | 17.65759 | 17.475730 | 17.39203 | |
| 13 | 184.160510 | 0.075645 | 19.57990 | 17.72815 | 16.98740 | 16.680760 | 16.50426 | |
| 14 | 184.189574 | 0.099482 | 19.25667 | 17.54869 | 16.63578 | 16.149220 | 15.76639 | |
| 15 | 184.350647 | 0.207230 | 18.73832 | 18.60962 | 18.39696 | 18.311740 | 17.97663 | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 16 | 184.221797 | 0.046955 | 19.05958 | 18.09512 | 17.92766 | 16.544191 | 17.90772 |
| 17 | 184.245664 | 0.198257 | 19.22143 | 19.30248 | 19.13823 | 19.113510 | 19.23454 |
| 18 | 184.388870 | 0.068287 | 19.04397 | 17.51106 | 16.87335 | 16.611140 | 16.48303 |
| 19 | 184.380919 | 0.174323 | 17.81661 | 16.86976 | 16.53884 | 16.195760 | 16.08668 |
| 20 | 184.466853 | 0.111965 | 19.39320 | 18.48274 | 18.16551 | 18.051220 | 18.04328 |
| 21 | 184.569411 | 0.137091 | 17.51339 | 16.41793 | 16.06695 | 15.937510 | 15.89478 |
| 22 | 184.654170 | 0.122673 | 19.07731 | 18.64518 | 18.49678 | 18.526770 | 18.45765 |
| 23 | 184.658795 | 0.159936 | 18.63032 | 17.23437 | 16.72749 | 16.517720 | 16.44039 |
| 24 | 184.516834 | 0.145550 | 19.21684 | 18.29956 | 18.19589 | 18.147580 | 18.26658 |
| 25 | 184.618563 | 0.080576 | 19.41270 | 17.47109 | 16.49870 | 16.065020 | 15.66848 |
| 26 | 184.627408 | 0.013234 | 18.48543 | 17.33900 | 16.90179 | 16.667120 | 16.48684 |
| 27 | 184.676762 | 0.059442 | 17.72835 | 16.19783 | 15.66073 | 15.515860 | 15.43615 |
| 28 | 184.711970 | 0.066572 | 17.70824 | 16.61069 | 16.19772 | 16.544191 | 16.00229 |
| 29 | 184.770204 | 0.198010 | 18.18974 | 17.23738 | 16.89113 | 16.757270 | 16.71052 |
| .. | ... | ... | ... | ... | ... | ... | ... |
| 569 | 244.284910 | -0.455477 | 18.63084 | 16.69326 | 15.67451 | 15.205570 | 14.82077 |
| 570 | 244.491831 | -0.470566 | 18.43597 | 17.47617 | 17.10257 | 16.909710 | 16.82976 |
| 571 | 244.496791 | -0.498398 | 17.98780 | 16.51811 | 15.74451 | 15.304330 | 14.98556 |
| 572 | 244.916397 | -0.523076 | 18.45084 | 17.39167 | 16.97855 | 16.825990 | 16.76456 |
| 573 | 244.925582 | -0.455836 | 19.34707 | 17.76923 | 16.92714 | 16.484860 | 16.14608 |
| 574 | 244.935545 | -0.479247 | 18.53797 | 17.16947 | 16.50210 | 16.189330 | 15.95892 |
| 575 | 245.128493 | -0.499415 | 19.18215 | 18.79711 | 18.81150 | 18.614920 | 18.51464 |
| 576 | 245.367353 | -0.457074 | 18.12316 | 16.31978 | 15.37991 | 14.910120 | 14.53670 |
| 577 | 245.376694 | -0.457008 | 19.46922 | 17.96987 | 17.31625 | 16.927290 | 16.71774 |
| 578 | 245.978479 | -0.483455 | 19.38072 | 17.38953 | 16.26663 | 15.756590 | 15.36963 |
| 579 | 246.034378 | -0.473284 | 19.51987 | 17.56103 | 16.44300 | 15.953020 | 15.57304 |
| 580 | 246.268157 | -0.426990 | 19.46025 | 18.28482 | 17.78494 | 17.503770 | 17.31497 |
| 581 | 246.695626 | -0.510111 | 18.01060 | 16.81185 | 16.32209 | 16.036530 | 15.92627 |
| 582 | 246.782081 | -0.492432 | 17.60994 | 15.90911 | 15.02090 | 14.549550 | 14.19971 |
| 583 | 247.021139 | -0.448677 | 19.23512 | 19.08207 | 18.78046 | 18.811240 | 18.99055 |
| 584 | 248.269016 | -0.488920 | 19.27635 | 17.99304 | 17.47953 | 17.290880 | 18.28560 |
| 585 | 248.894520 | -0.452329 | 18.73576 | 16.83826 | 15.94847 | 15.538030 | 15.15638 |
| 586 | 249.496551 | -0.498468 | 18.99369 | 17.31122 | 16.48666 | 16.053320 | 15.74858 |
| 587 | 189.429821 | -0.131042 | 17.83883 | 16.62422 | 16.84056 | 17.035840 | 17.15817 |
| 588 | 189.448544 | -0.052745 | 18.97562 | 17.84829 | 17.49412 | 17.374000 | 17.35097 |
| 589 | 189.453801 | -0.097313 | 17.66081 | 16.62739 | 16.28842 | 16.172150 | 16.14751 |
| 590 | 189.468747 | -0.036000 | 17.28231 | 16.20900 | 15.76480 | 15.513750 | 15.47968 |
| 591 | 189.522249 | -0.027031 | 16.32175 | 14.72385 | 13.88480 | 13.469230 | 13.15061 |
| 592 | 189.497150 | -0.110585 | 19.25978 | 18.10524 | 17.42960 | 17.089640 | 16.85153 |
| 593 | 225.529253 | -0.029788 | 19.21136 | 18.73800 | 18.76889 | 18.961950 | 19.18182 |
| 594 | 225.494356 | -0.114317 | 19.43974 | 18.09289 | 17.40579 | 17.008100 | 16.82082 |
| 595 | 145.288576 | 0.346512 | 19.22352 | 18.00232 | 17.37943 | 17.011100 | 16.76524 |
| 596 | 145.292821 | 0.376326 | 19.31302 | 18.13527 | 17.67289 | 17.499960 | 17.44187 |
| 597 | 145.593807 | 0.278308 | 19.10790 | 18.92201 | 18.59314 | 18.475730 | 18.40461 |
| 598 | 145.637275 | 0.373311 | 19.20753 | 17.87736 | 17.30583 | 17.086230 | 16.96132 |

| | run | camcol | field | specobjid | class | redshift | plate | mjd | \ |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 752.0 | 4 | 267 | 3.720000e+18 | STAR | -8.960000e-06 | 3306 | 54922 | |

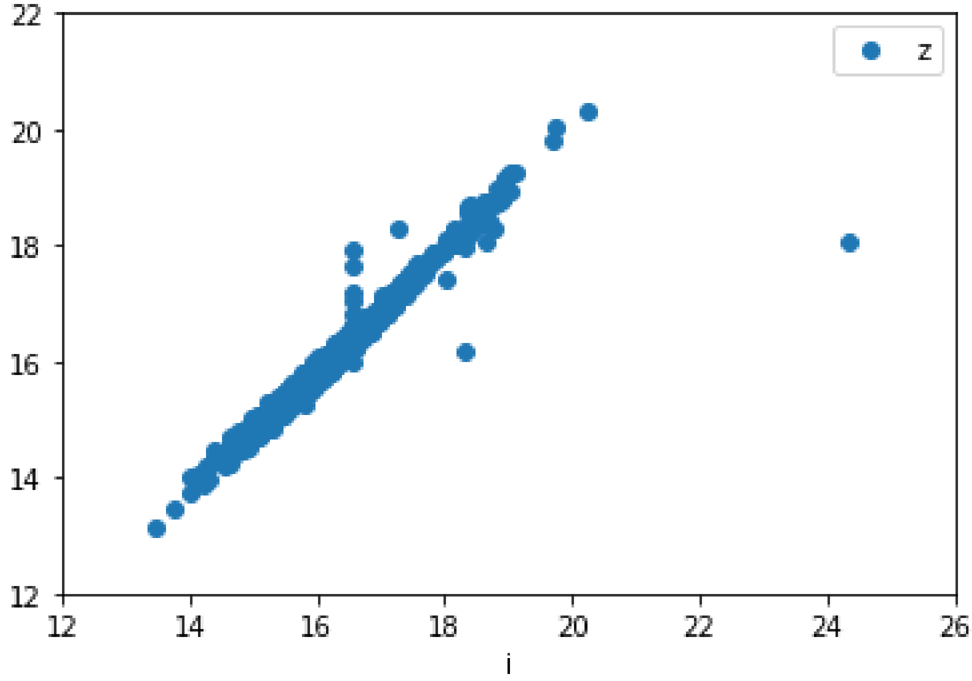
| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1 | 752.0 | 4 | 267 | 3.640000e+17 | STAR | -5.490000e-05 | 323 | 51615 |
| 2 | 752.0 | 4 | 268 | 3.230000e+17 | GALAXY | 1.231112e-01 | 287 | 52023 |
| 3 | 752.0 | 4 | 269 | 3.720000e+18 | STAR | -1.106160e-04 | 3306 | 54922 |
| 4 | 752.0 | 4 | 269 | 3.720000e+18 | STAR | 5.903570e-04 | 3306 | 54922 |
| 5 | 752.0 | 4 | 269 | 3.650000e+17 | STAR | 3.146030e-04 | 324 | 51666 |
| 6 | 752.0 | 4 | 269 | 3.230000e+17 | GALAXY | 1.002423e-01 | 287 | 52023 |
| 7 | 752.0 | 4 | 269 | 3.720000e+18 | STAR | 3.148480e-04 | 3306 | 54922 |
| 8 | 752.0 | 4 | 270 | 3.640000e+17 | STAR | 8.910000e-05 | 323 | 51615 |
| 9 | 752.0 | 4 | 270 | 3.240000e+17 | GALAXY | 4.050813e-02 | 288 | 52000 |
| 10 | 752.0 | 4 | 270 | 3.720000e+18 | STAR | -3.460000e-05 | 3306 | 54922 |
| 11 | 752.0 | 4 | 271 | 3.720000e+18 | STAR | 6.227740e-04 | 3306 | 54922 |
| 12 | 752.0 | 4 | 271 | 3.720000e+18 | STAR | 5.470000e-05 | 3306 | 54922 |
| 13 | 752.0 | 4 | 271 | 3.720000e+18 | STAR | 8.300000e-06 | 3306 | 54922 |
| 14 | 752.0 | 4 | 271 | 3.240000e+17 | GALAXY | 7.208736e-02 | 288 | 52000 |
| 15 | 752.0 | 4 | 272 | 3.230000e+17 | QSO | 2.719369e-01 | 287 | 52023 |
| 16 | 752.0 | 4 | 272 | 3.650000e+17 | STAR | 2.099740e-04 | 324 | 51666 |
| 17 | 752.0 | 4 | 272 | 3.230000e+17 | QSO | 1.178098e+00 | 287 | 52023 |
| 18 | 752.0 | 4 | 273 | 3.640000e+17 | STAR | 2.400980e-04 | 323 | 51615 |
| 19 | 752.0 | 4 | 273 | 3.230000e+17 | GALAXY | 7.277206e-02 | 287 | 52023 |
| 20 | 752.0 | 4 | 273 | 3.650000e+17 | STAR | -2.702180e-04 | 324 | 51666 |
| 21 | 752.0 | 4 | 274 | 3.240000e+17 | STAR | -3.550000e-05 | 288 | 52000 |
| 22 | 752.0 | 4 | 274 | 3.240000e+17 | QSO | 9.251733e-01 | 288 | 52000 |
| 23 | 752.0 | 4 | 274 | 2.880000e+18 | STAR | 4.310000e-06 | 2558 | 54140 |
| 24 | 752.0 | 4 | 274 | 3.720000e+18 | STAR | 8.800000e-05 | 3306 | 54922 |
| 25 | 752.0 | 4 | 274 | 3.240000e+17 | GALAXY | 1.167829e-01 | 288 | 52000 |
| 26 | 752.0 | 4 | 274 | 3.230000e+17 | GALAXY | 2.129653e-02 | 287 | 52023 |
| 27 | 752.0 | 4 | 275 | 2.880000e+18 | STAR | 1.753650e-04 | 2558 | 54140 |
| 28 | 752.0 | 4 | 275 | 3.720000e+18 | STAR | 1.723800e-04 | 3306 | 54922 |
| 29 | 752.0 | 4 | 275 | 2.880000e+18 | STAR | 5.198080e-04 | 2558 | 54140 |
| .. | ... | ... | ... | ... | ... | ... | ... | ... |
| 569 | 745.0 | 2 | 571 | 3.900000e+17 | GALAXY | 9.280799e-02 | 346 | 51693 |
| 570 | 745.0 | 2 | 573 | 3.900000e+17 | GALAXY | 1.660742e-02 | 346 | 51693 |
| 571 | 745.0 | 2 | 573 | 3.900000e+17 | GALAXY | 5.047690e-02 | 346 | 51693 |
| 572 | 745.0 | 2 | 576 | 3.900000e+17 | STAR | -3.407500e-04 | 346 | 51693 |
| 573 | 745.0 | 2 | 576 | 3.900000e+17 | GALAXY | 7.900908e-02 | 346 | 51693 |
| 574 | 745.0 | 2 | 576 | 3.900000e+17 | GALAXY | 5.936759e-02 | 346 | 51693 |
| 575 | 745.0 | 2 | 577 | 4.100000e+17 | QSO | 5.382954e-01 | 364 | 52000 |
| 576 | 745.0 | 2 | 579 | 4.100000e+17 | GALAXY | 5.909181e-02 | 364 | 52000 |
| 577 | 745.0 | 2 | 579 | 3.900000e+17 | GALAXY | 6.916405e-02 | 346 | 51693 |
| 578 | 745.0 | 2 | 583 | 4.100000e+17 | GALAXY | 9.362785e-02 | 364 | 52000 |
| 579 | 745.0 | 2 | 583 | 4.100000e+17 | GALAXY | 1.424379e-01 | 364 | 52000 |
| 580 | 745.0 | 2 | 585 | 4.100000e+17 | GALAXY | 4.653732e-02 | 364 | 52000 |
| 581 | 745.0 | 2 | 588 | 4.100000e+17 | GALAXY | 4.449500e-02 | 364 | 52000 |
| 582 | 745.0 | 2 | 588 | 4.100000e+17 | GALAXY | 4.619161e-02 | 364 | 52000 |
| 583 | 745.0 | 2 | 590 | 4.100000e+17 | QSO | 1.149604e+00 | 364 | 52000 |
| 584 | 745.0 | 2 | 598 | 3.920000e+17 | STAR | -7.320000e-05 | 348 | 51671 |
| 585 | 745.0 | 2 | 602 | 3.920000e+17 | GALAXY | 7.018682e-02 | 348 | 51671 |
| 586 | 745.0 | 2 | 606 | 3.920000e+17 | GALAXY | 5.894065e-02 | 348 | 51671 |

```
587  745.0        3    205  3.260000e+18    STAR  6.484750e-04   2895  54567
588  745.0        3    205  3.280000e+17    STAR  2.573870e-04    291  51928
589  745.0        3    205  3.260000e+18    STAR  5.910000e-07   2895  54567
590  745.0        3    205  3.260000e+18    STAR  5.791130e-04   2895  54567
591  745.0        3    205  3.270000e+17  GALAXY  1.251297e-02    290  51941
592  745.0        3    205  3.280000e+17  GALAXY  1.383053e-01    291  51928
593  745.0        3    446  4.520000e+18    STAR -1.976670e-04   4016  55632
594  745.0        3    446  3.490000e+17  GALAXY  4.019753e-02    310  51990
595  756.0        4    197  3.000000e+17  GALAXY  6.460119e-02    266  51630
596  756.0        4    197  3.000000e+17    STAR  2.133720e-04    266  51630
597  756.0        4    199  3.000000e+17     QSO  1.404700e+00    266  51630
598  756.0        4    199  3.000000e+17    STAR -3.480000e-05    266  51630

     fiberid
0        491
1        541
2        513
3        510
4        512
5        594
6        559
7        515
8        595
9        400
10       506
11       547
12       544
13       546
14       389
15       587
16        35
17       583
18       628
19       632
20       623
21       430
22       421
23       356
24       595
25       437
26       637
27       358
28       632
29       349
..       ...
569      162
570      148
571      146
```

```
572        81
573       106
574       108
575       307
576       302
577        67
578       221
579       190
580       161
581       146
582       107
583        74
584       236
585       163
586       106
587       116
588       310
589       117
590       119
591       516
592       318
593       396
594       224
595       359
596       346
597       396
598       390

[599 rows x 16 columns]
```

In [49]:
```
#Scatter Plot
df1.plot(x='i', y='z', style='o')
plt.xlim([12,26])
plt.ylim([12,22])
```

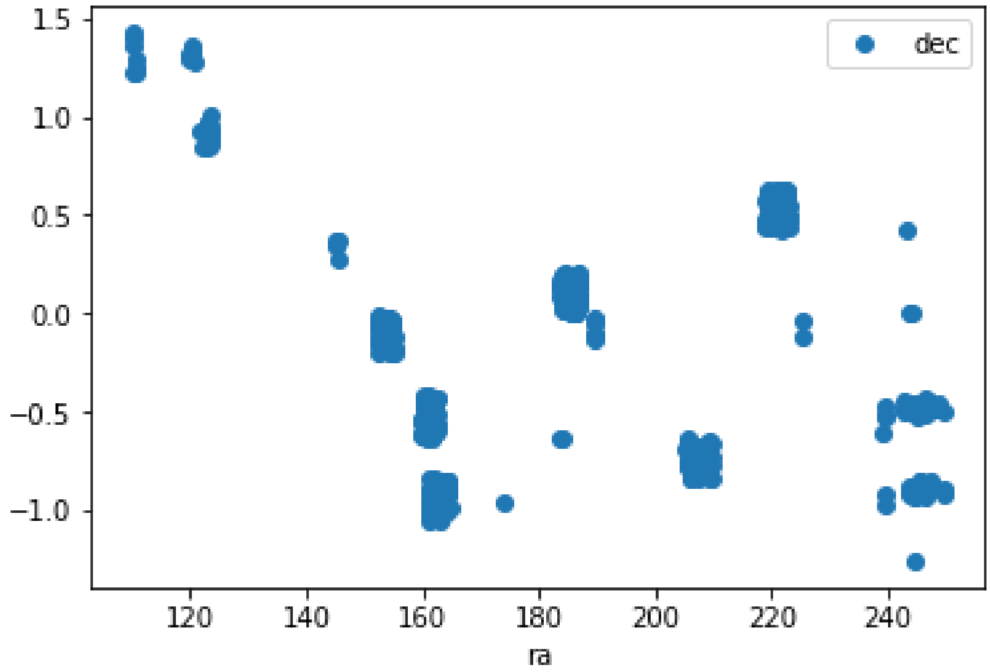
Out[49]:
```
(12, 22)
```



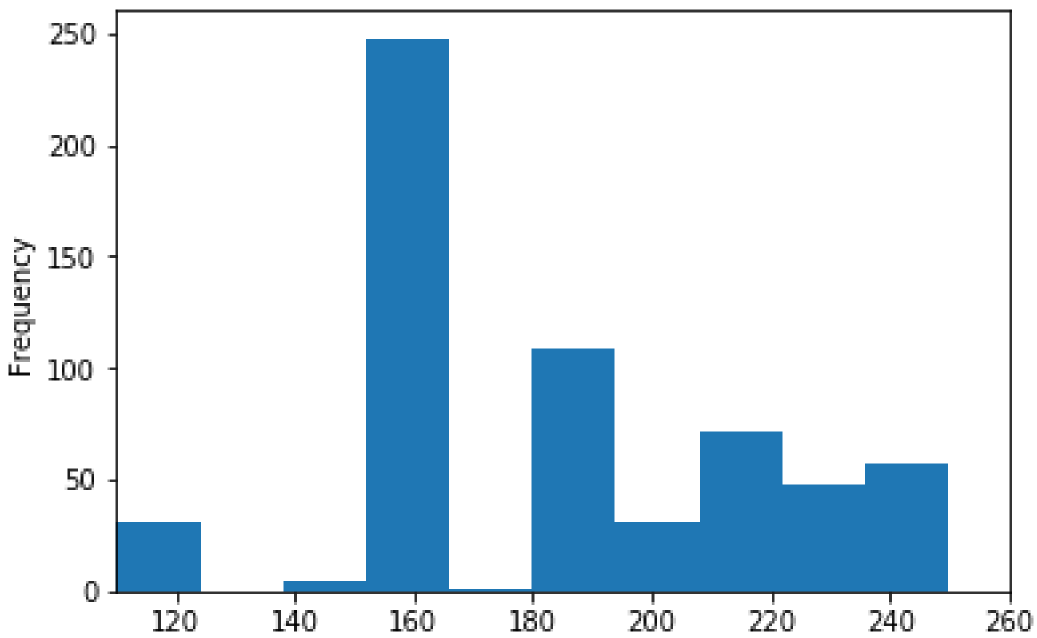

```
In [50]: #We see that scatterplot for infrared and 900nm response of the telescope are correlate
         #This is clear as infrared wavelenghts(700nm and above) and 900nm are almost near hence
```

```
In [51]: df1.plot(x='ra', y='dec', style='o')
```

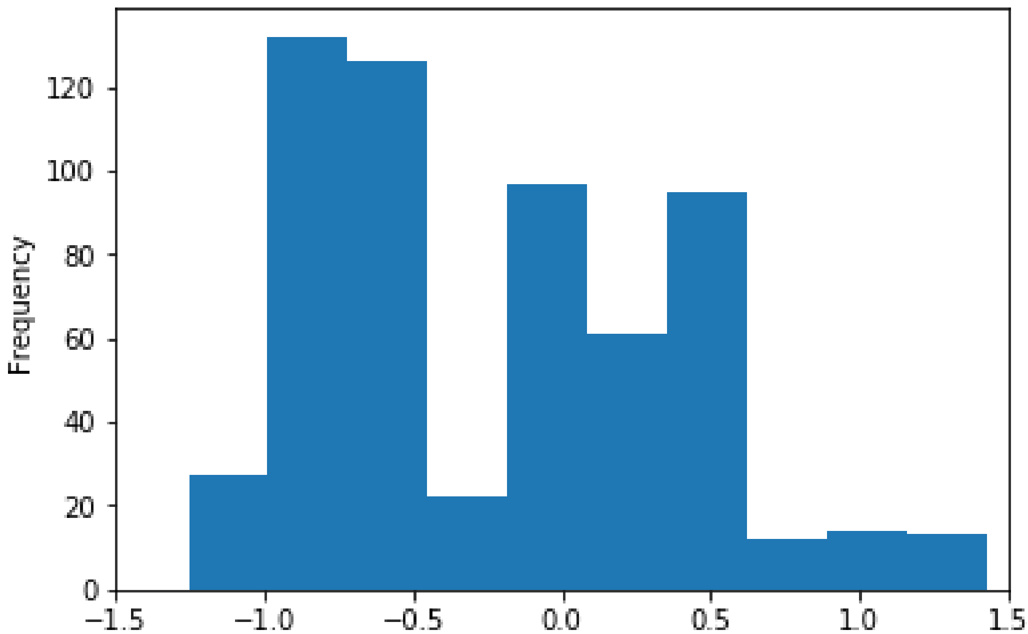
```
Out[51]: <matplotlib.axes._subplots.AxesSubplot at 0x243aff6ca90>
```



```
In [ ]: #We see that ra and dec are not correlated this is obvious as the right ascension angle

In [58]: #df1.hist(column='ra')
         df1['ra'].plot.hist(xlim = (110,260),ylim=(0,260))
         #plt.xlim([110,260])
         #plt.ylim([0,260])

Out[58]: <matplotlib.axes._subplots.AxesSubplot at 0x243b040fda0>
```



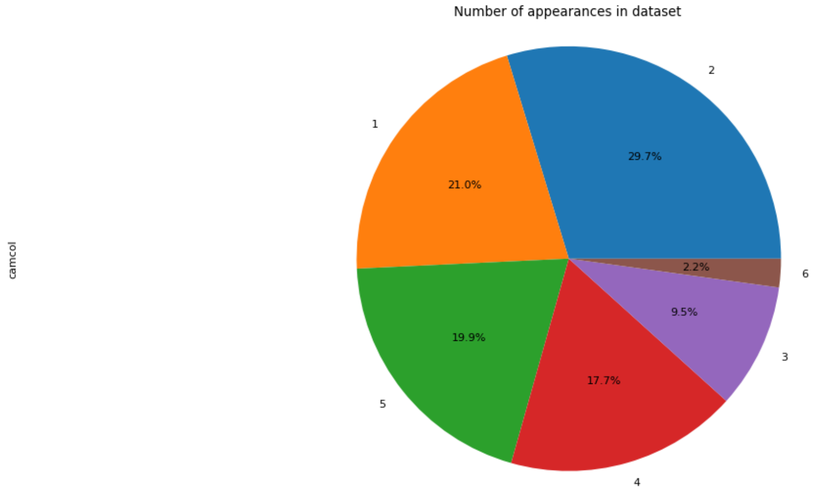

```
In [118]: # insights from this histogram plot shows that most of our data is taken from right as

In [63]: #df.hist(column='dec')
         df1['dec'].plot.hist(xlim=(-1.5,1.5))

Out[63]: <matplotlib.axes._subplots.AxesSubplot at 0x243b0923358>
```




```
In [41]: # insights from this histogram plot shows that declination angle of our data is little
```

```
In [42]: # from both of this histograms of ra and dec we can clearly see that most of our datase
```

```
In [40]: plt.figure(figsize=(18, 8), dpi=80, facecolor='w', edgecolor='k')
         df1['camcol'].value_counts().plot(kind='pie' ,autopct='%1.1f%%')
         plt.axis('equal')
         plt.title('Number of appearances in dataset')

         plt.show()
         df1['camcol'].value_counts()
```



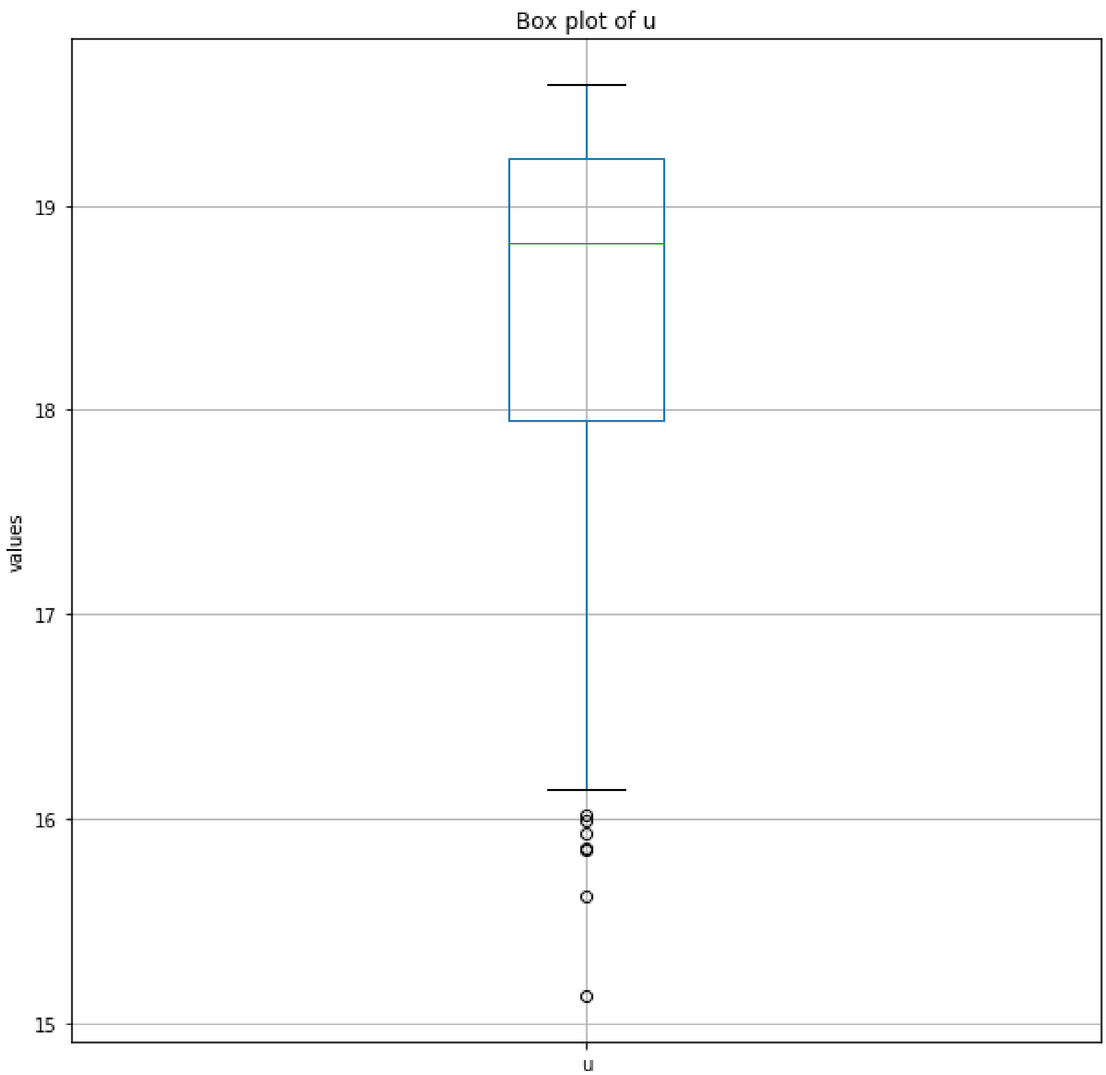
```
Out[40]: 2    178
         1    126
         5    119
         4    106
         3     57
         6     13
         Name: camcol, dtype: int64
```

```
In [44]: # We can see that most of our dataset contains Stars
```

```
In [45]: # very few 7% of our dataset contains quasars which says that quasars are very rare whe
```

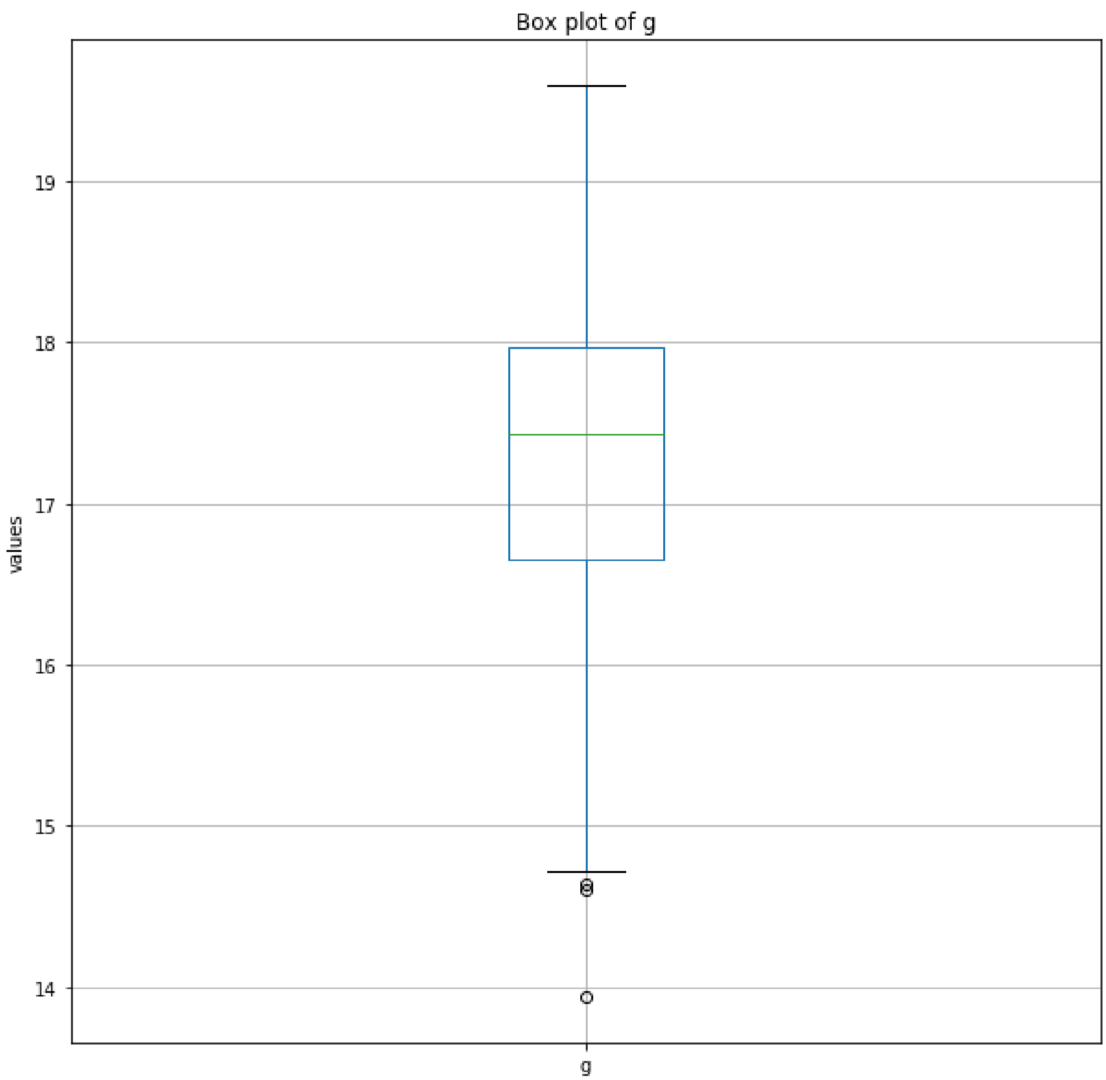
```
In [67]: df1.boxplot(column ='u',figsize=(10,10))
         pl.title("Box plot of u")
         pl.ylabel("values")
```

```
Out[67]: Text(0,0.5,'values')
```



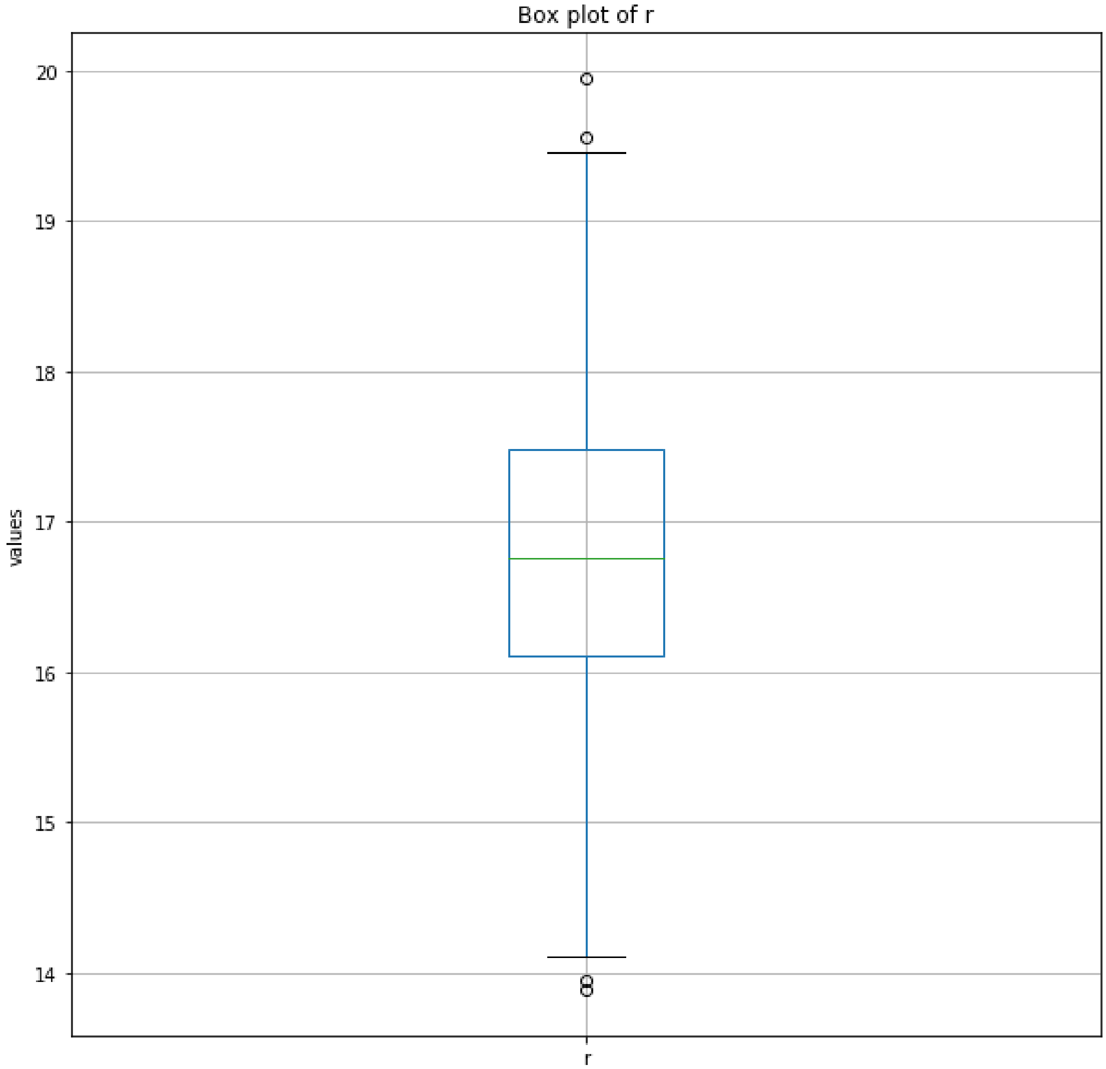
In [68]:
```
#We can see that the response of the telescope to ultraviolet band ranges from 16.2 to
df1.boxplot(column ='g',figsize=(10,10))
pl.title("Box plot of g")
pl.ylabel("values")
```

Out[68]:
```
Text(0,0.5,'values')
```



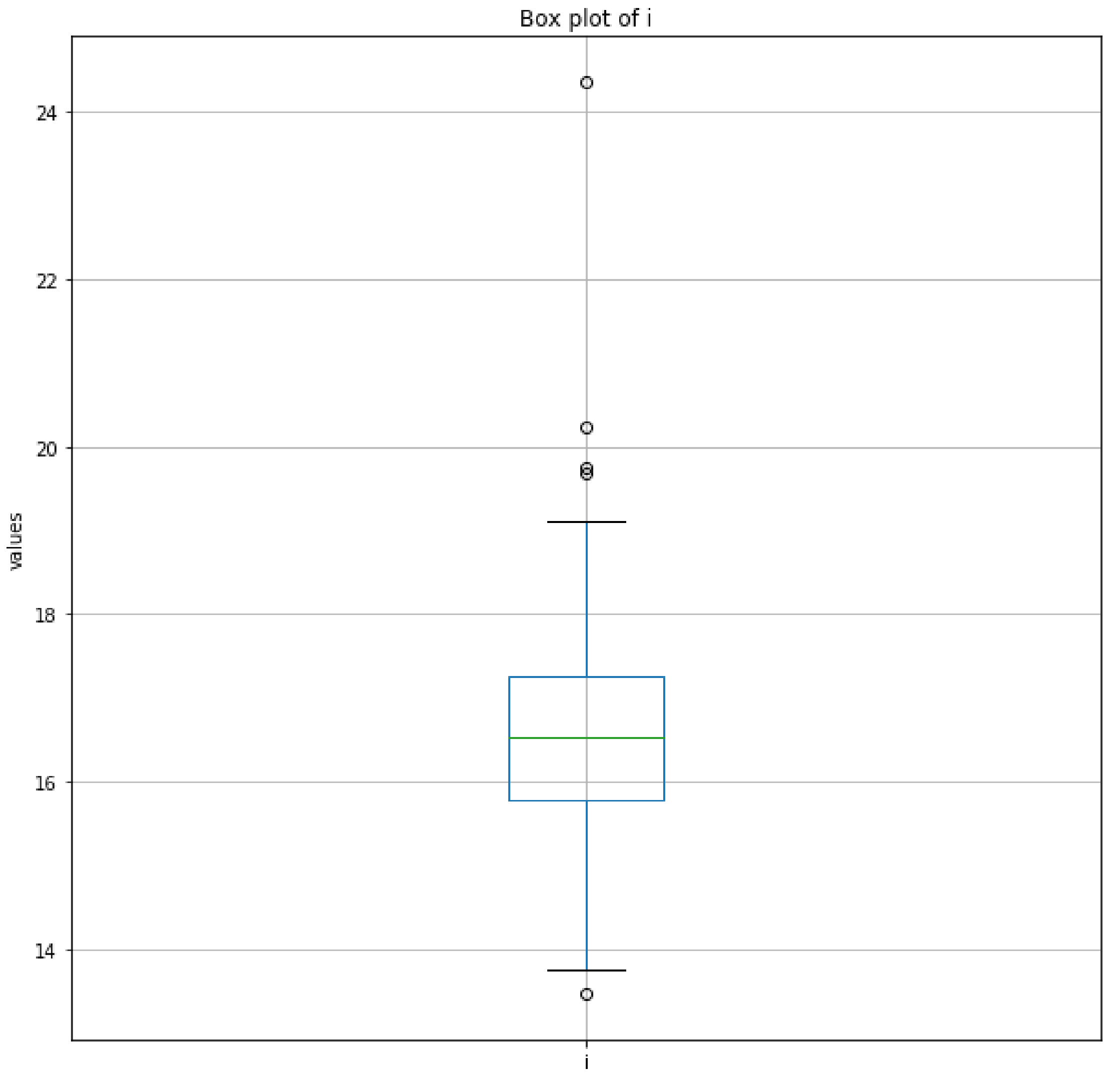

In [69]:
```
#We can see that the response of the telescope to green band ranges from 14.7 to 19.6 u
df1.boxplot(column ='r',figsize=(10,10))
pl.title("Box plot of r")
pl.ylabel("values")
```

Out[69]: Text(0,0.5,'values')

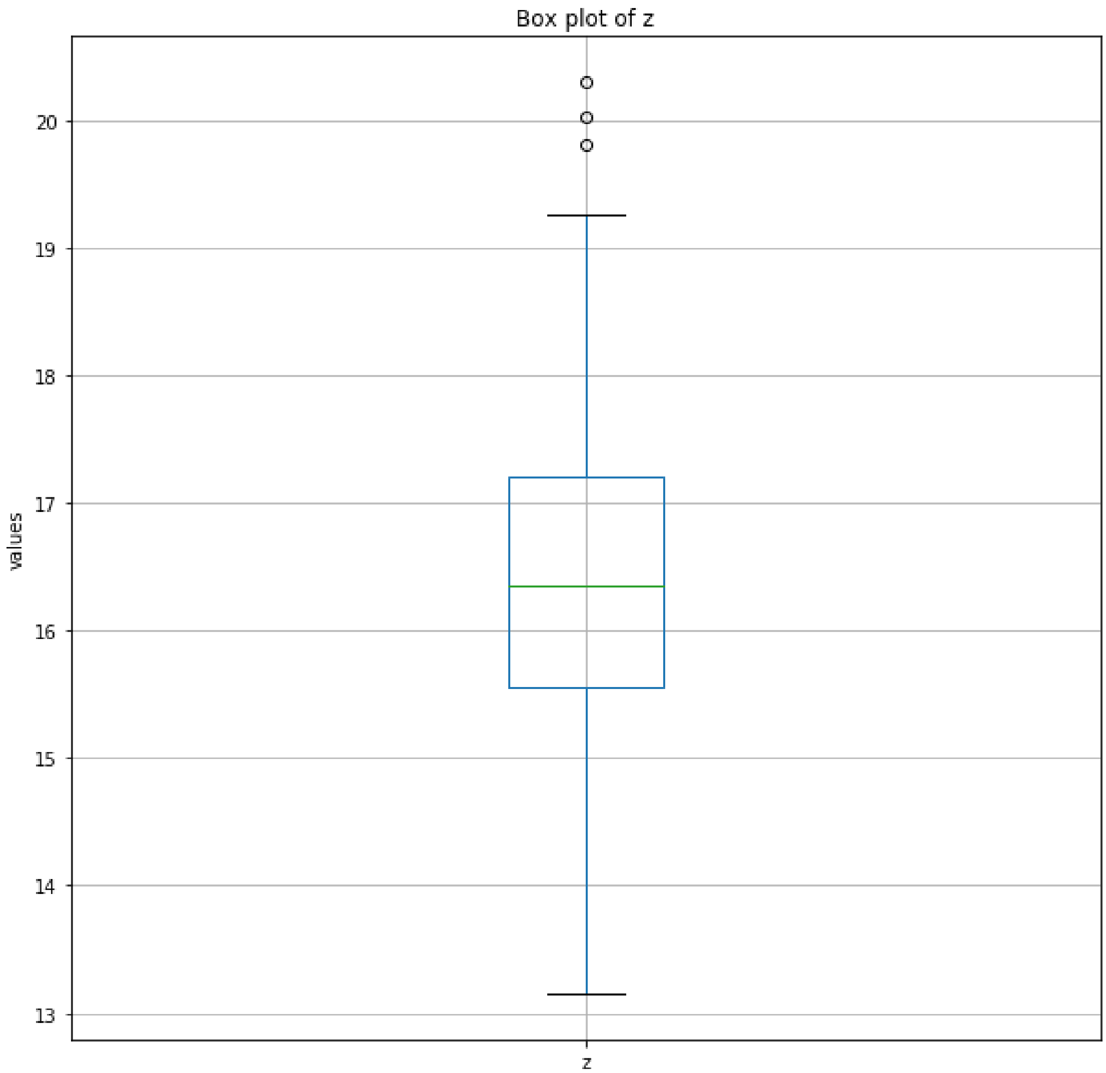



```
In [70]: #We can see that the response of the telescope to red band ranges from 14.1 to 19.5 wit
         df1.boxplot(column ='i',figsize=(10,10))
         pl.title("Box plot of i")
         pl.ylabel("values")
```

Out[70]: Text(0,0.5,'values')

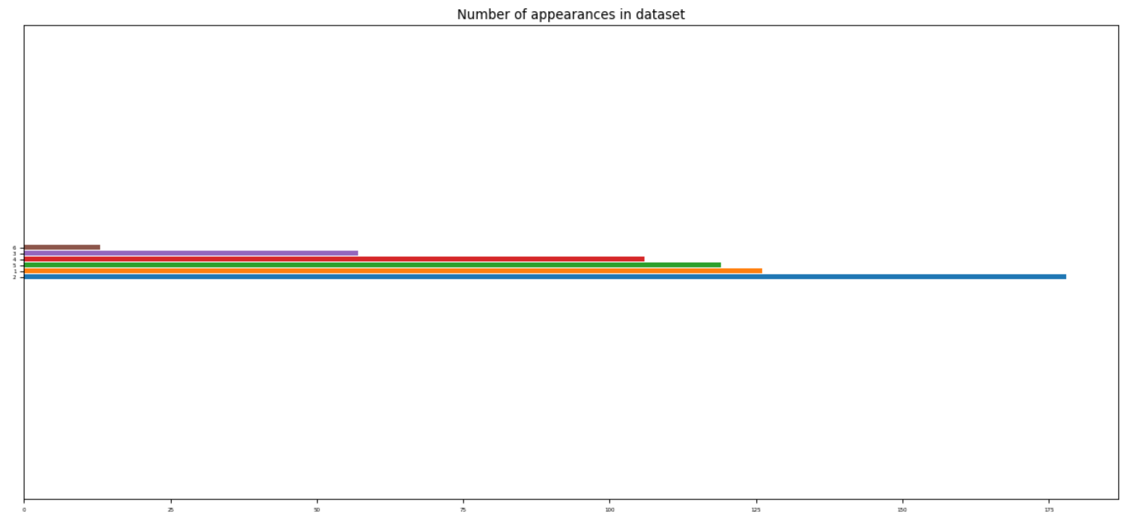



```
In [71]: #We can see that the response of the telescope to infrared band ranges from 13.9 to 19
         df1.boxplot(column ='z',figsize=(10,10))
         pl.title("Box plot of z")
         pl.ylabel("values")

Out[71]: Text(0,0.5,'values')
```




In [51]: #We can see that the response of the telescope to z(900nm) band ranges from 13.2 to 19.

In [76]:
```
plt.figure(num=None, figsize=(18, 8), dpi=80, facecolor='w', edgecolor='k')
df1['camcol'].value_counts().plot(kind='barh',fontsize = 5,width = 0.8,rot=0)
plt.axis('equal')
plt.title('Number of appearances in dataset')
plt.show()
```




```
In [53]: # We can see by this bar graph that camera scanline within in the run is used maximum f
```

```
In [97]: # 4) Hypothesis testing
```

```
In [101]: # picking a ramdom sample of size 100 from our dataset with replacement as a specific
```

```
In [99]: sample = df1.sample(n = 100,replace = True)
```

```
In [100]: sample
```

Out[100]:

| | ra | dec | u | g | r | i | z | \ |
|---|---|---|---|---|---|---|---|---|
| 585 | 248.894520 | -0.452329 | 18.73576 | 16.83826 | 15.94847 | 15.53803 | 15.15638 | |
| 89 | 160.996765 | -0.588245 | 18.38625 | 16.39808 | 15.40492 | 14.88012 | 14.50048 | |
| 544 | 222.955420 | 0.472939 | 18.31584 | 17.36814 | 17.06714 | 16.91268 | 16.87696 | |
| 464 | 162.964354 | -0.854905 | 15.92846 | 15.75834 | 15.68789 | 15.79105 | 15.25990 | |
| 105 | 161.303101 | -0.964276 | 19.01558 | 17.72445 | 17.18152 | 16.97813 | 16.85652 | |
| 288 | 209.090702 | -0.656281 | 19.23039 | 18.06944 | 17.59436 | 17.23812 | 17.09449 | |
| 182 | 221.163280 | 0.483958 | 17.88442 | 16.96119 | 16.64287 | 16.53269 | 16.48384 | |
| 472 | 163.162657 | -0.943445 | 18.98361 | 17.83656 | 17.32091 | 17.11787 | 17.00929 | |
| 156 | 219.581971 | 0.471853 | 18.56313 | 18.53411 | 18.38498 | 18.37235 | 18.36230 | |
| 218 | 110.246454 | 1.364872 | 18.58476 | 17.29450 | 16.42206 | 16.27426 | 15.89856 | |
| 108 | 161.350692 | -0.893119 | 18.12173 | 16.84951 | 16.38322 | 16.20272 | 16.10719 | |
| 447 | 162.670124 | -0.874412 | 19.07583 | 17.58162 | 16.95899 | 16.70053 | 16.53613 | |
| 243 | 244.949053 | -0.894235 | 19.15965 | 18.91703 | 18.61517 | 18.38544 | 18.42006 | |
| 51 | 159.797842 | -0.557196 | 19.19752 | 18.07471 | 17.76889 | 17.61480 | 17.49577 | |
| 543 | 222.933968 | 0.432157 | 18.23652 | 17.03079 | 16.52341 | 16.28072 | 16.18811 | |
| 578 | 245.978479 | -0.483455 | 19.38072 | 17.38953 | 16.26663 | 15.75659 | 15.36963 | |
| 239 | 243.847368 | -0.898343 | 19.41207 | 18.88776 | 18.56437 | 18.58996 | 18.39260 | |
| 163 | 220.122411 | 0.584089 | 16.93221 | 15.56652 | 14.92086 | 14.58185 | 14.32520 | |
| 297 | 209.951317 | -0.766163 | 19.15065 | 17.36285 | 16.52798 | 16.15387 | 15.80557 | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 178 | 221.041738 | 0.564546 | 18.77460 | 17.64604 | 17.22892 | 17.07624 | 17.01672 |
| 162 | 219.938725 | 0.531203 | 18.91460 | 17.67455 | 16.96098 | 16.57482 | 16.29057 |
| 532 | 222.509033 | 0.578781 | 17.78947 | 15.91691 | 15.07053 | 14.67859 | 14.36553 |
| 25 | 184.618563 | 0.080576 | 19.41270 | 17.47109 | 16.49870 | 16.06502 | 15.66848 |
| 414 | 153.449439 | -0.174722 | 19.33492 | 17.25626 | 16.32616 | 15.91549 | 15.54986 |
| 375 | 161.988413 | -0.455833 | 15.84520 | 14.78328 | 14.48869 | 14.40722 | 14.38870 |
| 29 | 184.770204 | 0.198010 | 18.18974 | 17.23738 | 16.89113 | 16.75727 | 16.71052 |
| 267 | 206.806474 | -0.671287 | 18.37380 | 16.84218 | 16.10014 | 15.71413 | 15.41793 |
| 467 | 163.063223 | -0.964999 | 17.85959 | 16.18301 | 15.56699 | 15.37634 | 15.28177 |
| 234 | 174.164488 | -0.956707 | 19.57255 | 18.59834 | 18.47911 | 18.45792 | 18.60459 |
| 27 | 184.676762 | 0.059442 | 17.72835 | 16.19783 | 15.66073 | 15.51586 | 15.43615 |
| .. | ... | ... | ... | ... | ... | ... | ... |
| 553 | 247.254489 | -0.854531 | 18.25633 | 16.75476 | 16.15203 | 15.92617 | 15.82208 |
| 21 | 184.569411 | 0.137091 | 17.51339 | 16.41793 | 16.06695 | 15.93751 | 15.89478 |
| 159 | 219.661166 | 0.569298 | 19.47595 | 18.61160 | 18.52926 | 18.76056 | 18.28581 |
| 134 | 162.246566 | -0.918542 | 17.71105 | 16.34725 | 15.83003 | 15.60209 | 15.47698 |
| 158 | 219.588635 | 0.510879 | 19.14528 | 18.19967 | 17.89295 | 17.78337 | 17.73488 |
| 499 | 221.766733 | 0.479062 | 17.68773 | 16.15192 | 15.37372 | 14.95320 | 14.64231 |
| 188 | 221.364993 | 0.454015 | 17.52671 | 16.10266 | 15.57270 | 15.35048 | 15.24569 |
| 222 | 110.469850 | 1.228435 | 19.42571 | 17.78678 | 17.11364 | 16.85967 | 16.67854 |
| 399 | 152.289487 | -0.063071 | 19.45274 | 17.62178 | 16.68552 | 16.27155 | 15.92060 |
| 53 | 159.847524 | -0.534266 | 19.38251 | 17.92302 | 17.11018 | 16.66492 | 16.34558 |
| 585 | 248.894520 | -0.452329 | 18.73576 | 16.83826 | 15.94847 | 15.53803 | 15.15638 |
| 301 | 185.379456 | 0.128519 | 19.25762 | 19.24938 | 18.97530 | 18.88568 | 18.91675 |
| 303 | 185.485219 | 0.163400 | 18.71479 | 16.97454 | 16.38521 | 16.12513 | 15.90649 |
| 564 | 242.954823 | -0.446457 | 18.90115 | 17.03906 | 16.03109 | 15.48150 | 15.06539 |
| 75 | 160.938240 | -0.423247 | 17.37104 | 15.78487 | 14.98982 | 14.58674 | 14.25872 |
| 221 | 110.455531 | 1.254180 | 19.58684 | 17.73149 | 16.96642 | 16.67645 | 16.49140 |
| 217 | 110.070039 | 1.218735 | 19.42591 | 18.42443 | 18.00135 | 17.79079 | 17.70945 |
| 509 | 221.984190 | 0.446387 | 17.50681 | 16.05808 | 15.52935 | 15.36461 | 15.28286 |
| 462 | 162.886970 | -1.002881 | 17.79835 | 16.35954 | 15.67576 | 15.28575 | 14.98440 |
| 453 | 162.766420 | -0.860514 | 17.67043 | 16.20217 | 15.63600 | 15.44809 | 15.32646 |
| 468 | 163.062060 | -0.955493 | 18.42401 | 16.48616 | 15.58818 | 15.16028 | 14.82537 |
| 85 | 161.094801 | -0.452050 | 17.82710 | 16.23793 | 15.67866 | 15.49427 | 15.41664 |
| 574 | 244.935545 | -0.479247 | 18.53797 | 17.16947 | 16.50210 | 16.18933 | 15.95892 |
| 434 | 154.611918 | -0.164200 | 18.79575 | 18.57321 | 18.55136 | 18.35209 | 18.30321 |
| 321 | 186.033836 | 0.113213 | 19.10102 | 17.89368 | 17.17626 | 16.78514 | 16.47559 |
| 446 | 155.248147 | -0.176081 | 18.43279 | 16.62750 | 15.71816 | 15.25928 | 14.88557 |
| 567 | 244.112957 | -0.492485 | 17.34144 | 17.66519 | 18.08743 | 18.37465 | 18.64286 |
| 170 | 220.711731 | 0.461339 | 19.26163 | 17.71671 | 17.12826 | 16.89513 | 16.79474 |
| 72 | 160.820289 | -0.522241 | 16.34400 | 14.85487 | 14.29812 | 14.12055 | 14.08315 |
| 277 | 207.456735 | -0.726105 | 19.19459 | 18.31113 | 18.09644 | 17.83205 | 17.75919 |

| | run | camcol | field | specobjid | class | redshift | plate | mjd | \ |
|---|---|---|---|---|---|---|---|---|---|
| 585 | 745.0 | 2 | 602 | 3.920000e+17 | GALAXY | 0.070187 | 348 | 51671 | |
| 89 | 756.0 | 2 | 302 | 3.100000e+17 | GALAXY | 0.060674 | 275 | 51910 | |
| 544 | 752.0 | 5 | 530 | 3.280000e+18 | STAR | 0.000026 | 2909 | 54653 | |
| 464 | 756.0 | 1 | 315 | 3.110000e+17 | QSO | 0.358947 | 276 | 51909 | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 105 | 756.0 | 1 | 304 | 2.690000e+18 | STAR | 0.000366 | 2389 | 54213 |
| 288 | 752.0 | 2 | 438 | 3.390000e+17 | GALAXY | 0.079088 | 301 | 51942 |
| 182 | 752.0 | 5 | 518 | 3.300000e+18 | STAR | 0.000032 | 2934 | 54626 |
| 472 | 756.0 | 1 | 316 | 3.110000e+17 | STAR | 0.000115 | 276 | 51909 |
| 156 | 752.0 | 5 | 508 | 3.460000e+17 | QSO | 1.257948 | 307 | 51663 |
| 218 | 308.0 | 6 | 26 | 8.390000e+18 | STAR | 0.000391 | 7450 | 56722 |
| 108 | 756.0 | 1 | 304 | 2.690000e+18 | STAR | 0.000239 | 2389 | 54213 |
| 447 | 756.0 | 1 | 313 | 2.880000e+18 | STAR | 0.001076 | 2559 | 54208 |
| 243 | 745.0 | 1 | 576 | 3.900000e+17 | QSO | 1.496161 | 346 | 51693 |
| 51 | 756.0 | 2 | 294 | 3.090000e+17 | GALAXY | 0.023351 | 274 | 51913 |
| 543 | 752.0 | 5 | 530 | 3.280000e+18 | STAR | 0.000123 | 2909 | 54653 |
| 578 | 745.0 | 2 | 583 | 4.100000e+17 | GALAXY | 0.093628 | 364 | 52000 |
| 239 | 745.0 | 1 | 568 | 3.900000e+17 | QSO | 1.062840 | 346 | 51693 |
| 163 | 752.0 | 5 | 511 | 3.460000e+17 | GALAXY | 0.037176 | 307 | 51663 |
| 297 | 752.0 | 2 | 443 | 3.390000e+17 | GALAXY | 0.106320 | 301 | 51942 |
| 178 | 752.0 | 5 | 518 | 3.460000e+17 | STAR | 0.000050 | 307 | 51663 |
| 162 | 752.0 | 5 | 510 | 3.460000e+17 | GALAXY | 0.116515 | 307 | 51663 |
| 532 | 752.0 | 5 | 527 | 3.480000e+17 | GALAXY | 0.040417 | 309 | 51994 |
| 25 | 752.0 | 4 | 274 | 3.240000e+17 | GALAXY | 0.116783 | 288 | 52000 |
| 414 | 756.0 | 3 | 251 | 3.040000e+17 | GALAXY | 0.088731 | 270 | 51909 |
| 375 | 756.0 | 2 | 308 | 2.880000e+18 | STAR | -0.000076 | 2559 | 54208 |
| 29 | 752.0 | 4 | 275 | 2.880000e+18 | STAR | 0.000520 | 2558 | 54140 |
| 267 | 752.0 | 2 | 422 | 3.380000e+17 | GALAXY | 0.047497 | 300 | 51943 |
| 467 | 756.0 | 1 | 315 | 2.880000e+18 | STAR | -0.000051 | 2559 | 54208 |
| 234 | 745.0 | 1 | 103 | 3.180000e+17 | STAR | 0.001135 | 282 | 51658 |
| 27 | 752.0 | 4 | 275 | 2.880000e+18 | STAR | 0.000175 | 2558 | 54140 |
| .. | ... | ... | ... | ... | ... | ... | ... | ... |
| 553 | 745.0 | 1 | 591 | 4.100000e+17 | STAR | 0.000009 | 364 | 52000 |
| 21 | 752.0 | 4 | 274 | 3.240000e+17 | STAR | -0.000036 | 288 | 52000 |
| 159 | 752.0 | 5 | 508 | 3.460000e+17 | STAR | -0.000385 | 307 | 51663 |
| 134 | 756.0 | 1 | 310 | 2.690000e+18 | STAR | -0.000456 | 2389 | 54213 |
| 158 | 752.0 | 5 | 508 | 4.530000e+18 | STAR | -0.000276 | 4025 | 55350 |
| 499 | 752.0 | 5 | 522 | 4.530000e+18 | GALAXY | 0.055807 | 4022 | 55352 |
| 188 | 752.0 | 5 | 520 | 3.280000e+18 | STAR | 0.000836 | 2909 | 54653 |
| 222 | 308.0 | 6 | 28 | 8.390000e+18 | STAR | 0.000182 | 7450 | 56722 |
| 399 | 756.0 | 3 | 244 | 3.040000e+17 | GALAXY | 0.095463 | 270 | 51909 |
| 53 | 756.0 | 2 | 294 | 3.090000e+17 | GALAXY | 0.114168 | 274 | 51913 |
| 585 | 745.0 | 2 | 602 | 3.920000e+17 | GALAXY | 0.070187 | 348 | 51671 |
| 301 | 752.0 | 4 | 279 | 3.240000e+17 | QSO | 1.335835 | 288 | 52000 |
| 303 | 752.0 | 4 | 280 | 3.240000e+17 | GALAXY | 0.106216 | 288 | 52000 |
| 564 | 745.0 | 2 | 563 | 3.880000e+17 | GALAXY | 0.080116 | 345 | 51690 |
| 75 | 756.0 | 2 | 301 | 3.100000e+17 | GALAXY | 0.039109 | 275 | 51910 |
| 221 | 308.0 | 6 | 28 | 8.390000e+18 | STAR | 0.000346 | 7450 | 56722 |
| 217 | 308.0 | 6 | 25 | 8.390000e+18 | STAR | 0.000104 | 7450 | 56722 |
| 509 | 752.0 | 5 | 524 | 3.280000e+18 | STAR | -0.000038 | 2909 | 54653 |
| 462 | 756.0 | 1 | 314 | 3.110000e+17 | GALAXY | 0.037305 | 276 | 51909 |
| 453 | 756.0 | 1 | 313 | 2.690000e+18 | STAR | 0.000153 | 2389 | 54213 |
| 468 | 756.0 | 1 | 315 | 3.110000e+17 | GALAXY | 0.048990 | 276 | 51909 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 85 | 756.0 | 2 | 302 | 2.880000e+18 | STAR | -0.000078 | 2559 | 54208 |
| 574 | 745.0 | 2 | 576 | 3.900000e+17 | GALAXY | 0.059368 | 346 | 51693 |
| 434 | 756.0 | 3 | 259 | 3.050000e+17 | QSO | 1.812651 | 271 | 51883 |
| 321 | 752.0 | 4 | 284 | 3.250000e+17 | GALAXY | 0.156098 | 289 | 51990 |
| 446 | 756.0 | 3 | 263 | 3.050000e+17 | GALAXY | 0.055945 | 271 | 51883 |
| 567 | 745.0 | 2 | 570 | 3.900000e+17 | STAR | -0.000005 | 346 | 51693 |
| 170 | 752.0 | 5 | 515 | 3.280000e+18 | STAR | -0.000233 | 2909 | 54653 |
| 72 | 756.0 | 2 | 300 | 2.880000e+18 | STAR | -0.000048 | 2559 | 54208 |
| 277 | 752.0 | 2 | 427 | 3.380000e+17 | GALAXY | 0.053106 | 300 | 51943 |

| | fiberid |
|---|---|
| 585 | 163 |
| 89 | 255 |
| 544 | 583 |
| 464 | 251 |
| 105 | 243 |
| 288 | 230 |
| 182 | 432 |
| 472 | 212 |
| 156 | 467 |
| 218 | 786 |
| 108 | 244 |
| 447 | 51 |
| 243 | 90 |
| 51 | 229 |
| 543 | 587 |
| 578 | 221 |
| 239 | 256 |
| 163 | 541 |
| 297 | 96 |
| 178 | 626 |
| 162 | 508 |
| 532 | 341 |
| 25 | 437 |
| 414 | 119 |
| 375 | 164 |
| 29 | 349 |
| 267 | 255 |
| 467 | 1 |
| 234 | 94 |
| 27 | 358 |
| .. | ... |
| 553 | 15 |
| 21 | 430 |
| 159 | 468 |
| 134 | 90 |
| 158 | 920 |
| 499 | 880 |

```
188      425
222      830
399      434
53       186
585      163
301      547
303      599
564      112
75       263
221      829
217      705
509      488
462      245
453       11
468      211
85       267
574      108
434      152
321      396
446       79
567      195
170      343
72       313
277      121

[100 rows x 16 columns]
```

In [103]:
```
# two sided Hypothesis test
from scipy.stats import norm
from math import sqrt


def two_sided_hypo(sample_mean, pop_mean, std_dev, sample_size, alpha):
    actual_z = abs(norm.ppf(alpha/2))
    hypo_z = (sample_mean - pop_mean) / (std_dev/sqrt(sample_size))
    print('actual z value :', actual_z)
    print('hypothesis z value :', hypo_z, '\n')
    if hypo_z >= actual_z or hypo_z <= -(actual_z):
        return True
    else:
        return False
```

In [104]:
```
#one sided hypothesis test(smaller than null hypothesis)
def one_sided_hypo(sample_mean, pop_mean, std_dev, sample_size, alpha):
    actual_z = abs(norm.ppf(alpha))
    hypo_z = (sample_mean - pop_mean) / (std_dev/sqrt(sample_size))
    print('actual z value :', actual_z)
    print('hypothesis z value :', hypo_z, '\n')
```

```
        if hypo_z >= actual_z:
            return True
        else:
            return False
```

In [105]:
```
# two sided hypothesis test
alpha = 0.05
sample_mean = sample['ra'].mean()
pop_mean = df1['ra'].mean()
sample_size =  100
std_dev = df1['ra'].std()

print('H0 : μ =', pop_mean)
print('H1 : μ !=', pop_mean)
print('alpha value is :', alpha, '\n')

reject = two_sided_hypo(sample_mean, pop_mean, std_dev, sample_size, alpha)
if reject:
    print('Reject NULL hypothesis')
else:
    print('Failed to reject NULL hypothesis')
```

```
H0 : μ = 185.01467401969944
H1 : μ != 185.01467401969944
alpha value is : 0.05

actual z value : 1.959963984540054
hypothesis z value : 2.1899914168152192

Reject NULL hypothesis
```

In [106]:
```
# one sided hypothesis test
```

In [107]:
```
alpha = 0.05
sample_mean = sample['u'].mean()
pop_mean = df1['u'].mean()
sample_size =  100
std_dev = df1['u'].std()

print('H0 : μ <=', pop_mean)
print('H1 : μ >', pop_mean)
print('alpha value is :', alpha, '\n')

reject = one_sided_hypo(sample_mean, pop_mean, std_dev, sample_size, alpha)
if reject:
    print('Reject NULL hypothesis')
else:
    print('Failed to reject NULL hypothesis')
```

```
H0 : μ <= 18.540124941569278
H1 : μ > 18.540124941569278
alpha value is : 0.05

actual z value : 1.6448536269514729
hypothesis z value : -0.9591187844229192

Failed to reject NULL hypothesis
```

In [108]:
```python
# 5) correlation
```

In [109]:
```python
#  columns which are most and least correlated  based on correlation co-efficient
```

In [113]:
```python
correlation_matrix = df1.corr(method='pearson', min_periods=2)
```

In [157]:
```python
#del correlation_matrix['specobjid']
correlation_matrix.drop(['specobjid'],inplace = True)
correlation_matrix
```

Out[157]:

| | ra | dec | u | g | r | i \ |
|---|---|---|---|---|---|---|
| ra | 1.000000 | 0.010176 | 0.015354 | 0.078481 | 0.102108 | 0.118357 |
| dec | 0.010176 | 1.000000 | -0.004955 | -0.028665 | -0.017594 | 0.014852 |
| u | 0.015354 | -0.004955 | 1.000000 | 0.870558 | 0.726617 | 0.634071 |
| g | 0.078481 | -0.028665 | 0.870558 | 1.000000 | 0.962313 | 0.887921 |
| r | 0.102108 | -0.017594 | 0.726617 | 0.962313 | 1.000000 | 0.958512 |
| i | 0.118357 | 0.014852 | 0.634071 | 0.887921 | 0.958512 | 1.000000 |
| z | 0.114417 | 0.015491 | 0.595036 | 0.882497 | 0.968352 | 0.963430 |
| run | 0.431633 | -0.511464 | 0.018972 | 0.073280 | 0.074058 | 0.065505 |
| camcol | 0.110580 | 0.978189 | -0.008179 | -0.012908 | 0.006059 | 0.040259 |
| field | 0.879565 | -0.093350 | -0.005122 | 0.042027 | 0.063001 | 0.082867 |
| redshift | 0.039125 | -0.081308 | 0.136255 | 0.350250 | 0.386455 | 0.366967 |
| plate | -0.260961 | 0.376920 | -0.156316 | -0.151644 | -0.090788 | -0.039513 |
| mjd | -0.245881 | 0.310098 | -0.207585 | -0.181252 | -0.104359 | -0.047218 |
| fiberid | 0.002694 | 0.712512 | 0.001515 | 0.010551 | 0.024853 | 0.049301 |

| | z | run | camcol | field | redshift | plate \ |
|---|---|---|---|---|---|---|
| ra | 0.114417 | 0.431633 | 0.110580 | 0.879565 | 0.039125 | -0.260961 |
| dec | 0.015491 | -0.511464 | 0.978189 | -0.093350 | -0.081308 | 0.376920 |
| u | 0.595036 | 0.018972 | -0.008179 | -0.005122 | 0.136255 | -0.156316 |
| g | 0.882497 | 0.073280 | -0.012908 | 0.042027 | 0.350250 | -0.151644 |
| r | 0.968352 | 0.074058 | 0.006059 | 0.063001 | 0.386455 | -0.090788 |
| i | 0.963430 | 0.065505 | 0.040259 | 0.082867 | 0.366967 | -0.039513 |
| z | 1.000000 | 0.058995 | 0.042313 | 0.073750 | 0.365302 | -0.001333 |
| run | 0.058995 | 1.000000 | -0.390660 | 0.460247 | 0.068438 | -0.454926 |
| camcol | 0.042313 | -0.390660 | 1.000000 | -0.023134 | -0.064613 | 0.344945 |
| field | 0.073750 | 0.460247 | -0.023134 | 1.000000 | 0.022774 | -0.256309 |
| redshift | 0.365302 | 0.068438 | -0.064613 | 0.022774 | 1.000000 | -0.150421 |
| plate | -0.001333 | -0.454926 | 0.344945 | -0.256309 | -0.150421 | 1.000000 |

```
mjd      -0.003059 -0.342423  0.292568 -0.223548 -0.159006  0.972875
fiberid   0.033802 -0.323926  0.721107 -0.110965 -0.034570  0.419519

               mjd   fiberid
ra       -0.245881  0.002694
dec       0.310098  0.712512
u        -0.207585  0.001515
g        -0.181252  0.010551
r        -0.104359  0.024853
i        -0.047218  0.049301
z        -0.003059  0.033802
run      -0.342423 -0.323926
camcol    0.292568  0.721107
field    -0.223548 -0.110965
redshift -0.159006 -0.034570
plate     0.972875  0.419519
mjd       1.000000  0.360050
fiberid   0.360050  1.000000
```

In [144]:
```
# Finding the most correlated values of each column based on pearson correlation coffi
correlation_matrix['ra'].nlargest(2)
```

Out[144]:
```
ra       1.000000
field    0.879565
Name: ra, dtype: float64
```

In [117]:
```
correlation_matrix['dec'].nlargest(2)
```

Out[117]:
```
dec       1.000000
camcol    0.978189
Name: dec, dtype: float64
```

In [118]:
```
correlation_matrix['u'].nlargest(2)
```

Out[118]:
```
u    1.000000
g    0.870558
Name: u, dtype: float64
```

In [119]:
```
correlation_matrix['g'].nlargest(2)
```

Out[119]:
```
g    1.000000
r    0.962313
Name: g, dtype: float64
```

In [120]:
```
correlation_matrix['r'].nlargest(2)
```

Out[120]:
```
r    1.000000
z    0.968352
Name: r, dtype: float64
```

In [121]:
```
correlation_matrix['i'].nlargest(2)
```

Out[121]:
```
i    1.00000
z    0.96343
Name: i, dtype: float64
```

In [122]:
```
correlation_matrix['z'].nlargest(2)
```

Out[122]:
```
z    1.000000
r    0.968352
Name: z, dtype: float64
```

In [123]:
```
correlation_matrix['run'].nlargest(2)
```

Out[123]:
```
run      1.000000
field    0.460247
Name: run, dtype: float64
```

In [124]:
```
correlation_matrix['camcol'].nlargest(2)
```

Out[124]:
```
camcol    1.000000
dec       0.978189
Name: camcol, dtype: float64
```

In [125]:
```
correlation_matrix['field'].nlargest(2)
```

Out[125]:
```
field    1.000000
ra       0.879565
Name: field, dtype: float64
```

In [131]:
```
correlation_matrix['redshift'].nlargest(2)
```

Out[131]:
```
redshift    1.000000
r           0.386455
Name: redshift, dtype: float64
```

In [132]:
```
# We can see that camcol and dec are most correlated
# The type of camera scanline used highly effects the declination angle to a large ext

# Second most correlated are g,r,i,z as all of this are color bands responsiveness of
# where g,r,i,z are the wavelenths g - green , r - red , i - infrared , z - 900nm
# hence the responsiveness of the telescope is same for all the wavelenghts measured
```

In [158]:
```
# Finding the least correlated values of each column based on pearson correlation coff
correlation_matrix['ra'].nsmallest(2)
```

Out[158]:
```
plate   -0.260961
mjd     -0.245881
Name: ra, dtype: float64
```

```
In [159]: correlation_matrix['dec'].nsmallest(2)

Out[159]: run      -0.511464
          field    -0.093350
          Name: dec, dtype: float64

In [160]: correlation_matrix['u'].nsmallest(2)

Out[160]: mjd      -0.207585
          plate    -0.156316
          Name: u, dtype: float64

In [161]: correlation_matrix['g'].nsmallest(2)

Out[161]: mjd      -0.181252
          plate    -0.151644
          Name: g, dtype: float64

In [162]: correlation_matrix['r'].nsmallest(2)

Out[162]: mjd      -0.104359
          plate    -0.090788
          Name: r, dtype: float64

In [135]: correlation_matrix['i'].nsmallest(2)

Out[135]: mjd          -0.047218
          specobjid    -0.039527
          Name: i, dtype: float64

In [163]: correlation_matrix['z'].nsmallest(2)

Out[163]: mjd      -0.003059
          plate    -0.001333
          Name: z, dtype: float64

In [164]: correlation_matrix['run'].nsmallest(2)

Out[164]: dec      -0.511464
          plate    -0.454926
          Name: run, dtype: float64

In [165]: correlation_matrix['camcol'].nsmallest(2)

Out[165]: run         -0.390660
          redshift    -0.064613
          Name: camcol, dtype: float64

In [166]: correlation_matrix['field'].nsmallest(2)
```

```
Out[166]: plate   -0.256309
          mjd     -0.223548
          Name: field, dtype: float64
```

```
In [167]: correlation_matrix['redshift'].nsmallest(2)
```

```
Out[167]: mjd     -0.159006
          plate   -0.150421
          Name: redshift, dtype: float64
```

```
In [168]: # We see that the least correlated features are the dec and the run
          # As the declination angle does not depend on the type of the scanline we use in the t

          # We see that the second least correlated features are the dec and the plate
          # As the declination angle does not depend on the plate used in the telescope
```