DATA WAREHOUSE AND DATA MINING

LAB DA – 5

NAME : HRITHIK HEM SUNDAR.B

REGNO : 19MID0021

K MEANS CLUSTERING WITHOUT LIBRARIES

CODE:

```
a1=(2,3,4,11,12,12,7,8,8)
a2=(3,1,2,5,4,6,5,4,6)
k=3                    #No.of.clusters
a=[(3,1),(11,5),(8,4)]    #Clusters
import pandas as pd
b=()
c=[]
d=[]
e=[]
f=()
g=[]
h=[]
p=[]
q=()
x=[]
s=1
y=1
a1=(2,3,4,11,12,12,7,8,8)
a2=(3,1,2,5,4,6,5,4,6)
k=3
a=[(3,1),(11,5),(8,4)]
flag=0
while(flag==0):
   print("ITERATION : ",s)
```

```
print("\n")
for i in range(len(a1)):
    for j in range(k):
        c.append((abs(a[j][0]-a1[i])+abs(a[j][1]-a2[i])))
    d.append(c)
    c=[]
for i in range(len(a1)):
    for j in range(k):
        if(d[i][j]==min(d[i])):
            break
    d[i].append(a[j])
    e.append(a[j])
df=pd.DataFrame(d)
sums=0
sums1=0
t=0
for l in range(len(a)):
    for m in range(len(a1)):
        if(e[m]==a[l]):
            t=t+1
            sums=sums+a1[m]
            sums1=sums1+a2[m]
    f=(sums,sums1)
    g.append(f)
    h.append(t)
    sums=0
    sums1=0
    t=0
for n in range(len(a)):
    mean1=g[n][0]/h[n]
    mean2=g[n][1]/h[n]
```

```
        q=(mean1,mean2)
        p.append(q)
    print("Clusters : ",a),print("\n")
    print(df),print("\n")
    print("Mean : ",p)
    print("\n")
    if(p==a):
        flag=1
        print("FINAL CLUSTERS : \n")
    else:
        a=p
        b=()
        c=[]
        d=[]
        e=[]
        f=()
        g=[]
        h=[]
        p=[]
        q=()
        s=s+1
        print("\n")
for u in range(k):
    for v in range(len(a1)):
        if(p[u]==e[v]):
            x.append((a1[v],a2[v]))
    print("CLUSTER  : ",y)
    print("\n")
    print(x)
    print("\n")
    y=y+1
```

x=[]

SCREENSHOTS WITH OUTPUT:

```
a1=(2,3,4,11,12,12,7,8,8)
a2=(3,1,2,5,4,6,5,4,6)
k=3                          #No.of.clusters
a=[(3,1),(11,5),(8,4)]       #Clusters
```

```
import pandas as pd
b=()
c=[]
d=[]
e=[]
f=()
g=[]
h=[]
p=[]
q=()
x=[]
s=1
y=1
a1=(2,3,4,11,12,12,7,8,8)
a2=(3,1,2,5,4,6,5,4,6)
k=3
a=[(3,1),(11,5),(8,4)]
flag=0
while(flag==0):
    print("ITERATION : ",s)
    print("\n")
    for i in range(len(a1)):
        for j in range(k):
             c.append((abs(a[j][0]-a1[i])+abs(a[j][1]-a2[i])))
        d.append(c)
        c=[]
    for i in range(len(a1)):
        for j in range(k):
            if(d[i][j]==min(d[i])):
                break
        d[i].append(a[j])
        e.append(a[j])
    df=pd.DataFrame(d)
    sums=0
    sums1=0
    t=0
```

```
    for l in range(len(a)):
        for m in range(len(a1)):
            if(e[m]==a[l]):
                t=t+1
                sums=sums+a1[m]
                sums1=sums1+a2[m]
        f=(sums,sums1)
        g.append(f)
        h.append(t)
        sums=0
        sums1=0
        t=0
    for n in range(len(a)):
        mean1=g[n][0]/h[n]
        mean2=g[n][1]/h[n]
        q=(mean1,mean2)
        p.append(q)
    print("Clusters : ",a),print("\n")
    print(df),print("\n")
    print("Mean : ",p)
    print("\n")
    if(p==a):
        flag=1
        print("FINAL CLUSTERS : \n")
    else:
        a=p
        b=()
        c=[]
        d=[]
        e=[]
        f=()
        g=[]
        h=[]
        p=[]
        q=()
        s=s+1
        print("\n")
```

```
for u in range(k):
    for v in range(len(a1)):
        if(p[u]==e[v]):
            x.append((a1[v],a2[v]))
    print("CLUSTER  : ",y)
    print("\n")
    print(x)
    print("\n")
    y=y+1
    x=[]
```

```
ITERATION :  1


Clusters :  [(3, 1), (11, 5), (8, 4)]


    0   1  2        3
0   3  11  7   (3, 1)
1   0  12  8   (3, 1)
2   2  10  6   (3, 1)
3  12   0  4  (11, 5)
4  12   2  4  (11, 5)
5  14   2  6  (11, 5)
6   8   4  2   (8, 4)
7   8   4  0   (8, 4)
8  10   4  2   (8, 4)


Mean :  [(3.0, 2.0), (11.666666666666666, 5.0), (7.666666666666667, 5.0)]
```

```
ITERATION :  2


Clusters :  [(3.0, 2.0), (11.666666666666666, 5.0), (7.666666666666667, 5.0)]


      0          1         2                          3
0   2.0  11.666667  7.666667                (3.0, 2.0)
1   1.0  12.666667  8.666667                (3.0, 2.0)
2   1.0  10.666667  6.666667                (3.0, 2.0)
3  11.0   0.666667  3.333333  (11.666666666666666, 5.0)
4  11.0   1.333333  5.333333  (11.666666666666666, 5.0)
5  13.0   1.333333  5.333333  (11.666666666666666, 5.0)
6   7.0   4.666667  0.666667   (7.666666666666667, 5.0)
7   7.0   4.666667  1.333333   (7.666666666666667, 5.0)
8   9.0   4.666667  1.333333   (7.666666666666667, 5.0)


Mean :  [(3.0, 2.0), (11.666666666666666, 5.0), (7.666666666666667, 5.0)]
```

```
FINAL CLUSTERS :

CLUSTER  :  1

[(2, 3), (3, 1), (4, 2)]

CLUSTER  :  2

[(11, 5), (12, 4), (12, 6)]

CLUSTER  :  3

[(7, 5), (8, 4), (8, 6)]
```