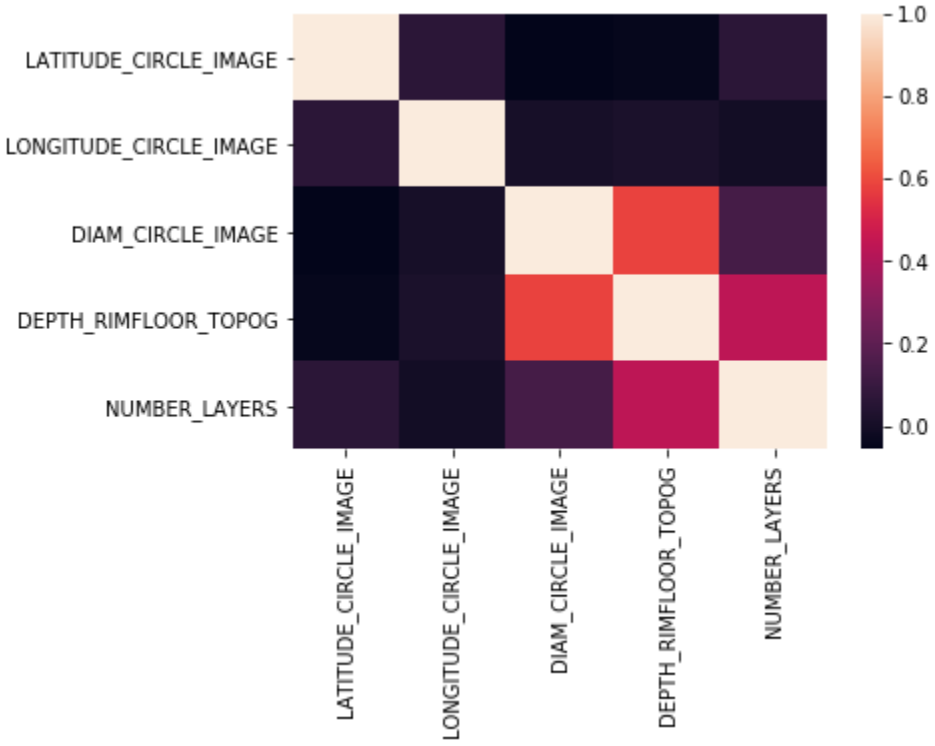
# Inferential Analysis
# Correlation Coefficient
***With***
***Python***

**AUTHOR:**
LUIS ORELLANA ALTAMIRANO

# Correlation Coefficient

The dataset used through this exercise is called mars crater (marscrater_pds.csv) which has information about “Heavily cratered terrain on Mars was created between 4.2 to 3.8 billion years ago”. Furthermore, it contains 378,540 craters statistically complete for diameters D ≥ 1 km. More information about this data could be found in the Ph.D. Thesis Planetary Surface Properties, Cratering Physics, and the Volcanic History of Mars from a New Global Martian Crater Database (2011) by Robbins, S.J., University of Colorado at Boulder[1]

It has 5 variables; therefore, a good first sight is to make a correlational matrix which is below:

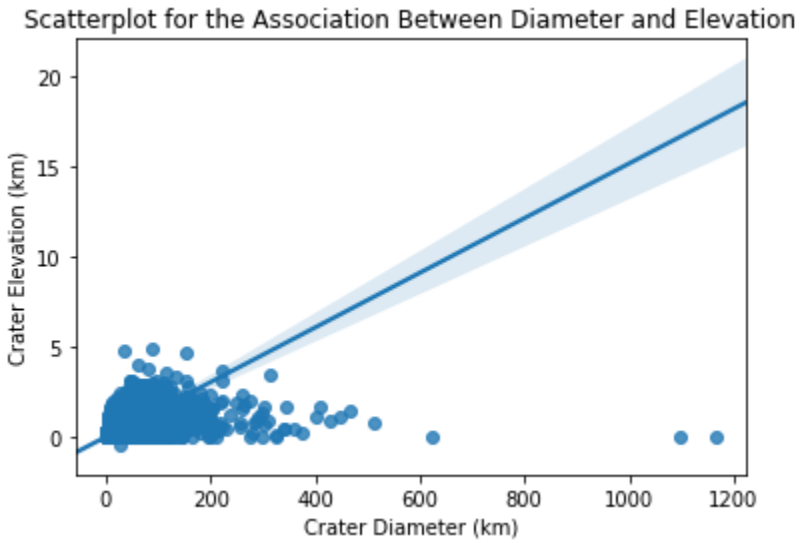



[1] http://about.sjrdesign.net/files/thesis/RobbinsThesis_LargeMB.pdf

It’s possible to see that the maximum correlation coefficient is between “DIAM_CIRCLE_IMAGE” and “DEPTH_RIMFLOOR_TOPOG”. The obtained coefficient is as next:

0.58671343794598463

This means, there is statistically meaningful relationship between diameter and elevation among surface craters. In addition, while craters are deeper, they have larger diameter. It could be checked through the next scatter plot:



The Python code is as follow:

```
import pandas
import numpy
import seaborn
import scipy
import matplotlib.pyplot as plt
import seaborn as sns

data = pandas.read_csv(‘marscrater_pds.csv’, low_memory=False)

data['DIAM_CIRCLE_IMAGE’] =
data['DIAM_CIRCLE_IMAGE’].convert_objects(convert_numeric=True)
data['DEPTH_RIMFLOOR_TOPOG’] =
data['DEPTH_RIMFLOOR_TOPOG’].convert_objects(convert_numeric=True)

data['DIAM_CIRCLE_IMAGE’]=data['DIAM_CIRCLE_IMAGE’].replace(’ ’, numpy.nan)
data['DEPTH_RIMFLOOR_TOPOG’]=data['DEPTH_RIMFLOOR_TOPOG’].replace(’ ’,
numpy.nan)

data_clean=data.dropna()

print ('association between urbanrate and internetuserate’)
print (scipy.stats.pearsonr(data_clean['DIAM_CIRCLE_IMAGE’],
data_clean['DEPTH_RIMFLOOR_TOPOG’]))

scat1 = seaborn.regplot(x=“DIAM_CIRCLE_IMAGE”, y=“DEPTH_RIMFLOOR_TOPOG”,
fit_reg=True, data=data_clean)
plt.xlabel('Crater Diameter (km)’)
plt.ylabel('Crater Elevation (km)’)
plt.title('Scatterplot for the Association Between Diameter and Elevation’)

corr = data_clean.corr()

sns.heatmap(corr,
       xticklabels=corr.columns,
       yticklabels=corr.columns)
```