## VIT-AP UNIVERSITY, ANDHRA PRADESH

Academic year: 2022-2023

Semester: Fall

Faculty Name: Dr. Aravapalli Rama Satish

Branch/ Class: B.Tech
Date: 16-12-2021
School: SCOPE

NAME:Majjiga Jaswanth REGNO:20BCD717I

1. Load the data from the following link:

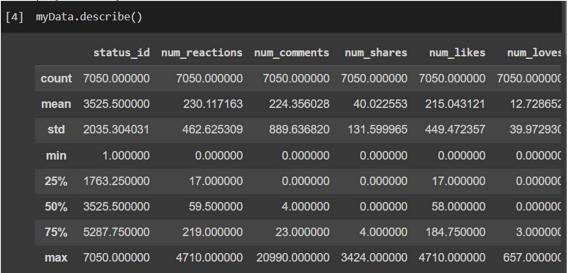
https://archive.ics.uci.edu/ml/datasets/Facebook+Live+Sellers+in+Thailand

```
import numpy as np
    import pandas as pd
   myData = pd.read csv("Live 20210128.csv")
   print(myData.head())
      status_id status_type status_published num_reactions num_comments
₽
                     video 4/22/2018 6:00
   0
                                                    529
                                                                 512
                     photo 4/21/2018 22:45
                                                    150
                                                                   0
                    video 4/21/2018 6:17
                                                    227
                                                                 236
                   photo 4/21/2018 2:29
                                                                   0
             4
                                                    111
                   photo 4/18/2018 3:22
                                                    213
      num shares num likes num loves num wows num hahas num sads \
            262
                      432
                      150
                                           0
                                                     0
             0
                                 0
                                                               0
                                 21
0
9
   2
             57
                       204
                                                               0
              0
                       111
                                          0
                                                    0
                                                               0
                                            0
   4
              0
                       204
                                                      0
      num angrys Column1 Column2 Column3 Column4
   0
              0
                    NaN
                             NaN
                                     NaN
                                              NaN
              0
                     NaN
                             NaN
                                     NaN
                                              NaN
   2
              0
                     NaN
                             NaN
                                     NaN
                                              NaN
              0
                     NaN
                             NaN
                                     NaN
                                              NaN
              0
                     NaN
                             NaN
                                     NaN
                                              NaN
```

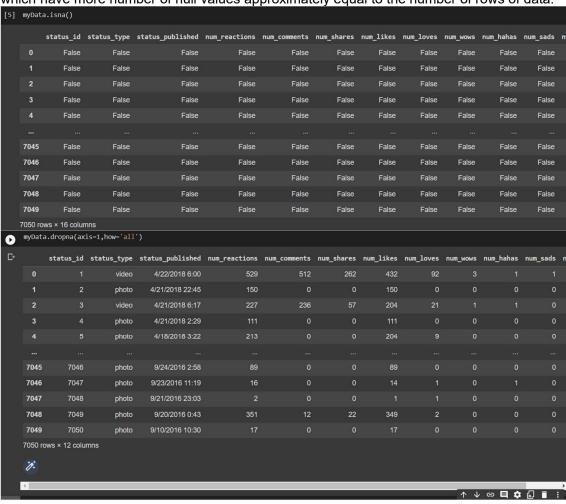
2. Display the number of rows and columns of loaded data set.

```
rows = len(myData.axes[0])
cols = len(myData.axes[1])
print(rows)
print(cols)
7050
16
```

3. Display summary of the loaded data set.



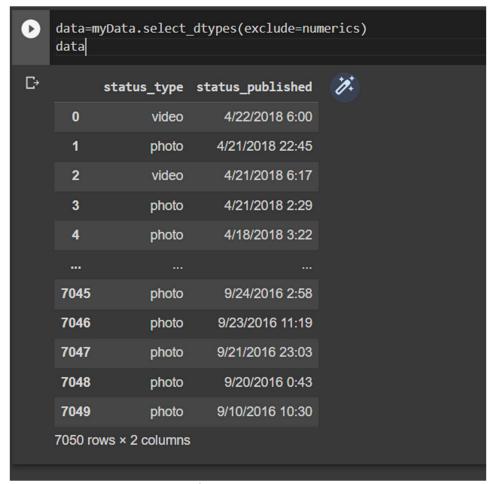
4. Count the number of missing values are there for each column and drop the columns which have more number of null values approximately equal to the number of rows of data.



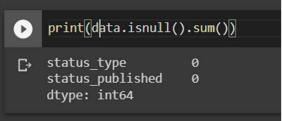
```
myData.dropna(axis=1,inplace=True)
[46] print(myData.isnull().sum())
     status_id
                        0
     status type
                        0
     status published
                        0
    num_reactions
                        0
    num_comments
                        0
    num shares
                       0
    num likes
                        0
    num loves
                      0
    num_wows
                      0
    num hahas
                        0
    num sads
                      0
    num_angrys
                        0
     dtype: int64
```

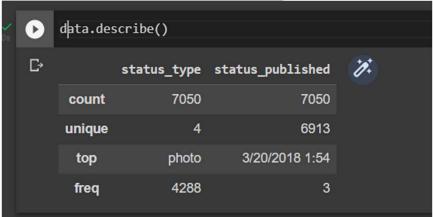
5.List out number of categorical variables and numerical variables are available in the data set.





6. display statistical summary of numerical attributes.

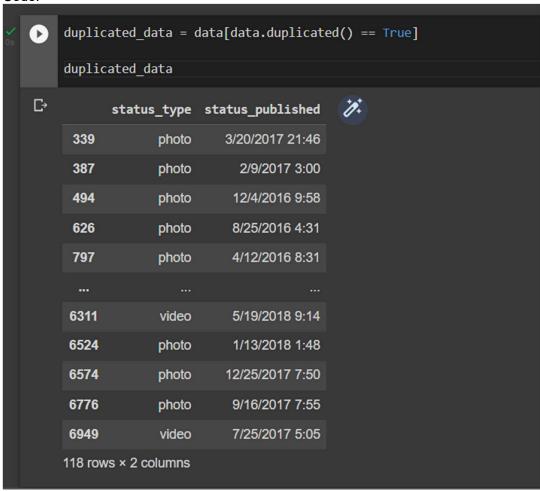




7. Count the number of unique values are existed in categorical attributes and drop the all columns which have number of unique values approximately equal to number of rows of

data.

Code:



- 8. Convert the remaining categorical variable into equivalent numerical form and perform Min-Max Scaling for all attributes.
- 9. Develop a python program to perform K-Means Clustering with number of clusters as K = 2, 3, 4, 5, 6 and display inter-cluster distances and sum of all inter-cluster distances. Identify at what K value the inter-cluster distance is minimal.

import pandas as pd

import numpy as np

import random as rd

import matplotlib.pyplot as plt

myData = pd.read\_csv("Live\_20210128.csv") print(myData.head())

diff = 1

```
j=0
while(diff!=0):
XD=X
i=1
for index1,row_c in Centroids.iterrows():
ED=[]
for index2,row_d in XD.iterrows():
d1=(row c["ApplicantIncome"]-row d["ApplicantIncome"])**2
d2=(row c["LoanAmount"]-row d["LoanAmount"])**2
d=np.sqrt(d1+d2)
ED.append(d)
X[i]=ED
i=i+1
C=[]
for index,row in X.iterrows():
min dist=row[1]
pos=1
for i in range(K):
if row[i+1] < min dist:
min dist = row[i+1]
pos=i+1
C.append(pos)
X["Cluster"]=C
Centroids new = X.groupby(["Cluster"]).mean()[["LoanAmount","ApplicantIncome"]]
if j == 0:
diff=1
j=j+1
else:
diff = (Centroids new['LoanAmount'] - Centroids['LoanAmount']).sum() +
(Centroids new['ApplicantIncome'] - Centroids['ApplicantIncome']).sum()
print(diff.sum())
Centroids = X.groupby(["Cluster"]).mean()[["LoanAmount","ApplicantIncome"]]
```

```
color=['blue','green','cyan']
for k in range(K):
data=X[X["Cluster"]==k+1]
plt.scatter(data["ApplicantIncome"],data["LoanAmount"],c=color[k])
plt.scatter(Centroids["ApplicantIncome"],Centroids["LoanAmount"],c='red')
plt.xlabel('Income')
plt.ylabel('Loan Amount (In Thousands)')
plt.show()
```