Assgn 4
DSTT

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression


from sklearn.model_selection import train_test_split
from sklearn.preprocessing import PolynomialFeatures
from sklearn.linear_model import LinearRegression
from sklearn.pipeline import make_pipeline


data= pd.read_csv('C:/Users/91882/Desktop/PYTHON/Exam 3/housingdata.csv',sep=',')

data=data[['PID','STATE','PRICE','NUM_BEDROOMS','NUM_BATH','SQ_FT']]

#1
data_top = data.head(6)
print(data_top)
#print(data[:6])

#2
column_names = data.columns
print(column_names)

ID='PID'
X=np.array(data.drop([ID],axis=1))


for index, value in enumerate(X):
    print("Index No :", index)


#3
print(data)
y=X[: , 0:1]
unique_state = np.unique(y)
print("Unique Strings:")
print(unique_state)
print('Total States = ', unique_state.size)
```

```
#4

print(X)

r,c=X.shape
print(r,c)
z=X.reshape(r*c)
dz=pd.DataFrame(z)
for c in range (dz.size):

    #print(z[c])
    if (z[c]=='na' or z[c]=='--' or z[c]=='nan' ):
      print(z[c])
    else:
        pass


df = pd.DataFrame(data)

# Dropping rows with N/A, NA, na values
df_cleaned = df.replace(['N/A', 'NA', 'na'], np.nan).dropna()

print("DataFrame after dropping rows with N/A, NA, na values:")
print(df_cleaned)
```

OUTPUT

```
runfile('C:/Users/91882/Assgn 4 DSTT.py', wdir='C:/Users/91882')
        PID       STATE     PRICE NUM_BEDROOMS NUM_BATH SQ_FT
0  100001000.0         MP    321654          3       1  1000
1  100002000.0  MAHARASHTRA     21325          3     1.5   --
2  100003000.0         AP    2541654        NaN       1   850
3  100004000.0         TN    321321          1     NaN   700
4         NaN         TN  589465645         3       2  1600
5  100006000.0         TN   65465466        NaN       1   800
Index(['PID', 'STATE', 'PRICE', 'NUM_BEDROOMS', 'NUM_BATH', 'SQ_FT'], dtype='object')
Index No : 0
Index No : 1
Index No : 2
Index No : 3
Index No : 4
Index No : 5
Index No : 6
Index No : 7
Index No : 8
        PID       STATE     PRICE NUM_BEDROOMS NUM_BATH SQ_FT
0  100001000.0         MP    321654          3       1  1000
1  100002000.0  MAHARASHTRA     21325          3     1.5   --
2  100003000.0         AP    2541654        NaN       1   850
3  100004000.0         TN    321321          1     NaN   700
4         NaN         TN  589465645         3       2  1600
5  100006000.0         TN   65465466        NaN       1   800
6  100007000.0      ASSAM    3222321          2   HURLEY   950
7  100008000.0         HP     23131          1       1   NaN
8  100009000.0         HP     21212         na       2  1800
Unique Strings:
['AP' 'ASSAM' 'HP' 'MAHARASHTRA' 'MP' 'TN']
Total States =  6
[['MP' 321654 '3' '1' '1000']
 ['MAHARASHTRA' 21325 '3' '1.5' '--']
 ['AP' 2541654 nan '1' '850']
 ['TN' 321321 '1' nan '700']
 ['TN' 589465645 '3' '2' '1600']
 ['TN' 65465466 nan '1' '800']
 ['ASSAM' 3222321 '2' 'HURLEY' '950']
 ['HP' 23131 '1' '1' nan]
 ['HP' 21212 'na' '2' '1800']]
9 5
--
na
DataFrame after dropping rows with N/A, NA, na values:
        PID       STATE    PRICE NUM_BEDROOMS NUM_BATH SQ_FT
```

0	100001000.0	MP	321654	3	1	1000	
1	100002000.0	MAHARASHTRA	21325	3	1.5	--	
6	100007000.0	ASSAM	3222321	2	HURLEY	950	