DSC540-Assignment-Topic1-Part1

August 8, 2021

Import Neccesary Packages

[12]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
```

Pandas Demo

[2]:
```
# Here we will use Pandas function to read a data file and display the␣
 ↪structure and header of the data file.
a_pd = pd.read_csv("H:/Krishna/GCU/DSC 530/Topic 7/Loans_Training",sep=',',␣
 ↪header=0)
```

[3]:
```
a_pd.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150302 entries, 0 to 150301
Data columns (total 5 columns):
 #   Column                Non-Null Count   Dtype
---  ------                --------------   -----
 0   Approval              150302 non-null  object
 1   Debt-to-Income Ratio  150302 non-null  float64
 2   FICO Score            150302 non-null  int64
 3   Request Amount        150302 non-null  int64
 4   Interest              150302 non-null  float64
dtypes: float64(2), int64(2), object(1)
memory usage: 5.7+ MB
```

[4]:
```
a_pd.head()
```

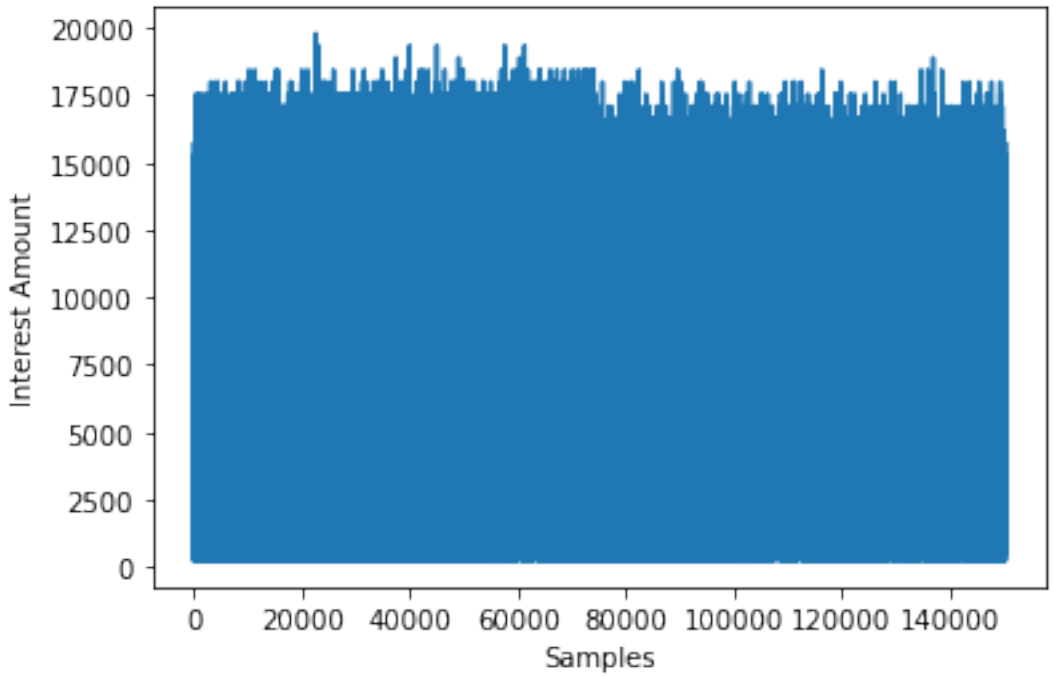
[4]:

	Approval	Debt-to-Income Ratio	FICO Score	Request Amount	Interest
0	F	0.0	397	1000	450.0
1	F	0.0	403	500	225.0
2	F	0.0	408	1000	450.0
3	F	0.0	408	2000	900.0
4	F	0.0	411	5000	2250.0

Pyplot - Matplotlib Demo

[8]:
```
# Here we will demonstrate the functionality of pyplot from Matplotlib package.
plt.plot(a_pd['Interest'])
plt.ylabel('Interest Amount')
plt.xlabel('Samples')
```

[8]:
```
Text(0.5, 0, 'Samples')
```



Numpy Demo

[9]:
```
# Here we will demonstrate the array creation using Numpy array function
b = np.array([1.2, 3.5, 5.1])
b.dtype
```

[9]:
```
dtype('float64')
```

[10]:
```
print(b)
```

```
[1.2 3.5 5.1]
```

SciKit-Learn Demo

[13]:
```
#We will demonstrate the Scikit-Learn packages functionality by using one of␣
 ↪the library to split the data into training set.

X_train, X_test = train_test_split(a_pd, test_size=0.2, random_state=123)
```

[15]:
```
X_train.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 120241 entries, 138437 to 15725
Data columns (total 5 columns):
 #   Column                Non-Null Count   Dtype
---  ------                --------------   -----
 0   Approval              120241 non-null  object
 1   Debt-to-Income Ratio  120241 non-null  float64
 2   FICO Score            120241 non-null  int64
 3   Request Amount        120241 non-null  int64
 4   Interest              120241 non-null  float64
dtypes: float64(2), int64(2), object(1)
memory usage: 5.5+ MB
```

[16]:
```
X_test.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 30061 entries, 54222 to 145215
Data columns (total 5 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   Approval              30061 non-null  object
 1   Debt-to-Income Ratio  30061 non-null  float64
 2   FICO Score            30061 non-null  int64
 3   Request Amount        30061 non-null  int64
 4   Interest              30061 non-null  float64
dtypes: float64(2), int64(2), object(1)
memory usage: 1.4+ MB
```

[ ]:
```

```