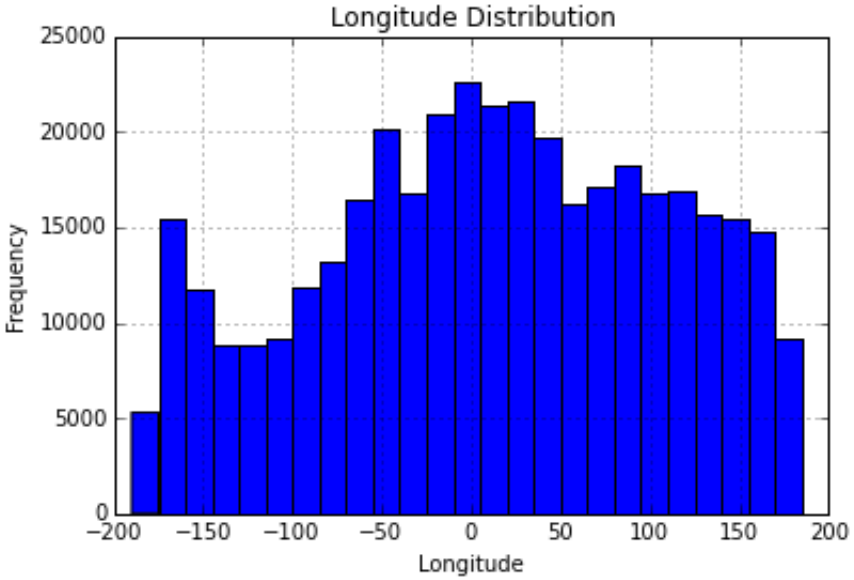
```
In [2]: runfile('C:/Data/Sandbox/Learn/Data Analysis and Interpretation/wk3/assignment_wk3.py',
wdir='C:/Data/Sandbox/Learn/Data Analysis and Interpretation/wk3')
========= 1. Data Exploration =========
---------- 1.1. General Dataset Characteristics ----------
Number of observations: 384343
Number of variables   : 10
Variables             :
    CRATER_ID
   CRATER_NAME
   LATITUDE_CIRCLE_IMAGE
   LONGITUDE_CIRCLE_IMAGE
   DIAM_CIRCLE_IMAGE
   DEPTH_RIMFLOOR_TOPOG
   MORPHOLOGY_EJECTA_1
   MORPHOLOGY_EJECTA_2
   MORPHOLOGY_EJECTA_3
   NUMBER_LAYERS

---------- 1.2. LONGITUDE_CIRCLE_IMAGE ----------
min:  -179.997
max:  179.997
Unique values           : 231245
Top 5 values and counts :
 -53.500     9
-163.016    8
 78.034     8
-3.987      8
-50.142     8
dtype: int64

---------- 1.3. LATITUDE_CIRCLE_IMAGE ----------
min:  -86.7
max:  85.70200000000001
Unique values           : 129197
Top 5 values and counts :
 -23.634    17
-2.572     16
-12.406    15
-22.340    15
-17.317    15
dtype: int64

---------- 1.4. DIAM_CIRCLE_IMAGE ----------
min:  1.0
```

```
max:  1164.22
Unique values           : 129197
Top 10 values and counts :
 1.01    6298
1.02    6077
1.03    6035
1.04    5941
1.05    5771
1.06    5556
1.07    5454
1.08    5418
1.09    5197
1.10    5088
dtype: int64
========= 2. Data Management =========

---------- 2.1. LONGITUDE_CIRCLE_IMAGE ----------
Frequences in bins with length 15:
-10     22587
 20     21624
 5      21339
-25     20902
-55     20190
 35     19694
 80     18286
 65     17066
 110    16935
-40     16782
 95     16764
-70     16465
 50     16250
 125    15677
 140    15479
-175    15431
 155    14796
-85     13176
-100    11836
-160    11751
 170     9202
-115     9180
-130     8841
-145     8802
-190     5288
dtype: int64
```

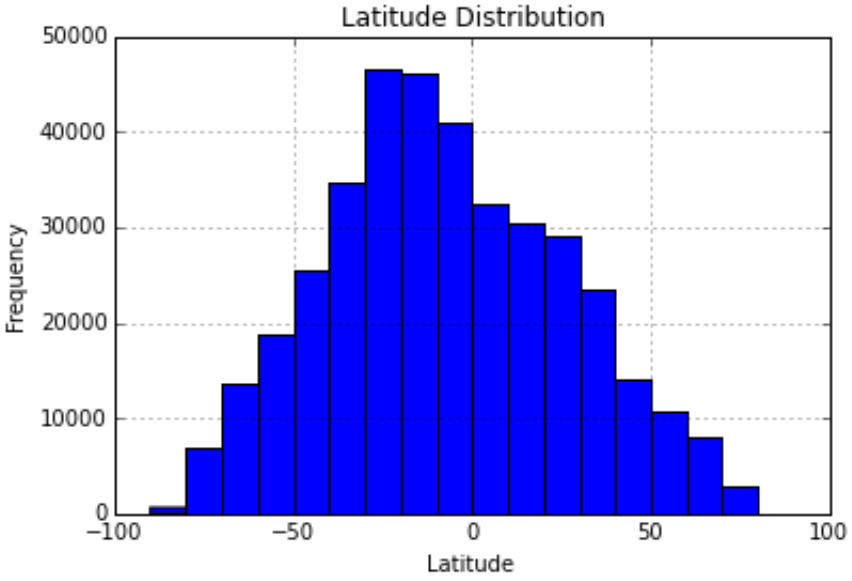
```
Group frequences, %:
(-10, 5]        5.876782
(20, 35]        5.626224
(5, 20]         5.552072
(-25, -10]      5.438371
(-55, -40]      5.253120
(35, 50]        5.124069
(80, 95]        4.757729
(65, 80]        4.440305
(110, 125]      4.406220
(-40, -25]      4.366412
(95, 110]       4.361729
(-70, -55]      4.283934
(50, 65]        4.227994
(125, 140]      4.078909
(140, 155]      4.027392
(-175, -160]    4.014903
(155, 170]      3.849686
(-85, -70]      3.428188
(-100, -85]     3.079541
(-160, -145]    3.057425
(170, 185]      2.394216
(-115, -100]    2.388492
(-130, -115]    2.300289
(-145, -130]    2.290142
(-190, -175]    1.375854
dtype: float64

After grouping the longitude values, we can observe that relatively uniform
distribution of the craters, although there are areas with smaller 'population'.
```




```
---------- 2.2. LATITUDE_CIRCLE_IMAGE ----------
Frequences in bins:
-30    46504
-20    46158
-10    40921
-40    34577
 0     32362
 10    30411
 20    28990
-50    25396
 30    23365
-60    18758
 40    14160
-70    13527
 50    10801
 60     7974
-80     6984
 70     2780
-90      631
 80       44
dtype: int64

Group frequences, %:
(-30, -20]    12.099609
```

```
(-20, -10]    12.009585
(-10, 0]      10.647000
(-40, -30]     8.996391
(0, 10]        8.420083
(10, 20]       7.912464
(20, 30]       7.542742
(-50, -40]     6.607640
(30, 40]       6.079205
(-60, -50]     4.880536
(40, 50]       3.684209
(-70, -60]     3.519513
(50, 60]       2.810250
(60, 70]       2.074709
(-80, -70]     1.817127
(70, 80]       0.723312
(-90, -80]     0.164176
(80, 90]       0.011448
dtype: float64
```

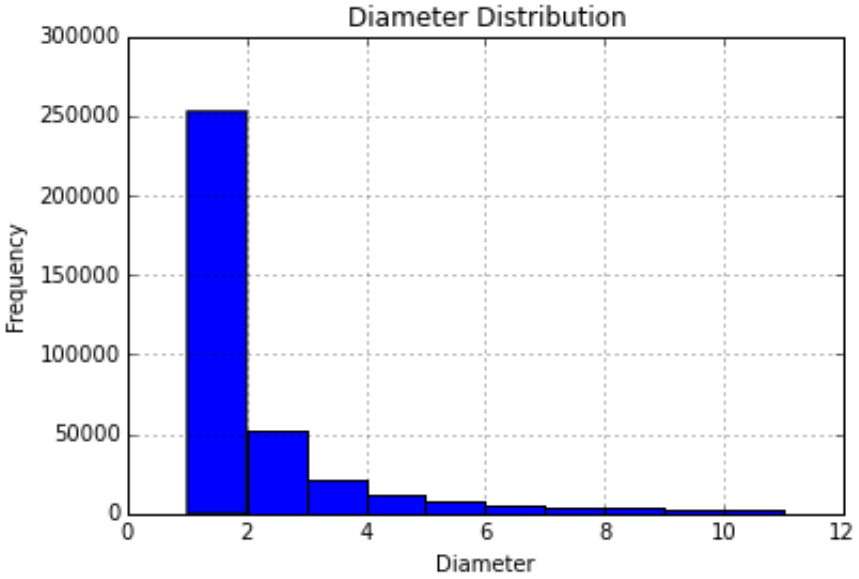
From the binned latitude values we can see that about 35% of the craters are located between 0 and -30 degrees.



---------- 2.3. DIAM_CIRCLE_IMAGE ----------
Frequences in bins range(0,2000, 100):

```
0       384039
100        255
200         28
300         12
400          5
500          1
1100         1
1000         1
600          1
700          0
1800         0
800          0
1700         0
1200         0
1300         0
1400         0
1500         0
1600         0
900          0
dtype: int64

Most of the diameter values are in the range (0,100].
Let's refine further the (0,100] range.

Frequences in bins range(0,100, 5):
0     336850
5      23135
10      8668
15      4802
20      3009
25      2060
30      1380
35      1015
40       788
45       566
50       442
55       315
60       254
65       206
70       128
75       126
80        94
85        80
90        73
```

```
dtype: int64

We can manually split into bins:
[0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 15, 20, 40, 60, 100, 500, 1000, 2000]
With following frequences (in %):
(1, 2]          65.177459
(2, 3]          13.303742
(3, 4]           5.418597
(4, 5]           2.929154
(20, 40]         1.942015
(5, 6]           1.907671
(6, 7]           1.419565
(15, 20]         1.249405
(12, 15]         1.161463
(7, 8]           1.101100
(8, 9]           0.880203
(0, 1]           0.814117
(9, 10]          0.710823
(10, 11]         0.596082
(40, 60]         0.549249
(11, 12]         0.497732
(60, 100]        0.262526
(100, 500]       0.078055
(500, 1000]      0.000520
(1000, 2000]     0.000520
dtype: float64
```


Diameter Distribution
300000
250000
200000
150000
100000
50000
0
Frequency
0
2
4
6
8
10
12
Diameter

In [3]: