

mall customer segmentation use for k means clustering

May 10, 2021

```
[1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import scipy.stats as stat
import warnings
warnings.filterwarnings('ignore')
```

```
C:\Users\Chandra Sekhar\Anaconda3\lib\site-
packages\statsmodels\tools\_testing.py:19: FutureWarning: pandas.util.testing is
deprecated. Use the functions in the public API at pandas.testing instead.
import pandas.util.testing as tm
```

```
[6]: df=pd.read_csv('Mall_Customers.csv',index_col='CustomerID')
```

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

```
[7]:
```

	Genre	Age	Annual_Income_(k\$)	Spending_Score
CustomerID				
1	Male	19	15	39
2	Male	21	15	81
3	Female	20	16	6
4	Female	23	16	77
5	Female	31	17	40

```
[8]: df.isnull().sum()
```

```
[8]: Genre          0
Age              0
Annual_Income_(k$)  0
Spending_Score    0
dtype: int64
```

```
[13]: x=df.iloc[:,[2, 3]].values
```

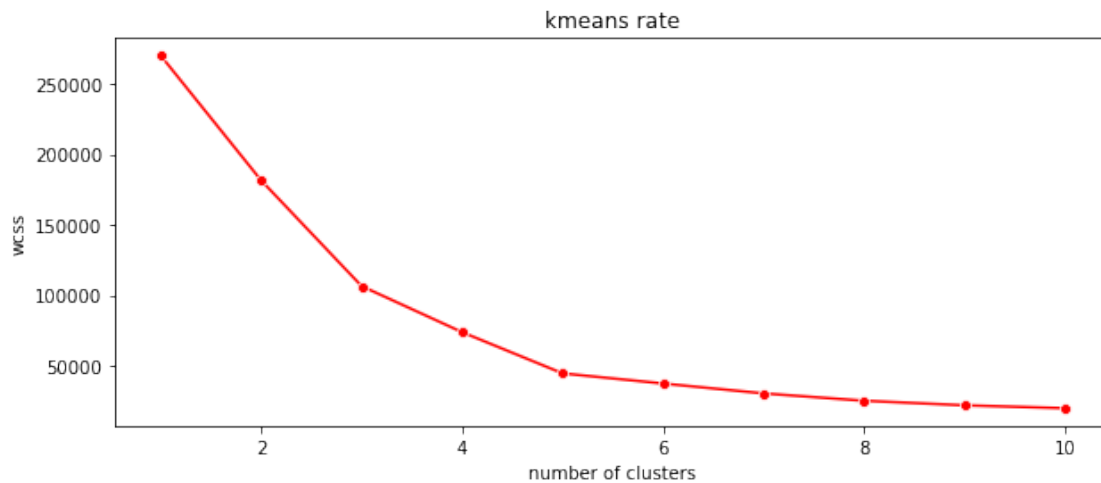
```
[15]: from sklearn.cluster import KMeans
```

```
[ ]: wcss = []
for i in range(1, 11):
    kmeans = KMeans(n_clusters = i, init = 'k-means++', random_state = 42)
```

```
kmeans.fit(X)
```

```
[36]: wcss = []  
  
for i in range(1,11):  
    kmeans=KMeans(n_clusters=i, init= 'k-means++', random_state=43)  
    kmeans.fit(x)  
  
    wcss.append(kmeans.inertia_)
```

```
[41]: plt.figure(figsize=(10,4))  
sns.lineplot(range(1,11),wcss,marker='o',color='red')  
plt.title('kmeans rate')  
plt.xlabel('number of clusters')  
plt.ylabel('wcss')  
plt.show()
```



```
[43]: km=KMeans(n_clusters=5)
```

```
[44]: km.fit(x)
```

```
[44]: KMeans(n_clusters=5)
```

```
[46]: clusters=km.predict(x)
```

```
[47]: df['clusters']=clusters
```

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

```
[48]:
```

	Genre	Age	Annual_Income_(k\$)	Spending_Score	clusters
CustomerID					
1	Male	19	15	39	1
2	Male	21	15	81	4
3	Female	20	16	6	1
4	Female	23	16	77	4

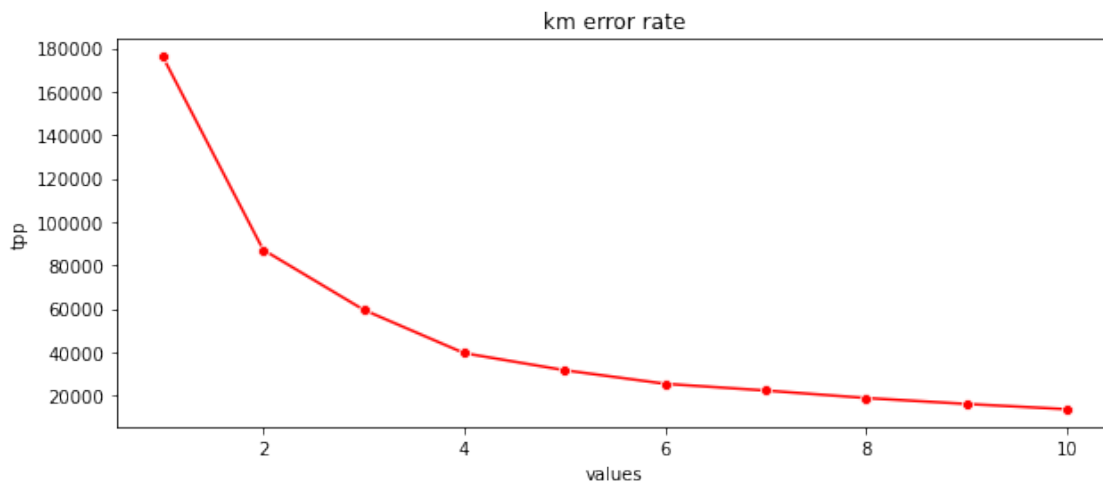
5	Female	31	17	40	1
---	--------	----	----	----	---

```
[53]: y=df.iloc[:,[1,2]].values
```

```
[54]: ttp=[]
```

```
for i in range(1,11):
    kmeans=KMeans(n_clusters=i, init='k-means++',random_state=42)
    kmeans.fit(y)
    ttp.append(kmeans.inertia_)
```

```
[56]: plt.figure(figsize=(10,4))
sns.lineplot(range(1,11),ttp,marker='o',color='red')
plt.title('km error rate')
plt.xlabel('values')
plt.ylabel('tpp')
plt.show()
```



```
[57]: kt=KMeans(n_clusters=5)
kt.fit(x)
```

```
[57]: KMeans(n_clusters=5)
```

```
[58]: clutr=kt.predict(x)
```

```
[60]: df['cluster1']=clutr
```

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

	Genre	Age	Annual_Income_(k\$)	Spending_Score	clusters	\
CustomerID						
1	Male	19	15	39	1	
2	Male	21	15	81	4	
3	Female	20	16	6	1	
4	Female	23	16	77	4	

5	Female	31	17	40	1
---	--------	----	----	----	---

cluster1

CustomerID

1	2
2	1
3	2
4	1
5	2

[]:

[]:

[]: