

DWDM SEM LAB

PRATHIKSHA M

18BCS101

SET 7

1. Download a sample dataset from any Repository. List the attributes and its type in a word Doc.

dataset:<https://www.kaggle.com/manugupta/road-accidents-in-india>

STATE/UT	YEAR	0-3 hrs	3-6 hrs	6-9 hrs	9-12 hrs	12-15 hrs	15-18 hrs	18-21 hrs	21-24 hrs	Total
A & N Islar	2001	2	6	29	40	39	40	18	7	181
A & N Islar	2002	2	6	22	41	33	33	23	8	168
A & N Islar	2003	2	8	31	35	28	36	25	15	180
A & N Islar	2004	2	5	29	42	43	43	37	14	215
A & N Islar	2005	0	8	27	28	38	42	50	13	206
A & N Islar	2006	1	3	17	33	33	29	38	7	155
A & N Islar	2007	2	5	20	30	30	27	31	7	152
A & N Islar	2008	3	7	33	24	40	31	40	13	191
A & N Islar	2009	2	6	35	41	64	54	50	19	271
A & N Islar	2010	2	10	36	45	64	57	53	18	285
A & N Islar	2011	4	3	36	27	46	50	51	18	235
A & N Islar	2012	4	6	26	29	54	45	58	14	236
A & N Islar	2013	2	3	22	31	40	40	37	25	200
A & N Islar	2014	4	6	25	28	55	36	41	23	218
Andhra Prc	2001	2239	3265	3198	3729	3604	3792	4098	3263	27188
Andhra Prc	2002	2931	3857	3671	4255	4153	4778	4844	4088	32577
Andhra Prc	2003	3158	4805	3749	4319	4266	4853	5218	4109	34537
Andhra Prc	2004	3191	4770	4598	5030	4033	4971	6031	4654	37078
Andhra Prc	2005	3826	6011	5002	4137	4261	4524	5096	4432	37289
Andhra Prc	2006	3635	5525	4270	5108	4918	6101	6757	5009	41323
Andhra Prc	2007	4054	4890	4748	5826	5547	6236	6719	5574	43594
Andhra Prc	2008	4051	4492	4532	5558	5620	6045	6453	5355	42106
Andhra Prc	2009	3718	4433	4569	5332	5320	6154	6302	6183	42011
Andhra Prc	2010	3822	5347	5553	5438	5176	5548	6363	5181	42428
Andhra Prc	2011	3577	4958	4494	5897	5420	5715	6418	4586	41066
Andhra Prc	2012	3026	3795	4413	5413	4978	5616	6877	5226	39344

This dataset contains the information of road accidents in Indian state and time of accident (0-3 hour, 3-6 hour ..)this information is obtained from the Ministry of Road Transport and Highways, Govt of India.

The main goal is to find some good patterns in the data set to help reduce accidents in India.

```

In [1]: #que 1 dataset and its attribute
import pandas as pd

In [3]: #data=pd.read_csv(r"C:\Users\prath\Documents\KCT 6TH SEM\only_road_accidents_data3.csv")
data.head()

Out[3]:
  STATE/UT  YEAR  0-3 hrs. (Night)  3-6 hrs. (Night)  6-9 hrs. (Day)  9-12 hrs. (Day)  12-15 hrs. (Day)  15-18 hrs. (Day)  18-21 hrs. (Night)  21-24 hrs. (Night)  Total
0  A & N Islands  2001         2         6         29         40         39         40         18         7        181
1  A & N Islands  2002         2         6        22        41        33        33        23         8        168
2  A & N Islands  2003         2         8        31        35        28        36        25        15        180
3  A & N Islands  2004         2         5        29        42        43        43        37        14        215
4  A & N Islands  2005         0         8        27        28        38        42        50        13        206

In [4]: #data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 490 entries, 0 to 489
Data columns (total 11 columns):
STATE/UT    490 non-null object
YEAR        490 non-null int64
0-3 hrs. (Night)  490 non-null int64
3-6 hrs. (Night)  490 non-null int64
6-9 hrs. (Day)   490 non-null int64
9-12 hrs. (Day)  490 non-null int64

```

Attribute and its type:

1. Nominal attribute -STATE/UT (categorical)
2. Numerical attribute -YEAR, 0-3 hrs. (Night), 3-6 hrs. (Night), 6-9 hrs (Day), 9-12 hrs (Day), 12-15 hrs (Day), 15-18 hrs (Day), 18-21 hrs (Night), 21-24 hrs (Night), Total.

2.Create a random dataset of 50 elements with x and y variables using random function between 30 to 100 integers for x and 60 to 150 integers for y. Apply K-means clustering to cluster the data into 3 clusters. Plot the graph and display the result. Use Tkinter GUI to Display the Results

Download/ X 18BCS101 PRATHIKSHA M DWDM SEM LAB Last Checkpoint a few seconds ago (autosaved) Logout

File Edit View Insert Cell Kernel Help Trusted Python 3

```
21-24 hrs (night) 490 non-null int64
Total 490 non-null int64
dtypes: int64(10), object(1)
memory usage: 42.2+ KB

In [5]: #que 2:
from pandas import DataFrame
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans

In [6]: # Creating a random dataset of 50 elements with
# x and y variables using random function
# between 30 to 100 integers for x and 60 to 150 integers for y
import random
import numpy
x_col=random.sample(range(30, 100), 50)
print("X:",x_col)
y_col=random.sample(range(60, 150), 50)
print("Y:",y_col)
values={"X":x_col,"Y":y_col}
dt=DataFrame(values)
dt.head()

X: [57, 68, 92, 36, 30, 45, 39, 94, 34, 42, 49, 99, 52, 82, 83, 93, 89, 44, 72, 86, 85, 66, 62, 91, 87, 56, 61, 65, 75, 78,
90, 51, 33, 58, 38, 50, 64, 43, 35, 98, 96, 88, 33, 69, 55, 40, 60, 46, 76, 50]
Y: [73, 65, 109, 139, 99, 62, 112, 128, 101, 69, 140, 77, 143, 124, 110, 131, 71, 88, 126, 130, 118, 72, 66, 125, 145, 90, 1
32, 127, 84, 138, 100, 116, 119, 105, 91, 129, 144, 136, 87, 122, 113, 89, 64, 134, 82, 141, 79, 74, 120, 142]

Out[6]:
   X  Y
0  57 73
1  68 65
2  92 109
3  36 139
4  30 99
```

Type here to search 20°C Partly cloudy 12:43 PM 6/23/2021

Download/ X 18BCS101 PRATHIKSHA M DWDM SEM LAB Last Checkpoint a minute ago (autosaved) Logout

File Edit View Insert Cell Kernel Help Trusted Python 3

```
1 68 65
2 92 109
3 36 139
4 30 99

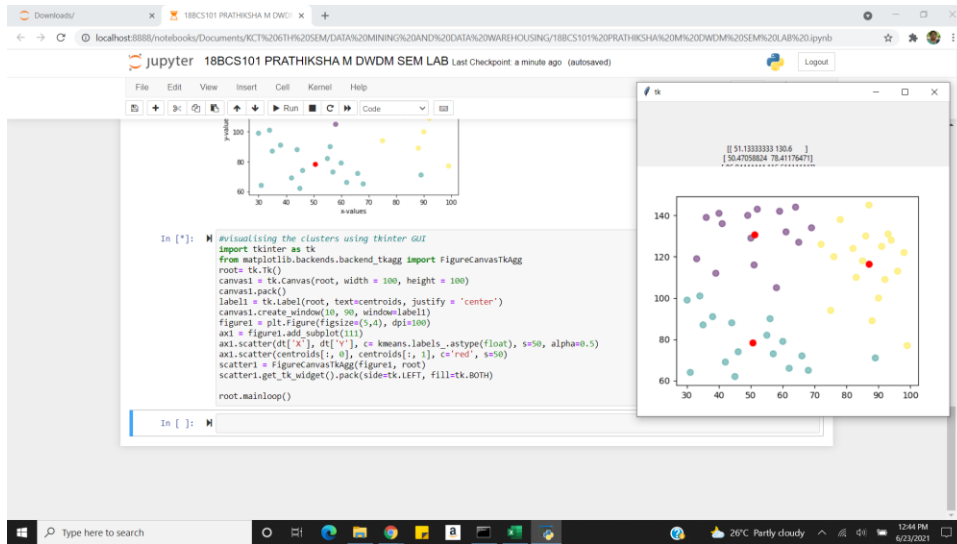
In [7]: #dividing the dataset into three clusters
kmeans = KMeans(n_clusters=3).fit(dt)
centroids = kmeans.cluster_centers_
print(centroids)

plt.scatter(dt['X'], dt['Y'], c=kmeans.labels_.astype(float), s=50, alpha=0.5)
plt.scatter(centroids[:, 0], centroids[:, 1], c='red', s=50)
plt.xlabel("x-values")
plt.ylabel("y-values")
plt.show()

[[ 51.13333333 130.6
   50.47058824 78.41176471
   86.94444444 116.61111111]]

140
120
100
80
60
30 40 50 60 70 80 90 100
x-values
y-values
```

Type here to search 20°C Partly cloudy 12:44 PM 6/23/2021



3. Upload in your GITHUB account. Provide the link for access.

<https://github.com/Prathiksha2000/end-sem-lab>