

```
In [27]: import pandas as pd
import numpy as np
```

```
In [28]: df = pd.read_csv(r'C:\Users\Churn_test.csv')
```

```
In [29]: import os
os.getcwd()
```

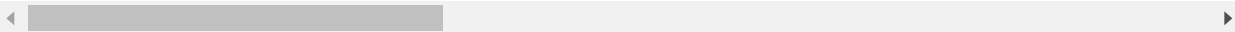
```
Out[29]: 'C:\\Users\\user'
```

```
In [30]: df
```

```
Out[30]:
```

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines
0	8879-ZKJOF	Female	0	No	No	41	Yes	No
1	0201-MIBOL	Female	1	No	No	66	Yes	Yes
2	1600-DILPE	Female	0	No	No	12	Yes	No
3	8601-QACRS	Female	0	No	No	5	Yes	Yes
4	7919-ZODZZ	Female	0	Yes	Yes	10	Yes	No
...
1404	5130-IEKQT	Male	1	No	No	25	Yes	Yes
1405	4452-ROHMO	Female	0	No	No	15	Yes	No
1406	6164-HAQTX	Male	0	No	No	71	No	No phone service
1407	3982-DQLUS	Male	1	Yes	Yes	65	Yes	Yes
1408	9874-QLCLH	Female	0	Yes	Yes	17	Yes	Yes

1409 rows × 20 columns



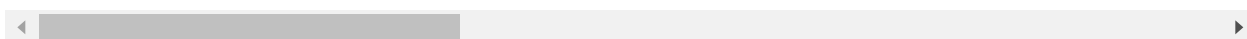
```
In [31]: df2 = pd.read_csv(r'C:\Users\Churn_train.csv')
```

In [32]: df2

Out[32]:

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines
0	5442-PPTJY	Male	0	Yes	Yes	12	Yes	No
1	6261-RCVNS	Female	0	No	No	42	Yes	No
2	2176-OSJUV	Male	0	Yes	No	71	Yes	Yes
3	6161-ERDGD	Male	0	Yes	Yes	71	Yes	Yes
4	2364-UFROM	Male	0	No	No	30	Yes	No
...
5629	0781-LKXBR	Male	1	No	No	9	Yes	Yes
5630	3507-GASNP	Male	0	No	Yes	60	Yes	No
5631	8868-WOZGU	Male	0	No	No	28	Yes	Yes
5632	1251-KRREG	Male	0	No	No	2	Yes	Yes
5633	5840-NVDCG	Female	0	Yes	Yes	16	Yes	No

5634 rows × 21 columns



In [33]: df2.describe()

Out[33]:

	SeniorCitizen	tenure	MonthlyCharges
count	5634.000000	5634.000000	5634.000000
mean	0.161874	32.277955	64.779127
std	0.368368	24.555211	30.104993
min	0.000000	0.000000	18.250000
25%	0.000000	9.000000	35.400000
50%	0.000000	29.000000	70.375000
75%	0.000000	55.000000	89.850000
max	1.000000	72.000000	118.650000

In [34]: `type(df2)`

Out[34]: `pandas.core.frame.DataFrame`

In [35]: `df2.dtypes`

Out[35]:

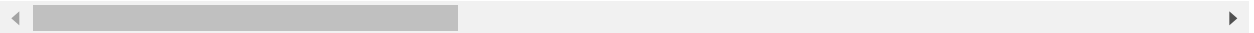
customerID	object
gender	object
SeniorCitizen	int64
Partner	object
Dependents	object
tenure	int64
PhoneService	object
MultipleLines	object
InternetService	object
OnlineSecurity	object
OnlineBackup	object
DeviceProtection	object
TechSupport	object
StreamingTV	object
StreamingMovies	object
Contract	object
PaperlessBilling	object
PaymentMethod	object
MonthlyCharges	float64
TotalCharges	object
Churn	object
dtype:	object

In [36]: `df2.head(4)`

Out[36]:

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	Ir
0	5442-PPTJY	Male	0	Yes	Yes	12	Yes	No	
1	6261-RCVNS	Female	0	No	No	42	Yes	No	
2	2176-OSJUV	Male	0	Yes	No	71	Yes	Yes	
3	6161-ERDGD	Male	0	Yes	Yes	71	Yes	Yes	

4 rows × 21 columns



In [37]: `df2.columns`

Out[37]: `Index(['customerID', 'gender', 'SeniorCitizen', 'Partner', 'Dependents', 'tenure', 'PhoneService', 'MultipleLines', 'InternetService', 'OnlineSecurity', 'OnlineBackup', 'DeviceProtection', 'TechSupport', 'StreamingTV', 'StreamingMovies', 'Contract', 'PaperlessBilling', 'PaymentMethod', 'MonthlyCharges', 'TotalCharges', 'Churn'], dtype='object')`

```
In [38]: df2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5634 entries, 0 to 5633
Data columns (total 21 columns):
 #   Column                Non-Null Count  Dtype  
---  -
 0   customerID            5634 non-null   object 
 1   gender                 5634 non-null   object 
 2   SeniorCitizen          5634 non-null   int64  
 3   Partner                5634 non-null   object 
 4   Dependents             5634 non-null   object 
 5   tenure                 5634 non-null   int64  
 6   PhoneService           5634 non-null   object 
 7   MultipleLines          5634 non-null   object 
 8   InternetService        5634 non-null   object 
 9   OnlineSecurity         5634 non-null   object 
10   OnlineBackup           5634 non-null   object 
11   DeviceProtection       5634 non-null   object 
12   TechSupport            5634 non-null   object 
13   StreamingTV            5634 non-null   object 
14   StreamingMovies        5634 non-null   object 
15   Contract               5634 non-null   object 
16   PaperlessBilling       5634 non-null   object 
17   PaymentMethod          5634 non-null   object 
18   MonthlyCharges         5634 non-null   float64 
19   TotalCharges           5634 non-null   object 
20   Churn                  5634 non-null   object 
dtypes: float64(1), int64(2), object(18)
memory usage: 924.5+ KB
```

```
In [39]: num_male=sum(df2['gender'] == 'Male')
```

```
In [40]: num_male
```

```
Out[40]: 2838
```

```
In [41]: num_female=sum(df2['gender'] == 'Female')
num_female
```

```
Out[41]: 2796
```

```
In [42]: num_equal=sum(df2['gender'] == 'Female')
num_equal
```

```
Out[42]: 2796
```

```
In [43]: df2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5634 entries, 0 to 5633
Data columns (total 21 columns):
 #   Column                Non-Null Count  Dtype
---  -
 0   customerID            5634 non-null   object
 1   gender                5634 non-null   object
 2   SeniorCitizen         5634 non-null   int64
 3   Partner               5634 non-null   object
 4   Dependents            5634 non-null   object
 5   tenure               5634 non-null   int64
 6   PhoneService          5634 non-null   object
 7   MultipleLines         5634 non-null   object
 8   InternetService       5634 non-null   object
 9   OnlineSecurity        5634 non-null   object
10   OnlineBackup          5634 non-null   object
11   DeviceProtection      5634 non-null   object
12   TechSupport           5634 non-null   object
13   StreamingTV           5634 non-null   object
14   StreamingMovies       5634 non-null   object
15   Contract              5634 non-null   object
16   PaperlessBilling      5634 non-null   object
17   PaymentMethod         5634 non-null   object
18   MonthlyCharges        5634 non-null   float64
19   TotalCharges          5634 non-null   object
20   Churn                 5634 non-null   object
dtypes: float64(1), int64(2), object(18)
memory usage: 924.5+ KB
```

```
In [44]: df2.columns
```

```
Out[44]: Index(['customerID', 'gender', 'SeniorCitizen', 'Partner', 'Dependents',
               'tenure', 'PhoneService', 'MultipleLines', 'InternetService',
               'OnlineSecurity', 'OnlineBackup', 'DeviceProtection', 'TechSupport',
               'StreamingTV', 'StreamingMovies', 'Contract', 'PaperlessBilling',
               'PaymentMethod', 'MonthlyCharges', 'TotalCharges', 'Churn'],
              dtype='object')
```

```
In [45]: max_monthlycharge=max(df2['MonthlyCharges'])
```

```
In [46]: max_monthlycharge
```

```
Out[46]: 118.65
```

```
In [47]: import matplotlib.pyplot as plt
```

```
In [48]: import seaborn as sns
```

```
In [49]: df2['Churn'].value_counts()
```

```
Out[49]: No      4113
        Yes      1521
        Name: Churn, dtype: int64
```

```
In [50]: df2.columns
```

```
Out[50]: Index(['customerID', 'gender', 'SeniorCitizen', 'Partner', 'Dependents',
               'tenure', 'PhoneService', 'MultipleLines', 'InternetService',
               'OnlineSecurity', 'OnlineBackup', 'DeviceProtection', 'TechSupport',
               'StreamingTV', 'StreamingMovies', 'Contract', 'PaperlessBilling',
               'PaymentMethod', 'MonthlyCharges', 'TotalCharges', 'Churn'],
              dtype='object')
```

```
In [26]: df2.columns('Contract')
```

TypeError

Traceback (most recent call last)

Input In [26], in <cell line: 1>()

```
----> 1 df2.columns('Contract')
```

TypeError: 'Index' object is not callable

```
In [ ]: df2.head(2)
```

```
In [ ]: min_tenure= min(df2['tenure'])
        max_tenure= max(df2['tenure'])
        min_tenure
        max_tenure
```

```
In [ ]: max_tenure
```

```
In [ ]: sns.pairplot(df2);
```

```
In [51]: df2.columns
```

```
Out[51]: Index(['customerID', 'gender', 'SeniorCitizen', 'Partner', 'Dependents',
               'tenure', 'PhoneService', 'MultipleLines', 'InternetService',
               'OnlineSecurity', 'OnlineBackup', 'DeviceProtection', 'TechSupport',
               'StreamingTV', 'StreamingMovies', 'Contract', 'PaperlessBilling',
               'PaymentMethod', 'MonthlyCharges', 'TotalCharges', 'Churn'],
              dtype='object')
```

```
In [56]: df4 = df2[['customerID', 'Churn', index= False] ]
```

Input In [56]

```
df4 = df2[['customerID', 'Churn', index= False] ]
```

SyntaxError: invalid syntax

In [57]: df4

Out[57]:

	customerID	Churn
0	5442-PPTJY	No
1	6261-RCVNS	Yes
2	2176-OSJUV	No
3	6161-ERDGD	No
4	2364-UFROM	No
...
5629	0781-LKXBR	Yes
5630	3507-GASNP	No
5631	8868-WOZGU	Yes
5632	1251-KRREG	Yes
5633	5840-NVDCG	No

5634 rows × 2 columns

In [64]: df5 = df4.head(1409)

In [65]: df5

Out[65]:

	customerID	Churn
0	5442-PPTJY	No
1	6261-RCVNS	Yes
2	2176-OSJUV	No
3	6161-ERDGD	No
4	2364-UFROM	No
...
1404	7838-LAZFO	No
1405	9975-SKRNR	No
1406	9483-GCPWE	No
1407	0674-EYYZV	No
1408	2511-ALLCS	No

1409 rows × 2 columns

In [66]: df5.to_csv('customer churn prediction.csv', index = False, header = False)

In [67]: `df2.head(4)`

Out[67]:

lineSecurity	...	DeviceProtection	TechSupport	StreamingTV	StreamingMovies	Contract	PaperlessB
No internet service	...	No internet service	No internet service	No internet service	No internet service	Two year	
Yes	...	Yes	Yes	No	Yes	One year	
Yes	...	No	Yes	No	No	Two year	
Yes	...	Yes	Yes	Yes	Yes	One year	



In [68]: `df2.shape`

Out[68]: (5634, 21)

In [70]: `df.dtypes`

Out[70]:

customerID	object
gender	object
SeniorCitizen	int64
Partner	object
Dependents	object
tenure	int64
PhoneService	object
MultipleLines	object
InternetService	object
OnlineSecurity	object
OnlineBackup	object
DeviceProtection	object
TechSupport	object
StreamingTV	object
StreamingMovies	object
Contract	object
PaperlessBilling	object
PaymentMethod	object
MonthlyCharges	float64
TotalCharges	object
dtype:	object


```
In [71]: def changeColumnsToString(df2):
          columnsNames=['Partner','Dependents','PhoneService','MultipleLines','OnlineSe
          for col in columnsNames:
              df2[col]=df2[col].astype('str').str.replace('Yes','1').replace('No','0').

          changeColumnsToString(df2)

          df2['SeniorCitizen']=df2['SeniorCitizen'].astype(bool)
          df2['TotalCharges']=pd.to_numeric(df2['TotalCharges'],errors='coerce')
```

```
In [72]: df2.head(4)
```

Out[72]:

lineSecurity	...	DeviceProtection	TechSupport	StreamingTV	StreamingMovies	Contract	PaperlessB
0	...	0	0	0	0	Two year	
1	...	1	1	0	1	One year	
1	...	0	1	0	0	Two year	
1	...	1	1	1	1	One year	

```
In [73]: df2.dtypes
```

```
Out[73]: customerID      object
gender      object
SeniorCitizen  bool
Partner      object
Dependents   object
tenure      int64
PhoneService  object
MultipleLines object
InternetService object
OnlineSecurity object
OnlineBackup  object
DeviceProtection object
TechSupport   object
StreamingTV   object
StreamingMovies object
Contract      object
PaperlessBilling object
PaymentMethod object
MonthlyCharges float64
TotalCharges  float64
Churn         object
dtype: object
```

In [74]: df2.columns

Out[74]: Index(['customerID', 'gender', 'SeniorCitizen', 'Partner', 'Dependents', 'tenure', 'PhoneService', 'MultipleLines', 'InternetService', 'OnlineSecurity', 'OnlineBackup', 'DeviceProtection', 'TechSupport', 'StreamingTV', 'StreamingMovies', 'Contract', 'PaperlessBilling', 'PaymentMethod', 'MonthlyCharges', 'TotalCharges', 'Churn'], dtype='object')

In [75]: `print("gender: ", df2.gender.unique())`
`print("Contract: ", df2.Contract.unique())`
`print("SeniorCitizen: ", df2.SeniorCitizen.unique())`
`print("InternetService: ", df2.InternetService.unique())`
`print("PaymentMethod: ", df2.PaymentMethod.unique())`

```
gender:  ['Male' 'Female']
Contract:  ['Two year' 'One year' 'Month-to-month']
SeniorCitizen:  [False  True]
InternetService:  ['No' 'DSL' 'Fiber optic']
PaymentMethod:  ['Mailed check' 'Credit card (automatic)' 'Bank transfer (automatic)'
                 'Electronic check']
```

In [77]: `df2['gender']=df2['gender'].astype('category')`
`df2['Contract']=df2['Contract'].astype('category')`
`df2['SeniorCitizen']=df2['SeniorCitizen'].astype('category')`
`df2['InternetService']=df2['InternetService'].astype('category')`
`df2['InternetService']=df2['InternetService'].astype('category')`
`df2.dtypes`

Out[77]:

customerID	object
gender	category
SeniorCitizen	category
Partner	object
Dependents	object
tenure	int64
PhoneService	object
MultipleLines	object
InternetService	category
OnlineSecurity	object
OnlineBackup	object
DeviceProtection	object
TechSupport	object
StreamingTV	object
StreamingMovies	object
Contract	category
PaperlessBilling	object
PaymentMethod	object
MonthlyCharges	float64
TotalCharges	float64
Churn	object
dtype:	object

```
In [78]: df2PaymentDummies = pd.get_dummies(df2['PaymentMethod'], prefix = 'payment')
df2ContractDummies = pd.get_dummies(df2['Contract'], prefix = 'contract')
df2GenderDummies = pd.get_dummies(df2['gender'], prefix = 'gender')
df2SeniorCitizenDummies = pd.get_dummies(df2['SeniorCitizen'], prefix = 'SC')
df2InternetServiceDummies = pd.get_dummies(df2['InternetService'], prefix = 'IS')

print(df2PaymentDummies.head(3))
print(df2ContractDummies.head(3))
print(df2GenderDummies.head(3))
print(df2SeniorCitizenDummies.head(3))
print(df2InternetServiceDummies.head(3))
```

```

    payment_Bank transfer (automatic)  payment_Credit card (automatic)  \
0                                     0                                0
1                                     0                                1
2                                     1                                0

    payment_Electronic check  payment_Mailed check
0                             0                     1
1                             0                     0
2                             0                     0
    contract_Month-to-month  contract_One year  contract_Two year
0                             0                 0                 1
1                             0                 1                 0
2                             0                 0                 1
    gender_Female  gender_Male
0                 0           1
1                 1           0
2                 0           1
    SC_False  SC_True
0           1       0
1           1       0
2           1       0
    IS_DSL  IS_Fiber optic  IS_No
0         0              0       1
1         1              0       0
2         1              0       0

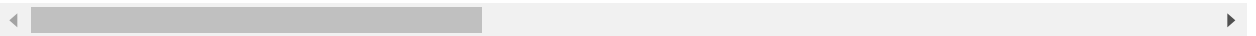
```

```
In [83]: df2 = pd.concat([df2, df2PaymentDummies], axis=1)
df2 = pd.concat([df2, df2ContractDummies], axis=1)
df2 = pd.concat([df2, df2GenderDummies], axis=1)
df2 = pd.concat([df2, df2SeniorCitizenDummies], axis=1)
df2 = pd.concat([df2, df2InternetServiceDummies], axis=1)
df2.head(2)
```

Out[83]:

	customerID	Partner	Dependents	tenure	PhoneService	MultipleLines	OnlineSecurity	OnlineBa
0	5442-PPTJY	1	1	12	1	0	0	
1	6261-RCVNS	0	0	42	1	0	1	

2 rows × 30 columns



```
In [85]: df2.columns = ['customerID', 'Partner', 'Dependents', 'tenure', 'PhoneService',
                        'MultipleLines', 'OnlineSecurity', 'OnlineBackup', 'DeviceProtection',
                        'TechSupport', 'StreamingTV', 'StreamingMovies', 'PaperlessBilling',
                        'MonthlyCharges', 'TotalCharges', 'Churn',
                        'payment_Bank transfer (automatic)', 'payment_Credit card (automatic)',
                        'payment_Electronic check', 'payment_Mailed check',
                        'contract_Month-to-month', 'contract_One year', 'contract_Two year',
                        'gender_Female', 'gender_Male', 'SC_False', 'SC_True', 'IS_DSL',
                        'IS_Fiber optic', 'IS_No']
```

In [82]: df2.head

```
Out[82]: <bound method NDFrame.head of
Service MultipleLines \
0      5442-PPTJY      1      1      12      1      0
1      6261-RCVNS      0      0      42      1      0
2      2176-OSJUV      1      0      71      1      1
3      6161-ERDGD      1      1      71      1      1
4      2364-UFROM      0      0      30      1      0
...      ...      ...      ...      ...      ...
5629    0781-LKXBR      0      0      9      1      1
5630    3507-GASNP      0      1      60      1      0
5631    8868-WOZGU      0      0      28      1      1
5632    1251-KRREG      0      0      2      1      1
5633    5840-NVDCG      1      1      16      1      0

      OnlineSecurity OnlineBackup DeviceProtection TechSupport StreamingTV \
0      0      0      0      0      0
1      1      1      1      1      1
2      1      1      0      1      1
3      1      0      1      1      1
4      1      1      0      1      1
...      ...      ...      ...      ...      ...
5629      0      0      1      0      1
5630      0      0      0      0      0
5631      0      1      1      0      1
5632      0      1      0      0      0
5633      1      1      0      1      0

      StreamingMovies PaperlessBilling MonthlyCharges TotalCharges Churn
0      0      0      19.70      258.35      0
1      1      0      73.90      3160.55      1
2      0      0      65.15      4681.75      0
3      1      0      85.45      6300.85      0
4      0      0      70.40      2044.75      0
...      ...      ...      ...      ...      ...
5629      1      1      100.50      918.60      1
5630      0      0      19.95      1189.90      0
5631      1      1      105.70      2979.50      1
5632      0      1      54.40      114.10      1
5633      1      0      68.25      1114.85      0

[5634 rows x 16 columns]>
```

```
In [89]: df2.columns = ['customerID', 'Partner', 'Dependents', 'tenure', 'PhoneService',
                        'MultipleLines', 'OnlineSecurity', 'OnlineBackup', 'DeviceProtection',
                        'TechSupport', 'StreamingTV', 'StreamingMovies', 'PaperlessBilling',
                        'MonthlyCharges', 'TotalCharges', 'Churn',
                        'payment_Bank transfer (automatic)', 'payment_Credit card (automatic)',
                        'payment_Electronic check', 'payment_Mailed check',
                        'contract_Month-to-month', 'contract_One year', 'contract_Two year',
                        'gender_Female', 'gender_Male', 'SC_False', 'SC_True', 'IS_DSL',
                        'IS_Fiber optic', 'IS_No']
```

```
In [91]: df2.columns
```

```
Out[91]: Index(['customerID', 'Partner', 'Dependents', 'tenure', 'PhoneService',  
               'MultipleLines', 'OnlineSecurity', 'OnlineBackup', 'DeviceProtection',  
               'TechSupport', 'StreamingTV', 'StreamingMovies', 'PaperlessBilling',  
               'MonthlyCharges', 'TotalCharges', 'Churn',  
               'payment_Bank transfer (automatic)', 'payment_Credit card (automatic)',  
               'payment_Electronic check', 'payment_Mailed check',  
               'contract_Month-to-month', 'contract_One year', 'contract_Two year',  
               'gender_Female', 'gender_Male', 'SC_False', 'SC_True', 'IS_DSL',  
               'IS_Fiber optic', 'IS_No'],  
              dtype='object')
```

```
In [92]: numericColumns=np.array(['Partner', 'Dependents', 'tenure', 'PhoneService',
    'MultipleLines', 'OnlineSecurity', 'OnlineBackup', 'DeviceProtection',
    'TechSupport', 'StreamingTV', 'StreamingMovies', 'PaperlessBilling',
    'MonthlyCharges', 'TotalCharges', 'Churn',
    'payment_Bank transfer (automatic)', 'payment_Credit card (automatic)',
    'payment_Electronic check', 'payment_Mailed check',
    'contract_Month-to-month', 'contract_One year', 'contract_Two year',
    'gender_Female', 'gender_Male', 'SC_False', 'SC_True', 'IS_DSL',
    'IS_Fiber optic', 'IS_No'])

for columnName in numericColumns:
    df2[columnName]=pd.to_numeric(df2[columnName],errors='coerce')
df2.dtypes
```

```
Out[92]: customerID      object
Partner                int64
Dependents             int64
tenure                 int64
PhoneService           int64
MultipleLines          int64
OnlineSecurity         int64
OnlineBackup           int64
DeviceProtection       int64
TechSupport            int64
StreamingTV            int64
StreamingMovies        int64
PaperlessBilling       int64
MonthlyCharges         float64
TotalCharges           float64
Churn                  int64
payment_Bank transfer (automatic)  uint8
payment_Credit card (automatic)    uint8
payment_Electronic check          uint8
payment_Mailed check              uint8
contract_Month-to-month          uint8
contract_One year                uint8
contract_Two year                uint8
gender_Female                   uint8
gender_Male                     uint8
SC_False                        uint8
SC_True                         uint8
IS_DSL                          uint8
IS_Fiber optic                  uint8
IS_No                           uint8
dtype: object
```

```
In [93]: modelData = df2.loc[:, df2.columns != 'customerID']
modelData.to_csv('modelData.csv')
```

```

In [95]: modelData=pd.read_csv('modelData.csv')

modelData[modelData==np.inf]=np.nan
modelData.fillna(modelData.mean(), inplace=True)

x=np.asarray(modelData.loc[:,modelData.columns != 'Churn'])
y=np.asarray(modelData['Churn'])

print(x[:2])
print(y[:2])

[[0.00000e+00  1.00000e+00  1.00000e+00  1.20000e+01  1.00000e+00  0.00000e+00
  0.00000e+00  0.00000e+00  0.00000e+00  0.00000e+00  0.00000e+00  0.00000e+00
  0.00000e+00  1.97000e+01  2.58350e+02  0.00000e+00  0.00000e+00  0.00000e+00
  1.00000e+00  0.00000e+00  0.00000e+00  1.00000e+00  0.00000e+00  1.00000e+00
  1.00000e+00  0.00000e+00  0.00000e+00  0.00000e+00  1.00000e+00]
[1.00000e+00  0.00000e+00  0.00000e+00  4.20000e+01  1.00000e+00  0.00000e+00
  1.00000e+00  1.00000e+00  1.00000e+00  1.00000e+00  0.00000e+00  1.00000e+00
  0.00000e+00  7.39000e+01  3.16055e+03  0.00000e+00  1.00000e+00  0.00000e+00
  0.00000e+00  0.00000e+00  1.00000e+00  0.00000e+00  1.00000e+00  0.00000e+00
  1.00000e+00  0.00000e+00  1.00000e+00  0.00000e+00  0.00000e+00]]
[0 1]

```

In []:

In []:

In []: