

Assignment_Day14_AH

January 15, 2022

1 Participant

Title= Mr.

Name= Abdul Haseeb

Email= haseebkhalilshaikh@gmail.com

```
[ ]: # import libraries
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
```

2 Question 1

How to add more than one columns in unique function?

3 Answer

```
[ ]: # when you want to find unique values of more than one column
kashti=sns.load_dataset("titanic")
kashti.head()
```

```
[ ]:  survived  pclass    sex   age  sibsp  parch    fare embarked  class  \
0         0         3   male  22.0     1     0   7.2500         S   Third
1         1         1  female  38.0     1     0  71.2833         C   First
2         1         3  female  26.0     0     0   7.9250         S   Third
3         1         1  female  35.0     1     0  53.1000         S   First
4         0         3   male  35.0     0     0   8.0500         S   Third
```

```
      who  adult_male deck  embark_town  alive  alone
0    man          True  NaN  Southampton    no  False
1  woman         False    C   Cherbourg   yes  False
2  woman         False  NaN  Southampton   yes   True
3  woman         False    C   Southampton   yes  False
4    man          True  NaN  Southampton    no   True
```

```
[ ]: np.unique(kashti[["sex", "class", "who"]])
```

```
[ ]: array(['First', 'Second', 'Third', 'child', 'female', 'male', 'man',
          'woman'], dtype=object)
```

4 Question 2

Take log of fare add in the column and observe the difference between fare and log fare

5 Answer

```
[ ]: ks_clean=sns.load_dataset("titanic")
      ks_clean=ks_clean[ks_clean["fare"]<300]
```

```
[ ]: ks_clean["farelog"]=np.log(ks_clean["fare"])
      ks_clean.head()
```

C:\Users\Abdul Haseeb\AppData\Local\Programs\Python\Python310\lib\site-packages\pandas\core\arraylike.py:364: RuntimeWarning: divide by zero encountered in log

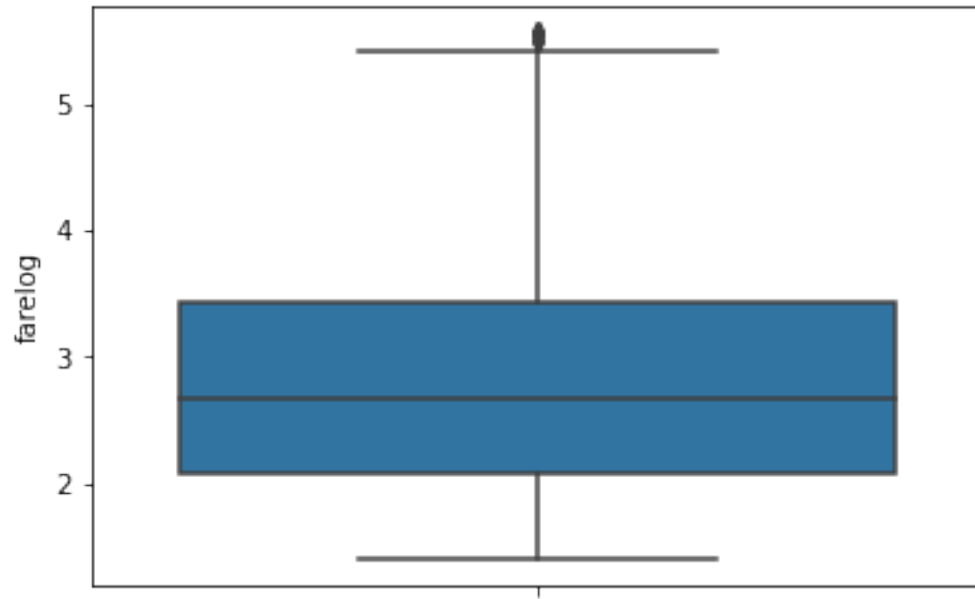
```
result = getattr(ufunc, method)(*inputs, **kwargs)
```

```
[ ]:   survived  pclass    sex  age  sibsp  parch    fare embarked  class \
0         0        3   male  22.0     1     0   7.2500         S  Third
1         1        1  female  38.0     1     0  71.2833         C  First
2         1        3  female  26.0     0     0   7.9250         S  Third
3         1        1  female  35.0     1     0  53.1000         S  First
4         0        3   male  35.0     0     0   8.0500         S  Third
```

```
      who  adult_male  deck  embark_town  alive  alone  farelog
0   man         True  NaN  Southampton    no  False  1.981001
1 woman        False   C   Cherbourg   yes  False  4.266662
2 woman        False  NaN  Southampton   yes  True   2.070022
3 woman        False   C   Southampton   yes  False  3.972177
4   man         True  NaN  Southampton    no  True   2.085672
```

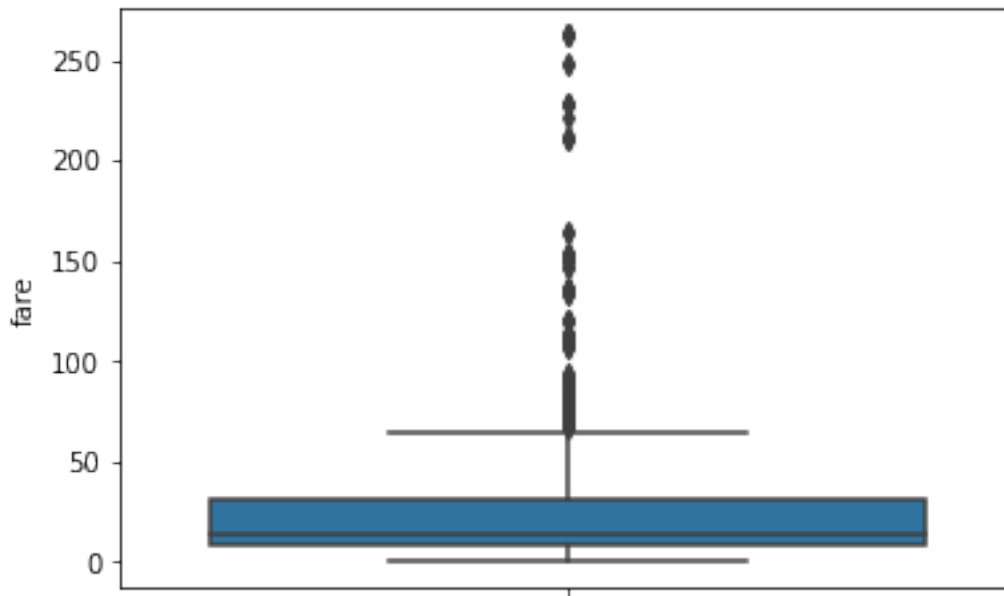
```
[ ]: sns.boxplot(y="farelog", data=ks_clean)
```

```
[ ]: <AxesSubplot:ylabel='farelog'>
```



```
[ ]: sns.boxplot(y="fare", data=ks_clean)
```

```
[ ]: <AxesSubplot:ylabel='fare'>
```



The main difference in both fare and fare log is that farelog has quite small values while these values are more in fare. This is the beauty of log it makes the data understandable through reducing

its values. In this way the outlayers in farelog are closer as due to log these digits are close to each other hence these can easily be removed and data can be brought to original shape of fare through taking antilog.