

---

# **ml4rocks**

***Release 0.0.1***

**Apr 21, 2020**



<b>1</b>	<b>Installation</b>	<b>3</b>
1.1	Requirements . . . . .	3
1.2	Package managers . . . . .	3
1.3	From source . . . . .	3
1.4	Testing . . . . .	4
<b>2</b>	<b>Quickstart</b>	<b>5</b>
2.1	Importing data . . . . .	5
2.2	Data distributions . . . . .	6
2.3	The feature space . . . . .	7



This site aims to teach the basics of exploratory data analysis and machine learning modeling using real examples of mineral exploration data sources. Geologists, geophysicists and geochemists may find the contents of this site useful to get familiar with the concepts used in data science. Read the **'tutorials'** to get started, or go to the **'installation'** page to see the packages needed to do the tutorials yourself.



BISIP is compatible with Python 3.6+.

## 1.1 Requirements

The following packages are required and should be installed automatically on setup:

- [numpy](#)
- [cython](#)
- [matplotlib](#)
- [emcee](#)

These optional packages are used for progress bars and corner plots:

- [tqdm](#)
- [corner](#)

## 1.2 Package managers

TODO: Add BISIP to conda-forge.

## 1.3 From source

BISIP is developed on [GitHub](#). Clone the repository to your computer. Then navigate to the bisip directory. Finally run the setup.py script with Python.

```
git clone https://github.com/clberube/ml4rocks
cd bisip2
python setup.py install -f
```

## 1.4 Testing



In this first tutorial we load the petrophysical properties dataset and use matplotlib to visualize the data space. start by importing the numpy, pandas and matplotlib libraries.

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

## 2.1 Importing data

```
[4]: url = 'https://gist.githubusercontent.com/clberube/786a5774acfc3137039288d1a6d84fc0/
↳ raw/dc59ee0104b02141a9ca000ef52faf3d854387a1/MalarticRockPhysics.csv'
df = pd.read_csv(url)
```

First let's inspect some basic properties of this simple data file, like its shape (n\_rows, n\_columns), its index and its column names.

```
[10]: print(df.shape)
```

```
(845, 3)
```

```
[12]: df.head()
```

```
[12]:
```

	Lithology	RockDensity	log10_MagSusceptibility
0	Meta-sedimentary_rock	2.755	-3.573489
1	Meta-sedimentary_rock	2.742	-3.131356
2	Meta-sedimentary_rock	2.829	-2.058489
3	Felsic-intermediate_intrusive_rock	2.674	-4.359519
4	Meta-sedimentary_rock	2.786	-3.434152

The data set contains 845 rows and 3 columns. The first column `Lithology` is a description of each sample's rock type. The second column `RockDensity` is the skeletal density of the rock samples expressed in  $\text{g/cm}^3$ . The third column `log10_MagSusceptibility` is the logarithm base 10 value of their magnetic susceptibility (measured in SI units). Each rock sample is labeled with an index number (from 0 to 844).

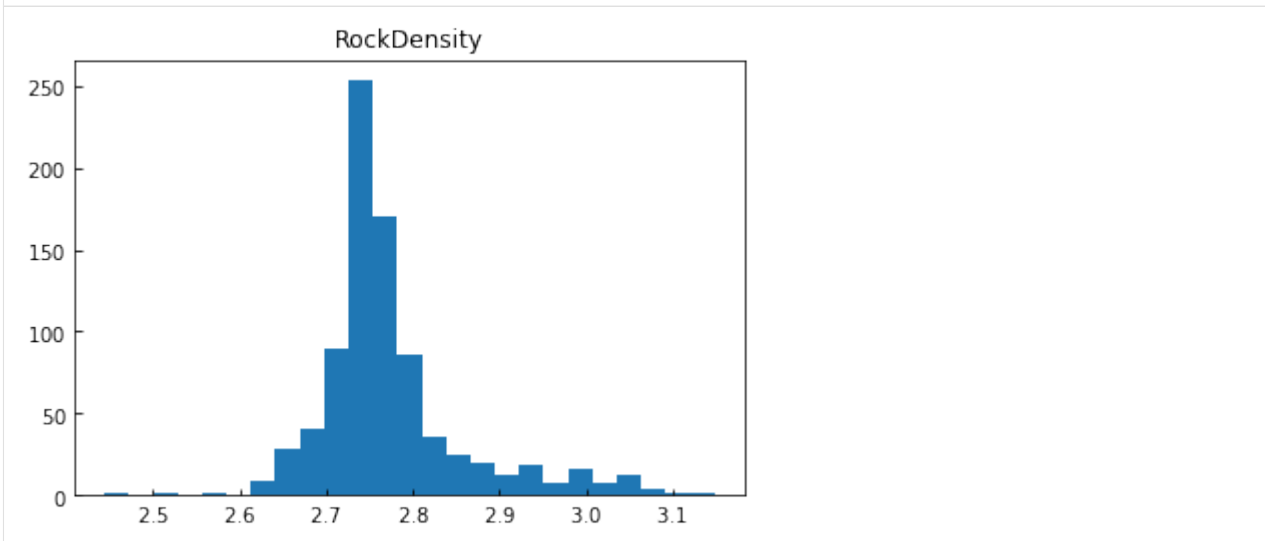
## 2.2 Data distributions

This dataset contains three features, or properties, which have been observed on each rock sample. The first feature is the `Lithology` column. It contains text values that correspond to categories of rock types. Let's find out how each category of rock type is represented in the dataset.

```
[15]: df['Lithology'].value_counts()
[15]: Meta-sedimentary_rock      583
      Mafic_dyke                140
      Felsic-intermediate_intrusive_rock  122
      Name: Lithology, dtype: int64
```

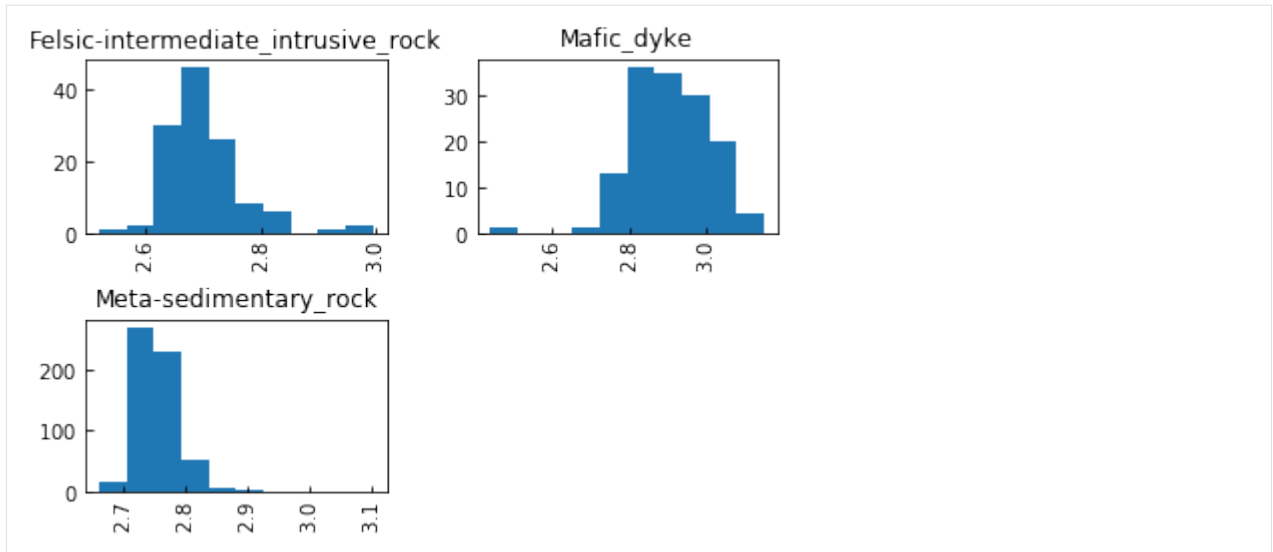
The *Meta-sedimentary\_rock* category is the most common one, with 583 rock samples in it. Other rock types include *Mafic\_dyke* and *Felsic-intermediate\_intrusive\_rock*. The second feature is `RockDensity`. This is a continuous variable, we can visualize its distribution using a histogram.

```
[98]: df.hist(column='RockDensity', bins=25, grid=False)
[98]: array([[<matplotlib.axes._subplots.AxesSubplot object at 0x11c9a28d0>]],
      dtype=object)
```



We can also visualize the rock density distribution after grouping the dataset by rock type.

```
[52]: df.hist(column='RockDensity', by='Lithology')
[52]: array([[<matplotlib.axes._subplots.AxesSubplot object at 0x11a674320>,
      <matplotlib.axes._subplots.AxesSubplot object at 0x11499cba8>],
      [<matplotlib.axes._subplots.AxesSubplot object at 0x114a75320>,
      <matplotlib.axes._subplots.AxesSubplot object at 0x114917e80>]],
      dtype=object)
```

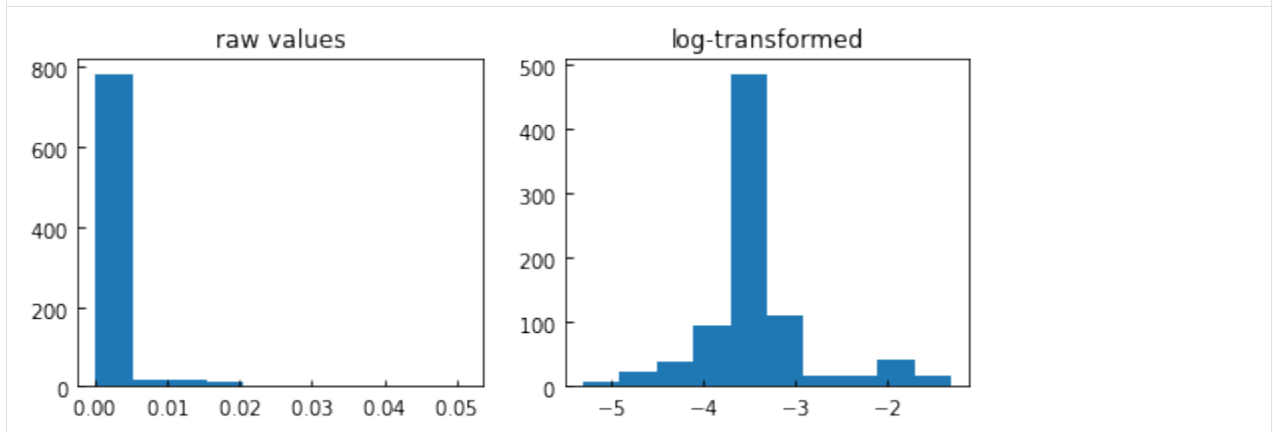


The third feature `log10_MagSusceptibility` is log-transformed. This is a useful data transformation when the values of a variable can vary over several orders of magnitude. Let's visualize the distribution of the rock samples magnetic susceptibility to understand what happens when this property is log-transformed.

```
[99]: log_ms = df['log10_MagSusceptibility']

fig, ax = plt.subplots(1, 2, figsize=(8, 3))
ax[0].hist(10*log_ms)
ax[0].set_title('raw values')
ax[1].hist(log_ms)
ax[1].set_title('log-transformed')

[99]: Text(0.5, 1.0, 'log-transformed')
```



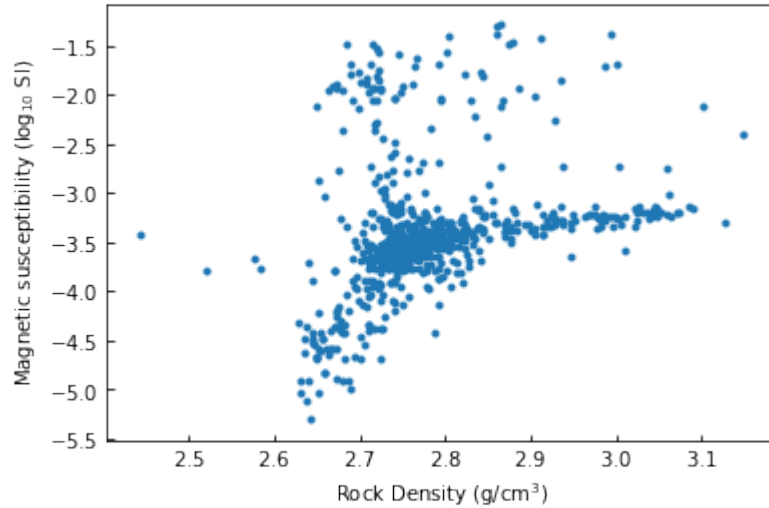
Performing a log-transform allows the magnetic susceptibility data to be spread out more evenly in its range of possible values.

## 2.3 The feature space

The feature space, or data space, is the 2D space in which our `RockDensity` and `log10_MagSusceptibility` measurements lie. It is straightforward to visualize this space with a simple scatter plot.

```
[79]: fig, ax = plt.subplots()
      ax.scatter(df['RockDensity'], df['log10_MagSusceptibility'], marker='.')
      ax.set_xlabel('Rock Density (g/cm3)')
      ax.set_ylabel('Magnetic susceptibility ($\log_{10}$ SI)')
```

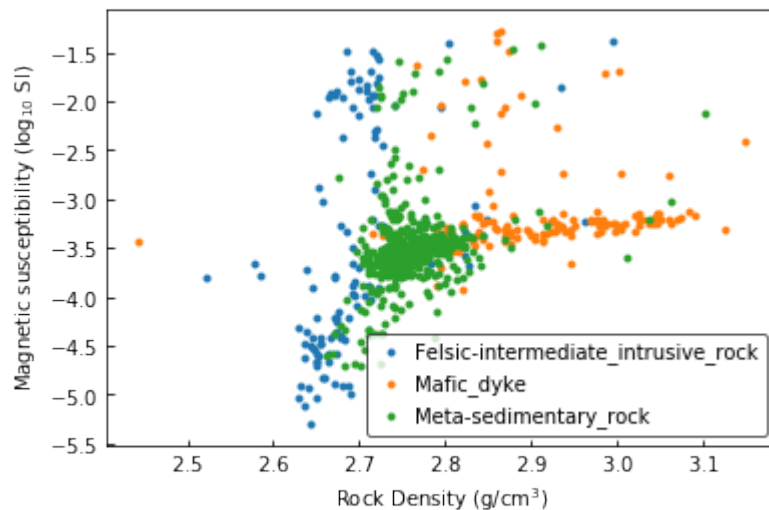
```
[79]: Text(0, 0.5, 'Magnetic susceptibility ($\log_{10}$ SI)')
```



Even better, we could use a color code to identify each point in the scatter plot by its lithology.

```
[101]: fig, ax = plt.subplots()
      groups = df.groupby('Lithology')
      for name, group in groups:
          ax.scatter(group['RockDensity'],
                     group['log10_MagSusceptibility'],
                     marker='.',
                     label=name)
      ax.legend()
      ax.set_xlabel('Rock Density (g/cm3)')
      ax.set_ylabel('Magnetic susceptibility ($\log_{10}$ SI)')
```

```
[101]: Text(0, 0.5, 'Magnetic susceptibility ($\log_{10}$ SI)')
```



It is clear from the previous figure that the various rock types have contrasting physical properties. Navigate through the various tutorials on this site to further explore the relationships between lithology, magnetic susceptibility and rock density.