

Homework_7

April 22, 2018

```
In [10]: import numpy as np
         %matplotlib inline
         import matplotlib.pyplot as plt
         import pandas as pd
         import seaborn as sns
         import scipy as scipy
```

Question_1:

```
In [47]: def f(t):
         return np.exp(-t) * np.cos(2*np.pi*t)

         # TODO:
         # create two random series using np.arange

         t1 = np.arange(15)
         t2 = np.random.rand(15,1)
         t3 = np.random.rand(15,1)

         # plot two using plt.subplot
         # for each plot define a different color

         #plot 1: x-axis = t1, y-axis = f(t1)

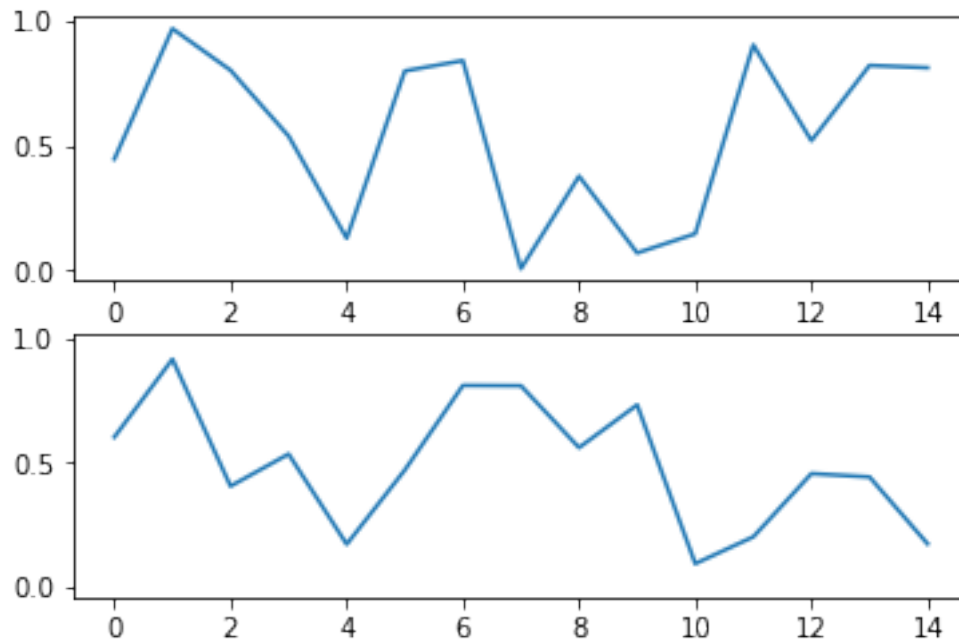
         #plot 2: x-axis = t2, y-axis = f(t2)
         f, (ax1, ax2) = plt.subplots(2, 1, sharey=True)
         plt.plot([1,2,3])
         # now create a subplot which represents the top plot of a grid
         # with 2 rows and 1 column. Since this subplot will overlap the
         # first, the plot (and its axes) previously created, will be removed

         ax1.plot(t1,t2)
         ax2.plot(t1,t3)

         #plt.subplot(211)
```

```
#plt.plot(range(12))
#plt.subplot(212, facecolor='y')
```

Out[47]: [<matplotlib.lines.Line2D at 0x11b2117b8>]



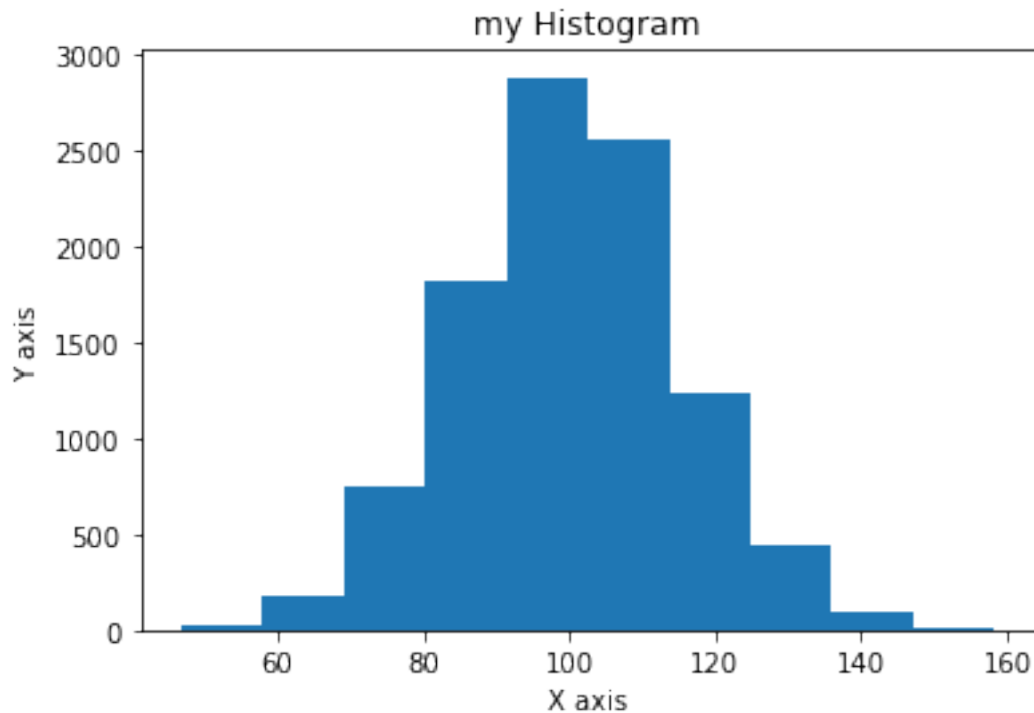
Question_2:

```
In [48]: # Fixing random state for reproducibility
np.random.seed(19680801)

mu, sigma = 100, 15
x = mu + sigma * np.random.randn(10000)

# the histogram of the data
# plot a histogram using plt.hist() function
# define axis labels and add title for the plot
mu, sigma = 100, 15
x = mu + sigma * np.random.randn(10000)
plt.hist(x,10)
plt.xlabel('X axis')
plt.ylabel('Y axis')
plt.title(r'my Histogram')
# the histogram of the data
# plot a histogram using plt.hist() function
# define axis labels and add title for the plot
```

Out[48]: <matplotlib.text.Text at 0x11b3bfa90>



Question_3:

```
In [66]: from matplotlib.ticker import NullFormatter # useful for `logit` scale
```

```
# make up some data in the interval ]0, 1[
y = np.random.normal(loc=0.5, scale=0.4, size=10000)
y = y[(y > 0) & (y < 1)]
y.sort()
x = np.arange(len(y))

# plot 4 plots in same plot, remember to define the location of each plot using plt.s

# create a linear plot of (x and y)
plt.figure(figsize=(18,18))
plt.subplot(221)
plt.plot(x, y)
plt.yscale('linear')
plt.ylabel('linear')
plt.title('linear')
plt.grid(True)
plt.gca().xaxis.grid(True, which='minor') # minor grid on too
```

```

#create a log plot of (x and y)
plt.subplot(222)
plt.plot(x, y)
plt.yscale('log')
plt.ylabel('log')
plt.title('log')

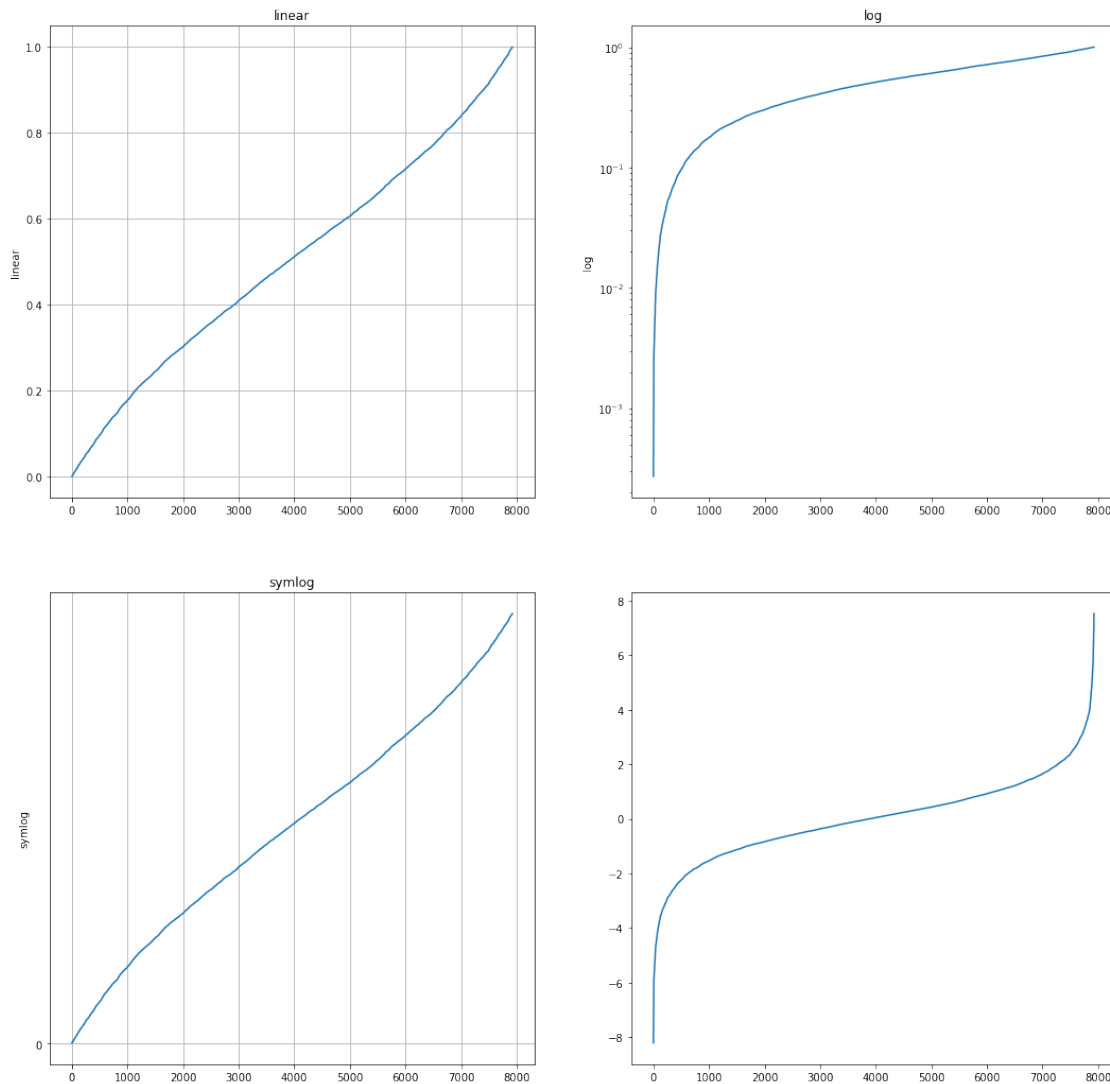
#create a symetric log of (x and y)
plt.subplot(223)
plt.plot(x, y)
plt.yscale('symlog')
plt.ylabel('symlog')
plt.title('symlog')
plt.grid(True)

#create a logit plot of (x and y)

plt.subplot(224)
plt.plot(x, scipy.special.logit(y))

```

Out[66]: [<matplotlib.lines.Line2D at 0x11eb75898>]



1 Question_4:

Find a dataset of your own * plot the correlation between at least three columns. * Plot the density histogram of at least two columns. * plot the correlation matrix (colorful) of your variables.

1.1 1. Plot the correlation between at least three columns

```
In [4]: hotel=pd.read_csv('hotel_energy.csv')
```

```
In [5]: hotel.head()
```

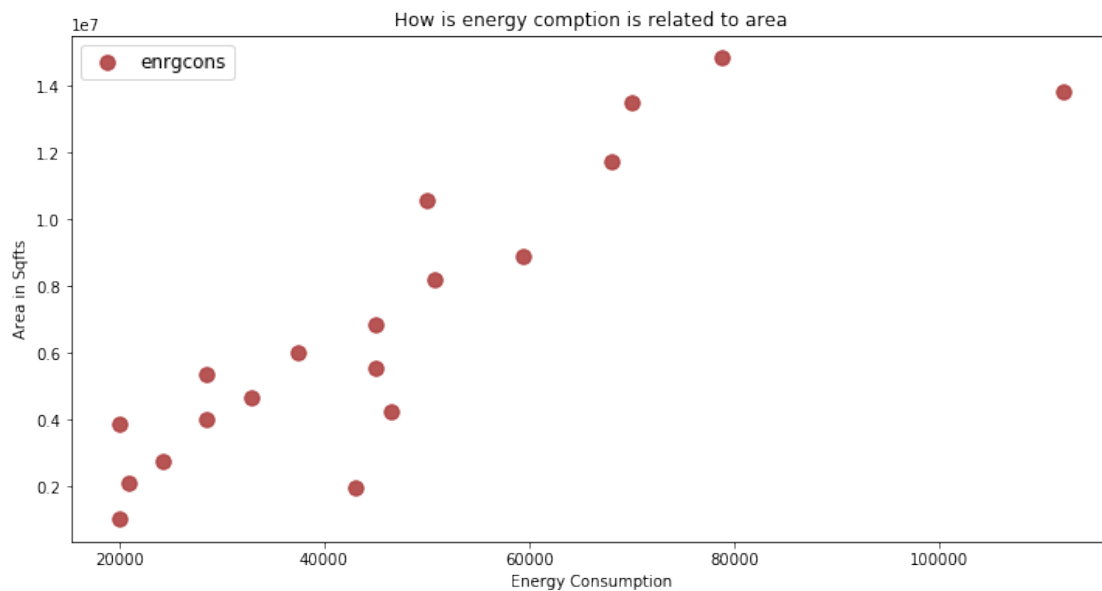
```
Out[5]:
```

	hotel	enrgcons	area	age	numrooms	occrate	effrooms
0	1	1953916	43000	6	420	0.3260	136.92
1	2	1045555	19979	16	215	0.6300	135.45

2	3	4245313	46529	7	273	0.6505	177.59
3	4	2126199	20962	6	222	0.7050	156.51
4	5	2785958	24212	5	474	0.6970	330.38

```
In [12]: plt.figure(figsize=(12,6))
plt.scatter(x=hotel['area'],y=hotel['enrgcons'],color=('brown'),s=90,alpha=0.8)
plt.legend(loc=2,fontsize='large')
plt.xlabel('Area in Sqfts')
plt.ylabel('Energy Consumption')
plt.title('How is energy comption is related to area')
```

Out[12]: <matplotlib.text.Text at 0x113366b00>

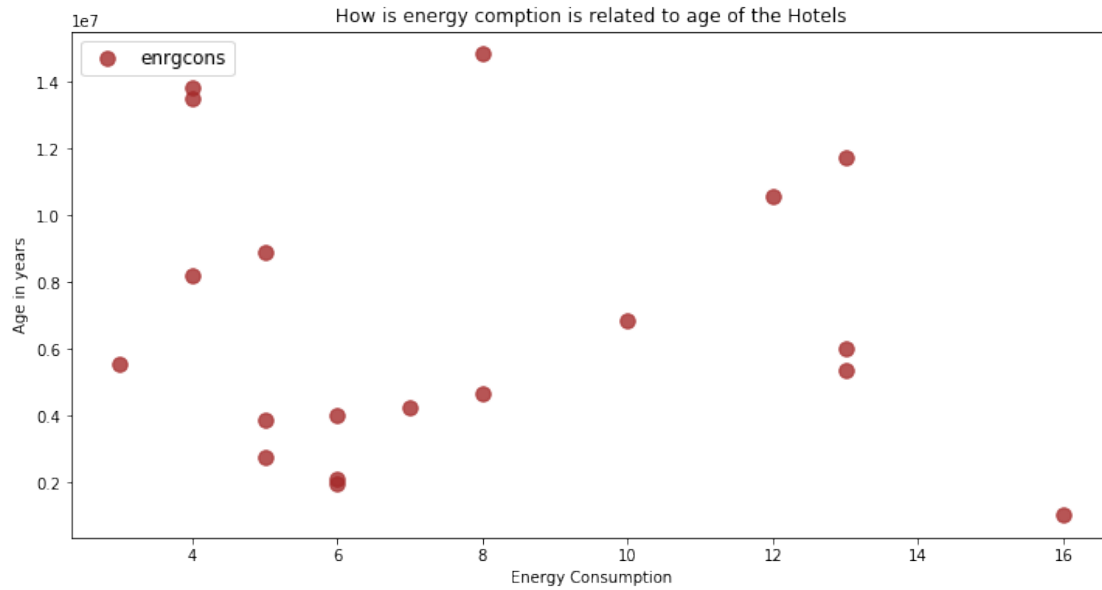


```
In [19]: np.corrcoef(hotel.area,hotel.enrgcons)
```

Out[19]: array([[1. , 0.8789621],
[0.8789621, 1.]])

```
In [13]: plt.figure(figsize=(12,6))
plt.scatter(x=hotel['age'],y=hotel['enrgcons'],color=('brown'),s=90,alpha=0.8)
plt.legend(loc=2,fontsize='large')
plt.xlabel('Age in years')
plt.ylabel('Energy Consumption')
plt.title('How is energy comption is related to age of the Hotels')
```

Out[13]: <matplotlib.text.Text at 0x118e8c710>

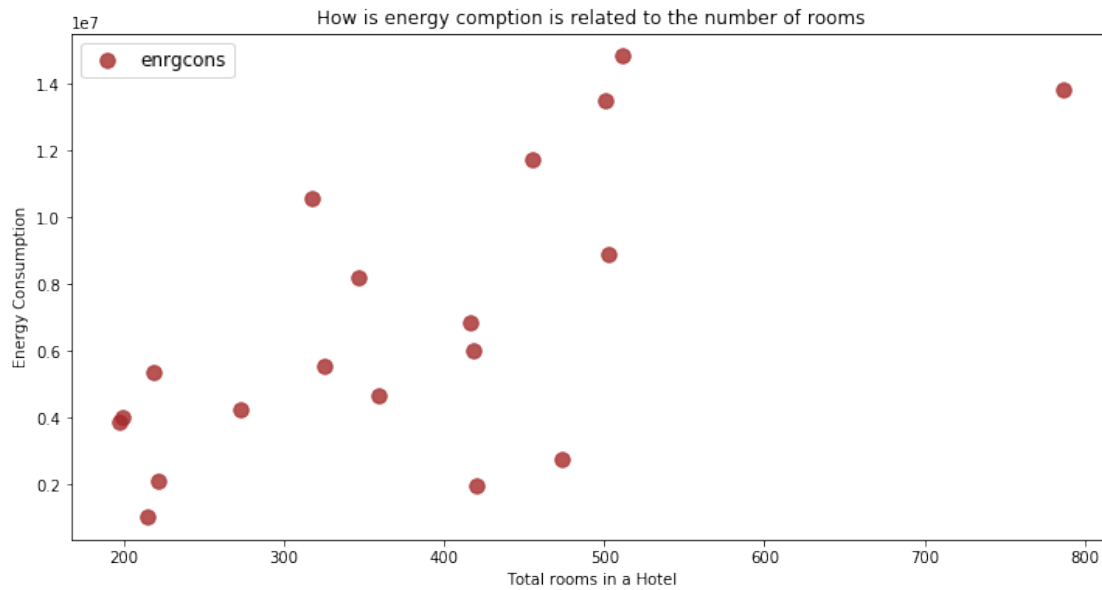


```
In [18]: np.corrcoef(hotel.age,hotel.enrgcons)
```

```
Out[18]: array([[ 1.          , -0.09942026],
                [-0.09942026,  1.          ]])
```

```
In [15]: plt.figure(figsize=(12,6))
plt.scatter(x=hotel['numrooms'],y=hotel['enrgcons'],color=('brown'),s=90,alpha=0.8)
plt.legend(loc=2,fontsize='large')
plt.xlabel('Total rooms in a Hotel')
plt.ylabel('Energy Consumption')
plt.title('How is energy comption is related to the number of rooms')
```

```
Out[15]: <matplotlib.text.Text at 0x119671048>
```



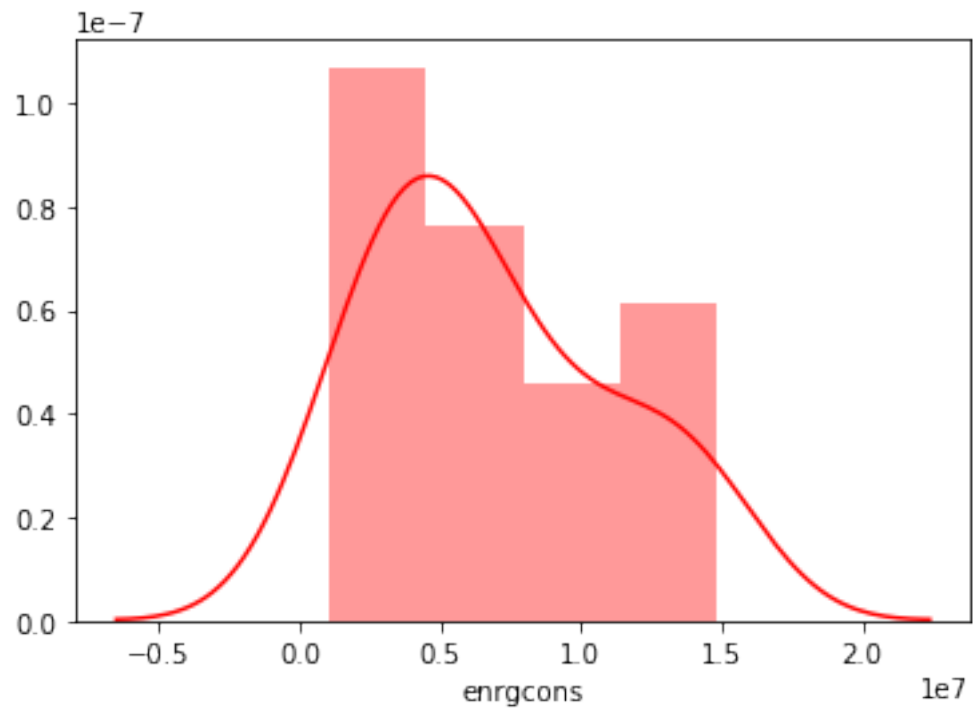
```
In [17]: np.corrcoef(hotel.numrooms,hotel.enrgcons)
```

```
Out[17]: array([[ 1.          ,  0.68509097],
                [ 0.68509097,  1.          ]])
```

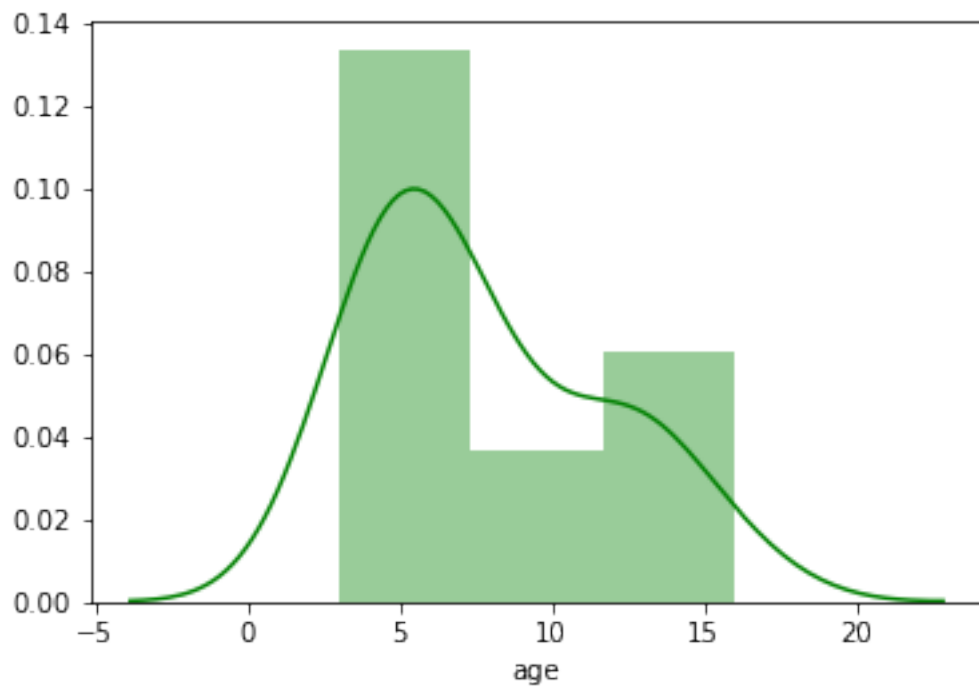
1.2 2. Plot the density histogram of at least two columns

```
In [22]: #sns.kdeplot(, shade=True, )
sns.distplot(hotel.enrgcons,color="r")
```

```
Out[22]: <matplotlib.axes._subplots.AxesSubplot at 0x119af5438>
```

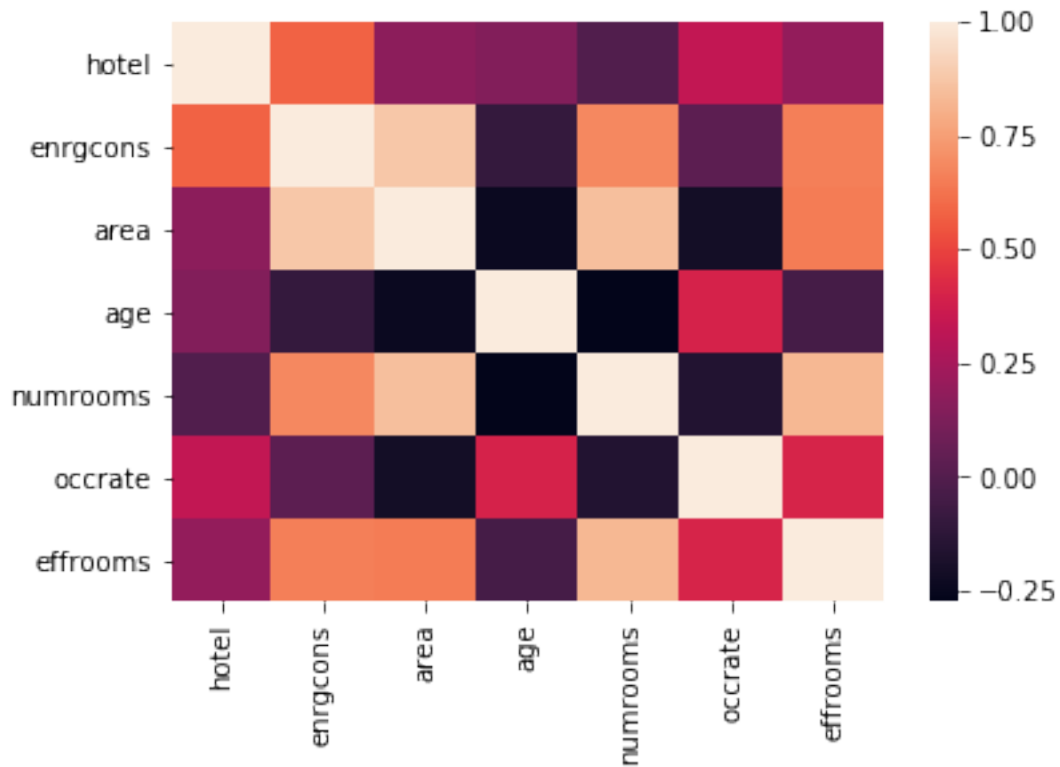
```
In [24]: ax = sns.distplot(hotel.age, color="g")
```



1.3 3. Plot the correlation matrix (colorful) of your variables

```
In [46]: corr = hotel.corr()  
         sns.heatmap(corr)
```

```
Out[46]: <matplotlib.axes._subplots.AxesSubplot at 0x11b0df208>
```



```
In [ ]:
```

```
In [ ]:
```