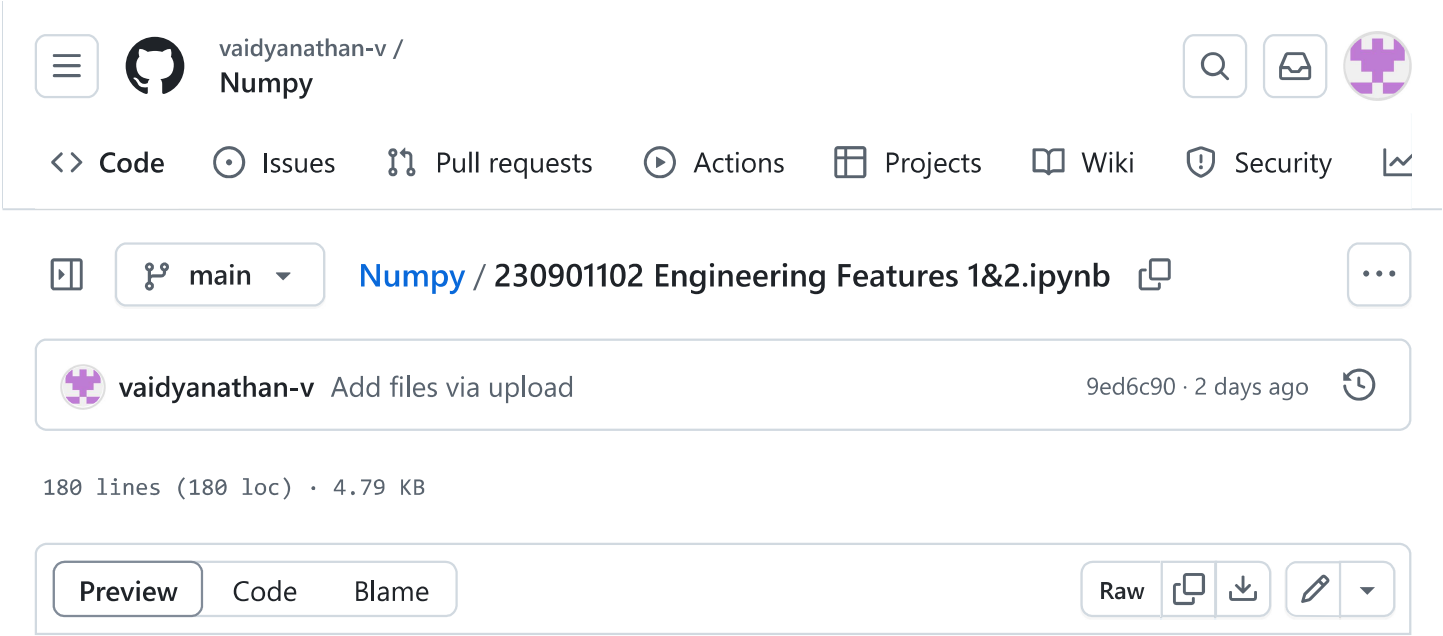
vaidyanathan-v /
Numpy
Code
Issues
Pull requests
Actions
Projects
Wiki
Security
main
Numpy / 230901102 Engineering Features 1&2.ipynb
vaidyanathan-v Add files via upload
9ed6c90 · 2 days ago
180 lines (180 loc) · 4.79 KB
Preview
Code
Blame
Raw

In [1]:

```
import pandas as pd
import numpy as np
#creating a dummy dataframe of 15 numbers randomly
#ranging from 1-100 for age
df=pd.DataFrame({'Age':[42,15,67,55,1,29,75,89,4,10,15,38,22,77]})
#printing DataFrame before sorting Continuous
#to Categories
print(df)
#A column of name 'Label' is created in DataFrame
#Categorizing Age into 4 Categories
#Baby/Toddler:(0,3),0 is excluded & 63 included
#Elderly:[63,99],63 is excluded & 99 is included
df['Label']=pd.cut(x=df['Age'],bins=[0,3,17,63,99],labels=['Baby/Toddler',
#Printing Dataframe after sorting continuous to
#Categories
print("After:\n")
print(df)
#check the number of values in each bin
print("Categories:\n")
print(df['Label'].value_counts())
```

```
    Age
0    42
1    15
2    67
3    55
4     1
5    29
6    75
7    89
8     4
9    10
10   15
11   38
12   22
13   77
After:

    Age         Label
0    42         Adult
1    15         child
2    67       Elderly
3    55         Adult
4     1  Baby/Toddler
5    29         Adult
6    75       Elderly
7    89       Elderly
8     4         child
9    10         child
10   15         child
11   38         Adult
12   22         Adult
13   77       Elderly
Categories:

Label
Adult          5
child          4
Elderly        4
Baby/Toddler   1
Name: count, dtype: int64
```

In [2]:

```
#importing pandas and numpy libraries
import pandas as pd
import numpy as np
#creating a dummy DataFrame of 12 numbers randomly
#ranging from 150-180 for height
df=pd.DataFrame({'Height':[150.4,157.6,170,176,164.2,155,159.2,175,162.4,17
#Printing DataFrame beofre Sorting Continuous to Categories
print("Before:")
print(df)
# A column of name 'Label' is created in DataFrame
# Categorizing height into 3 categories
# Short:(150,157),150 is excluded & 157 is included
# Average:(157,169),157 is excluded & 169 is included
# Tall:(169,180),169 is excluded & 180 is included
df['Label']=pd.cut(x=df['Height'],bins=[150,157,169,180],labels=['Short','A
#printing dataframe after sorting continuous to categorie
print("After:")
print(df)
#Checking the number of values in each bin
print("Categories:")
print(df['Label'].value_counts())
```

```
Before:
    Height
0    150.4
1    157.6
2    170.0
3    176.0
4    164.2
5    155.0
6    159.2
7    175.0
8    162.4
9    176.0
10   153.0
11   170.9
After:
    Height    Label
0    150.4    Short
1    157.6  Average
2    170.0     Tall
3    176.0     Tall
4    164.2  Average
5    155.0    Short
6    159.2  Average
7    175.0     Tall
8    162.4  Average
9    176.0     Tall
10   153.0    Short
11   170.9     Tall
Categories:
Label
Tall       5
Average    4
Short      3
Name: count, dtype: int64
```

In [ ]: