In [1]:
```
import pandas as pd
import numpy as np
import numpy as np
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
import seaborn as sns
from sklearn.decomposition import PCA
from sklearn import datasets
```

In [2]:
```
#Reading dataset
data=pd.read_csv(r"C:\\Neema Updated\competitor_influencer_similarity.csv", low_memory=
```

In [3]:
```
data['rank'] = data.groupby(['influencer_name'])['similarity_score'].rank(ascending=Fal
```

In [4]:
```
a1=data[(data['influencer_name']=='aaroncgshore')]
sorted_df = a1.sort_values(by='rank')
sorted_df.head(10)
```

Out[4]:

	competitor_name	influencer_name	similarity_score	rank
150	neat_nutrition	aaroncgshore	0.054774	1.0
0	bulk	aaroncgshore	0.052173	2.0
120	motionnutrition	aaroncgshore	0.049527	3.0
240	liveinnermost	aaroncgshore	0.047144	4.0
270	puresport	aaroncgshore	0.043901	5.0
30	formnutrition	aaroncgshore	0.041036	6.0
60	indisupplements	aaroncgshore	0.038453	7.0
210	thenue_co	aaroncgshore	0.033165	8.0
90	medterra.international	aaroncgshore	0.031670	9.0
180	neurohacker	aaroncgshore	0.031141	10.0

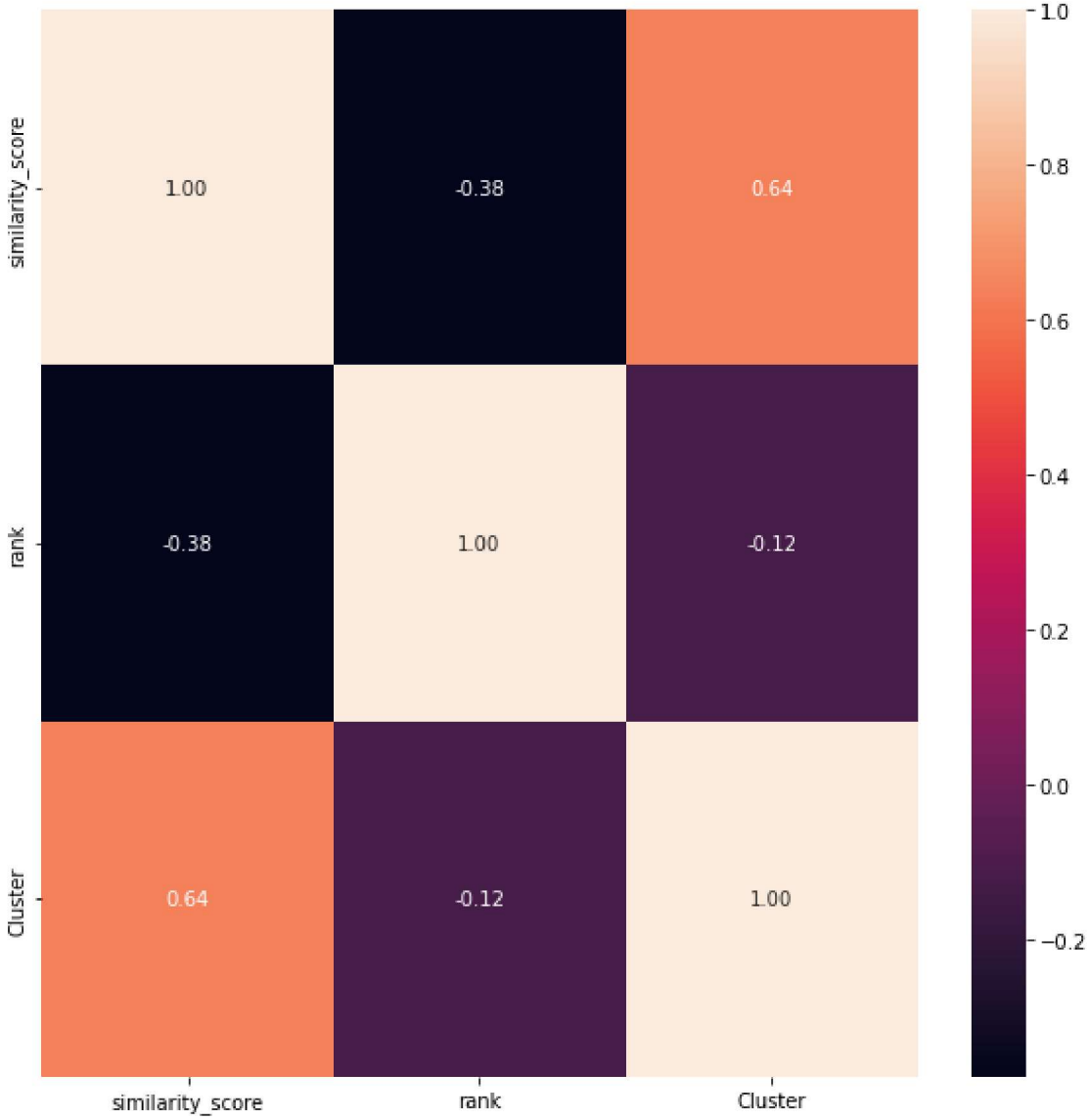
In [5]:
```
x = data.iloc[:, [2]]
```

In [6]:
```
x  = np.array(x)
```

In [7]:
```
#K-means clustering
kmeans_model = KMeans(n_clusters=3,  random_state=32932)
# Fit into our dataset fit
kmeans_predict = kmeans_model.fit_predict(x)
```

In [8]:
```
#Forming the cluster column
data['Cluster'] = kmeans_predict
```

In [9]:
```
#Correlation matrix
f,axs = plt.subplots(figsize=(10, 10))
sns.heatmap(data.corr(), annot = True, fmt= '.2f')
plt.show()
```



In [10]:
```
#Percentage
data['Brand_Percentage']=data['similarity_score'] / data.groupby('influencer_name')['si
```

In [11]:
```
print(data.head(20))
```

```
   competitor_name     influencer_name  similarity_score  rank  Cluster  \
0             bulk        aaroncgshore          0.052173   2.0        0
1             bulk         adamcollard          0.112365   3.0        2
2             bulk        aliceliveing          0.105055   3.0        2
3             bulk          brown.elle          0.031806   5.0        0
4             bulk      charlottedawsy          0.096758   5.0        2
5             bulk         chessieking          0.086953   8.0        2
6             bulk          chloe.khan          0.055821   4.0        0
7             bulk      courtneydblack          0.101296   3.0        2
```

```
8            bulk  danosborneofficial          0.076872   3.0        2
9            bulk      gabbydawnallen          0.096660   5.0        2
10           bulk           ini.helen          0.032789   4.0        0
11           bulk       itsalwayshana          0.071972   8.0        2
12           bulk         jamesgshore          0.066546   2.0        0
13           bulk        jamessmithpt          0.064629   2.0        0
14           bulk           jesshunt2          0.047472   3.0        0
15           bulk     jessica_rose_uk          0.050291   2.0        0
16           bulk        jesswright77          0.059196   4.0        0
17           bulk         katiepiper_          0.084433   9.0        2
18           bulk         korisampson          0.054819   3.0        0
19           bulk           lucymeck1          0.078999   7.0        2

    Brand_Percentage
0           0.123345
1           0.108590
2           0.100250
3           0.099872
4           0.099523
5           0.089869
6           0.107274
7           0.104530
8           0.110017
9           0.101088
10          0.124270
11          0.089664
12          0.115957
13          0.112817
14          0.114522
15          0.119257
16          0.102149
17          0.088318
18          0.107003
19          0.097050
```

In [12]:

```
a2=data[(data['influencer_name']=='adamcollard')]
a2.head(20)
```

Out[12]:

	competitor_name	influencer_name	similarity_score	rank	Cluster	Brand_Percentage
1	bulk	adamcollard	0.112365	3.0	2	0.108590
31	formnutrition	adamcollard	0.098637	6.0	2	0.095324
61	indisupplements	adamcollard	0.146757	1.0	1	0.141827
91	medterra.international	adamcollard	0.084716	9.0	2	0.081870
121	motionnutrition	adamcollard	0.070313	10.0	0	0.067951
151	neat_nutrition	adamcollard	0.100602	5.0	2	0.097223
181	neurohacker	adamcollard	0.090497	8.0	2	0.087457
211	thenue_co	adamcollard	0.103609	4.0	2	0.100128
241	liveinnermost	adamcollard	0.096969	7.0	2	0.093711
271	puresport	adamcollard	0.130297	2.0	1	0.125919

In [ ]: