In [70]:
```
#Author:Nichole Etienne
#Date: Wednesday September 23, 2021
#This code is meant to apply the unsupervised K-means Algorithm to the famous
```

In [2]:
```
#import the required Libraries
## matplotlib.pyplot: a collection of command style functions that make matpl
#pandas: data analysis tool kit
#Seaborn: library for making statistical graphics in Python
import time
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns
%matplotlib inline

#ignore warnings
import warnings
warnings.filterwarnings('ignore')
```

In [3]:
```
# Import the Facebook Live Sellers in the old Faithful Geyser Dataset
data= pd.read_csv('./Desktop/OFData.csv')
```

In [4]:
```
# the top 5 observations in the dataset
data.head()
```

Out[4]:

	eruptions	waiting
0	3.600	79
1	1.800	54
2	3.333	74
3	2.283	62
4	4.533	85

In [5]:
```
#.info prints information about a DataFrame including the index dtype and col
data.info
```

Out[5]:
```
<bound method DataFrame.info of      eruptions  waiting
0        3.600       79
1        1.800       54
2        3.333       74
3        2.283       62
4        4.533       85
..         ...      ...
267      4.117       81
268      2.150       46
269      4.417       90
270      1.817       46
271      4.467       74

[272 rows x 2 columns]>
```

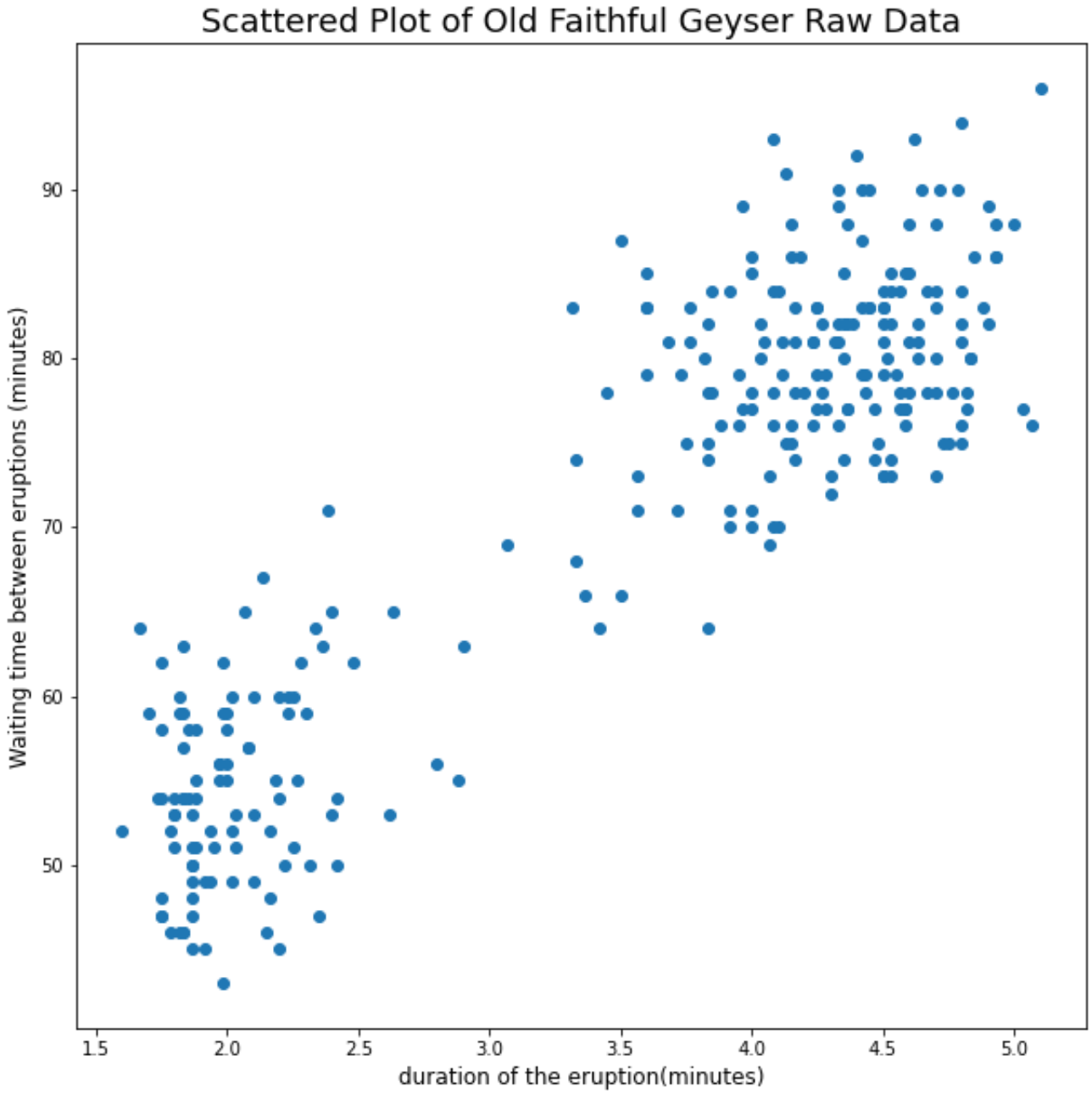
In [6]:
```
#return the missing variables within the data set
data.isnull().sum()
```

Out[6]:
```
eruptions    0
waiting      0
dtype: int64
```

In [7]:
```
#Since there are no missing , we may proceeed to analysis.
#That is first plotting the raw data
```

In [9]:
```
#plot the raw data
plt.figure(figsize=(10, 10))
plt.scatter(data.iloc[:, 0], data.iloc[:, 1])
plt.xlabel('duration of the eruption(minutes)', fontsize=12)
plt.ylabel('Waiting time between eruptions (minutes)', fontsize=12)
plt.title('Scattered Plot of Old Faithful Geyser Raw Data', fontsize= 18)
```

Out[9]: Text(0.5, 1.0, 'Scattered Plot of Old Faithful Geyser Raw Data')



In [ ]:

In [ ]:

In [ ]: