exercicios-29.05.2017

June 13, 2017

0.0.1 Exercicio 1

In [0]:

In [3]:
```
import numpy as np
import matplotlib.pyplot as plt
```

In [4]:
```
file=open("livro.txt","r")
```

In [5]:
```
texto=file.readlines()
```

In [6]:
```
dict={}
for line in texto:
    for word in line.split():
        if word in dict:
            dict [word]=dict[word]+1
        else:
            dict[word]=1
```

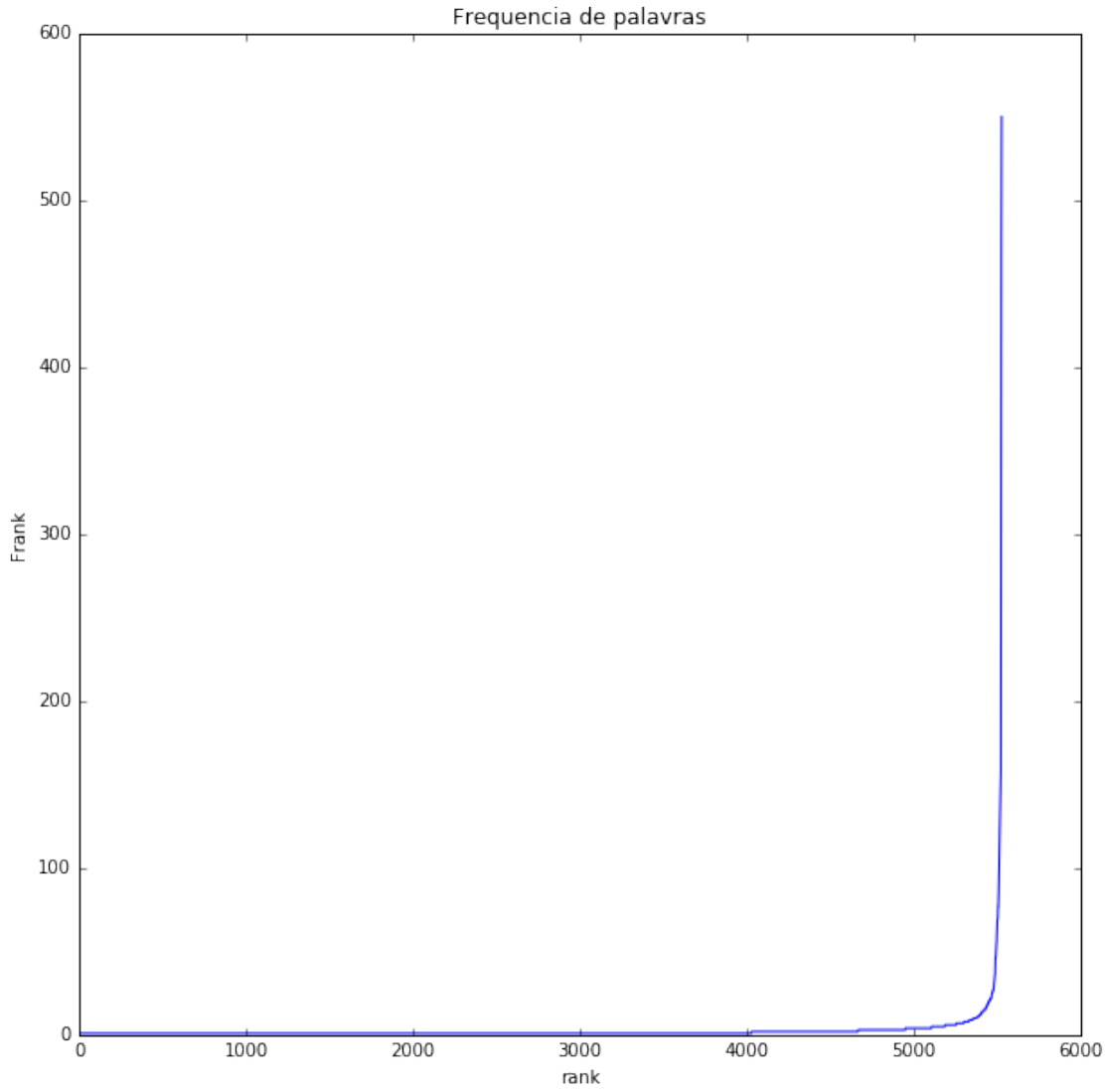
In [0]:
```
dict
```

In [8]:
```
import numpy as np
oco=dict.values()
oco.sort()
x= np.array(range(len(oco)))+1
y = np.array(oco)
```

In [9]:
```
import matplotlib.pyplot as plt
fig = plt.figure(figsize=(10,10))
ax = fig.gca()
x = fig.gca()
plt.plot(oco)
ax.set_title('Frequencia de palavras')
ax.set_xlabel('rank')
ax.set_ylabel('Frank')
```

Out[9]:
```
<matplotlib.text.Text at 0x7f5c5e82de50>
```

Out[9]:




```
In [10]: oco=dict.values()
         oco.sort()
         oco.reverse()
         x=np.array(range(len(oco)))+1
         y=np.array(oco)

In [12]: logx=np.log(x)
         logy=np.log(y)
         fig = plt.figure(figsize=(10, 10)) # define plot area
         ax = fig.gca() # define axis
         x=fig.gca()
         plt.plot(logx,logy)
         ax.set_title('frequencia de palavras das Lusiadas') # Título Principal
```

```
ax.set_xlabel('rank') # Eixo x
ax.set_ylabel('Freq')# Eixo y
```

```
---------------------------------------------------------------------------

AttributeError                            Traceback (most recent call last)

<ipython-input-12-f456f762c9f0> in <module>()
----> 1 logx=np.log(x)
      2 logy=np.log(y)
      3 fig = plt.figure(figsize=(10, 10)) # define plot area
      4 ax = fig.gca() # define axis
      5 x=fig.gca()


AttributeError: 'AxesSubplot' object has no attribute 'log'
```

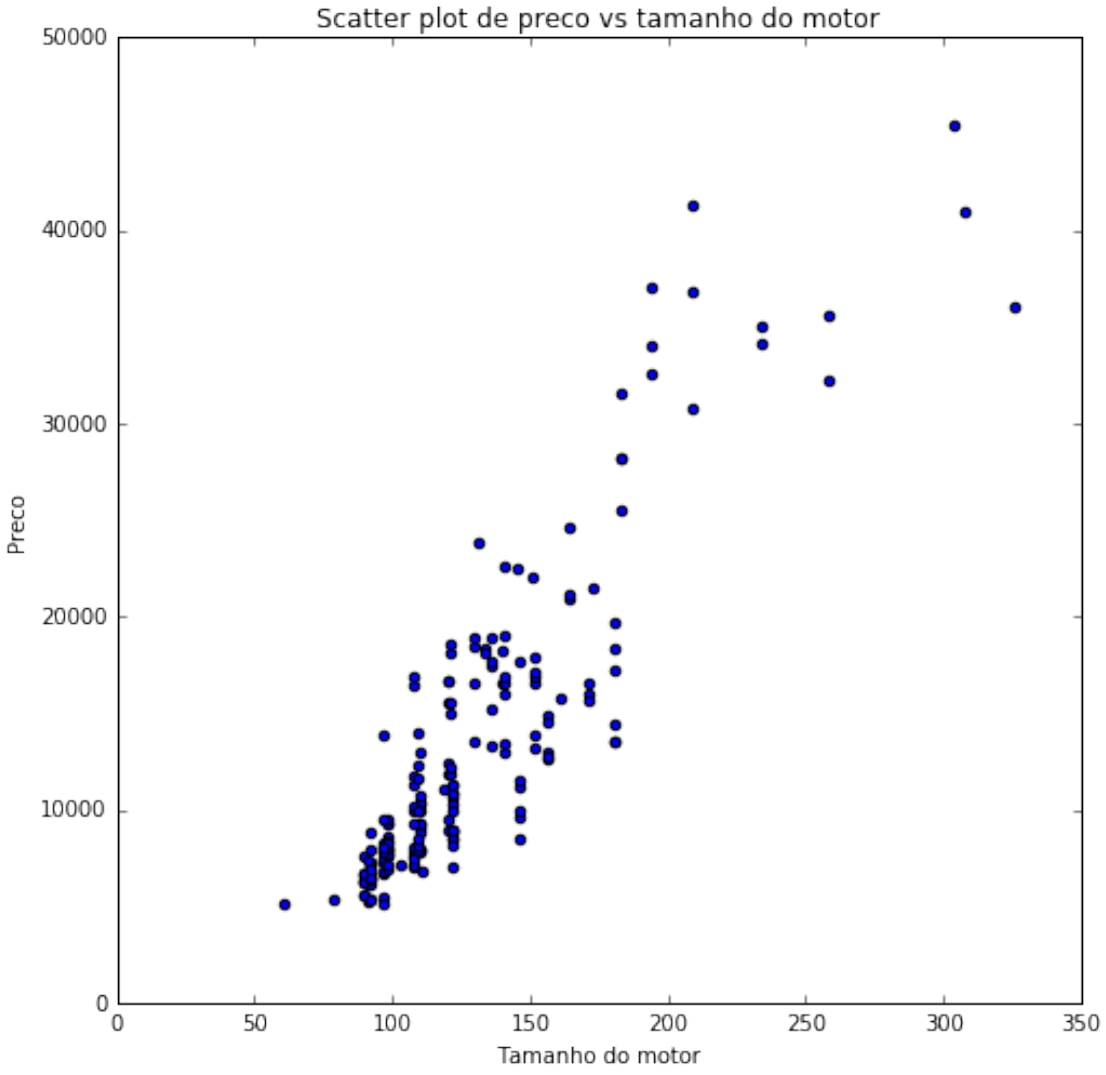
0.0.2 Exercicio 2

```
In [13]: def read_auto_data(fileName = "Automobile price data.csv"):
             'Function to load the auto price data set from a .csv file'
             import pandas as pd
             import numpy as np

             ## Read the .csv file with the pandas read_csv method
             auto_prices = pd.read_csv(fileName)

             ## Remove rows with missing values, accounting for mising values coded as '?'
             cols = ['price', 'bore', 'stroke',
                   'horsepower', 'peak-rpm']
             for column in cols:
                 auto_prices.loc[auto_prices[column] == '?', column] = np.nan
             auto_prices.dropna(axis = 0, inplace = True)

             ## Convert some columns to numeric values
             for column in cols:
                 auto_prices[column] = pd.to_numeric(auto_prices[column])
         #    auto_prices[cols] = auto_prices[cols].as_type(int64)

             return auto_prices
         auto_prices = read_auto_data()

In [16]: import matplotlib.pyplot as plt
         fig = plt.figure(figsize=(8,8)) # define plot area
         ax = fig.gca() # define axis
         auto_prices.plot(kind = 'scatter', x = 'engine-size', y = 'price', ax = ax)
```

```
        ax.set_title('Scatter plot de preco vs tamanho do motor') # Give the plot a main title
        ax.set_xlabel('Tamanho do motor') # Set text for the x axis
        ax.set_ylabel('Preco')# Set text for y axis
```

Out[16]: <matplotlib.text.Text at 0x7f5c54582ed0>

Out[16]:



In [0]: