	K-means			Set Nyasha M, 28 A	Apr 2021	hotos, status ui		
	pages of 10 Thai fash Each Facebook post "sad" and "angry") re The dataset came fro https://archive.ics.uci	hion and cosmetics t had its engagement ecorded into the dat om the following we	retail sellers betw nt metrics (coonsis aset, along with a ebpage at the Univ	veen March 2012 sting of shares, o timestamp of wh versity of Californ	2 and June 2018 comments, likes, ien each engage ia Irvine (UCI):	or emoji reactio	onsnamely, "lov	
	Objective: The goal categories, such as be that the posts were n	of this notebook was by the type of media made.	as to determine if t a they posted, the	the Facebook po types of engage	sts of the sellers ment they gener	ated from Face	book users, or b	y the year/month
	Citation: Nassim De 2018: The 20th Confe	ference of the Austr	ralian and New Ze	•		•		•
In [2]:	<pre>import matplot1: import pandas a: import numpy as import seaborn import scipy as # Modeling</pre>	s pd np as sns	lt					
T [4].	import sklearn.pfrom sklearn.com from sklearn.clufrom sklearn.meefrom yellowbrick	mpose import m uster import K trics import s k.cluster.silh	ake_column_tra Means as kmean ilhouette_sco	ansformer ns re	isualizer, s	ilhouette_v	Lsualizer	
In [4]: In [26]:	plt.style.use('I	bmh ') review our dataset.						
Out[26]:	<pre>df = pd.read_cs df.head(3) status_id status_</pre>	_type status_publis video 4/22/2018	shed num_reaction		nts num_shares	num_likes nu	m_loves num_w 92	rows num_hahas
In [92]: Out[92]:	2 3	photo 4/21/2018 2 video 4/21/2018		50 27 2	0 0 36 57	150 204	21	0 0 1
In [27]:	The last 4 columns to the last 4 columns.					the rows where	e the data is NO ⁻	T missing, within
Out[27]: In [28]:			they only con	_	_	num_likes nun	n_loves num_wo	ows num_hahas
Out[28]:	status_id status_ 0 1 1 2	_type status_publis video 4/22/2018 photo 4/21/2018 2 video 4/21/2018	6:00 52 2:45 15	29 5 50	nts num_shares 12 262 0 0 36 57	432 150	m_loves num_w 92 0 21	rows num_hahas 3 1 0 0 1 1
In [7]:		ize=(5, 3))	tegorical var			111 tle('status	0 _type (our ou	0 0
	photo 4288 video 2334 status 365 link 63 Name: status_typ							
	status_typ	pe (our outcome va	ariable)					
	1000 video pr	hoto link	status					
In [96]:		shed.astype('d).hist(color	='dodgerblue	', edgecolon	r='black'), p	olt.title(' <mark>dat</mark>
	3500 - 3000 - 2500 - 2000 - 1500 -							
In [97]:	# How many column len (df.iloc[:,3			first two c	olumns (stat	us and date,	?	
Out[97]: In [98]:	Checking the distribu	:]].hist(figsi			ue', edgecol	or='black')		
	plt.show() num_ar 6000			nents 6000 4000		ahas		
	2000 - 10 num_li	2000 20 30 0 likes	0 5000 10000 1 num_lov	es 6000	num_rea	100 150 ctions		
	4000 - 2000 - 0 2000 num_s	4000 2000 4000 sads	0 200 400 num_sha		0 2000 num_w	4000 rows		
	6000 - 4000 - 2000 - 0 20	4000 2000 40) -	2000	, -	200		
<pre>In [8]: Out[8]:</pre>				s num_comment	s num_shares	num_likes nun	n_loves_ num_wo	ows num_hahas
	Let's remove the st interesting data that whith problematic/cumb might provide some i	catus_id variable we could try feeding persome since 12 m	since it seems to	be just an arbitra	ary feature. sta	tus_publish	ed seems like it	might have some
	df = df.drop([': Extracting the years to Let's check the formal	from status_puk at:	olished returne d	d an error, becau	se it appears tha	at this variable i	sn't actually in a	datetime format.
	dtype('0') Odd. What happens	if we try changing i	t into a datetime o	•				
	<pre>df['date'] = pd # It did, so no o now drop # the old times df['year'] = df</pre>	w let's extrac	t the years f	rom our new (datetime col		tus_published	d'. We can als
In [38]: Out[38]:	status_type num	n_reactions num_co	omments num_sha	ares num_likes				
	0 video1 photo2 video3 photo4 photo	529 150 227 111 213	512 0 236 0	 262 432 0 150 57 204 0 111 0 204 	92 0 21 0 9	3 0 1 0	1 1 0 0 1 0 0 0 0 0 0 0 0	0 20 0 20 0 20
	Now, K-means can o OneHotEncode) both 'adversely affect', I m won't get an accurate	only work on intege h our status_tyr nean that our categ	rs, so let's change pe and year fea ories cannot simpl	status_type atures so that the ly be in the form	into an int-type by don't adverse of ordinal values	feature. We'll a y affect our K-n (i.e., 1, 2, 3 r	lso have to 'dum neans algorithm. n), because if the	nmify' (or When I say
In [34]: In [35]:	<pre>trans = enc.fit trans_df = pd.Da df_merged = pd.a</pre>	_transform(df[vataFrame(trans concat([df, tr	['status_type , columns = en	nc.get_featu:		tatus_type',	'year']))	
Out[35]:		1) n_reactions num_co 529 150		num_likes 262 432 0 150	num_loves num	n_wows num_h 3	ahas num_sads 1 1 0 0	0
In [36]:	<pre>2 video 3 photo 4 rows × 22 columns # df['status li.</pre>		236 0	57 204 0 111 f['status sta	21 0 atus'l, df['	1 0 status video	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	0
	ns[1], df_trans # df['year_2012 rans[6], df_tra. # df['year_2016	[2], df_trans['], df['year_2 ns[7]] ['], df['year_2	[3]] [013'], df['yea [017'], df['yea	ar_2014'], d ar_2018'] =	f['year_2015 [df_trans[8	'] = [df_t:], df_trans	_ rans[4], df_t [9], df_trans	_ trans[5], df_t s[10]]
In [37]:	<pre># Now standardi we have counts clust_array = p: df_merged.loc[: df_merged = df_r df merged.head()</pre>	reprocess.Stan , 'num_reactio merged.drop(['	dardScaler().: ns':'num_angr	fit_transform ys'] = clust	m(df_merged. _array			
Out[37]:	num_reactions n 0 0.646104 1 -0.173192		1.686879 0.4827 -0.304144 -0.1447	727 1.983266 720 -0.318454	0.196196 0	n_hahas num_s .076713 0.473 .176010 -0.152	570 -0.155748	B 0.0
	2 -0.006738 Model Deploym	0.013089 nent	0.129017 -0.0245	71 0.206938				R Of
	Let's run our clusterir	ng model on the da	ita across a range	of test/dummy c		.076713 -0.152		
	Let's run our clustering best division of data/out will plot the sum of determine the ideal number that the clusters are to	of squared errors a number of clusters t	ne dataset. after each cluster (i for our dataset. Th	i.e., use the elbo	luster numbers a w method) as w re ranges from -	.076713 -0.152 and record whic ell as use the si I to 1; values cl	h number of clus lhouette method	sters lead to the to help us
In [39]:	best division of data/o We will plot the sum of determine the ideal in that the clusters are for the silhouette_avg of the silhouet	of squared errors and another of clusters of fitting the outlines of the continuous	ne dataset. Ifter each cluster (iffor our dataset. The offor the data (if clusted to determine are set), init = 1.	i.e., use the elbone silhouette scorers are indeed powers are indeed powers are clusted where clusted at the cluster in the cl	w method) as we re ranges from - resent) the best.	and record whice ell as use the sind to 1; values classed are placed are placed 12, random_s	th number of cluss thouette method osest to 1 are be the control of the control o	to help us est and suggest
In [39]:	best division of data/o We will plot the sum of determine the ideal in that the clusters are to the silhound that the clusters are to the clusters are the clusters are to the clusters are the clu	of squared errors and any squared errors and squared e	ne dataset. Ifter each cluster (ifter each cluster (ifter our dataset. The of the data (if cluster) If to determine and the data, with ertia_ From each attemption of the data and the data, with ertia_	i.e., use the elbone silhouette scorers are indeed powers are indeed powers are indeed powers. **where cluste* - 'k-means++ ch cluster i hoout overfi pted cluster	w method) as we re ranges from - resent) the best. er centroids ', n_init = teration. Fi tting.	and record whice ell as use the sind to 1; values classed are placed are placed 12, random_s	th number of cluss thouette method osest to 1 are be the control of the control o	to help us est and suggest
In [39]:	best division of data/e We will plot the sum of determine the ideal in that the clusters are for that the clusters are for k in range (if the sum of the	of squared errors and an	ne dataset. Infter each cluster (infor our dataset. The of the data (if clusted at a determine are set as a determine at a data, with the da	i.e., use the elboore silhouette score ers are indeed properties are indeed properties. where cluster in the c	w method) as we re ranges from - resent) the best. er centroids ', n_init = teration. Fi tting. number labels) tte_avg[k-2] .73430147218 .60459182979 .60711577550	and record whice ell as use the single to 1; values classed are placed are placed are placed are number of the num	th number of cluss thouette method osest to 1 are be the control of the control o	to help us est and suggest
In [39]:	best division of data/e We will plot the sum of determine the ideal in that the clusters are for that the clusters are for k in range (if the sum of the	of squared errors and the continuous of squared errors and the continuous of squared errors and the continuous of squared errors of squared errors of squared errors of squared errors of squared errors. In the continuous errors of squared errors. In the continuous errors errors of squared errors. In the continuous errors errors of squared errors. In the continuous errors	ne dataset. after each cluster (ifer our dataset. The for our dataset. The for the data (if clusted at the data, with ertia_ com each attempted at the score is esilhouette_se silhouette_se esilhouette_se esilhouett	i.e., use the elboore silhouette score ers are indeed properties are indeed properties. where cluster in the c	w method) as we re ranges from - resent) the best. er centroids ', n_init = teration. Fi tting. number labels) tte_avg[k-2] .73430147218 .60459182979 .60459182979 .6071577550 .60955468611 .48931081109 .48615034379 .29486087907 .47573260612 0.2733538708	and record whice ell as use the sill to 1; values class of the number of	th number of cluss thouette method osest to 1 are be the control of the control o	to help us est and suggest
	best division of data/e We will plot the sum of determine the ideal in that the clusters are for that the clusters are for k in range (in the first substitution of the first	distinct groups in the of squared errors and the of clusters of fitting the outlines of squared errors and the control of the clusters of the control of the clusters of the control of the clusters of the cl	ne dataset. after each cluster (ifer our dataset. The for our dataset. The first of the data (if clusters) at the data (if clusters) at the data, with ertia_ com each attempth ouette_score is esilhouette_se silhouette_se silhouette_se silhouette_se esilhouette_se silhouette_se	i.e., use the elboure silhouette score ers are indeed proposed by the score is an elboure score is an elboure is a core is an elboure is an elbo	w method) as we re ranges from - resent) the best. er centroids ', n_init = teration. Fi tting. number labels) tte_avg[k-2] .73430147218 .60459182979 .60459182979 .60711577550 .60955468611 .48931081109 .48615034379 .29486087907 .47573260612 0.2733538708 0.3027844002 e'.', marker abel('Sum of	and record whice ell as use the six is to 1; values class are placed are placed 12, random_s and the number 13, random_s and the number 14, random_s are placed 14, random_s and the number 15, random_s are placed 15, random_s are placed 12, random_s are placed 13, random_s are placed 14, random_s are placed 14, random_s are placed 15, random_s are placed 14, random_s are placed 14, random_s are placed 15, random_s are placed 15, random_s are placed 14, random_s are placed 14, random_s are placed 15, random	th number of clus thoughte method cosest to 1 are be state = 2) er of cluster white', marke	to help us est and suggest reproducible rs that explai
	We will plot the sum of determine the ideal in that the clusters are form that the clusters are form in the clusters. The clusters are form in the clusters. The clusters are form in the clusters are form in the clusters are form in the clusters. The clusters are form in the clusters are form in the clusters are form in the clusters. The clusters are form in the clusters are form in the clusters are form in the clusters. The clusters are form in the clusters are form in the clusters are form in the clusters. The clusters are form in the clusters are form in the clusters are form in the clusters. The clusters are form in the clusters are form in the clusters are form in the clusters. The clusters are form in the clusters are form in the clusters are form in the clusters. The clusters are form in the clusters. The clusters are form in th	distinct groups in the of squared errors and the of clusters in the fitting the outlines of the squared errors and the squared errors in the squared error	ne dataset. Infter each cluster (infor our dataset. The for our dataset. The for the data (if clusters) If the data (if cluster	i.e., use the elboore silhouette score ers are indeed properties a	w method) as we re ranges from - resent) the best. er centroids ', n_init = teration. Fi tting. number labels) tte_avg[k-2] .73430147218 .60459182979 .60459182979 .60711577550 .60955468611 .48931081109 .48615034379 .29486087907 .47573260612 0.2733538708 0.3027844002 e'.', marker abel('Sum of marker='.', in the content of the	and record which are placed are p	thoughte method posest to 1 are best to 1 are best to 1 are best to 2 are allows for state = 2) For of cluster white', market cor') color='white',	to help us est and suggest reproducible rs that explai
	We will plot the sum determine the ideal in that the clusters are form that the clusters are form in the clusters are for	distinct groups in the of squared errors and the continuous of clusters in fitting the outlines of the continuous of the	ne dataset. Infter each cluster (infor our dataset. The for our dataset. The for the data (if clusters) If the data (if cluster	i.e., use the elboure silhouette score ers are indeed properties are indeed properties are indeed properties. It is a second and the second are indeed as a seco	w method) as we re ranges from - resent) the best. er centroids ', n_init = teration. Fi tting. number labels) tte_avg[k-2] .73430147218 .60459182979 .60711577550 .60955468611 .48931081109 .48615034379 .29486087907 .47573260612 0.2733538708 0.3027844002 e'.', marker abel('Sum of marker='.', in the content of th	and record which are placed are p	though method object to 1 are best to 1 are best to 1 are best to 1 are best to 2 are of cluster of	to help us est and suggest reproducible rs that explai
	We will plot the sum of determine the ideal in that the clusters are for it is a many to the sum of that the clusters are for it is a many to the sum of t	distinct groups in the of squared errors and the continuous of clusters in fitting the outlines of the continuous of the	ne dataset. Infter each cluster (infor our dataset. The for our dataset. The for the data (if clusters) If the data (if cluster	i.e., use the elboure silhouette score ers are indeed properties are indeed properties are indeed properties. I where cluster in the cluster is the cluster in the cluster	w method) as we re ranges from - resent) the best. er centroids ', n_init = teration. Fi tting. number labels) tte_avg[k-2] .73430147218 .60459182979 .60459182979 .60955468611 .48931081109 .48615034379 .29486087907 .47573260612 0.2733538708 0.3027844002 e'.', marker abel('Averag	and record which are placed are p	though method object to 1 are best to 1 are best to 1 are best to 1 are best to 2 are of cluster of	to help us est and suggest reproducible rs that explai
	We will plot the sum determine the ideal in that the clusters are form in the clusters are form	distinct groups in the of squared errors and the of clusters in the fitting the outlines of the fitting the outlin	ne dataset. Ifter each cluster (ifter our dataset. The for our data (if clusted at a constant of the data, with ertia from each attempt the data, with ertia from each attempt the score is estable at a constant of the data, with ertia from each attempt the score is estable attempt the silhouette from each attempt the	i.e., use the elboure silhouette score ers are indeed properties are indeed properties are indeed properties. It is a second and the second are indeed cluster (df_merged, indeed clust	w method) as we ranges from - resent) the best. er centroids ', n_init = teration. Fi tting. number labels) tte_avg[k-2] .73430147218 .60459182979 .60459182979 .60955468611 .48931081109 .48615034379 .29486087907 .47573260612 0.2733538708 0.3027844002 e'.', marker abel('Sum of marker='.', sabel('Average) abel('Average) abel('Average) abel('Average)	and record which are placed are p	th number of clus thoughte method cosest to 1 are be state = 2) er of cluster white', market cor') clor='white', e score') houette score	eredgewidth=1, markeredgewi no markeredgewi
In [46]:	We will plot the sum determine the ideal in that the clusters are in the greatest are in the greatest amount of the clusters are in the greatest amount of the clusters are in the clusters are	distinct groups in the of squared errors and the of clusters in the outlines of the squared errors and the outlines of the squared errors and the outlines of the squared errors and the squared errors and the squared errors are squared errors. The squared errors are squared errors are squared errors. The squared errors are squared errors. The average and the squared errors are squared errors. The average errors are squared errors are squared errors. The average errors are squared errors are squared errors. The average errors are squared errors are squared errors. The average errors are squared errors are squared errors. The average errors are squared errors are squared errors are squared errors. The average errors are squared errors are squared errors are squared errors. The squared errors are squared errors are squared errors are squared errors. The squared errors are squared errors are squared errors are squared errors are squared errors. The squared errors are squared errors are squared errors are squared errors are squared errors. The squared errors are squared errors. The squared errors are squ	figsize=(14, ist, color='b: clusters'), as figsize=(12, r(kmeans(2, n)) for our dataset. The figsize=(14, ist, color='b: clusters'), as figsize=(12, r(kmeans(2, n)))	i.e., use the elboure silhouette score ers are indeed properties a	w method) as we reanges from resent) the best. er centroids ', n_init = teration. Fi tting. number labels) tte_avg[k-2] .73430147218 .60459182979 .60711577550 .60955468611 .48931081109 .48615034379 .29486087907 .47573260612 0.2733538708 0.3027844002 e'.', marker abel('Sum of marker='.', see the contours of the con	and record which are placed are p	th number of clus Ihouette method best to 1 are best atate = 2) ar of cluster white', market cor') clor='white', escore') houette score archive score coring to see the score of clusters specifically to see	eredgewidth=1, markeredgewi ehow each
In [46]:	We will plot the sum of determine the ideal in that the clusters are in the	distinct groups in the of squared errors and the of clusters of the fitting the outlines of the squared errors and the outlines of the squared errors and the squared errors and the squared errors and the squared errors and the squared errors are squared errors and the squared errors are squared errors. In the squared errors are squared errors are squared errors are squared errors and the squared errors are squared errors are squared errors are squared errors. In the average errors are squared errors are squared errors are squared errors are squared errors. In the average errors are squared errors are squared errors are squared errors. The squared errors are squared errors are squared errors are squared errors. In the average errors are squared errors are squared errors. In the average errors are squared errors are squared errors. In the average errors are squared errors are squared errors are squared errors. In the squared errors are squared errors are squared errors are squared errors. In the squared errors are squared errors are squared errors are squared errors. In the squared errors are square	figsize=(14, ist, color='b' dplot') clusters'), as for 7050 samples for 70	i.e., use the elbo le silhouette score lers are indeed properties are indeed properties. where cluster i.e., use the elbo le silhouette score lers are indeed properties. 'k-means++ ch cluster indeed cluster (df_merged, indeed cluster (df	w method) as we reanges from resent) the best. er centroids ', n_init = teration. Fi tting. number labels) tte_avg[k-2] .73430147218 .60459182979 .60711577550 .60955468611 .48931081109 .48615034379 .29486087907 .47573260612 0.2733538708 0.3027844002 e'.', marker abel('Sum of marker='.', in abel('Average) of the contours of th	and record which are placed are p	h number of clus thouette method basest to 1 are be state = 2) er of cluster white', market cor') clor='white', ascore') houette score merged, ax =	eredgewidth=1, markeredgewi above ach action help us est and suggest reproducible as that explai are deredgewidth=1, and how each
In [46]:	We will plot the sum of determine the ideal in that the clusters are in the	distinct groups in the of squared errors and the of clusters in the fitting the outlines of an] *10 = [np.nan] *10 = [np.nan] *10 2, 12): ate = set.seed means (n_cluster (df_merged)) mof squared extends in [and the content of squared ex	figsize=(14, ist, color='b' dplot') clusters'), as for 7050 samples for 70	i.e., use the elbo we silhouette score ers are indeed processore indeed processore is a consistent of the cluster of the clust	w method) as we reanges from resent) the best. er centroids ', n_init = teration. Fi tting. number labels) tte_avg[k-2] .73430147218 .60459182979 .60711577550 .60955468611 .48931081109 .48615034379 .29486087907 .47573260612 0.2733538708 0.3027844002 e'.', marker abel('Sum of marker='.', in abel('Average) of the contours of th	and record which all as use the sil to 1; values class of each cluster serious are placed are place	th number of clus thouette method best to 1 are be thoughter for state = 2) the of cluster white', market cor') thouette score merged, ax = axs[1], colo	eredgewidth=1, markeredgewi above ach action help us est and suggest reproducible as that explai are deredgewidth=1, and how each
In [46]:	We will plot the sum determine the ideal in that the clusters are in th	distinct groups in the of squared errors and the of clusters in the fitting the outlines of an] *10 = [np.nan] *10 = [np.nan] *10 2, 12): ate = set.seed means (n_cluster (df_merged)) mof squared extends in [and the content of squared ex	figsize=(14, ist, color='b' dplot') clusters'), as for 7050 samples for 70	i.e., use the elbo le silhouette score lers are indeed propers are indeed propers. where cluster i.e., use the elbo le silhouette score lers are indeed propers. where cluster i.e., we are indeed propers. i.e., means++ ch cluster i hoout overfi pted cluster (df_merged, i i.e., silhouette score is : 0 score is : 0	w method) as we reanges from resent) the best. er centroids ', n_init = teration. Fi tting. number labels) tte_avg[k-2] .73430147218 .60459182979 .60711577550 .60955468611 .48931081109 .48615034379 .29486087907 .47573260612 0.2733538708 0.3027844002 e'.', marker abel('Sum of marker='.', in abel('Average) of the contours of th	and record which are placed are p	th number of clus thouette method best to 1 are be thoughter for state = 2) the of cluster white', market cor') thouette score merged, ax = axs[1], colo	eredgewidth=1, markeredgewi above ach action help us est and suggest reproducible as that explai are defended and the explain are defended and the explain are defended and the
In [46]:	We will plot the sum determine the ideal in that the clusters are in th	distinct groups in the of squared errors and the of clusters in the fitting the outlines of an] *10 = [np.nan] *10 = [np.nan] *10 2, 12): ate = set.seed means (n_cluster (df_merged)) mof squared extends in [and the content of squared ex	ne dataset. Inter each cluster (iffor our dataset. The for our dataset. The for our dataset. The for our dataset. The for our dataset. The form the data (if cluster is a substitute of the data, with t	i.e., use the elbo le silhouette score les silhouette score lers are indeed properties where cluster 'k-means++ ch cluster i hoout overfi pted cluster (df_merged, :", silhouet score is: 0 sco	w method) as we ranges from resent) the best. er centroids ', n_init = teration. Fi tting. number labels) tte_avg[k-2] .73430147218 .60459182979 .60711577550 .60955468611 .48931081109 .48615034379 .29486087907 .475732508612 0.2733538708 0.3027844002 e'.', marker abel('Sum of marker='.', sabel('Average Siles) random_state amples, \n2 C random_state e Plot of KMear Average Siles	and record which are placed are p	th number of clus thouette method best to 1 are be allows for state = 2) er of cluster are cor') blor='white', e score') houette score merged, ax = axs[1], colo	eredgewidth=1, markeredgewi ers in 5 Centers sters lead to the to help us est and suggest reproducible rs that explai eredgewidth=1,
In [46]: In [52]:	We will plot the sum determine the ideal in that the clusters are in th	distinct groups in the of squared errors and squared errors and squared errors are fitting the outlines of the squared	ne dataset. Inter each cluster (ifter each cluster (ifter each cluster (ifter each cluster (ifter each cluster)) If to determine (ifter each cluster) If the data (ifter each each each each each each each each	i.e., use the elbo ie silhouette score iers are indeed processed indeed processed indeed processed indeed processed indeed processed indeed processed indeed in the second in the	duster numbers as we method) as we ranges from resent) the best. ar centroids der centroids d	and record which are the sill to 1; values classed are placed are	thouette method osest to 1 are becomes for clusters white', marked of clusters and clusters specifically to see the core of clusters specifically to see the core of clusters axs[1], cold or 7050 Sample axs[1], cold or 7050 Sample	to help us est and suggest reproducible rs that explain the end end end end end end end end end en
In [46]: Out [52]:	best division of data/ We will plot the sum determine the ideal in that the clusters are in the greatest are in the greatest amount of so in the greatest amount of the gre	distinct groups in the of squared errors and square	ne dataset. Inter each cluster (inter each cluster) Inter each cluster (inter our dataset. The first of the data (if clusters) Inter the data (if clusters) Inter the data, with the d	i.e., use the elbo le silhouette score lers are indeed process are indeed process are indeed process. where cluster if the means are indeed process. It is the cluster in the cluster i	duster numbers as we method) as we ranges from	and record which are sile as use the sile in to 1; values classed are placed	thouette method obsest to 1 are be obsest to 1 are	to help us est and suggest reproducible rs that explain the end end end end end end end end end en
In [52]: In [54]: In [55]:	best division of data/ We will plot the sum determine the ideal in that the clusters are in the cluster in the cluster are in the cluster are in the cluster are in the clusters are	distinct groups in the of squared errors as number of clusters in fitting the outlines of an	ne dataset. Inter each cluster (ifor our dataset. The of the data (if clusters) and the data (if clusters) are some ach attempted to the data, with the dat	i.e., use the elbo le silhouette score les silhouette score les are indeed proposed in the state of the state	duster numbers a w method) as we re ranges from resent) the best. er centroids ', n_init = teration. Fi tting. number labels) tte_avg[k-2] .73430147218 .60459182979 .60711577550 .60955468611 .48931081109 .48615034379 .29486087907 .47573260612 0.2733538708 0.3027844002 e'.', marker abel ('Sum of marker='.', sabel ('Average abel ('Average amples, \n2 C random_state amples, \n2 C random_state e Plot of KMear Average Sil	and record which are considered as use the single s	th number of clus thoughte method disest to 1 are be the control of the control of clusters white', market for') for of clusters specifically to see merged, ax = axs[1], colo or 7050 Sample for 7050 Sample axs[1], colo or 7050 Sample	to help us set and suggest reproducible rs that explai eredgewidth=1, markeredgewi ors='Dark2').f
In [46]: In [52]: In [54]:	best division of data/ We will plot the sum determine the ideal in that the clusters are in the cluster in the cluster are in the cluster are in the cluster are in the clusters are	distinct groups in the of squared errors and the property of clusters in the fitting the outlines of the property of the prope	ne dataset. Inter each cluster (if for our dataset. The for our dataset. The first data (if cluster data (if cluster data (if cluster data) and the data, with the data, w	i.e., use the elbo le silhouette score les silhouette score les are indeed proposed in the state of the state	duster numbers a w method) as we re ranges from resent) the best. er centroids ', n_init = teration. Fi tting. number labels) tte_avg[k-2] .73430147218 .60459182979 .60711577550 .60955468611 .48931081109 .48615034379 .29486087907 .47573260612 0.2733538708 0.3027844002 e'.', marker abel ('Sum of marker='.', sabel ('Average abel ('Average amples, \n2 C random_state amples, \n2 C random_state e Plot of KMear Average Sil	and record which are considered as use the single s	th number of clus thoughte method disest to 1 are be the control of the control of clusters white', market for') for of clusters specifically to see merged, ax = axs[1], colo or 7050 Sample for 7050 Sample axs[1], colo or 7050 Sample	eters lead to the stand suggest reproducible reproducible rs that explair rendered ready and the standard ready an
In [52]: In [54]: In [56]:	best division of data/ We will plot the sum determine the ideal in that the clusters are in the clusters are in the cluster are in the clusters are in the clusters are in the clusters are in the cluster are in the clusters ar	distinct groups in the of squared errors and the squared errors and	ne dataset. Infer each cluster (ifor our dataset. The for the data (if cluster is a consistent of the data (if cluster is a consistent in the consistent in the consistent is a consistent in the consistent in	i.e., use the elbo le silhouette score ers are indeed process are indeed process are indeed process. I where cluster I hoout overfice I hoout overfice I wilhoue	duster numbers a w method) as we re ranges from resent) the best. er centroids ', n_init = teration. Fi tting. number labels) tte_avg[k-2] .73430147218 .60459182979 .60711577550 .60955468611 .48931081109 .48615034379 .29486087907 .47573260612 0.2733538708 0.3027844002 e'', marker abel ('Sum of marker='.', see abel ('Average Siles) abel ('Average Siles) t_num, n_ini I dataset. e Plot of KMear Average Siles t_num, n_ini I dataset.	and record which are considered as use the single to 1; values classes are placed are pl	th number of clus thoughter method dosest to 1 are be thoughter method dosest to 1 are be thoughter for state = 2) the rof cluster the score of cluster the	exters lead to the stocked here to help us est and suggest and suggest are reproducible as that explain that explain the explain the explain that explain the explain the explain that explain the explain that explain the e
In [52]: In [54]: In [56]:	best division of data/ We will plot the sund determine the ideal in that the clusters are in th	distinct groups in the of squared errors and the squared errors and	ne dataset. Infer each cluster (ifor our dataset. The for the data (if cluster is a consistent of the data (if cluster is a consistent in the consistent in the consistent is a consistent in the consistent in	i.e., use the elbo le silhouette score ers are indeed process are indeed process are indeed process. I where cluster I hoout overfice I hoout overfice I wilhoue	duster numbers a w method) as we re ranges from resent) the best. er centroids ', n_init = teration. Fi tting. number labels) tte_avg[k-2] .73430147218 .60459182979 .60711577550 .60955468611 .48931081109 .48615034379 .29486087907 .47573260612 0.2733538708 0.3027844002 e'', marker abel ('Sum of marker='.', see abel ('Average Siles) abel ('Average Siles) t_num, n_ini I dataset. e Plot of KMear Average Siles t_num, n_ini I dataset.	and record which are considered as use the single to 1; values classes are placed are pl	th number of clus thoughter method dosest to 1 are be thoughter method dosest to 1 are be thoughter for state = 2) the rof cluster the score of cluster the	eters lead to the stand suggest reproducible reproducible rs that explair redgewidth=1, markeredgewidth=1, markeredgewidth=1, res in 5 Centers num_angrys ye 0 20 0 20
In [52]: In [54]: In [56]: Out [56]:	best division of data/ We will plot the sum determine the ideal in that the clusters are in the clusters are in the clusters. Sound in the cluster are in the clusters are in the clusters. For n_clusters are in the clusters are in the clusters are in the clusters are in the clusters are in the clusters. For n_clusters are in the clusters are in the clusters are in the clusters. For n_clusters are in the clusters are in the clusters. For n_clusters are in the clusters are in the clusters. For n_clusters are in the clusters are in the clusters. For n_clusters are in the clusters. For n_clusters are in the clusters. For n_clusters are in the clusters are in the clusters. For n_clusters are in the clusters are in the clusters. For n_clusters are in the clusters are in the clusters. For n_clusters are in the clusters are in the clusters. For n_clusters are in the clusters are in the clusters. For n_clusters are in the clusters are in the clusters. For n_clusters are in the clusters are in the clusters. For n_clusters are in the clusters are in the clusters. For n_clusters are in the clusters are in the clusters are in the clusters are in the clusters. For n_clusters are in the clusters are in the clusters are in	distinct groups in the of squared errors a number of clusters fitting the outlines of an	me dataset. Intereach cluster (intereach cluster) (intereach clus	i.e., use the elbo ie silhouette score ers are indeed property in the score is an inde	duster numbers a w method) as w eranges from resent) the best. er centroids ', n_init = teration. Fi tting. number labels) tte_avg[k-2] .73430147218 .60459182979 .6071577550 .60955468110 .48861034379 .29486087907 .47573260612 0.2733538708 .3027844002 e'.', marker abel ('Sum of marker='.', i abel ('Average of the contours of marker='.', i abel ('Average abel ('Average of the contours of t	and record which and record which are placed are placed are placed 12, random 12, random 12, random 12, random 13, random 14 16 17 18 18 19 19 19 19 19 19 19 19 19 19 19 19 19	th number of clus thought method control of the second of	eters lead to the to help us est and suggest reproducible rs that explai eredgewidth=1, markeredgewi a how each ests. no.7 0.8 mmands to see if ests.
<pre>In [52]: In [54]: In [56]: Out [56]:</pre>	best division of data/ We will plot the sum determine the ideal in that the clusters are division that the clusters are division that the clusters are division to the clusters are for n_clusters are f	distinct groups in the of squared errors a number of clusters in the of squared errors a number of clusters in the other state of the outlines of the clusters in the clusters	ne dataset. Inter each cluster (ifter each cluster (ifter each cluster (ifter each cluster (ifter each cluster) and the data (ifter each attential each each attential each each each each each each each each	i.e., use the elbo e silhouette score ers are indeed properties are indeed properties. where cluster 'k-means++ ch cluster i hoout overfi pted cluster (df_merged, : :", silhoue score is: 0 scor	without as we method) as we ranges from resent) the best. er centroids ', n_init = teration. Fi tting. number labels) tte_avg[k-2] .73430147218 .60459182979 .6071577550 .60955468611 .48931081109 .486150347907 .47573260612 .2733538708 .3027844002 e'.', marker abel('Sum of marker='.', state of the contours of t	and record which are placed are p	th number of clus thouette method best to 1 are be the sest to 1 are be	to help us set and suggest reproducible rs that explai eredgewidth=1, markeredgewi and suggest reproducible rs that explai credgewidth=1, markeredgewi and suggest reproducible rs that explai reproducible reproducible rs that explai reproducib
In [52]: In [54]: In [56]:	best division of data/ We will plot the sum- determine the ideal of that the clusters are in the cluster are in the cluster are in the cluster are in the clusters are in the cluste	distinct groups in the of squared errors a fitting the outlines of clusters in fitting the outlines of an interest of the outlines of an interest of the outlines of an interest of an int	ne dataset. Inter each cluster (ifter each cluster) (for our dataset. The of the data (if cluster) (ifter data (ifter da	i.e., use the elbo le silhouette score ers are indeed property in the silhouette score 'k-means++ ch cluster i	w method) as we ranges from the centroids resent) the best. er centroids ', n_init = teration. Fi tting. number labels) tte_avg[k-2] .73430147218 .60459182979 .6071577550 .60955468611 .48931081109 .48615034379 .24957326061 .2733538708 .3027844002 .2733538708 .3027844002 e'.', marker abel('Sum of marker='.', iiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiii	and record which and record which and record which are placed are placed 12, random_s and the number 12, random_s and the number 13, random_s and the number 14, random_s and the number 15, then do s are silhouette Average sil	th number of clus thought method consent to 1 are be the sest to 1 are b	to help us set and suggest reproducible rs that explai eredgewidth=1, markeredgewi ors='Dark2').f es in 5 Centers num_angrys ye 0 20 0 20 0 20 0 20 0 20
In [52]: In [52]: In [54]: Out [55]: Out [56]: Out [57]:	best division of data/ We will plot the summare determine the ideal that the clusters are in that the clusters are in the clu	distinct groups in the of squared errors and the squared errors and	ne dataset. Infer each cluster (iffor our dataset. The for the data (if cluster) is to determine as a second at the data, with the data, wit	i.e., use the elbo e silhouette score ers are indeed process are indeed process are indeed process are indeed process. where cluster 'k-means++ ch cluster i hoout overfi pted cluster (df_merged, :", silhouete score is: 0 score is	duster numbers a w method) as w we ranges from - resent) the best. er centroids ', n_init = teration. Fi tting. number labels) tte_avg[k-2] .73430147218 .60459182279 .6071577550 .60955468611 .48931081109 .426180343797 .44573260612 .2733538708 .3027844002 e'.', marker abel('Sum of marker='.', i abel('Average of the contours of th	and record which and record which and record which are placed are placed are placed 12, random_s and the number 12, random_s 12, random_s 12, random_s 12, random_s 12, random_s 12, random_s 14, random_s 15, random_s 16, random_s 17, random_s 18, random_s 19, random_s 19, random_s 19, random_s 10, random_s 10, random_s 10, random_s 11, random	th number of clus thouette method cosest to 1 are be thouette for state = 2) ar of cluster white', marke cor') thouette score merged, ax = axs[1], colo for 7050 Sample cor	enters lead to the to help us est and suggest reproducible rs that explair reproducible rs that explain reproducible rs that explair
Out [52]: In [54]: In [56]: Out [56]: Out [57]:	best division of data/ We will plot the sum- determine the ideal in that the clusters are in ss_dist = [np.n. silhouette_avg: for k in range(): # random_st. # results # Record sums the greatest # amount of ss_dist[-2 # Get silho labels = k_g silhouette_print("For: # Get silho labels = k_g silhouette_print("For: # The in strong clusters = For n_clusters = For	distinct groups in the of squared errors and square	ne dataset. Interest cluster (interest cluster) Interest cluster (interest cluster)	i.e., use the elbo i.e., use the	duster numbers a w method) as w w method) as w we ranges from resent) the best. er centroids ', n_init = teration. Fi tting. number labels) tte_avg[k-2] .73430147218 .60459182979 .6071157750 .60955468611 .48931081109 .29486087907 .4757326067 .2733538708 .23027844002 e'.', marker abel ('Sum of marker='.', marker abel ('Average a	and record which are placed are placed 12, random_s 10, random_s 11, values class are placed 12, random_s 12, random_s 13, random_s 14, random_s 15, random_s 16, random_s 17, random_s 18, random_s 19, random_s 10, random_s 10, random_s 10, random_s 11,	th number of clus thouette method best to 1 are be allows for state = 2) er of cluster white', market for') cor', white', escore') thouette score merged, ax = axs[1], colo or 7050 Sample axs[1], colo or 7050 Sample dom_state=2) and and num_sads num_sads num or 1 or 2 for 1 or 2 for 1 or 2 for 3 or 4 or 3 or 4 or 5 or 6 dom_state=2) and and num_sads or 1 or 2 for 3 or 4 or 7	num_angrys year_(magests. num_angrys year_(magests.) num_angrys year_(magests.) num_angrys year_(magests.) num_angrys year_(magests.) num_angrys year_(magests.) num_angrys year_(magests.)
In [54]: In [54]: In [56]: Out [55]: Out [57]:	best division of data/ We will plot the sum- determine the ideal in that the clusters are in ss_dist = [np.m. silhouette_avg. for k in range() # random st. **Example of the st. **Exam	distinct groups in the of squared errors and the outlines of squared errors of squar	ne dataset. Interest dataset.	i.e., use the elboore is.e., use the elboore silhouette score ers are indeed provided and the score is are indeed provided cluster in the court overfil pred cluster in the	w method) as we method) as we ranges from - resent) the best. w method) as we ranges from - resent) the best. er centroids ', n_init = teration. Fi tting. number labels) tte_avg[k-2] .73430147218 .60459182979 .2948608790 .4057548661 .4893108109 .48615034379 .2948608790 .4757328708 .3027844002 e'-', marker abel('Sum of marker='.', sabel('Sum of marker='.', sabel('Average Silver abels) t_num, n_ini dataset. e Plot of KMear Average Silver abels t_num, n_ini dataset. features except to the contours of the	and record which and record which ell as use the si I to 1; values cla are placed 12, random_s 12, random_s 12, random_s 14, random_s 15, then do s 16, squared error 16, squared error 17, squared error 18, the squared error 19, ax = squared error 19, ax = squared error 10, ax = squared error 11, ax = squared error 12, random 13, squared error 14, squared error 15, then do s 16, squared error 16, squared error 17, squared error 18, squared error 19, ax = squared error 19, ax = squared error 10, squared error 11, squared error 12, random 13, squared error 14, squared error 15, squared error 16, squared error 17, squared error 18, squared error 19, squared error 19, squared error 10, squared error 10, squared error 11, squared error 11, squared error 12, squared error 13, squared error 14, squared error 15, squared error 16, squared error 17, squared error 18, squared error 19, squared error 19, squared error 10, squared error 10, squared error 10, squared error 11, squared error 12, squared error 13, squared error 14, squared error 15, squared error 16, squared error 17, squared error 18, squared error 19, squared error 19, squared error 19, squared error 10, squared error 10, squared error 11, squared error 12, squared error 13, squared error 14, squared error 15, squared error 16, squared error 16, squared error 17, squared error 18, squared error 19, squared error 19, squared error 19, squared error 10, squared error 10, squared error 10, squared error 11, squared error 11, squared error 12, squared error 13, squared error 14, squared error 15, squared error 16, squared error 17, squared error 18, squared error 19, squared error 19, squared error 10, squared error 10, squared error 10, squared error 10, squared error 11, squared error 11, squared error 12, squared error 13, squared error 14, squared error 15, squared error 16, squared error 16, squared error 17, squared error 18, squared error 19, squared err	th number of clus thouette method best to 1 are be allows for state = 2) er of cluster white', market for') cor', white', escore') thouette score merged, ax = axs[1], colo or 7050 Sample axs[1], colo or 7050 Sample dom_state=2) and and num_sads num_sads num or 1 or 2 for 1 or 2 for 1 or 2 for 3 or 4 or 3 or 4 or 5 or 6 dom_state=2) and and num_sads or 1 or 2 for 3 or 4 or 7	ters lead to the to help us sist and suggest reproducible rs that explai are dependent explai are dependen
In [54]: In [54]: In [56]: Out [55]: Out [57]:	best division of data/ We will plot the sum- determine the ideal in that the clusters are in schools are in the clusters are in schools are in the cluster are schools are are are schools are are are schools are are are schoo	distinct groups in the of squared errors and the outlines of squared errors of squar	ne dataset. Interest dataset.	i.e., use the elboore is.e., use the elboore silhouette score ers are indeed provided and the score is are indeed provided cluster in the court overfil pred cluster in the	w method) as we method) as we ranges from - resent) the best. w method) as we ranges from - resent) the best. er centroids ', n_init = teration. Fi tting. number labels) tte_avg[k-2] .73430147218 .60459182979 .2948608790 .40575486611 .48931081109 .48615034379 .2948608790 .4757328708 .3027844002 e'-', marker abel('Sum of marker='.', si abel('Average of the contours of marker='.', si abel('Average) of the contou	and record which and record which ell as use the si I to 1; values cla are placed 12, random_s 12, random_s 12, random_s 14, random_s 15, then do s 16, squared error 16, squared error 17, squared error 18, the squared error 19, ax = squared error 19, ax = squared error 10, ax = squared error 11, ax = squared error 12, random 13, squared error 14, squared error 15, then do s 16, squared error 16, squared error 17, squared error 18, squared error 19, ax = squared error 19, ax = squared error 10, squared error 11, squared error 12, random 13, squared error 14, squared error 15, squared error 16, squared error 17, squared error 18, squared error 19, squared error 19, squared error 10, squared error 10, squared error 11, squared error 11, squared error 12, squared error 13, squared error 14, squared error 15, squared error 16, squared error 17, squared error 18, squared error 19, squared error 19, squared error 10, squared error 10, squared error 10, squared error 11, squared error 12, squared error 13, squared error 14, squared error 15, squared error 16, squared error 17, squared error 18, squared error 19, squared error 19, squared error 19, squared error 10, squared error 10, squared error 11, squared error 12, squared error 13, squared error 14, squared error 15, squared error 16, squared error 16, squared error 17, squared error 18, squared error 19, squared error 19, squared error 19, squared error 10, squared error 10, squared error 10, squared error 11, squared error 11, squared error 12, squared error 13, squared error 14, squared error 15, squared error 16, squared error 17, squared error 18, squared error 19, squared error 19, squared error 10, squared error 10, squared error 10, squared error 10, squared error 11, squared error 11, squared error 12, squared error 13, squared error 14, squared error 15, squared error 16, squared error 16, squared error 17, squared error 18, squared error 19, squared err	th number of clus thouette method best to 1 are be allows for state = 2) er of cluster white', market for') cor', white', escore') thouette score merged, ax = axs[1], colo or 7050 Sample axs[1], colo or 7050 Sample dom_state=2) and and num_sads num_sads num or 1 or 2 for 1 or 2 for 1 or 2 for 3 or 4 or 3 or 4 or 5 or 6 dom_state=2) and and num_sads or 1 or 2 for 3 or 4 or 7	ters lead to the to help us set and suggest reproducible rest that explai are dispersion of the set of the
In [54]: In [54]: In [57]: In [57]: In [57]: In [57]:	best division of data/ We will plot the sum- determine the ideal in that the clusters are in that the clusters are in silhouette_arg; for k in range(; # random_st. Record sum sthe greatest # Record sum is the greatest # amount of ss_dist[k-2 # Get silho labels = k, silhouette_ print("For inclusters for	distinct groups in the of squared errors a sumber of clusters in fitting the outlines of an	me dataset. Inter each cluster (if or our dataset. Inter each cluster (if or our dataset. The first of the data (if cluster if the data (if cluster if the data (if cluster if the data), with the data, with ertia	i.e., use the elbo i.e., use the elbo i.e. also the solo e silhouette score ers are indeed pi where cluster i hoout overfi pted cluster (df_merged, is not over is not not over is not	w method) as we retained as we method) as we retained as we retained as we read the formation of the contours	and record which are placed are p	ch number of clus chouette method chouette method chouette state = 2) cr of cluster choice and cluster choic	enters lead to the to help us set and suggest reproducible rs that explai enters eredgewidth=1, markeredgewi and so see if ests. node) n_angrys year_(m o column o
<pre>In [46]: In [52]: In [54]: In [56]: Out [56]: Out [57]:</pre>	best division of data/ We will plot the sum- determine the ideal in that the clusters are in that the clusters are in silhouette_arg; for k in range(; # random_st. Record sum sthe greatest # Record sum is the greatest # amount of ss_dist[k-2 # Get silho labels = k, silhouette_ print("For inclusters for	distinct groups in the of squared errors a sumber of clusters in fitting the outlines of all the squared errors and in the	ne dataset. Inter each cluster (inter each cluster) Inter each cluster (inter each cluster) Inter our dataset. The first of the data (if cluster) Inter our dataset. The first out the data, with each attem,	i.e., use the elbo i.e.,	w method) as we retained as we method) as we retained as we retained as we read the formation of the contours	and record which are placed are p	ch number of clus chouette method chouette method chouette state = 2) cr of cluster choice and cluster choic	enters lead to the to help us set and suggest reproducible rs that explai are degewidth=1, markeredgewi enters are dog on the set of the
In [54]: In [56]: In [56]: In [57]: In [94]:	best division of data/ We will plot the sum- determine the ideal in that the clusters are in so dist = [np.n. silhouette_avg; for k in range() # random_st. k_means = kin k_means.fit # Record sun is the greatest # amount of ss_dist[k-2 # Get silho labels = k_, silhouette, print("For; print("For; print("For; print("For; print("The; For n_clusters: For n_clusters	distinct groups in the of squared errors a sumber of clusters in fitting the outlines of an interest in the control of squared errors and in	ne dataset. Inter each cluster (if or our dataset. In our data, with errors from each attem, with errors from each attem, whouette_score is esilhouette_score clusters'), an ouette_avg, conducte_avg, conducte_score clusters'), and our dataset. In our algorithm back is essentially a conducted in our dataset. In our	i.e., use the elbo i.e., use the elbo i.e. illower to establish out to score i.e. indeed property to the score is and to th	w method) as we retained as we method) as we retained as we retained as we read the formation of the contours	and record which all as use the sile to 1; values classes are placed are plac	h number of clus h number of clus houette method sest to 1 are be at at a = 2) ar of cluster at at = 2) ar of cluster at at = 2, ar of cluster at at = 3, and of cluster at at a = 3, and of cluster at ax = 1, ax =	ters lead to the to help us set and suggest reproducible rs that explai are degewidth=1, are degewidth=1, are markeredgewi a how each a how eac
In [54]: In [54]: In [57]: Out [57]: In [89]:	best division of data/ We will plot the sum- determine the ideal in that the clusters are in ss dist = [np.n. stihoue the angle; for kinanders, for kinanders, kmeans = kinanders, kmeans = kinanders, kmeans = kinanders, kmeans = kinanders, for still planters, for nclusters,	distinct groups in the of squared errors a sumber of clusters in the fitting the outlines of an angle in the control of squared errors a sumber of clusters in the squared errors are in the average err	ne dataset. Inter each cluster (if or our dataset. In our data, with errors from each attem, with errors from each attem, whouette_score is esilhouette_score clusters'), an ouette_avg, conducte_avg, conducte_score clusters'), and our dataset. In our algorithm back is essentially a conducted in our dataset. In our	i.e., use the elbo e silhouette score ers are indeed process are indeed process are indeed process. where cluster 'k-means++ ch cluster i hoout overfi pted cluster (df_merged, i ", silhouette score is: 0 sc	Juster numbers a w method) as w w method) as w w reanges from w ersent) the best. er centroids ', n_init = teration. Fi ting. number labels) tte_avg[k-2] .73430147218 .60459182979 .60715745610 .48931081109 .48615034379 .2948603290 .2733538708 .3027844002 e'.', marker abel('Sum of marker='.', n abel('Averag of the contours of markers and markers and markers e Plot of KMear and m_state amples, \n2 C random_state amples,	and record which are placed and record which are placed	ahas num_sads of clusters pecifically to see merged, ax = axs[1], colo or 7050 Sample merged, ax = axs[1], colo or 7050 Sample for 7050 Sample	enters lead to the to help us set and suggest reproducible rest that explai are that explai are degewidth=1, markeredgewi enters and suggest reproducible rest that explai are degewidth=1, are degewi
In [54]: In [54]: In [57]: Out [57]: In [89]:	best division of data// We will plot the sum- determine the ideal in that the clusters are in sa dist = [np.in- sin kinet and go in for in and go in for in and go in in sum in and go i	distinct groups in the of squared errors a number of clusters in fitting the outlines of an	me dataset. Inter each cluster (ifter each cluster) (ifter each each each each each each each each	i.e., use the elbo e silhouette score es are indeed process are indeed process are indeed process are indeed process. I hout overfi bred cluster (df_merged, :", silhouette score is: 0 score is:	wmethod) as we range from resent) the best. er centroids ', n_init = teration. Fi tting. number labels) tte_avg[k-2] .73430147218 .6045192979 .60453408107 .92946087907 .4573538708 .3027844002 e', marker abel ('Average of the contours of the co	and record which and record which and record which and the suse the si I to 1; values ch are placed are place	ahas num_sads of clusters of clusters of clusters as a 0.4 0.5 0.6 ar 7050 Sample or 7050 Sample	sters lead to the to help us set and suggest reproducible rest that explai eredgewidth=1, markeredgewi a how each eredgewidth=1, conderedgewidth=1, conderedg
In [54]: In [54]: In [57]: Out [57]: In [89]:	best division of data/ We will plot the sum- determine the ideal in that the clusters are solist = (np. in sin with a ready of the same o	distinct groups in the of squared errors as sumber of clusters iffitting the outlines of squared errors and in the state of the squared errors and in the squared errors and i	ne dataset. Inter each cluster (if or our dataset. Inter our data (if cluster our data) with the data, with the dat	i.e., use the elbo e silhouette score es ilhouete score ers are indeed processore indeed processore where cluster elbout overfi pted cluster (df_merged, :", silhouete score is: 0 score	wmethod) as we method) as we ranges from recent per centroids "a ranges from recent) the best. "a range from recent per centroids "a ra	and record which is a set of the	ch number of clus chouette method cosest to 1 are be consisted and session of clusters cor') consisted and session of clusters cor') consisted and session of clusters core and session of clu	atters lead to the to help us to help us to the lay us to
In [52]: In [54]: In [57]: Out [57]: In [79]:	best division of data/ We will plot the sum- determine the ideal in that the clusters are so dist = [np.in silhousters are so n clusters are so n clusters are so n clusters are for n clus	distinct groups in the of squared errors and	the dataset. Infer each cluster (if for our dataset. Infer each cluster. Infer each cluster. Infer our dataset. Infer each cluster. Infer each data (if cluster. Infer each attempt the data (if cluster. Infer each attempt each each attempt each each each each each each each each	i.e., use the elbo e silhouette score es ilhouete score ers are indeed processore indeed processore where cluster elbout overfi pted cluster (df_merged, :", silhouete score is: 0 score	wmethod) as we method) as we ranges from recent per centroids "a ranges from recent) the best. "a range from recent per centroids "a ra	and record which is a set of the	ch number of clus chouette method cosest to 1 are be consisted and session of clusters cor') consisted and session of clusters cor') consisted and session of clusters core and session of clu	atters lead to the to help us to help us to the lay us to
In [52]: In [54]: In [57]: Out [57]: In [79]:	best division of data/ We will plot the sum- devill plot the sum- devill plot the sum- devill plot the sum- determine the idea in that the clusters are in state and in state	distinct groups in the	ne dataset. Inter each cluster (inter each cluster) Inter each cluster (inter each cluster) Inter each cluster. Inter each cluster.	i.e., use the elbo e silhouette score is silhouette score ers are indeed processore indeed processore is consisted to cluster indeed processore is consisted to cluster indeed processore is consisted to consore is consore	we method) as we ranges from resent) the best. er centroids er cen	and record which are placed and in the number of each cluster in the sile of each clus	ahas num_sads of clusters specifically to see axs[1], color or 7050 Sample	a how each a how
In [52]: In [54]: In [57]: Out [57]: In [79]:	best division of data/ We will plot the sum- determine the idea in that the cluster are so distress and so di	default') subplots (1, 2, 1, 2, 1, 2, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,	ne dataset. Inter each cluster (inter each cluster) Inter each cluster (inter each cluster) Inter each cluster. Inter each each cluster. Inter each each cluster. Inter each each cluster. Inter each each cluster. Inter each each each each each each each each	i.e., use the elbo e silhouette soor es are indeed proposed and in	we method) as we ranges from resent) the best. Proceeding from resent) the best. Proceding from resent) the bes	and record which are placed as the sile to 1; values of the number of the number of the sile to 1; values of the number of the sile to 1; values of the number of the number of the sile to 1; values of the number	ahas num_sads of clusters specifically to see axs[1], color or 7050 Sample	num_angrys year_(magest) num_angrys year_(m
In [54]: In [54]: In [56]: In [57]: In [79]: In [79]:	best division of data/ We will plot the sund determine the ideal of the sund	default') subplots (1, 2, 1, 2, 1, 2, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,	me dataset. Inter each cluster (if or our dataset. Inter each cluster (if or our dataset. The for our dataset. Th	i.e., use the elbo e silhouette score es are indeed pr where cluster ers are indeed pr where cluster is are indeed pr where cluster is hoout overfi pted cluster if hoout overfi pted cluster if hoout overfi pted cluster is 10 score is 10 sco	wethod) as we ranges from we and store sent) the best. ar centroids ye ranges from resent) the best. ar centroids ye ranges from resent) the best. ar centroids ye ranges from resent) the best. number labels) tte_avg[k-2] 73430147218 604791827950 604918297950 604918297950 605918297950 606918297950 607918297950	and record which are placed as the sile to 1; values of the number of the number of the sile to 1; values of the number of the sile to 1; values of the number of the number of the sile to 1; values of the number	ahas num_sads of clusters specifically to see axs[1], color or 7050 Sample	num_angrys year_(mode) num_an