dbscan-11-12-02-pdf

February 13, 2025

[ ]:
```
from sklearn.cluster import DBSCAN
from sklearn.preprocessing import StandardScaler
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

[ ]:
```
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

[ ]:
```
#impoer csv file and convert it into a dataframe
import pandas as pd
file_path = '/content/Wholesale customers data.csv'
df = pd.read_csv(file_path)
```

[ ]:
```
df.head()
```

[ ]:

	Channel	Region	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicassen
0	2	3	12669	9656	7561	214	2674	1338
1	2	3	7057	9810	9568	1762	3293	1776
2	2	3	6353	8808	7684	2405	3516	7844
3	1	3	13265	1196	4221	6404	507	1788
4	2	3	22615	5410	7198	3915	1777	5185

[ ]:
```
df
```

[ ]:

	Channel	Region	Fresh	Milk	Grocery	Frozen	Detergents_Paper	\
0	2	3	12669	9656	7561	214	2674	
1	2	3	7057	9810	9568	1762	3293	
2	2	3	6353	8808	7684	2405	3516	
3	1	3	13265	1196	4221	6404	507	
4	2	3	22615	5410	7198	3915	1777	
..	…	…	…	…	…	…	…	
435	1	3	29703	12051	16027	13135	182	
436	1	3	39228	1431	764	4510	93	
437	2	3	14531	15488	30243	437	14841	
438	1	3	10290	1981	2232	1038	168	

```
439        1       3   2787   1698     2510      65                477

     Delicassen
0          1338
1          1776
2          7844
3          1788
4          5185
..          …
435        2204
436        2346
437        1867
438        2125
439          52

[440 rows x 8 columns]
```

[ ]:
```
print(df.info)
```

```
<bound method DataFrame.info of      Channel  Region  Fresh   Milk  Grocery
Frozen  Detergents_Paper  \
0          2       3  12669   9656     7561     214              2674
1          2       3   7057   9810     9568    1762              3293
2          2       3   6353   8808     7684    2405              3516
3          1       3  13265   1196     4221    6404               507
4          2       3  22615   5410     7198    3915              1777
..       …     …    …     …      …      …                 …
435        1       3  29703  12051    16027   13135               182
436        1       3  39228   1431      764    4510                93
437        2       3  14531  15488    30243     437             14841
438        1       3  10290   1981     2232    1038               168
439        1       3   2787   1698     2510      65               477

     Delicassen
0          1338
1          1776
2          7844
3          1788
4          5185
..          …
435        2204
436        2346
437        1867
438        2125
439          52

[440 rows x 8 columns]>
```

```
[ ]: df.drop('Channel', axis=1, inplace=True)
     df.drop('Region', axis=1, inplace=True )
```

```
[ ]: array=df.values
```

```
[ ]: array
```

```
[ ]: array([[12669,  9656,  7561,   214,  2674,  1338],
            [ 7057,  9810,  9568,  1762,  3293,  1776],
            [ 6353,  8808,  7684,  2405,  3516,  7844],
            …,
            [14531, 15488, 30243,   437, 14841,  1867],
            [10290,  1981,  2232,  1038,   168,  2125],
            [ 2787,  1698,  2510,    65,   477,    52]])
```

```
[ ]: stscaler = StandardScaler().fit(array)
     X=stscaler.transform(array)
```

```
[ ]: dbscan = DBSCAN(eps=0.8, min_samples=6)
     dbscan.fit(X)
```

```
[ ]: DBSCAN(eps=0.8, min_samples=6)
```

```
[ ]: cl=pd.DataFrame(dbscan.labels_,columns=['cluster'])
```

```
[ ]: cl
```

```
[ ]:      cluster
     0          0
     1          0
     2         -1
     3          0
     4         -1
     ..       …
     435       -1
     436        0
     437       -1
     438        0
     439        0

     [440 rows x 1 columns]
```

```
[ ]: pd.Series(dbscan.labels_).value_counts()
```

```
[ ]:  0    360
     -1     80
     Name: count, dtype: int64
```

[ ]:
```
print(df.info)
```

```
<bound method DataFrame.info of      Fresh   Milk  Grocery  Frozen
Detergents_Paper  Delicassen
0    12669   9656     7561     214              2674        1338
1     7057   9810     9568    1762              3293        1776
2     6353   8808     7684    2405              3516        7844
3    13265   1196     4221    6404               507        1788
4    22615   5410     7198    3915              1777        5185
..     …     …       …       …                 …           …
435  29703  12051    16027   13135              182        2204
436  39228   1431      764    4510               93        2346
437  14531  15488    30243     437             14841        1867
438  10290   1981     2232    1038               168        2125
439   2787   1698     2510      65               477          52

[440 rows x 6 columns]>
```

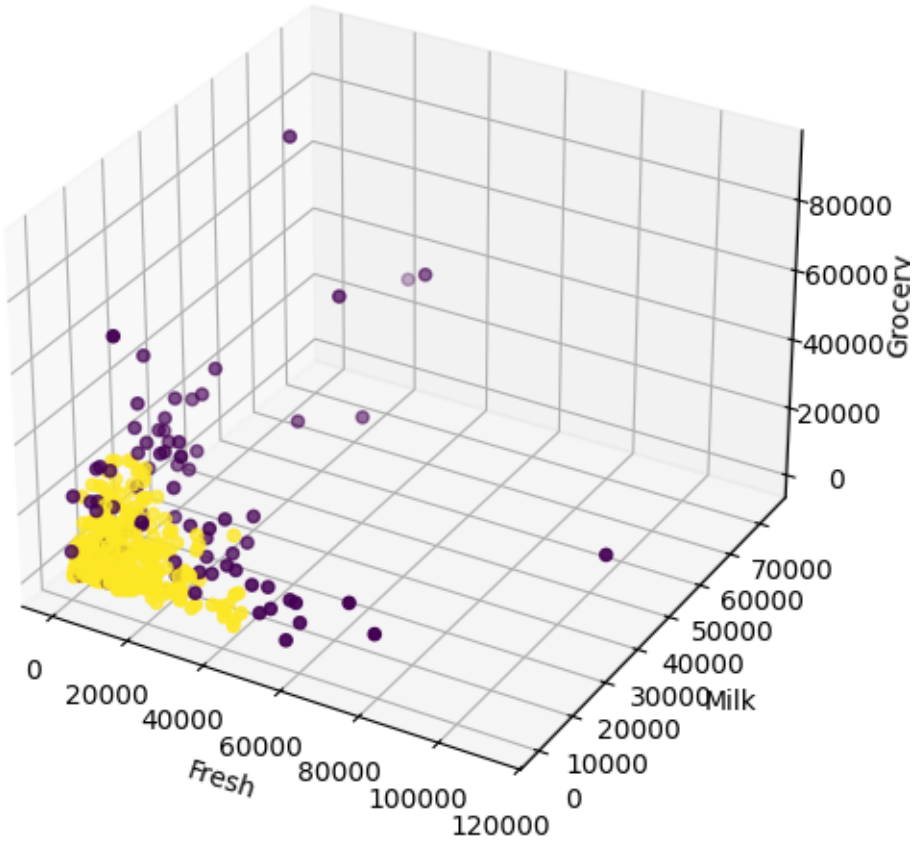
[ ]:
```
pd.concat([df,cl],axis=1)
```

[ ]:

	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicassen	cluster
0	12669	9656	7561	214	2674	1338	0
1	7057	9810	9568	1762	3293	1776	0
2	6353	8808	7684	2405	3516	7844	-1
3	13265	1196	4221	6404	507	1788	0
4	22615	5410	7198	3915	1777	5185	-1
..	…	…	…	…	…	…	…
435	29703	12051	16027	13135	182	2204	-1
436	39228	1431	764	4510	93	2346	0
437	14531	15488	30243	437	14841	1867	-1
438	10290	1981	2232	1038	168	2125	0
439	2787	1698	2510	65	477	52	0

[440 rows x 7 columns]

[ ]:
```
fig = plt.figure(figsize=(10, 6))
ax = fig.add_subplot(111,projection='3d')
ax.scatter(df['Fresh'], df['Milk'], df['Grocery'], c=dbscan.labels_)
ax.set_xlabel('Fresh')
ax.set_ylabel('Milk')
ax.set_zlabel('Grocery')
ax.set_title('DBSCAN Clustering')
plt.show()
```


DBSCAN Clustering
Grocery
Milk
Fresh
0
20000
40000
60000
80000
100000
120000
10000
30000
50000
70000

[ ]:

[ ]: