Lab10 Exercises

Q1) Fill the empty data with the median of the whole data frame:

Input:

```
df = pd.DataFrame(np.random.rand(7,3))
df.iloc[:4,1] = NA
df.iloc[:2,2] = NA
```

Q2) group the following data by the key1 and key2 depending on the mean then print the data in the data1 column and their key is one

Input:

```
df = pd.DataFrame({'key1' : ['a', 'a', 'b', 'b', 'a'],
 'key2' : ['one', 'two', 'one', 'two', 'one'],
 'data1' : np.random.randn(5),
 'data2' : np.random.randn(5)})
```

output:

```
key1
a   -1.303812
b   -1.179970
```

Q3) Create function that get the difference between the mean and the median then use this function to group the data based on it then print the data1 after grouping (the input is the same as previous) :

Output:

key1								
a	3.0	0.825578	0.772160	0.370610	0.379803	0.388996	1.053063	1.717129
b	2.0	-0.166921	1.004261	-0.877041	-0.521981	-0.166921	0.188139	0.543198

Q4) for the following dataset check and handle the missing then remove duplicate records if it is exist then find the correlation and remove one of the columns that causes the high correlation(greater than 0.8) depends on your understand for the dataset