```
In [1]:
               import pandas as pd
               import numpy as np
               import matplotlib.pyplot as plt
               %matplotlib inline
            5
               import warnings
               warnings.filterwarnings('ignore')
In [2]:
               df1 = pd.read_csv("Documents/Reactions.csv",index_col = 0)
In [3]:
               df1
Out[3]:
                                       Content ID
                                                                           User ID
                                                                                        Type
                                                                                                    Datetime
                         97522e57-d9ab-4bd6-97bf-
                                                                                                  2021-04-22
               0
                                                                              NaN
                                                                                        NaN
                                    c24d952602d2
                                                                                                    15:17:15
                         97522e57-d9ab-4bd6-97bf-
                                                          5d454588-283d-459d-915d-
                                                                                                  2020-11-07
               1
                                                                                      disgust
                                    c24d952602d2
                                                                      c48a2cb4c27f
                                                                                                    09:43:50
                         97522e57-d9ab-4bd6-97bf-
                                                           92b87fa5-f271-43e0-af66-
                                                                                                  2021-06-17
               2
                                                                                       dislike
                                    c24d952602d2
                                                                      84fac21052e6
                                                                                                    12:22:51
                         97522e57-d9ab-4bd6-97bf-
                                                           163daa38-8b77-48c9-9af6-
                                                                                                  2021-04-18
               3
                                                                                      scared
                                    c24d952602d2
                                                                     37a6c1447ac2
                                                                                                    05:13:58
                                                          34e8add9-0206-47fd-a501-
                         97522e57-d9ab-4bd6-97bf-
                                                                                                  2021-01-06
               4
                                                                                      disgust
                                    c24d952602d2
                                                                     037b994650a2
                                                                                                    19:13:01
                                                                                                  2020-06-27
                         75d6b589-7fae-4a6d-b0d0-
                                                           80c9ce48-46f9-4f5e-b3ca-
           25548
                                                                                       dislike
                                    752845150e56
                                                                      3b698fc2e949
                                                                                                    09:46:48
                         75d6b589-7fae-4a6d-b0d0-
                                                          2bd9c167-e06c-47c1-a978-
                                                                                                  2021-02-16
           25549
                                                                                     intrigued
                                    752845150e56
                                                                     3403d6724606
                                                                                                    17:17:02
                         75d6b589-7fae-4a6d-b0d0-
                                                                                                  2020-09-12
           25550
                                                                              NaN
                                                                                   interested
                                    752845150e56
                                                                                                    03:54:58
                         75d6b589-7fae-4a6d-b0d0-
                                                           5ffd8b51-164e-47e2-885e-
                                                                                                  2020-11-04
           25551
                                                                                      worried
                                                                     8b8c46eb63ed
                                                                                                    20:08:31
                                    752845150e56
                         75d6b589-7fae-4a6d-b0d0-
                                                          4edc3d1a-a7d9-4db6-89c3-
                                                                                                  2021-01-04
           25552
                                                                                      cherish
                                    752845150e56
                                                                     f784d9954172
                                                                                                    04:55:11
          25553 rows × 4 columns
```

df2 = pd.read csv('Documents/ReactionTypes.csv',index col = 0)

In [4]:

In [5]: 1 df2

Out[5]:

	Туре	Sentiment	Score
0	heart	positive	60
1	want	positive	70
2	disgust	negative	0
3	hate	negative	5
4	interested	positive	30
5	indifferent	neutral	20
6	love	positive	65
7	super love	positive	75
8	cherish	positive	70
9	adore	positive	72
10	like	positive	50
11	dislike	negative	10
12	intrigued	positive	45
13	peeking	neutral	35
14	scared	negative	15
15	worried	negative	12

In [6]: 1 df3 = pd.read_csv('Documents/Content.csv',index_col=0)

In [7]: 1 df3

Out[7]:

	Content ID	User ID	Туре	Category	URL
0	97522e57- d9ab-4bd6- 97bf- c24d952602d2	8d3cd87d- 8a31-4935- 9a4f- b319bfe05f31	photo	Studying	https://socialbuzz.cdn.com/content/storage/975
1	9f737e0a- 3cdd-4d29- 9d24- 753f4e3be810	beb1f34e-7870- 46d6-9fc7- 2e12eb83ce43	photo	healthy eating	https://socialbuzz.cdn.com/content/storage/9f7
2	230c4e4d- 70c3-461d- b42c- ec09396efb3f	a5c65404- 5894-4b87- 82f2- d787cbee86b4	photo	healthy eating	https://socialbuzz.cdn.com/content/storage/230
3	356fff80-da4d- 4785-9f43- bc1261031dc6	9fb4ce88-fac1- 406c-8544- 1a899cee7aaf	photo	technology	https://socialbuzz.cdn.com/content/storage/356
4	01ab84dd- 6364-4236- abbb- 3f237db77180	e206e31b-5f85- 4964-b6ea- d7ee5324def1	video	food	https://socialbuzz.cdn.com/content/storage/01a
995	b4cef9ef-627b- 41d7-a051- 5961b0204ebb	5b62e10e- 3c19-4d28- a57c- e9bdc3d6758d	video	public speaking	NaN
996	7a79f4e4- 3b7d-44dc- bdef- bc990740252c	4fe420fa-a193- 4408-bd5d- 62a020233609	GIF	technology	https://socialbuzz.cdn.com/content/storage/7a7
997	435007a5- 6261-4d8b- b0a4- 55fdc189754b	35d6a1f3-e358- 4d4b-8074- 05f3b7f35c2a	audio	veganism	https://socialbuzz.cdn.com/content/storage/435
998	4e4c9690- c013-4ee7- 9e66- 943d8cbd27b7	b9bcd994-f000- 4f6b-87fc- caae08acfaa1	GIF	culture	https://socialbuzz.cdn.com/content/storage/4e4
999	75d6b589- 7fae-4a6d- b0d0- 752845150e56	b8c653b5- 0118-4d7e- 9bde- 07c2de90f0ff	audio	technology	https://socialbuzz.cdn.com/content/storage/75d

1000 rows × 5 columns

In [8]: 1 df1.shape #Reactions

Out[8]: (25553, 4)

```
In [9]: 1 df2.shape #Reactiontypes
Out[9]: (16, 3)
In [10]: 1 df3.shape #Content
Out[10]: (1000, 5)
In [11]: 1 df4 = df1.merge(df3,how = 'left',on =['Content ID'])
```

In [12]: 1 df4

Out[12]:

	Content ID	User ID_x	Type_x	Datetime	User ID_y	Type_y	Category	
0	97522e57- d9ab-4bd6- 97bf- c24d952602d2	NaN	NaN	2021-04- 22 15:17:15	8d3cd87d- 8a31-4935- 9a4f- b319bfe05f31	photo	Studying	http:
1	97522e57- d9ab-4bd6- 97bf- c24d952602d2	5d454588- 283d-459d- 915d- c48a2cb4c27f	disgust	2020-11- 07 09:43:50	8d3cd87d- 8a31-4935- 9a4f- b319bfe05f31	photo	Studying	http:
2	97522e57- d9ab-4bd6- 97bf- c24d952602d2	92b87fa5- f271-43e0- af66- 84fac21052e6	dislike	2021-06- 17 12:22:51	8d3cd87d- 8a31-4935- 9a4f- b319bfe05f31	photo	Studying	http:
3	97522e57- d9ab-4bd6- 97bf- c24d952602d2	163daa38- 8b77-48c9- 9af6- 37a6c1447ac2	scared	2021-04- 18 05:13:58	8d3cd87d- 8a31-4935- 9a4f- b319bfe05f31	photo	Studying	http:
4	97522e57- d9ab-4bd6- 97bf- c24d952602d2	34e8add9- 0206-47fd- a501- 037b994650a2	disgust	2021-01- 06 19:13:01	8d3cd87d- 8a31-4935- 9a4f- b319bfe05f31	photo	Studying	http:
		***			***			
25548	75d6b589- 7fae-4a6d- b0d0- 752845150e56	80c9ce48- 46f9-4f5e- b3ca- 3b698fc2e949	dislike	2020-06- 27 09:46:48	b8c653b5- 0118-4d7e- 9bde- 07c2de90f0ff	audio	technology	http:
25549	75d6b589- 7fae-4a6d- b0d0- 752845150e56	2bd9c167- e06c-47c1- a978- 3403d6724606	intrigued	2021-02- 16 17:17:02	b8c653b5- 0118-4d7e- 9bde- 07c2de90f0ff	audio	technology	http:
25550	75d6b589- 7fae-4a6d- b0d0- 752845150e56	NaN	interested	2020-09- 12 03:54:58	b8c653b5- 0118-4d7e- 9bde- 07c2de90f0ff	audio	technology	http:
25551	75d6b589- 7fae-4a6d- b0d0- 752845150e56	5ffd8b51- 164e-47e2- 885e- 8b8c46eb63ed	worried	2020-11- 04 20:08:31	b8c653b5- 0118-4d7e- 9bde- 07c2de90f0ff	audio	technology	http:
25552	75d6b589- 7fae-4a6d- b0d0- 752845150e56	4edc3d1a- a7d9-4db6- 89c3- f784d9954172	cherish	2021-01- 04 04:55:11	b8c653b5- 0118-4d7e- 9bde- 07c2de90f0ff	audio	technology	http:

```
In [13]: 1 df4.shape
```

Out[13]: (25553, 8)

```
In [14]: 1 df4.drop('User ID_x',axis = 1,inplace = True)
```

In [15]:	1 df4						
Out[15]:	Co	ontent ID	Type x	Datetime	User ID v 1	Tvpe v	Category

:		Content ID	Type_x	Datetime	User ID_y	Type_y	Category	
-	0	97522e57- d9ab-4bd6- 97bf- c24d952602d2	NaN	2021-04- 22 15:17:15	8d3cd87d- 8a31-4935- 9a4f- b319bfe05f31	photo	Studying	https://socialbuzz.cdn.con
	1	97522e57- d9ab-4bd6- 97bf- c24d952602d2	disgust	2020-11- 07 09:43:50	8d3cd87d- 8a31-4935- 9a4f- b319bfe05f31	photo	Studying	https://socialbuzz.cdn.con
	2	97522e57- d9ab-4bd6- 97bf- c24d952602d2	dislike	2021-06- 17 12:22:51	8d3cd87d- 8a31-4935- 9a4f- b319bfe05f31	photo	Studying	https://socialbuzz.cdn.con
	3	97522e57- d9ab-4bd6- 97bf- c24d952602d2	scared	2021-04- 18 05:13:58	8d3cd87d- 8a31-4935- 9a4f- b319bfe05f31	photo	Studying	https://socialbuzz.cdn.con
	4	97522e57- d9ab-4bd6- 97bf- c24d952602d2	disgust	2021-01- 06 19:13:01	8d3cd87d- 8a31-4935- 9a4f- b319bfe05f31	photo	Studying	https://socialbuzz.cdn.con
		•••						
	25548	75d6b589- 7fae-4a6d- b0d0- 752845150e56	dislike	2020-06- 27 09:46:48	b8c653b5- 0118-4d7e- 9bde- 07c2de90f0ff	audio	technology	https://socialbuzz.cdn.con
	25549	75d6b589- 7fae-4a6d- b0d0- 752845150e56	intrigued	2021-02- 16 17:17:02	b8c653b5- 0118-4d7e- 9bde- 07c2de90f0ff	audio	technology	https://socialbuzz.cdn.con
	25550	75d6b589- 7fae-4a6d- b0d0- 752845150e56	interested	2020-09- 12 03:54:58	b8c653b5- 0118-4d7e- 9bde- 07c2de90f0ff	audio	technology	https://socialbuzz.cdn.con
	25551	75d6b589- 7fae-4a6d- b0d0- 752845150e56	worried	2020-11- 04 20:08:31	b8c653b5- 0118-4d7e- 9bde- 07c2de90f0ff	audio	technology	https://socialbuzz.cdn.con
	25552	75d6b589- 7fae-4a6d- b0d0- 752845150e56	cherish	2021-01- 04 04:55:11	b8c653b5- 0118-4d7e- 9bde- 07c2de90f0ff	audio	technology	https://socialbuzz.cdn.con
		_						

25553 rows × 7 columns

In [16]: 1 df4.rename(columns = {'Type_x':'Reaction_Type','Type_y':'Content_type','User In [17]: 1 df4

Out[17]:

	Content ID	Reaction_Type	Datetime	User ID	Content_type	Category	
0	97522e57- d9ab-4bd6- 97bf- c24d952602d2	NaN	2021-04- 22 15:17:15	8d3cd87d- 8a31-4935- 9a4f- b319bfe05f31	photo	Studying	https://social
1	97522e57- d9ab-4bd6- 97bf- c24d952602d2	disgust	2020-11- 07 09:43:50	8d3cd87d- 8a31-4935- 9a4f- b319bfe05f31	photo	Studying	https://sociall
2	97522e57- d9ab-4bd6- 97bf- c24d952602d2	dislike	2021-06- 17 12:22:51	8d3cd87d- 8a31-4935- 9a4f- b319bfe05f31	photo	Studying	https://sociall
3	97522e57- d9ab-4bd6- 97bf- c24d952602d2	scared	2021-04- 18 05:13:58	8d3cd87d- 8a31-4935- 9a4f- b319bfe05f31	photo	Studying	https://sociall
4	97522e57- d9ab-4bd6- 97bf- c24d952602d2	disgust	2021-01- 06 19:13:01	8d3cd87d- 8a31-4935- 9a4f- b319bfe05f31	photo	Studying	https://sociall
25548	75d6b589- 7fae-4a6d- b0d0- 752845150e56	dislike	2020-06- 27 09:46:48	b8c653b5- 0118-4d7e- 9bde- 07c2de90f0ff	audio	technology	https://social
25549	75d6b589- 7fae-4a6d- b0d0- 752845150e56	intrigued	2021-02- 16 17:17:02	b8c653b5- 0118-4d7e- 9bde- 07c2de90f0ff	audio	technology	https://social
25550	75d6b589- 7fae-4a6d- b0d0- 752845150e56	interested	2020-09- 12 03:54:58	b8c653b5- 0118-4d7e- 9bde- 07c2de90f0ff	audio	technology	https://social
25551	75d6b589- 7fae-4a6d- b0d0- 752845150e56	worried	2020-11- 04 20:08:31	b8c653b5- 0118-4d7e- 9bde- 07c2de90f0ff	audio	technology	https://social
25552	75d6b589- 7fae-4a6d- b0d0- 752845150e56	cherish	2021-01- 04 04:55:11	b8c653b5- 0118-4d7e- 9bde- 07c2de90f0ff	audio	technology	https://sociall

25553 rows × 7 columns

```
In [18]: 1 df5 = df4.merge(df2,how = 'left',left_on='Reaction_Type',right_on='Type')
In [19]: 1 df5.drop('Type',inplace = True,axis = 1)
```

In [20]: 1 df5

Out[20]:		Content ID	Reaction_Type	Datetime	User ID	Content_type	Category	
	0	97522e57- d9ab-4bd6- 97bf- c24d952602d2	NaN	2021-04- 22 15:17:15	8d3cd87d- 8a31-4935- 9a4f- b319bfe05f31	photo	Studying	https://social
	1	97522e57- d9ab-4bd6- 97bf- c24d952602d2	disgust	2020-11- 07 09:43:50	8d3cd87d- 8a31-4935- 9a4f- b319bfe05f31	photo	Studying	https://social
	2	97522e57- d9ab-4bd6- 97bf- c24d952602d2	dislike	2021-06- 17 12:22:51	8d3cd87d- 8a31-4935- 9a4f- b319bfe05f31	photo	Studying	https://social
	3	97522e57- d9ab-4bd6- 97bf- c24d952602d2	scared	2021-04- 18 05:13:58	8d3cd87d- 8a31-4935- 9a4f- b319bfe05f31	photo	Studying	https://social
	4	97522e57- d9ab-4bd6- 97bf- c24d952602d2	disgust	2021-01- 06 19:13:01	8d3cd87d- 8a31-4935- 9a4f- b319bfe05f31	photo	Studying	https://social
	25548	75d6b589- 7fae-4a6d- b0d0- 752845150e56	dislike	2020-06- 27 09:46:48	b8c653b5- 0118-4d7e- 9bde- 07c2de90f0ff	audio	technology	https://social
	25549	75d6b589- 7fae-4a6d- b0d0- 752845150e56	intrigued	2021-02- 16 17:17:02	b8c653b5- 0118-4d7e- 9bde- 07c2de90f0ff	audio	technology	https://social
	25550	75d6b589- 7fae-4a6d- b0d0- 752845150e56	interested	2020-09- 12 03:54:58	b8c653b5- 0118-4d7e- 9bde- 07c2de90f0ff	audio	technology	https://social
	25551	75d6b589- 7fae-4a6d- b0d0- 752845150e56	worried	2020-11- 04 20:08:31	b8c653b5- 0118-4d7e- 9bde- 07c2de90f0ff	audio	technology	https://social
	25552	75d6b589- 7fae-4a6d- b0d0- 752845150e56	cherish	2021-01- 04 04:55:11	b8c653b5- 0118-4d7e- 9bde- 07c2de90f0ff	audio	technology	https://social
	25553 r	rows × 9 colum	ns					
	1	OWO O COIGITI						•
				_				
In [21]:	1 df	-5.drop('URL'	<pre>,inplace = T</pre>	rue,axis	=1)			
In [22]:	1 df	5.shape						
Out[22]:	(25553	, 8)						

In [23]: 1 finalised_dataset = df5.dropna(subset=['Score']) #dropping rows where values
In [24]: 1 finalised_dataset.shape

Out[24]: (24573, 8)

In [25]:	1	1 finalised_dataset						
Out[25]:	Content ID		Reaction_Type	Datetime	User ID	Content_type	Category	Sentiment
		97522e57- d9ab-4bd6- 97bf- c24d952602d2	disgust	2020-11- 07 09:43:50	8d3cd87d- 8a31-4935- 9a4f- b319bfe05f31	photo	Studying	negative
		97522e57- d9ab-4bd6- 97bf- c24d952602d2	dislike	2021-06- 17 12:22:51	8d3cd87d- 8a31-4935- 9a4f- b319bfe05f31	photo	Studying	negative
	3 d9ab-4bd6 97b c24d952602d 97522e57 d9ab-4bd6 97b	97522e57- d9ab-4bd6- 97bf- c24d952602d2	scared	2021-04- 18 05:13:58	8d3cd87d- 8a31-4935- 9a4f- b319bfe05f31	photo	Studying	negative
		4 designate disgust		2021-01- 06 19:13:01	8d3cd87d- 8a31-4935- 9a4f- b319bfe05f31	photo	Studying	negative
		97522e57- d9ab-4bd6- 97bf- c24d952602d2	interested	2020-08- 23 12:25:58	8d3cd87d- 8a31-4935- 9a4f- b319bfe05f31	photo	Studying	positive
								
	2554	75d6b589- 7fae-4a6d- b0d0- 752845150e56	dislike	2020-06- 27 09:46:48	b8c653b5- 0118-4d7e- 9bde- 07c2de90f0ff	audio	technology	negative
	25550 - - 25551	75d6b589- 7fae-4a6d- b0d0- 752845150e56	intrigued	2021-02- 16 17:17:02	16 0118-407e-		technology	positive
		75d6b589- 7fae-4a6d- b0d0- 752845150e56	interested	2020-09- 12 03:54:58	b8c653b5- 0118-4d7e- 9bde- 07c2de90f0ff	audio	technology	positive
		75d6b589- 7fae-4a6d- b0d0- 752845150e56	worried	2020-11- 04 20:08:31	b8c653b5- 0118-4d7e- 9bde- 07c2de90f0ff	audio	technology	negative
	255	75d6b589- 7fae-4a6d- b0d0- 752845150e56	cherish	2021-01- 04 04:55:11	b8c653b5- 0118-4d7e- 9bde- 07c2de90f0ff	audio	technology	positive

07c2de90f0ff

24573 rows × 8 columns

```
In [26]:
              1 finalised dataset['Category'].unique()
Out[26]: array(['Studying', 'healthy eating', 'technology', 'food', 'cooking',
                    'dogs', 'soccer', 'public speaking', 'science', 'tennis', 'travel',
                    'fitness', 'education', 'studying', 'veganism', 'Animals', 'animals', 'culture', '"culture"', 'Fitness', '"studying"',
                    'Veganism', '"animals"', 'Travel', '"soccer"', 'Education',
'"dogs"', 'Technology', 'Soccer', 'Culture', '"food"', 'Food',
'"technology"', 'Healthy Eating', '"cooking"', 'Science',
'"public speaking"', '"veganism"', 'Public Speaking', '"science"'],
                   dtype=object)
In [27]:
                finalised dataset['Category'] = finalised dataset['Category'].str.upper()
              1 finalised dataset['Category'].unique()
In [28]:
Out[28]: array(['STUDYING', 'HEALTHY EATING', 'TECHNOLOGY', 'FOOD', 'COOKING',
                    'DOGS', 'SOCCER', 'PUBLIC SPEAKING', 'SCIENCE', 'TENNIS', 'TRAVEL',
                    'FITNESS', 'EDUCATION', 'VEGANISM', 'ANIMALS', 'CULTURE',
                    '"CULTURE"', '"STUDYING"', '"ANIMALS"', '"SOCCER"', '"DOGS"', '"FOOD"', '"TECHNOLOGY"', '"COOKING"', '"PUBLIC SPEAKING"',
                    '"VEGANISM"', '"SCIENCE"'], dtype=object)
In [29]:
                finalised dataset['Category'] = finalised dataset['Category'].str.replace('"
              2 finalised dataset['Category'].unique()
Out[29]: array(['STUDYING', 'HEALTHY EATING', 'TECHNOLOGY', 'FOOD', 'COOKING',
                     'DOGS', 'SOCCER', 'PUBLIC SPEAKING', 'SCIENCE', 'TENNIS', 'TRAVEL',
                    'FITNESS', 'EDUCATION', 'VEGANISM', 'ANIMALS', 'CULTURE'],
                   dtype=object)
                finalised dataset['Category'].nunique()
In [30]:
Out[30]: 16
In [31]:
                 finalised dataset['Category'].value counts().head()
Out[31]: ANIMALS
                                  1897
           SCIENCE
                                  1796
           HEALTHY EATING
                                  1717
           FOOD
                                  1699
           TECHNOLOGY
                                  1698
           Name: Category, dtype: int64
```

In [32]: 1 finalised_dataset

Out[32]:

	Content ID	Reaction_Type	Datetime	User ID	Content_type	Category	Sentim
1	97522e57- d9ab-4bd6- 97bf- c24d952602d2	disgust	2020-11- 07 09:43:50	8d3cd87d- 8a31-4935- 9a4f- b319bfe05f31	photo	STUDYING	negat
2	97522e57- d9ab-4bd6- 97bf- c24d952602d2	dislike	2021-06- 17 12:22:51	8d3cd87d- 8a31-4935- 9a4f- b319bfe05f31	photo	STUDYING	negat
3	97522e57- d9ab-4bd6- 97bf- c24d952602d2	scared	2021-04- 18 05:13:58	8d3cd87d- 8a31-4935- 9a4f- b319bfe05f31	photo	STUDYING	negat
4	97522e57- d9ab-4bd6- 97bf- c24d952602d2	disgust	2021-01- 06 19:13:01	8d3cd87d- 8a31-4935- 9a4f- b319bfe05f31	photo	STUDYING	negat
5	97522e57- d9ab-4bd6- 97bf- c24d952602d2	interested	2020-08- 23 12:25:58	8d3cd87d- 8a31-4935- 9a4f- b319bfe05f31	photo	STUDYING	posit
25548	75d6b589- 7fae-4a6d- b0d0- 752845150e56	dislike	2020-06- 27 09:46:48	b8c653b5- 0118-4d7e- 9bde- 07c2de90f0ff	audio	TECHNOLOGY	negat
25549	75d6b589- 7fae-4a6d- b0d0- 752845150e56	intrigued	2021-02- 16 17:17:02	b8c653b5- 0118-4d7e- 9bde- 07c2de90f0ff	audio	TECHNOLOGY	posit
25550	75d6b589- 7fae-4a6d- b0d0- 752845150e56	interested	2020-09- 12 03:54:58	b8c653b5- 0118-4d7e- 9bde- 07c2de90f0ff	audio	TECHNOLOGY	posit
25551	75d6b589- 7fae-4a6d- b0d0- 752845150e56	worried	2020-11- 04 20:08:31	b8c653b5- 0118-4d7e- 9bde- 07c2de90f0ff	audio	TECHNOLOGY	negat
25552	75d6b589- 7fae-4a6d- b0d0- 752845150e56	cherish	2021-01- 04 04:55:11	b8c653b5- 0118-4d7e- 9bde- 07c2de90f0ff	audio	TECHNOLOGY	posit

24573 rows × 8 columns

In [33]: 1 finalised_dataset.to_csv('Finalised_dataset.csv')

```
In [34]:
           1 finalised_dataset.groupby('Category').sum().sort_values(by = 'Score',ascendi
Out[34]:
                           Score
                 Category
                 ANIMALS 74965.0
                 SCIENCE 71168.0
          HEALTHY EATING 69339.0
             TECHNOLOGY 68738.0
                   FOOD 66676.0
In [35]:
              case study = finalised dataset.groupby('Category').sum().sort values(by =
In [36]:
           1 type(case_study)
Out[36]: pandas.core.frame.DataFrame
In [37]:
           1 case_study.columns
Out[37]: Index(['Score'], dtype='object')
In [38]:
              case_study.reset_index(inplace= True)
In [39]:
              case_study
Out[39]:
                    Category
                              Score
                   ANIMALS 74965.0
                   SCIENCE 71168.0
             HEALTHY EATING 69339.0
          3
               TECHNOLOGY 68738.0
                      FOOD 66676.0
In [40]:
              case_study.columns
Out[40]: Index(['Category', 'Score'], dtype='object')
```

```
In [41]:
           1 category = case_study['Category']
             category
Out[41]: 0
                      ANIMALS
                      SCIENCE
         1
         2
              HEALTHY EATING
         3
                   TECHNOLOGY
         4
                         FOOD
         Name: Category, dtype: object
In [42]:
              case_study.to_csv('Top 5 Category.csv')
In [43]:
           1 x = list(category)
In [44]:
              score = case_study['Score']
In [45]:
              score
Out[45]: 0
              74965.0
         1
              71168.0
         2
              69339.0
              68738.0
         3
         4
              66676.0
         Name: Score, dtype: float64
           1 y = list(score)
In [46]:
In [48]:
           1 print(x)
           2 print(y)
          ['ANIMALS', 'SCIENCE', 'HEALTHY EATING', 'TECHNOLOGY', 'FOOD']
         [74965.0, 71168.0, 69339.0, 68738.0, 66676.0]
In [49]:
           1 xpos = np.arange(len(x))
           2 xpos
Out[49]: array([0, 1, 2, 3, 4])
```

```
In [50]: 1 plt.figure(figsize=(5,5))
2 plt.xlabel('Sum of Score in each category')
3 plt.ylabel("Category")
4 plt.title('Case study of top 5 categories')
5 plt.yticks(xpos,x)
6 plt.barh(xpos,y,height=0.5,color = 'green')
7 plt.savefig('Barchart.png',bbox_inches="tight", pad_inches=1, transparent=Tr plt.show()
```





