

Data Description

Research question

We are studying projects on the platform gofundme and the diverse factors that lead a project to be found. In our data set, we have extract the name, the date of the project, the description, the category, the pourcentage raised, the number of donors etc...

Here are the questions we want to answer :

- Does the category have an influence on the founding ?
- Does the pourcentage raised intuitively increase with time ?
- Do Keywords in the description make a project more attractive ?
- Does the donation per user influence the foundation ?

In [428...

```
#Useful imports
import time
time_duration = 0.5
import requests
from bs4 import BeautifulSoup

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt
```

In [604...

```
df=pd.read_csv("DataFinalfromWebScraping2.csv")
```

II/CLEANING DATA

II.1/We transform Creation Date to a datetime

For this we will tokenize the str sentence Creation Date, create a dictionary for month and transform Created 2 days ago in 2022-01-14 for example with the function Convertdate.

Then, we transform it in a proper DateTime

In [605...

```
from datetime import date

#We use nltk to tokenize the str of Creation Date
import nltk
nltk.download('punkt')
from nltk.tokenize import word_tokenize
```

```
[nltk_data] Downloading package punkt to
[nltk_data]      /Users/albandhauthuille/nltk_data...
[nltk_data] Package punkt is already up-to-date!
```

In [606...

```
def monthToNum(shortMonth):
    return {
        'January': 1,
        'February': 2,
        'March': 3,
        'April': 4,
        'May': 5,
        'June': 6,
        'July': 7,
        'August': 8,
        'September': 9,
        'October': 10,
        'November': 11,
        'December': 12
    }[shortMonth]
```

In [607...

```
def Convertdate():
    new_Date=[]
    #print(new_Date)
    for i in range (df['Creation Date'].size):
        #print(i)
        Creation_date = word_tokenize(df['Creation Date'][i])
        #print(Creation_date)

        if 'now' in Creation_date or 'hours' in Creation_date:
            new_Date.append(str(str(2022)+'-'+str('01')+'-'+str(24)))
        elif 'ago' not in Creation_date:
            #print('no')
            new_Date.append(str(Creation_date[4])+'-'+str(monthToNum(Creation

        elif 'ago' in Creation_date:
            new_Date.append(str(str(2022)+'-'+str('01')+'-'+str(24-int(Creation_da
    return new_Date
```

In [608...

```
df['Creation Date']=Convertdate()
```

In [609...

```
pd.to_datetime(df['Creation Date'])
df['Creation Date']=pd.to_datetime(df['Creation Date'])
```

II.2/ Retiring the '%' of Pourcentage Raised to plot properly

In [610...

```
pd.options.mode.chained_assignment = None # default='warn'
for i in range(len(df['Pourcentage Raised'])):
    df['Pourcentage Raised'][i]=df['Pourcentage Raised'][i][::-1]
```

In [611...

```
#For example:
df['Pourcentage Raised'][4]
```

```
Out[611... '1.0'
```

II.3/ Convert to proper type

```
In [624... df=df.convert_dtypes()
```

```
In [625... df['Pourcentage Raised']=df['Pourcentage Raised'].astype(float)
df
```

```
Out[625...
```

	Unnamed: 0	title	Categorie	collect	town	Pourcentage Raised	Creation Date
0	0	aider arta se soigner du cancer	Medical, Illness & Healing	€1,457 raised of €50,000	Annecy	2.9	2022-01-21
1	1	soutien ffs pour simone	Medical, Illness & Healing	€1,952 raised of €1,900	Aubervilliers	100.0	2022-01-08
2	2	help a mom in coma wake up again	Medical, Illness & Healing	€6,065 raised of €1	Metz	100.0	2022-01-11
3	3	aidez gladys pour un ultime traitement	Medical, Illness & Healing	€5,207 raised of €100,000	Héricy	5.2	2022-01-14
4	4	aidez donner naima une autre chance de vie	Medical, Illness & Healing	€630 raised of €110,000	Marseille	1.0	2022-01-14
...
1723	1723	mvfdp4 i just wanna be a millionaire	Dreams, Hopes & Wishes	€0 raised of €1.0M	Nice	1.0	2022-01-14
1724	1724	bring a smile to the unfortunate	Dreams, Hopes & Wishes	€0 raised of €100,000	Sainte- Rose, La Réunion	1.0	2022-01-14
1725	1725	testv55	Dreams, Hopes & Wishes	€0 raised of €200	Sully-sur- Loire	1.0	2022-01-14
1726	1726	weekend final	Dreams, Hopes & Wishes	€0 raised of €300	Poitiers	1.0	2022-01-13

	Unnamed: 0	title	Categorie	collect	town	Pourcentage Raised	Creation Date
1727	1727	acheter la doudoune de ses rves lola	Dreams, Hopes & Wishes	€0 raised of €475	Lille	1.0	2022-01-13

II.4/ Duration of a collect

In [626...

```

Duration =[]
for i in range(len(df['Creation Date'])):
    Duration.append((pd.Timestamp.today()-df['Creation Date'][i]).days)
df["Duration in days"]=Duration

```

II.5/We only study the keywords, so we reduce the 'Description' to a 'Short_description' with with words of lenght>4

In [627...

```

Short_description=[]
for i in range(len(df['Description'])):
    list=[]
    a=word_tokenize(df['Description'][i])
    for word in a:
        if len(word)>4:
            list.append(word)
    Short_description.append(list)

df['Short_description']=Short_description
df

```

Out[627...

	Unnamed: 0	title	Categorie	collect	town	Pourcentage Raised	Creation Date
	0	aider arta se soigner du cancer	Medical, Illness & Healing	€1,457 raised of €50,000	Annecy	2.9	2022-01-21
	1	soutien ffs pour simone	Medical, Illness & Healing	€1,952 raised of €1,900	Aubervilliers	100.0	2022-01-08
	2	help a mom in coma wake up again	Medical, Illness & Healing	€6,065 raised of €1	Metz	100.0	2022-01-11
	3	aidez gladys pour un ultime traitement	Medical, Illness & Healing	€5,207 raised of €100,000	Héricy	5.2	2022-01-14
	4	aidez donner naima une autre chance de	Medical, Illness & Healing	€630 raised of €110,000	Marseille	1.0	2022-01-14

	Unnamed: 0	title	Categorie	collect	town	Pourcentage Raised	Creation Date
		vie					
...
1723	1723	mvfdp4 i just wanna be a millionaire	Dreams, Hopes & Wishes	€0 raised of €1.0M	Nice	1.0	2022-01-14
1724	1724	bring a smile to the unfortunate	Dreams, Hopes & Wishes	€0 raised of €100,000	Sainte-Rose, La Réunion	1.0	2022-01-14
1725	1725	testv55	Dreams, Hopes & Wishes	€0 raised of €200	Sully-sur-Loire	1.0	2022-01-14
1726	1726	weekend final	Dreams, Hopes & Wishes	€0 raised of €300	Poitiers	1.0	2022-01-13
1727	1727	acheter la doudoune de ses rves lola	Dreams, Hopes & Wishes	€0 raised of €475	Lille	1.0	2022-01-13

II.6.1/ Creation of column 'Amount Collected'

In [638...

```

Amount_Collected=[]
#print(df)
for i in range(len(df['collect'])):
    #word=(df['collect'][i])[1:]
    word=word_tokenize(df['collect'][i][1:])
    #If the amount is about millions:

    if 'M' in word[0]:
        Amount_Collected.append(float(word[0].replace(',','.'))[:-1])*1000000
    #print(word)
    else:
        if ',' in word[0]:
            Amount_Collected.append(float(word[0].replace(',','.'))*1000)
        else:
            Amount_Collected.append(float(word[0].replace(',','.')))

df['Amount collected ']=Amount_Collected

```

II.6.2/ Creation of Column 'Amount targeted'

In [639...

```

Amount_Targeted=[]
for i in range(len(df['collect'])):
    word=word_tokenize(df['collect'][i])
    #print(word)
    #If the amount is about millions:
    if 'M' in word[-1]:
        Amount_Targeted.append(float(((word[-1].replace('M', '')).replace('€', ''))))
    #print(word)
    elif 'B' in word[-1]:
        Amount_Targeted.append(float(((word[-1].replace('B', '')).replace('€', ''))))
    else:
        if ',' in word[-1]:
            Amount_Targeted.append(float(((word[-1].replace(',', '.')).replace('€', ''))))
        else:
            Amount_Targeted.append(float(((word[-1].replace(',', '.')).replace('€', ''))))

#len(Amount_Targeted)
df['Amount targeted']=Amount_Targeted

```

II.7/ MeanDonation

In [657...

```

MeanDonation =[]
for i in range(len(df['NumberDonors'])):
    if df['NumberDonors'][i]>0:
        MeanDonation.append((df['Amount collected '][i])/(df['NumberDonors'][i]))
    else : MeanDonation.append(0)
df['MeanDonation']=MeanDonation

```

II.8/ Categories and conversion to numeric to plot and corr properly

In [658...

```

catégorie = ["Medical, Illness & Healing", "Funerals & Memorials", "Accidents"]
#print(catégorie)

```

In [659...

```

df['categoryNumeric'] = pd.factorize(df['Categorie'])[0] + 1
df['townNumeric'] = pd.factorize(df['town'])[0] + 1

```

In [660...

```
df.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1728 entries, 0 to 1727
Data columns (total 16 columns):
 #   Column                Non-Null Count  Dtype
---  -
 0   Unnamed: 0            1728 non-null   Int64
 1   title                 1728 non-null   string
 2   Categorie             1728 non-null   string
 3   collect               1728 non-null   string
 4   town                 1728 non-null   string
 5   Pourcentage Raised    1728 non-null   float64
 6   Creation Date         1728 non-null   datetime64[ns]
 7   Description           1728 non-null   string
 8   NumberDonors          1728 non-null   Int64

```

```

9   Duration in days      1728 non-null    int64
10  Short_description     1728 non-null    object
11  Amount collected      1728 non-null    float64
12  Amount targeted      1728 non-null    float64
13  MeanDonation          1728 non-null    float64
14  categoryNumeric       1728 non-null    int64
15  townNumeric           1728 non-null    int64
dtypes: Int64(2), datetime64[ns](1), float64(4), int64(3), object(1), string
(5)
memory usage: 219.5+ KB

```

III/ First Analysis

III.1. First look

In [661]

```
df.describe()
```

Out[661]

	Unnamed: 0	Pourcentage Raised	NumberDonors	Duration in days	Amount collected	Amount targeted
count	1728.000000	1728.000000	1728.000000	1728.000000	1.728000e+03	1.728000e+03
mean	863.500000	22.10787	85.384259	55.33044	7.450310e+03	3.493280e+03
std	498.974949	32.37733	1174.531989	69.87727	9.808869e+04	5.711015e+04
min	0.000000	1.000000	0.000000	0.000000	0.000000e+00	1.000000e+00
25%	431.750000	1.000000	0.000000	11.000000	0.000000e+00	1.000000e+00
50%	863.500000	2.950000	3.000000	36.000000	9.000000e+01	3.000000e+01
75%	1295.250000	30.450000	16.000000	67.000000	6.887500e+02	1.000000e+02
max	1727.000000	100.000000	44392.000000	365.000000	3.300000e+06	1.000000e+06

In [693]

```
df.corr()
```

Out[693]

	Unnamed: 0	Pourcentage Raised	NumberDonors	Duration in days	Amount collected	Amount targeted
Unnamed: 0	1.000000	-0.236096	-0.045198	0.053928	-0.057343	0.075547
Pourcentage Raised	-0.236096	1.000000	0.112374	0.022336	0.117099	-0.039527
NumberDonors	-0.045198	0.112374	1.000000	0.008824	0.922988	-0.002119
Duration in days	0.053928	0.022336	0.008824	1.000000	0.021402	-0.025243
Amount collected	-0.057343	0.117099	0.922988	0.021402	1.000000	-0.002209
Amount targeted	0.075547	-0.039527	-0.002119	-0.025243	-0.002209	1.000000
MeanDonation	-0.126093	0.247021	0.033937	0.082175	0.121436	-0.020222
categoryNumeric	0.998385	-0.227938	-0.050347	0.044506	-0.059323	0.075311
townNumeric	0.448352	-0.091195	0.008070	-0.053951	0.000662	0.008321
ScoreWord	-0.151570	0.058123	0.008432	-0.014183	0.017715	-0.026141

In [696...

```

for e in categorie:
    print(e)
    print('\n')
    print(df.loc[df['Categorie']==e].describe())
    print('\n')
    print('\n')

```

Medical, Illness & Healing

	Unnamed: 0	Pourcentage Raised	NumberDonors	Duration in days	\
count	97.000000	97.000000	97.000000	97.000000	
mean	50.072165	26.856701	182.432990	39.474227	
std	37.545597	30.696284	1516.492457	33.596284	
min	0.000000	1.000000	1.000000	0.000000	
25%	24.000000	3.800000	8.000000	13.000000	
50%	48.000000	13.200000	15.000000	39.000000	
75%	72.000000	41.000000	29.000000	54.000000	
max	297.000000	100.000000	14960.000000	252.000000	

	Amount collected	Amount targeted	MeanDonation	categoryNumeric	\
count	97.000000	9.700000e+01	97.000000	97.0	
mean	9054.515464	4.442945e+04	54.148241	1.0	
std	73222.610124	3.041569e+05	58.439691	0.0	
min	10.000000	1.000000e+00	7.500000	1.0	
25%	235.000000	2.000000e+03	21.333333	1.0	
50%	610.000000	5.000000e+03	33.942308	1.0	
75%	1575.000000	1.000000e+04	56.419355	1.0	
max	722188.000000	3.000000e+06	357.500000	1.0	

	townNumeric	ScoreWord
count	97.000000	97.000000
mean	35.762887	1.680412
std	27.740529	2.049118
min	1.000000	0.000000
25%	19.000000	0.000000
50%	31.000000	1.000000
75%	51.000000	2.000000
max	219.000000	11.000000

Funerals & Memorials

	Unnamed: 0	Pourcentage Raised	NumberDonors	Duration in days	\
count	96.000000	96.000000	96.000000	96.000000	
mean	143.500000	40.238542	163.791667	34.145833	
std	27.856777	41.376671	366.811924	34.713255	
min	96.000000	1.000000	0.000000	0.000000	
25%	119.750000	1.000000	0.000000	4.000000	
50%	143.500000	22.550000	8.500000	27.000000	
75%	167.250000	85.950000	154.500000	52.000000	
max	191.000000	100.000000	2698.000000	166.000000	

	Amount collected	Amount targeted	MeanDonation	categoryNumeric	\
--	------------------	-----------------	--------------	-----------------	---

count	96.000000	96.000000	96.000000	96.0
mean	10928.322917	14275.510417	45.504453	2.0
std	26972.548926	28918.311402	55.069801	0.0
min	0.000000	1.000000	0.000000	2.0
25%	0.000000	2000.000000	0.000000	2.0
50%	417.500000	4125.000000	38.133583	2.0
75%	7827.250000	11250.000000	64.375000	2.0
max	191220.000000	200000.000000	314.000000	2.0

	townNumeric	ScoreWord
count	96.000000	96.000000
mean	97.947917	1.125000
std	40.770312	1.502629
min	2.000000	0.000000
25%	81.750000	0.000000
50%	105.500000	1.000000
75%	129.250000	1.250000
max	152.000000	9.000000

Accidents & Emergencies

	Unnamed: 0	Pourcentage Raised	NumberDonors	Duration in days	\
count	96.000000	96.000000	96.000000	96.000000	
mean	239.500000	36.888542	28.833333	32.239583	
std	27.856777	32.991804	42.675191	22.523319	
min	192.000000	1.000000	1.000000	1.000000	
25%	215.750000	9.900000	6.000000	15.750000	
50%	239.500000	26.100000	18.000000	29.500000	
75%	263.250000	65.025000	33.000000	39.500000	
max	287.000000	100.000000	327.000000	90.000000	

	Amount collected	Amount targeted	MeanDonation	categoryNumeric	\
count	96.000000	9.600000e+01	96.000000	96.0	
mean	1869.843750	5.465234e+04	55.471360	3.0	
std	4303.372262	4.586876e+05	51.524959	0.0	
min	10.000000	5.000000e+01	5.000000	3.0	
25%	308.750000	1.475000e+03	29.411765	3.0	
50%	680.000000	3.000000e+03	42.250000	3.0	
75%	1982.500000	8.500000e+03	66.128788	3.0	
max	27990.000000	4.500000e+06	420.000000	3.0	

	townNumeric	ScoreWord
count	96.000000	96.000000
mean	125.177083	1.416667
std	74.659225	2.115938
min	3.000000	0.000000
25%	42.250000	0.000000
50%	163.500000	1.000000
75%	185.250000	2.000000
max	209.000000	14.000000

Non-Profits & Charities

	Unnamed: 0	Pourcentage Raised	NumberDonors	Duration in days	\
count	95.000000	95.000000	95.000000	95.000000	
mean	335.905263	61.660000	455.189474	28.410526	
std	27.718599	34.48446	1196.307051	36.895625	
min	288.000000	2.900000	22.000000	0.000000	
25%	312.500000	31.050000	48.000000	6.000000	
50%	336.000000	65.800000	89.000000	14.000000	
75%	359.500000	100.000000	217.000000	37.000000	
max	383.000000	100.000000	7010.000000	190.000000	

	Amount collected	Amount targeted	MeanDonation	categoryNumeric	\
count	9.500000e+01	9.500000e+01	95.000000	95.0	
mean	6.804401e+04	1.220816e+05	133.068633	4.0	
std	2.249932e+05	3.788881e+05	301.858306	0.0	
min	6.040000e+02	5.000000e+02	12.384615	4.0	
25%	3.659000e+03	5.000000e+03	49.719904	4.0	
50%	7.535000e+03	2.000000e+04	80.308219	4.0	
75%	1.982950e+04	6.000000e+04	130.805060	4.0	
max	1.600000e+06	3.000000e+06	2950.962963	4.0	

	townNumeric	ScoreWord
count	95.000000	95.000000
mean	250.263158	2.747368
std	30.097434	3.473206
min	148.000000	0.000000
25%	227.500000	1.000000
50%	251.000000	2.000000
75%	274.500000	3.000000
max	298.000000	22.000000

Education & Learning

	Unnamed: 0	Pourcentage Raised	NumberDonors	Duration in days	\
count	96.000000	96.000000	96.000000	96.000000	
mean	431.500000	16.614583	8.760417	39.031250	
std	27.856777	28.755503	15.814547	28.720231	
min	384.000000	1.000000	0.000000	0.000000	
25%	407.750000	1.000000	1.000000	8.000000	
50%	431.500000	2.250000	2.500000	41.500000	
75%	455.250000	14.675000	9.250000	63.000000	
max	479.000000	100.000000	93.000000	89.000000	

	Amount collected	Amount targeted	MeanDonation	categoryNumeric	\
count	96.000000	96.000000	96.000000	96.0	
mean	477.697917	14236.572917	54.407731	5.0	
std	953.332715	54739.571287	174.138079	0.0	
min	0.000000	1.000000	0.000000	5.0	
25%	18.750000	1500.000000	10.000000	5.0	
50%	89.500000	3885.000000	24.000000	5.0	
75%	486.750000	8725.000000	45.684524	5.0	
max	5033.000000	500000.000000	1677.666667	5.0	

	townNumeric	ScoreWord
count	96.000000	96.000000
mean	192.260417	2.427083

std	143.004545	4.062006
min	5.000000	0.000000
25%	22.750000	0.000000
50%	301.500000	1.000000
75%	323.250000	2.000000
max	347.000000	19.000000

Animals & Pets

	Unnamed: 0	Pourcentage Raised	NumberDonors	Duration in days	\
count	96.000000	96.000000	96.000000	96.000000	
mean	527.500000	32.623958	14.062500	40.927083	
std	27.856777	32.127554	16.498046	27.135234	
min	480.000000	1.000000	1.000000	0.000000	
25%	503.750000	7.075000	5.000000	17.750000	
50%	527.500000	22.400000	7.500000	40.500000	
75%	551.250000	48.350000	16.000000	63.250000	
max	575.000000	100.000000	86.000000	89.000000	

	Amount collected	Amount targeted	MeanDonation	categoryNumeric	\
count	96.000000	96.000000	96.000000	96.0	
mean	364.114583	1764.927083	27.361165	6.0	
std	400.721609	1677.284034	14.756126	0.0	
min	5.000000	200.000000	5.000000	6.0	
25%	113.750000	696.750000	17.458333	6.0	
50%	220.000000	1260.000000	24.722222	6.0	
75%	445.000000	2125.000000	32.337912	6.0	
max	2425.000000	11000.000000	78.000000	6.0	

	townNumeric	ScoreWord
count	96.000000	96.000000
mean	308.270833	2.552083
std	133.972934	3.361922
min	5.000000	0.000000
25%	324.500000	0.000000
50%	367.500000	2.000000
75%	391.250000	3.250000
max	415.000000	18.000000

Environment

	Unnamed: 0	Pourcentage Raised	NumberDonors	Duration in days	\
count	96.000000	96.000000	96.000000	96.000000	
mean	623.500000	10.713542	17.656250	127.802083	
std	27.856777	21.816602	96.313723	106.910451	
min	576.000000	1.000000	0.000000	2.000000	
25%	599.750000	1.000000	0.000000	37.500000	
50%	623.500000	1.000000	1.000000	83.000000	
75%	647.250000	10.200000	10.000000	215.000000	
max	671.000000	100.000000	937.000000	363.000000	

	Amount collected	Amount targeted	MeanDonation	categoryNumeric	\
--	------------------	-----------------	--------------	-----------------	---

count	96.0000	96.000000	96.000000	96.0
mean	919.2500	47903.802083	28.388342	7.0
std	4529.7205	151224.936173	41.209149	0.0
min	0.0000	10.000000	0.000000	7.0
25%	0.0000	2000.000000	0.000000	7.0
50%	10.0000	10000.000000	10.000000	7.0
75%	517.5000	25000.000000	41.132812	7.0
max	43262.0000	1000000.000000	215.666667	7.0

	townNumeric	ScoreWord
count	96.000000	96.000000
mean	256.625000	1.375000
std	193.630427	2.699903
min	8.000000	0.000000
25%	29.500000	0.000000
50%	371.000000	0.000000
75%	437.250000	1.250000
max	461.000000	12.000000

Business & Entrepreneurs

	Unnamed: 0	Pourcentage Raised	NumberDonors	Duration in days	\
count	96.000000	96.000000	96.000000	96.000000	
mean	719.500000	6.917708	2.593750	24.760417	
std	27.856777	18.888632	4.359088	25.536659	
min	672.000000	1.000000	0.000000	0.000000	
25%	695.750000	1.000000	0.000000	4.000000	
50%	719.500000	1.000000	1.000000	10.000000	
75%	743.250000	1.125000	2.000000	45.250000	
max	767.000000	100.000000	18.000000	90.000000	

	Amount collected	Amount targeted	MeanDonation	categoryNumeric	\
count	96.000000	96.000000	96.000000	96.0	
mean	117.739583	63338.239583	19.363348	8.0	
std	248.157982	195077.633370	26.499041	0.0	
min	0.000000	1.000000	0.000000	8.0	
25%	0.000000	2000.000000	0.000000	8.0	
50%	6.500000	8000.000000	5.000000	8.0	
75%	75.000000	25000.000000	30.000000	8.0	
max	1197.000000	1000000.000000	133.000000	8.0	

	townNumeric	ScoreWord
count	96.000000	96.000000
mean	290.427083	1.020833
std	210.197388	2.200379
min	3.000000	0.000000
25%	39.000000	0.000000
50%	406.500000	0.000000
75%	483.250000	1.000000
max	507.000000	12.000000

Community & Neighbors

	Unnamed: 0	Pourcentage Raised	NumberDonors	Duration in days	\
count	96.000000	96.000000	96.000000	96.000000	
mean	815.500000	23.842708	588.281250	91.687500	
std	27.856777	35.317575	4549.443047	101.224951	
min	768.000000	1.000000	0.000000	1.000000	
25%	791.750000	1.000000	0.000000	26.000000	
50%	815.500000	1.900000	2.500000	53.500000	
75%	839.250000	31.650000	20.750000	95.750000	
max	863.000000	100.000000	44392.000000	361.000000	

	Amount collected	Amount targeted	MeanDonation	categoryNumeric	\
count	9.600000e+01	9.600000e+01	96.000000	96.0	
mean	3.790868e+04	6.293976e+04	30.000488	9.0	
std	3.366242e+05	4.188204e+05	44.578314	0.0	
min	0.000000e+00	5.000000e+00	0.000000	9.0	
25%	0.000000e+00	1.000000e+03	0.000000	9.0	
50%	6.750000e+01	4.250000e+03	17.045940	9.0	
75%	9.737500e+02	1.500000e+04	34.218421	9.0	
max	3.300000e+06	4.000000e+06	222.500000	9.0	

	townNumeric	ScoreWord
count	96.000000	96.000000
mean	306.291667	0.947917
std	232.267990	1.970144
min	1.000000	0.000000
25%	31.000000	0.000000
50%	401.000000	0.000000
75%	528.000000	1.000000
max	551.000000	9.000000

Competitions & Pageants

	Unnamed: 0	Pourcentage Raised	NumberDonors	Duration in days	\
count	96.000000	96.000000	96.000000	96.000000	
mean	911.500000	14.933333	5.958333	142.156250	
std	27.856777	29.067984	14.113761	104.221356	
min	864.000000	1.000000	0.000000	1.000000	
25%	887.750000	1.000000	0.000000	50.750000	
50%	911.500000	1.000000	0.000000	133.500000	
75%	935.250000	6.575000	4.250000	220.500000	
max	959.000000	100.000000	80.000000	365.000000	

	Amount collected	Amount targeted	MeanDonation	categoryNumeric	\
count	96.000000	9.600000e+01	96.000000	96.0	
mean	291.697917	1.857182e+05	19.373279	10.0	
std	790.342779	1.176324e+06	43.395885	0.0	
min	0.000000	1.000000e+00	0.000000	10.0	
25%	0.000000	4.000000e+02	0.000000	10.0	
50%	0.000000	1.500000e+03	0.000000	10.0	
75%	120.000000	5.000000e+03	21.267857	10.0	
max	5100.000000	1.000000e+07	316.666667	10.0	

	townNumeric	ScoreWord
count	96.000000	96.000000
mean	347.677083	0.864583

std	257.290599	1.658279
min	5.000000	0.000000
25%	23.000000	0.000000
50%	541.000000	0.000000
75%	574.250000	1.000000
max	598.000000	9.000000

Creative Arts, Music & Film

	Unnamed: 0	Pourcentage Raised	NumberDonors	Duration in days	\
count	96.000000	96.000000	96.000000	96.000000	
mean	1007.500000	13.879167	5.416667	31.125000	
std	27.856777	23.658050	7.849461	26.437712	
min	960.000000	1.000000	0.000000	0.000000	
25%	983.750000	1.000000	0.000000	6.000000	
50%	1007.500000	1.500000	1.000000	27.000000	
75%	1031.250000	14.550000	7.000000	53.250000	
max	1055.000000	100.000000	40.000000	89.000000	

	Amount collected	Amount targeted	MeanDonation	categoryNumeric	\
count	96.000000	96.000000	96.000000	96.0	
mean	237.947917	4668.645833	25.451557	11.0	
std	421.217918	9058.767328	31.252427	0.0	
min	0.000000	10.000000	0.000000	11.0	
25%	0.000000	1000.000000	0.000000	11.0	
50%	33.500000	2000.000000	19.583333	11.0	
75%	286.250000	5000.000000	38.107143	11.0	
max	2549.000000	80000.000000	175.000000	11.0	

	townNumeric	ScoreWord
count	96.000000	96.000000
mean	291.927083	0.68750
std	273.952214	1.71257
min	1.000000	0.00000
25%	22.750000	0.00000
50%	172.500000	0.00000
75%	610.250000	0.25000
max	633.000000	11.00000

Celebrations & Events

	Unnamed: 0	Pourcentage Raised	NumberDonors	Duration in days	\
count	96.000000	96.000000	96.000000	96.000000	
mean	1103.500000	15.157292	2.916667	28.666667	
std	27.856777	29.147529	5.654373	19.967693	
min	1056.000000	1.000000	0.000000	0.000000	
25%	1079.750000	1.000000	0.000000	13.500000	
50%	1103.500000	1.000000	0.000000	25.000000	
75%	1127.250000	10.000000	3.000000	41.250000	
max	1151.000000	100.000000	34.000000	85.000000	

	Amount collected	Amount targeted	MeanDonation	categoryNumeric	\
--	------------------	-----------------	--------------	-----------------	---

count	96.000000	9.600000e+01	96.000000	96.0
mean	58.572917	8.360915e+06	12.385275	12.0
std	102.649198	8.164694e+07	21.584858	0.0
min	0.000000	4.000000e+00	0.000000	12.0
25%	0.000000	2.000000e+02	0.000000	12.0
50%	0.000000	6.500000e+02	0.000000	12.0
75%	65.250000	3.625000e+03	20.000000	12.0
max	475.000000	8.000000e+08	124.500000	12.0

	townNumeric	ScoreWord
count	96.000000	96.000000
mean	355.718750	0.333333
std	285.347386	1.193550
min	1.000000	0.000000
25%	29.500000	0.000000
50%	388.000000	0.000000
75%	647.250000	0.000000
max	670.000000	9.000000

Missions, Faith & Church

	Unnamed: 0	Pourcentage Raised	NumberDonors	Duration in days \
count	96.000000	96.000000	96.000000	96.000000
mean	1199.500000	15.747917	17.625000	112.322917
std	27.856777	27.105016	52.170419	103.284634
min	1152.000000	1.000000	0.000000	1.000000
25%	1175.750000	1.000000	0.000000	27.750000
50%	1199.500000	1.000000	1.000000	78.500000
75%	1223.250000	18.300000	17.500000	171.250000
max	1247.000000	100.000000	446.000000	365.000000

	Amount collected	Amount targeted	MeanDonation	categoryNumeric \
count	96.000000	9.600000e+01	96.000000	96.0
mean	1582.666667	1.046996e+07	60.075620	13.0
std	3353.647060	1.020569e+08	145.503516	0.0
min	0.000000	3.000000e+00	0.000000	13.0
25%	0.000000	1.612500e+03	0.000000	13.0
50%	17.500000	5.000000e+03	10.000000	13.0
75%	1607.750000	2.000000e+04	63.005952	13.0
max	20902.000000	1.000000e+09	1229.529412	13.0

	townNumeric	ScoreWord
count	96.000000	96.000000
mean	376.500000	0.968750
std	311.294614	2.231134
min	2.000000	0.000000
25%	23.000000	0.000000
50%	486.000000	0.000000
75%	688.250000	1.000000
max	712.000000	12.000000

Babies, Kids & Family

	Unnamed: 0	Pourcentage Raised	NumberDonors	Duration in days	\
count	96.000000	96.000000	96.000000	96.000000	
mean	1295.500000	12.653125	2.802083	27.541667	
std	27.856777	25.624680	5.258817	27.290992	
min	1248.000000	1.000000	0.000000	0.000000	
25%	1271.750000	1.000000	0.000000	5.750000	
50%	1295.500000	1.000000	1.000000	12.500000	
75%	1319.250000	9.275000	3.000000	50.000000	
max	1343.000000	100.000000	30.000000	90.000000	

	Amount collected	Amount targeted	MeanDonation	categoryNumeric	\
count	96.000000	96.000000	96.000000	96.0	
mean	218.760417	37710.322917	50.338971	14.0	
std	550.873885	152326.708273	114.139291	0.0	
min	0.000000	1.000000	0.000000	14.0	
25%	0.000000	1000.000000	0.000000	14.0	
50%	20.000000	2750.000000	18.571429	14.0	
75%	176.250000	10000.000000	40.250000	14.0	
max	3475.000000	100000.000000	680.000000	14.0	

	townNumeric	ScoreWord
count	96.000000	96.000000
mean	474.729167	1.708333
std	302.153690	3.586280
min	2.000000	0.000000
25%	126.250000	0.000000
50%	668.000000	1.000000
75%	732.250000	2.000000
max	755.000000	27.000000

Sports, Teams & Clubs

	Unnamed: 0	Pourcentage Raised	NumberDonors	Duration in days	\
count	96.000000	96.000000	96.000000	96.000000	
mean	1391.500000	24.410417	26.479167	90.583333	
std	27.856777	32.835419	54.950117	92.277005	
min	1344.000000	1.000000	0.000000	0.000000	
25%	1367.750000	1.000000	0.000000	27.000000	
50%	1391.500000	4.050000	4.500000	62.000000	
75%	1415.250000	37.900000	23.500000	110.000000	
max	1439.000000	100.000000	330.000000	355.000000	

	Amount collected	Amount targeted	MeanDonation	categoryNumeric	\
count	96.000000	96.000000	96.000000	96.0	
mean	1782.541667	11405.500000	35.659510	15.0	
std	3700.069222	37046.562825	43.559059	0.0	
min	0.000000	100.000000	0.000000	15.0	
25%	0.000000	1075.000000	0.000000	15.0	
50%	120.000000	2334.500000	20.405405	15.0	
75%	1322.750000	7200.000000	54.769097	15.0	
max	15620.000000	300000.000000	200.000000	15.0	

	townNumeric	ScoreWord
count	96.000000	96.000000
mean	459.947917	0.416667

std	316.208471	1.319224
min	1.000000	0.000000
25%	106.000000	0.000000
50%	510.500000	0.000000
75%	773.250000	0.000000
max	797.000000	11.000000

Travel & Adventure

	Unnamed: 0	Pourcentage Raised	NumberDonors	Duration in days	\
count	96.000000	96.000000	96.000000	96.000000	
mean	1487.500000	9.035417	3.708333	44.250000	
std	27.856777	24.962595	13.415558	41.021945	
min	1440.000000	1.000000	0.000000	0.000000	
25%	1463.750000	1.000000	0.000000	14.000000	
50%	1487.500000	1.000000	0.000000	43.000000	
75%	1511.250000	1.000000	0.000000	62.500000	
max	1535.000000	100.000000	81.000000	323.000000	

	Amount collected	Amount targeted	MeanDonation	categoryNumeric	\
count	96.000000	9.600000e+01	96.000000	96.0	
mean	262.895833	1.106615e+07	11.883829	16.0	
std	994.711463	1.021786e+08	40.988366	0.0	
min	0.000000	1.000000e+00	0.000000	16.0	
25%	0.000000	1.000000e+03	0.000000	16.0	
50%	0.000000	2.350000e+03	0.000000	16.0	
75%	0.000000	5.000000e+03	0.000000	16.0	
max	6077.000000	1.000000e+09	223.764706	16.0	

	townNumeric	ScoreWord
count	96.000000	96.000000
mean	481.583333	0.84375
std	360.233365	1.44607
min	5.000000	0.000000
25%	65.000000	0.000000
50%	680.000000	0.000000
75%	817.250000	1.000000
max	841.000000	8.000000

Volunteer & Service

	Unnamed: 0	Pourcentage Raised	NumberDonors	Duration in days	\
count	96.000000	96.000000	96.000000	96.000000	
mean	1583.500000	22.482292	9.958333	38.927083	
std	27.856777	28.305447	17.038604	26.884644	
min	1536.000000	1.000000	0.000000	0.000000	
25%	1559.750000	1.000000	1.000000	12.000000	
50%	1583.500000	5.350000	3.500000	41.000000	
75%	1607.250000	36.850000	10.000000	58.000000	
max	1631.000000	100.000000	105.000000	90.000000	

	Amount collected	Amount targeted	MeanDonation	categoryNumeric	\
--	------------------	-----------------	--------------	-----------------	---

count	96.000000	9.600000e+01	96.000000	96.0
mean	442.260417	1.046534e+06	33.677183	17.0
std	939.134980	1.020571e+07	34.952945	0.0
min	0.000000	2.000000e+01	0.000000	17.0
25%	8.750000	5.000000e+02	6.875000	17.0
50%	135.000000	2.000000e+03	25.000000	17.0
75%	384.500000	5.000000e+03	47.857143	17.0
max	6165.000000	1.000000e+08	153.000000	17.0

	townNumeric	ScoreWord
count	96.000000	96.000000
mean	426.447917	1.322917
std	363.005027	2.430052
min	3.000000	0.000000
25%	44.500000	0.000000
50%	352.500000	0.000000
75%	851.250000	2.000000
max	874.000000	16.000000

Dreams, Hopes & Wishes

	Unnamed: 0	Pourcentage Raised	NumberDonors	Duration in days \
count	96.000000	96.000000	96.000000	96.000000
mean	1679.500000	13.648958	3.291667	21.781250
std	27.856777	28.972085	7.471936	23.718171
min	1632.000000	1.000000	0.000000	0.000000
25%	1655.750000	1.000000	0.000000	4.750000
50%	1679.500000	1.000000	0.500000	10.000000
75%	1703.250000	6.350000	3.000000	36.500000
max	1727.000000	100.000000	50.000000	89.000000

	Amount collected	Amount targeted	MeanDonation	categoryNumeric \
count	96.000000	9.600000e+01	96.000000	96.0
mean	158.531250	3.127117e+07	14.689226	18.0
std	546.355233	1.749022e+08	23.731492	0.0
min	0.000000	4.000000e+00	0.000000	18.0
25%	0.000000	5.000000e+02	0.000000	18.0
50%	2.500000	1.250000e+03	2.500000	18.0
75%	61.250000	7.250000e+03	20.000000	18.0
max	4245.000000	1.000000e+09	124.852941	18.0

	townNumeric	ScoreWord
count	96.000000	96.000000
mean	486.750000	0.750000
std	380.657354	2.348572
min	2.000000	0.000000
25%	62.250000	0.000000
50%	508.500000	0.000000
75%	887.250000	1.000000
max	911.000000	17.000000

Other

	Unnamed: 0	Pourcentage Raised	NumberDonors	Duration in days	\
count	0	0.0	0	0.0	
mean	<NA>	NaN	<NA>	NaN	
std	<NA>	NaN	<NA>	NaN	
min	<NA>	NaN	<NA>	NaN	
25%	<NA>	NaN	<NA>	NaN	
50%	<NA>	NaN	<NA>	NaN	
75%	<NA>	NaN	<NA>	NaN	
max	<NA>	NaN	<NA>	NaN	

	Amount collected	Amount targeted	MeanDonation	categoryNumeric	\
count	0.0	0.0	0.0	0.0	
mean	NaN	NaN	NaN	NaN	
std	NaN	NaN	NaN	NaN	
min	NaN	NaN	NaN	NaN	
25%	NaN	NaN	NaN	NaN	
50%	NaN	NaN	NaN	NaN	
75%	NaN	NaN	NaN	NaN	
max	NaN	NaN	NaN	NaN	

	townNumeric	ScoreWord
count	0.0	0.0
mean	NaN	NaN
std	NaN	NaN
min	NaN	NaN
25%	NaN	NaN
50%	NaN	NaN
75%	NaN	NaN
max	NaN	NaN

III.2/Let's look at the collect of the last month, we will only work on DECEMBER

In [697...

```
mask_december = (df['Creation Date'] > '2022-1-1') & (df['Creation Date'] <=
mask_SmallCap = (df['Amount targeted']<5000)
mask_MidCap = (df['Amount targeted']>5000)& (df['Amount targeted']<50000)
mask_BigCap=(df['Amount targeted']>50000)

#Collects of december
december=df.loc[mask_december]
```

III.3/ Relation Duration of the collect and Pourcentage Raised

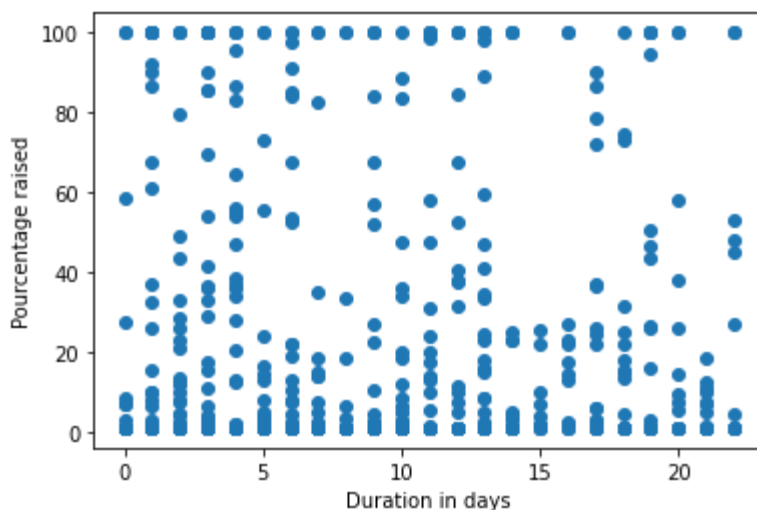
In [664...

```
#For all categories
plt.figure()

x=december['Duration in days']
y=december['Pourcentage Raised']

fig=plt.scatter(x, y )
plt.xlabel('Duration in days')
plt.ylabel('Pourcentage raised')
```

Out[664... Text(0, 0.5, 'Pourcentage raised')

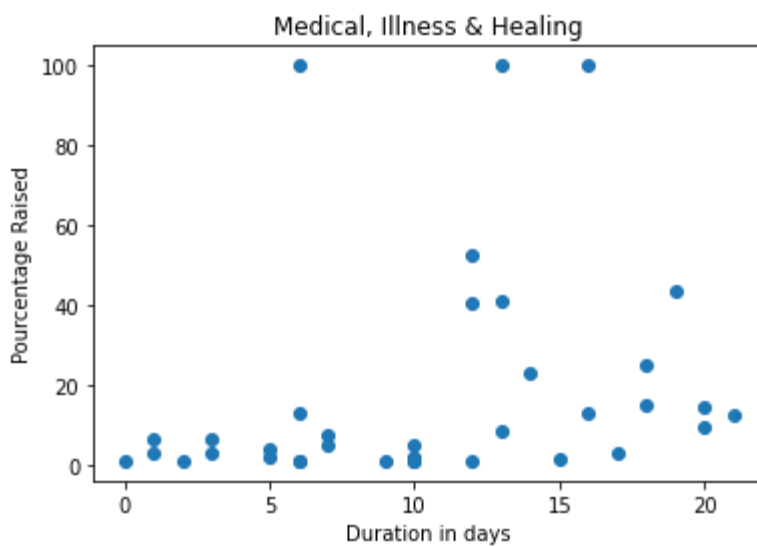


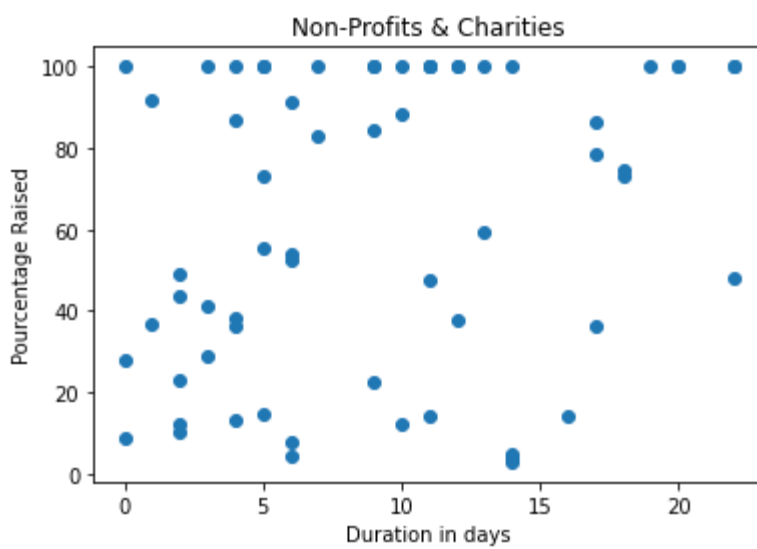
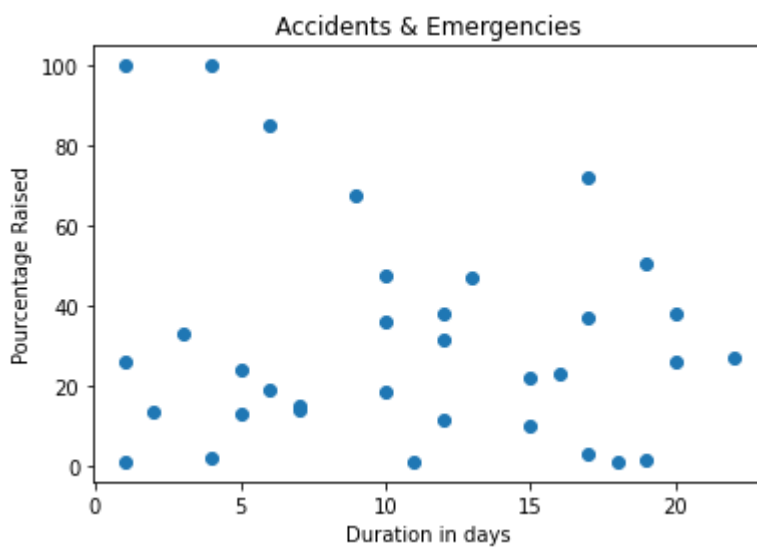
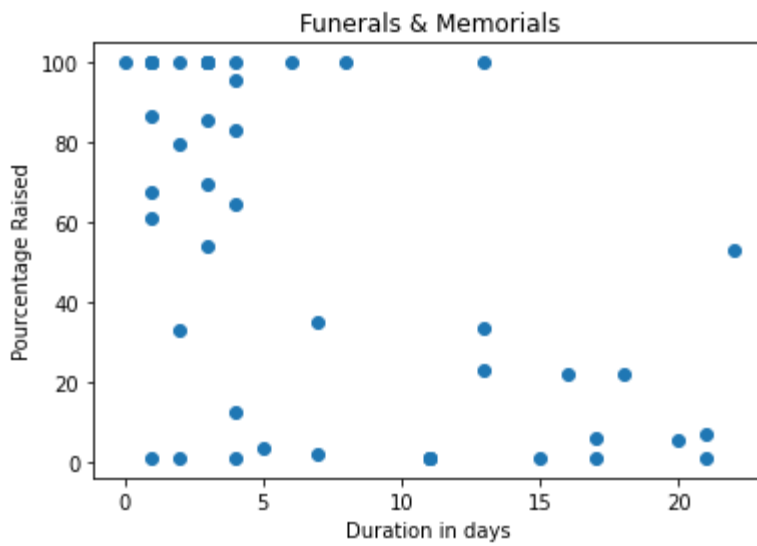
In [665...

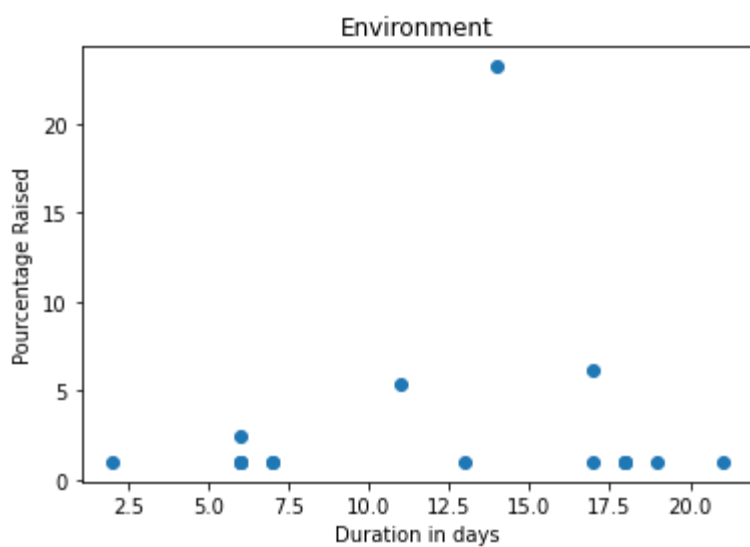
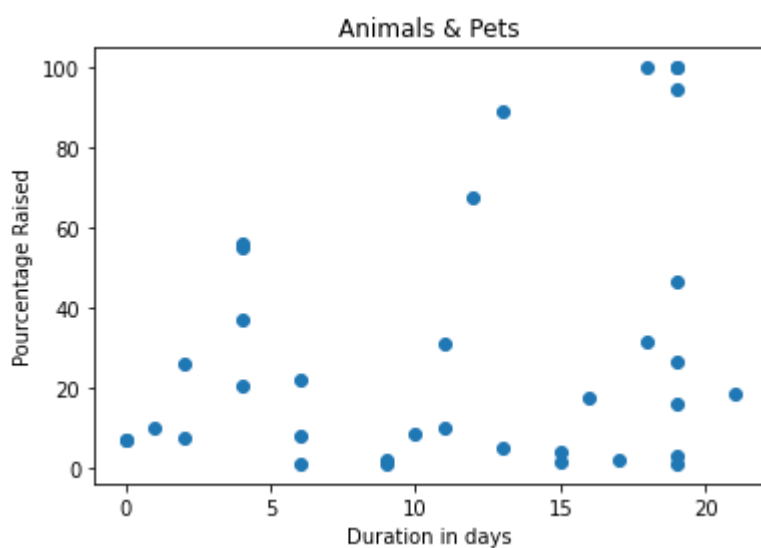
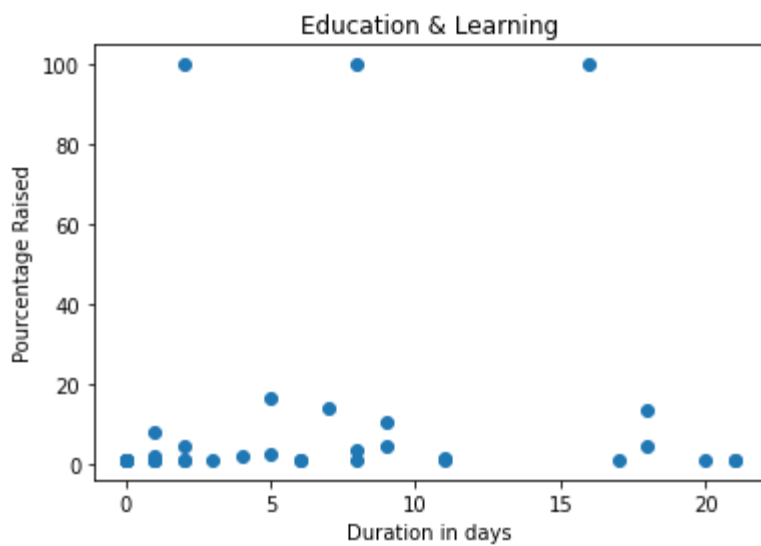
```
#For each category
for cat in catégorie:
    plt.figure()

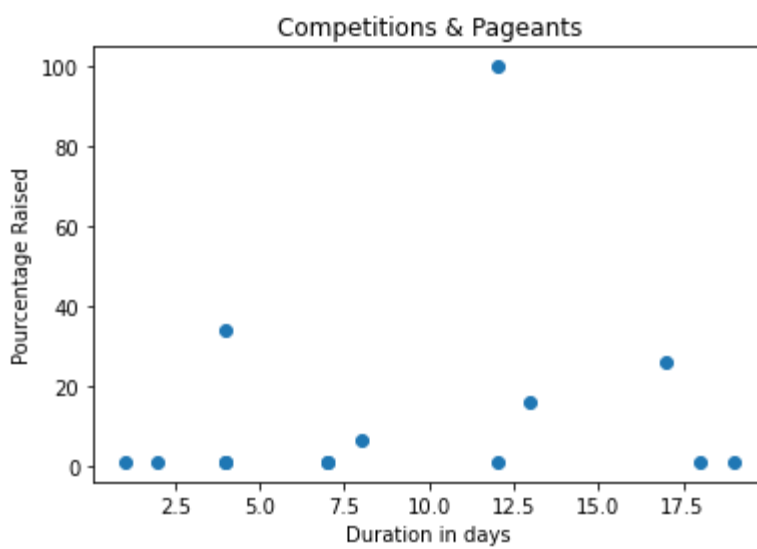
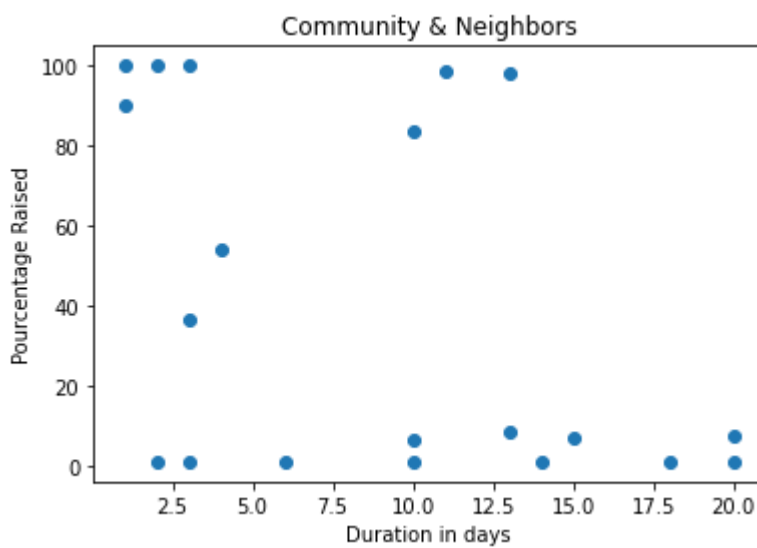
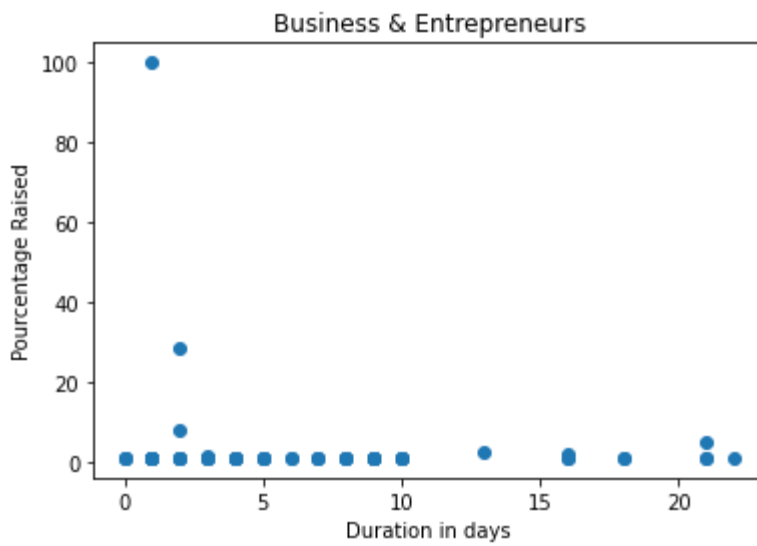
    x=december.loc[(df['Catégorie'] ==cat)]['Duration in days']
    y=december.loc[(df['Catégorie'] ==cat)]['Pourcentage Raised']

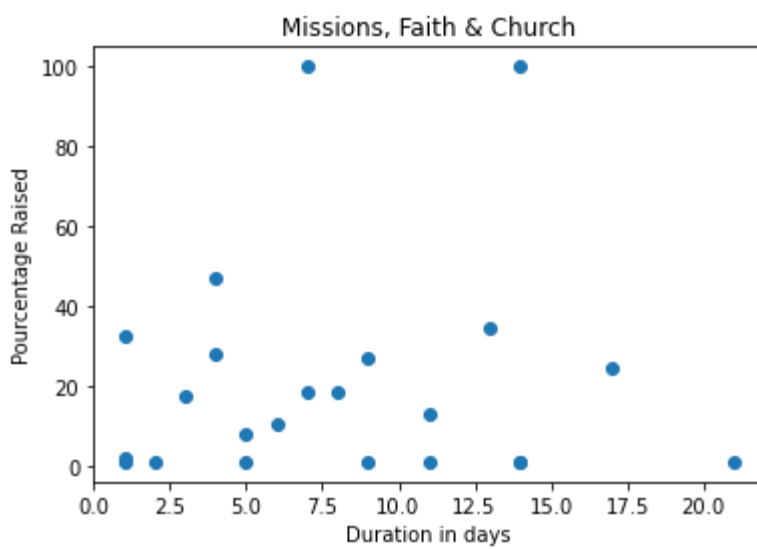
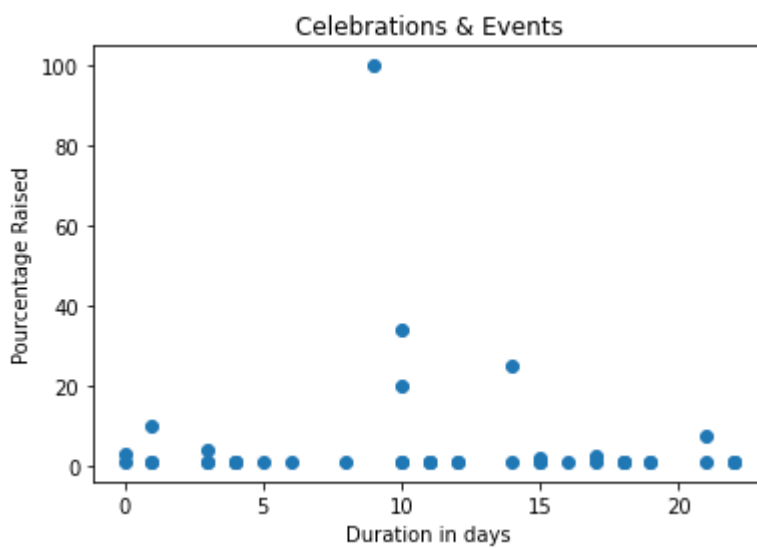
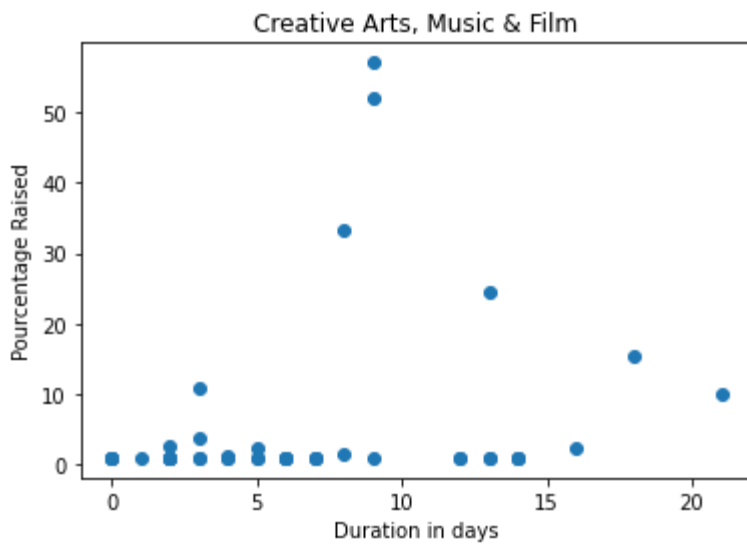
    fig=plt.scatter(x, y )
    plt.title(cat)
    plt.xlabel('Duration in days')
    plt.ylabel('Pourcentage Raised')
```

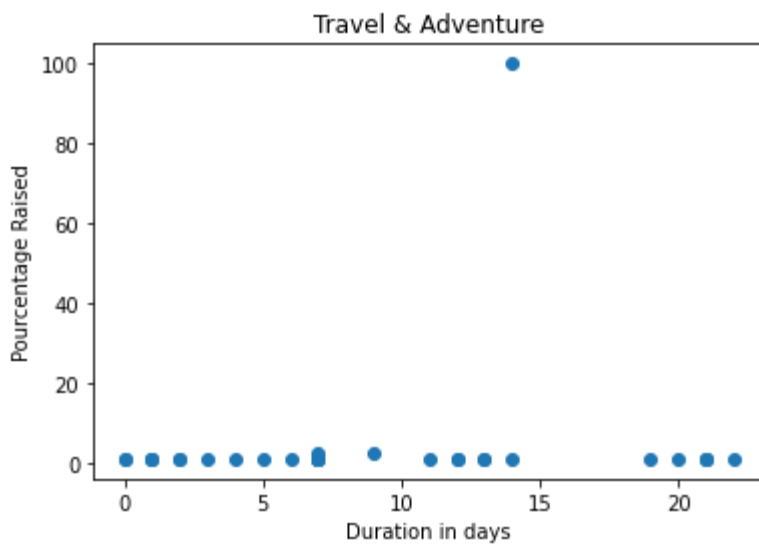
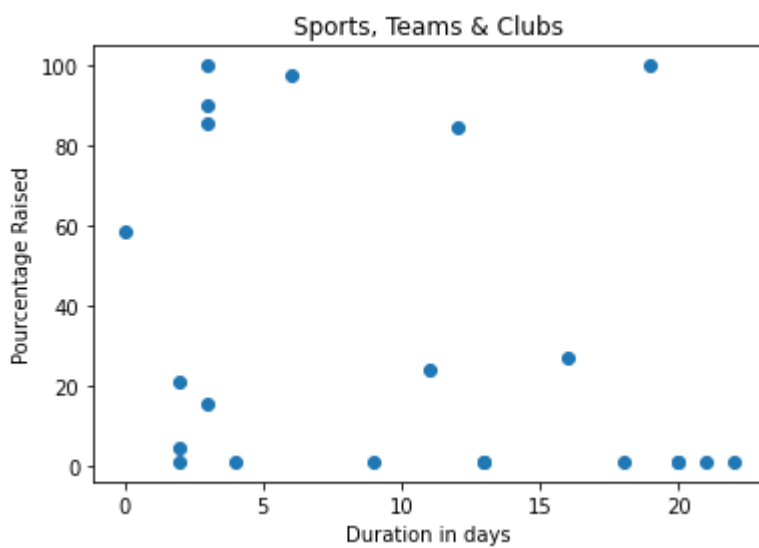
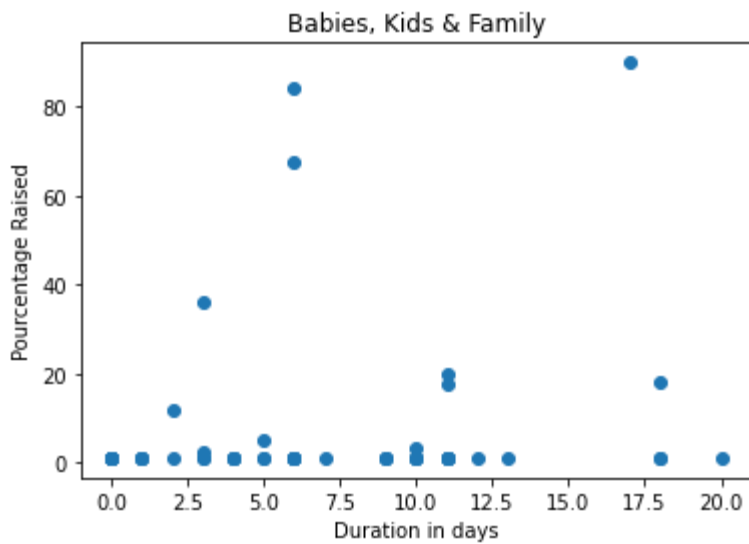


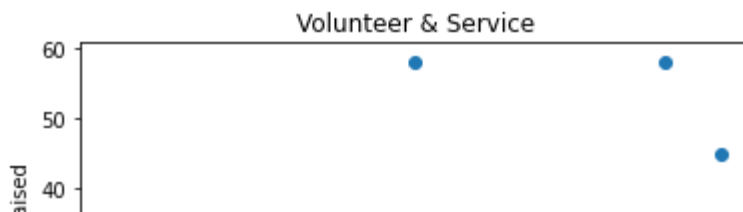












=>No real model between Duration of the collect and Pourcentage Raised

III.4/ Relation between descriptions of projects and Pourcentage Raised

III.4.1/ We make a dictionary of keywords for each category we will use to study descriptions

In [666...

```
#We look at the words with the biggest occurrences for each category (open 'oc
from collections import Counter
txt=[]
for cat in categorie:
    list1=[]
    for e in df.loc[(df['Categorie'] ==cat)][['Short_description']]:
        list1=list1+e
    counts = Counter(list1)
    txt.append(cat)
    txt.append(" ")
    txt.append(counts)

with open('occurences.txt', 'w') as f:
    for item in txt:
        f.write("%s\n" % item)
```

In [667...

```
#We list the words of occurrences that are keywords, which means the ones with

motsclés1=[[' douleur, association, funds, générosité, surgery, financier, re
[' family, obsèques, children, soutien, remercie, enterrement, hopital, assoc
[' typhon, community, home, treatment, soutien, reconstruire, destroyed, mais
[' support, community, together, children, foundation, students, animal, orga
[' school, France, children, formation, yhèse, étude, construction, associati
[' vétérinaire, opération, chien, frais, traitements, sauver, chimiothérapie,
[' association, environnement, société, développement, qualité, paris, clima
[' agricole, communication, restaurant, investissement, agriculture, business
[' vaccination, family, community, water, business, needs, health '],
[' compétition, championnat, international, opportunity, formation, passion,
[' association, computer, vidéo, ordinateur, album, montage, formation, docu
[' festival, conference, transpersonnalité, musique, climate, justice, agree
[' mission, community, christ, church, évangélisation, voyage, prayer, chris
[' famille, maison, besoin, place, cancer, association, logement, travaux, c
[' coach, tennis, équipements, soccer, Olympic, academy, club, basketball, soc
[' voyage, dream, equipment, découvrir, trail, aventure, vidéo, professionne
[' enfants, villages, typhon, association, education, maisons, acheter, chri
[' refugee, voyage, meditation, pédagogique, besoin, school, program '],
[]

]
motsclés=[]
```

```
for e in motsclés1:
    motsclés.append(word_tokenize(str(e).replace(',', ' ').replace('[', ' ').repl
```

In [668...

```
#We create a dictionary linking a category and the keywords linked with it
motclé_dict = []

motclé_dict=dict()
#catégorie is defined in II.8/
for i in range(len(catégorie)):
    motclé_dict[catégorie[i]]=motsclés[i]
```

III.4.2/ We give for each project a "scoreword": its the occurrences of the keywords in its description

In [669...

```
#This function counts the occurrences of the words of a list1 in another list2
def countword(list1, list2):
    occu=dict()
    for word in list1:
        #print(word)
        occu[word]=list2.count(word)
    return occu
```

In [670]...

```
#This function counts the numbers of matching words from the description (lis
def scoreWord(index):
    sum=0
    count=countword(motclé_dict[df['Catégorie'][index]],df['Short_description']
    for e in count:
        sum+=count[e]
    return sum
```

In [671]...

```
#We create a column scoreword
scoreword=[]
for i in range(len(df['Short_description'])):
    scoreword.append(int(scoreWord(i)))
df['ScoreWord']=scoreword
```

III.4.3/ Relation between ScoreWord and Pourcentage Raised

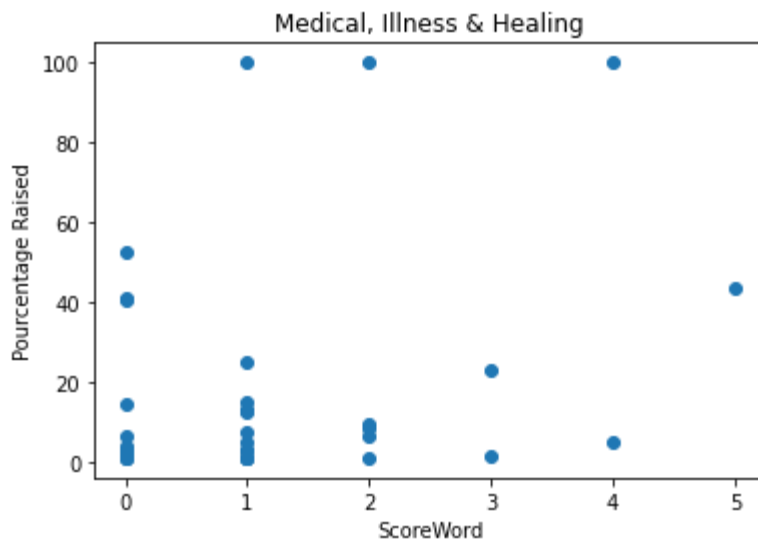
In [672]...

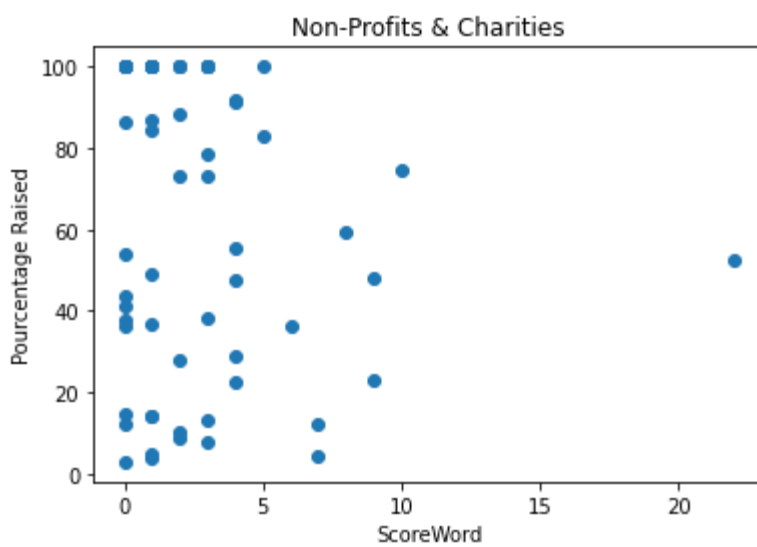
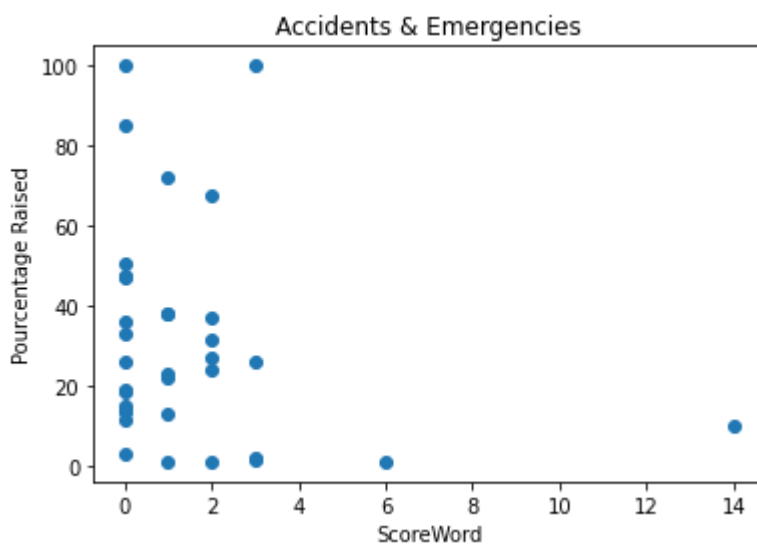
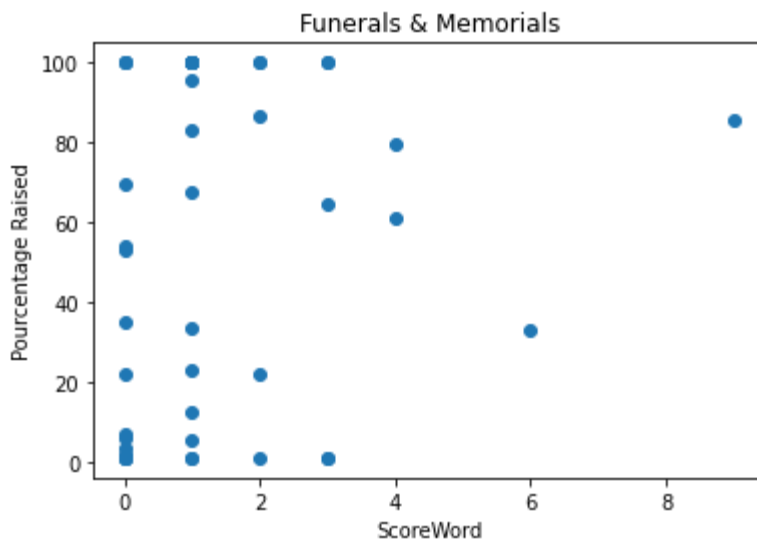
```
december=df.loc[mask_december]

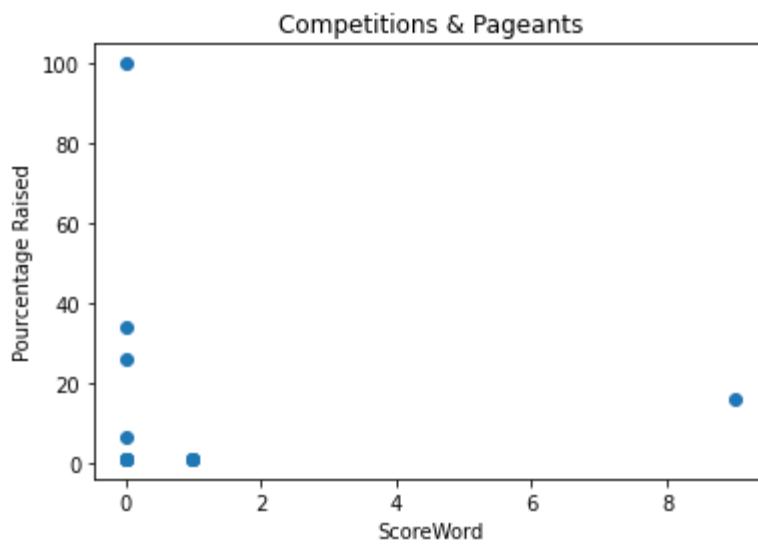
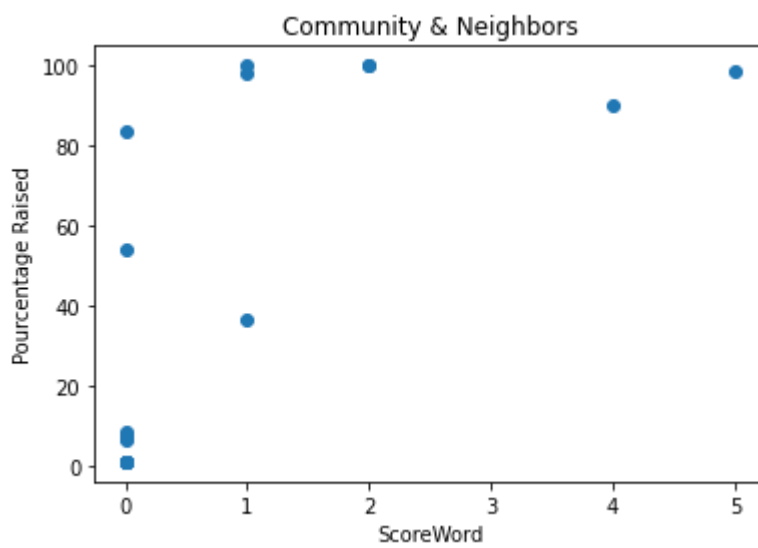
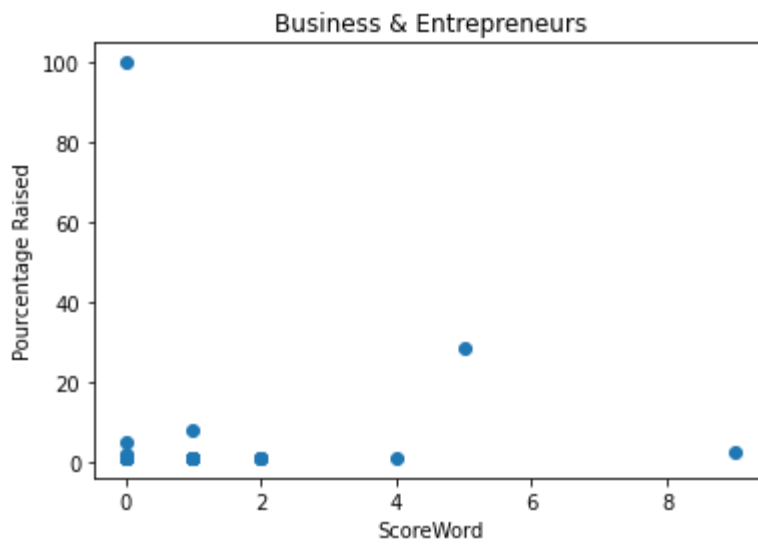
for cat in catégorie:
    plt.figure()

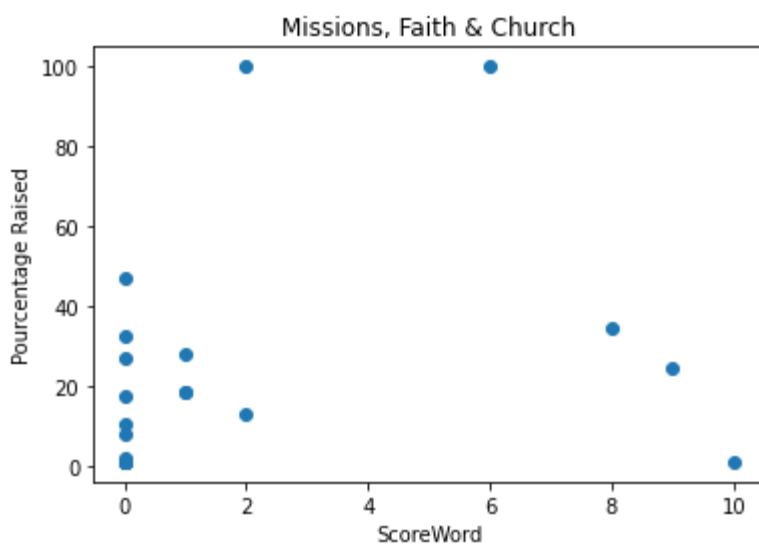
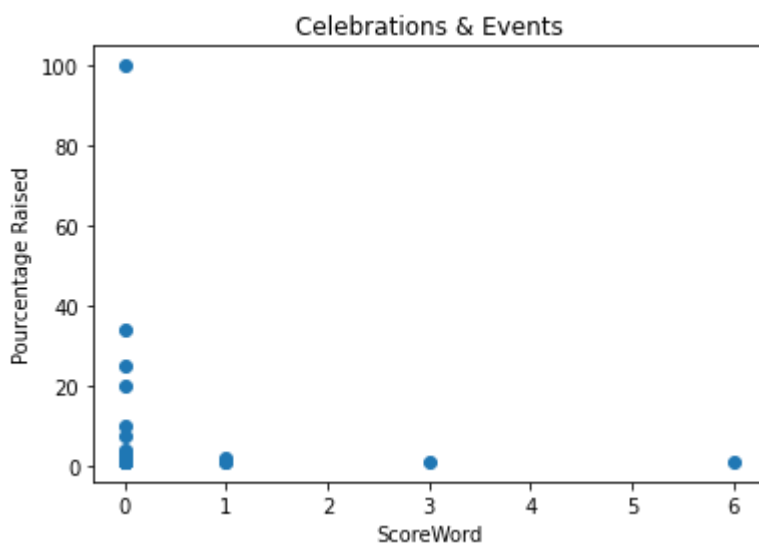
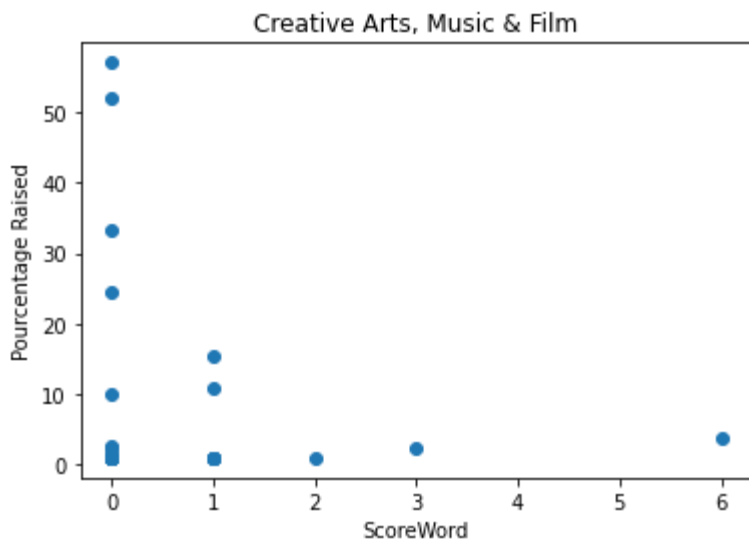
    x=december.loc[(december['Catégorie'] ==cat)][ 'ScoreWord' ]
    y=december.loc[(december['Catégorie'] ==cat)][ 'Pourcentage Raised' ]

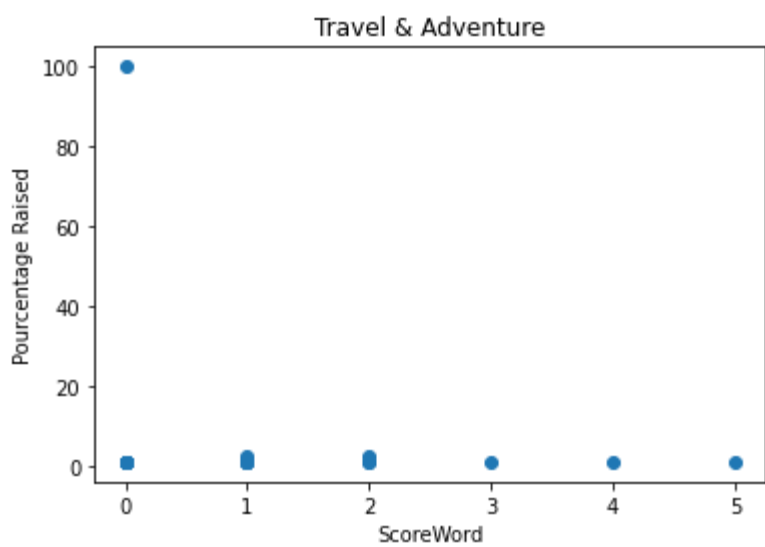
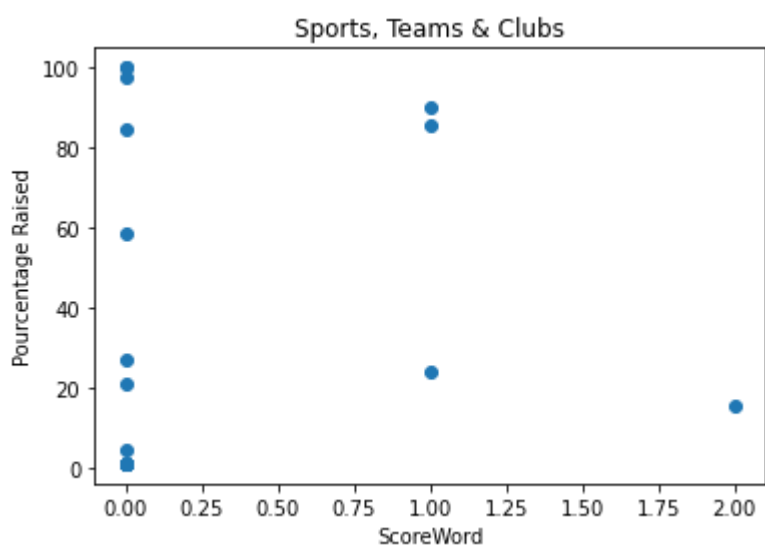
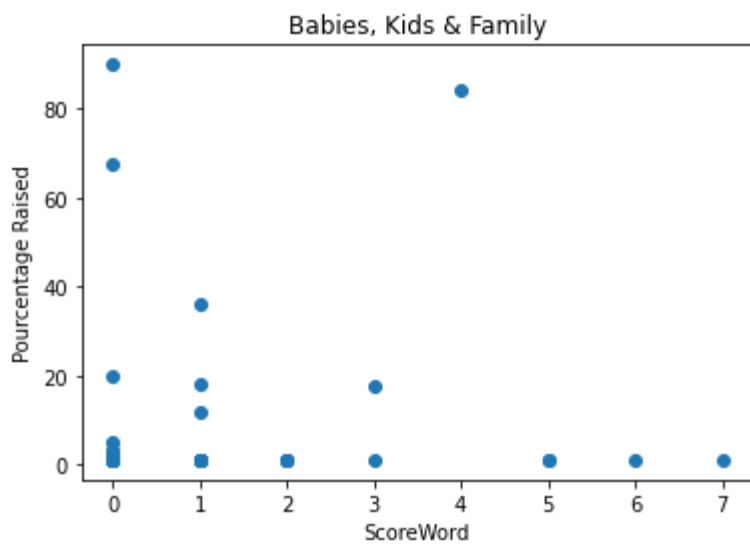
    fig=plt.scatter(x, y )
    plt.title(cat)
    plt.xlabel('ScoreWord')
    plt.ylabel('Pourcentage Raised')
```

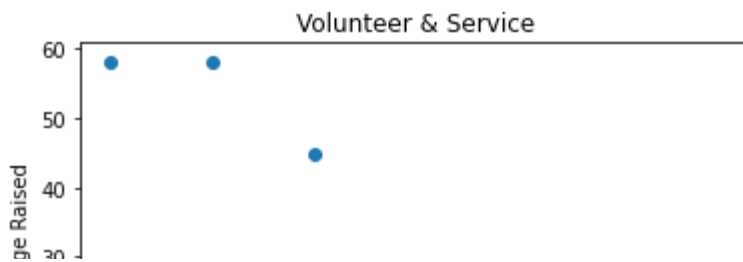












Relation between Number of Donors and ScoreWord for each category

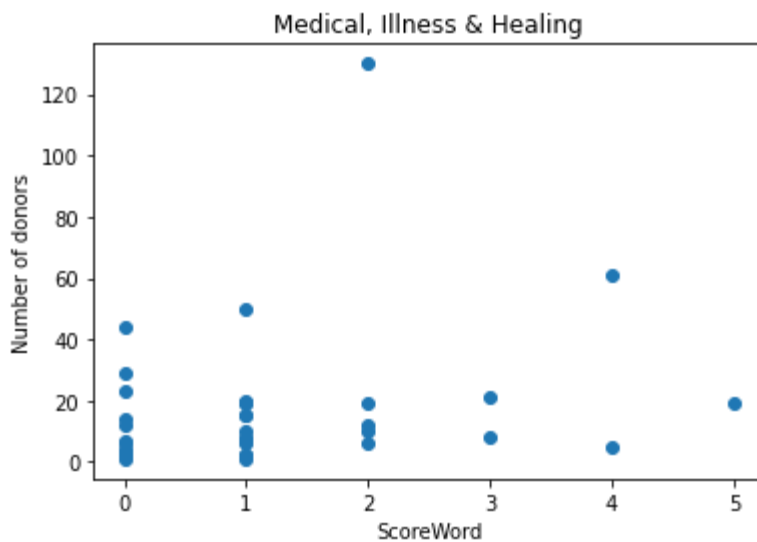
In [676...

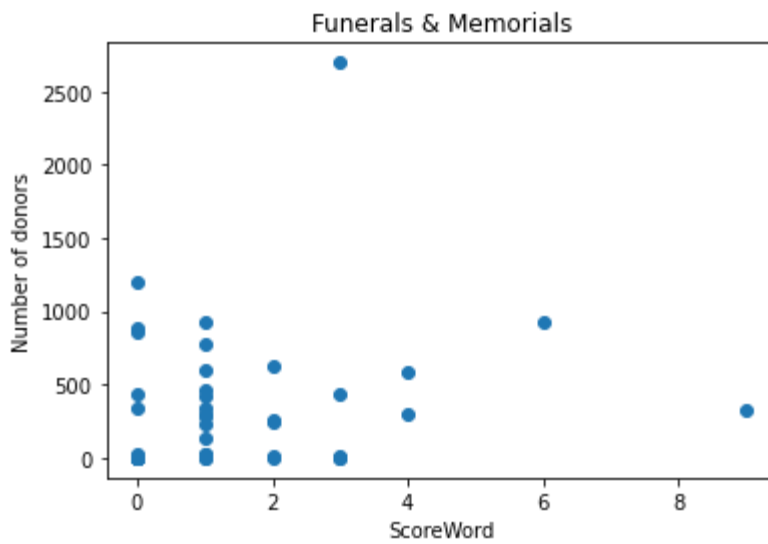
```
december=df.loc[mask_december]

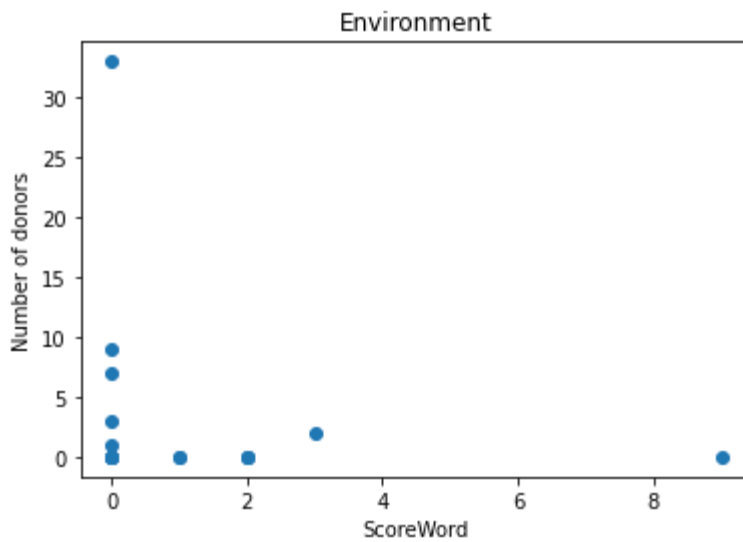
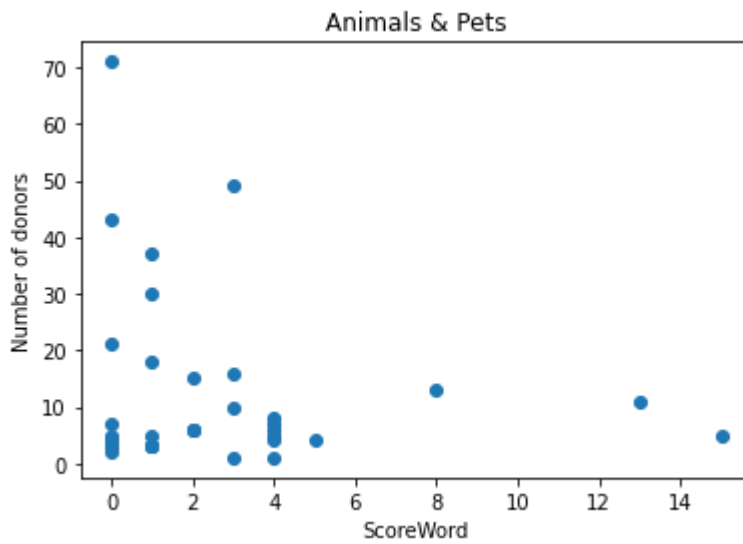
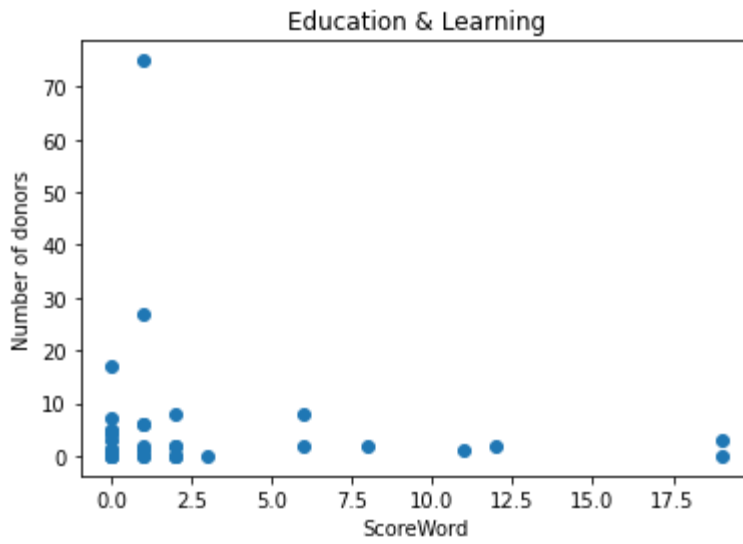
for cat in categorie:
    plt.figure()

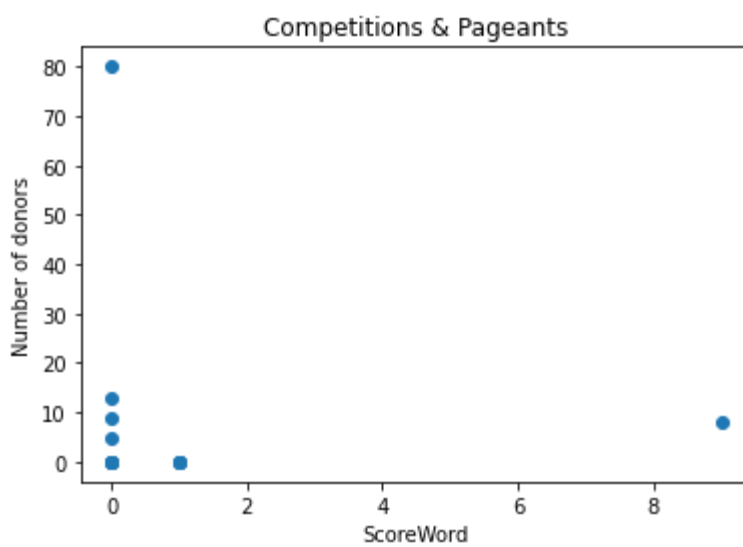
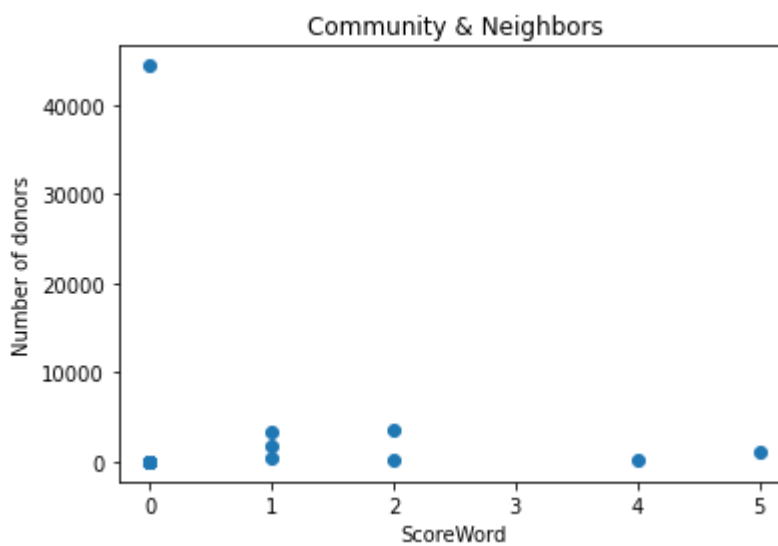
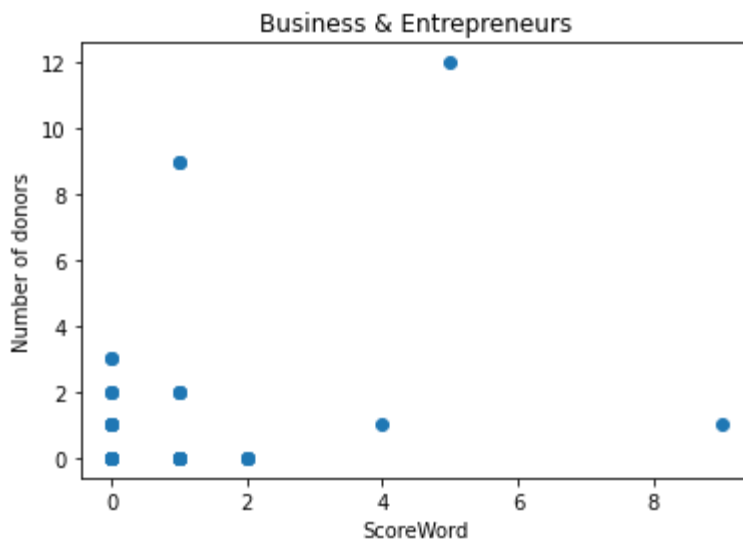
    x=december.loc[(december['Categorie'] ==cat)][ 'ScoreWord' ]
    y=december.loc[(december['Categorie'] ==cat)][ 'NumberDonors' ]

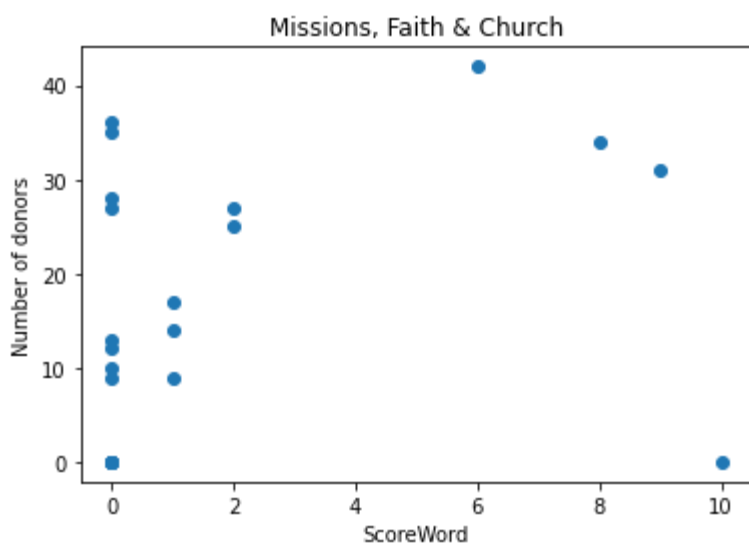
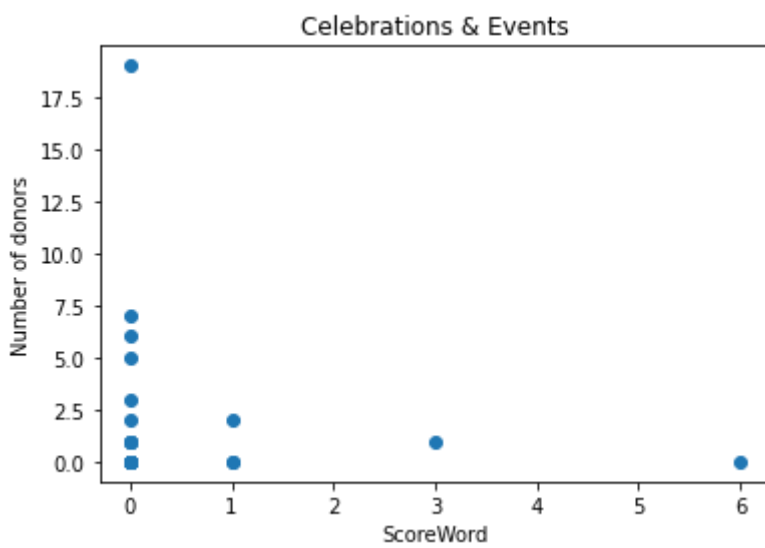
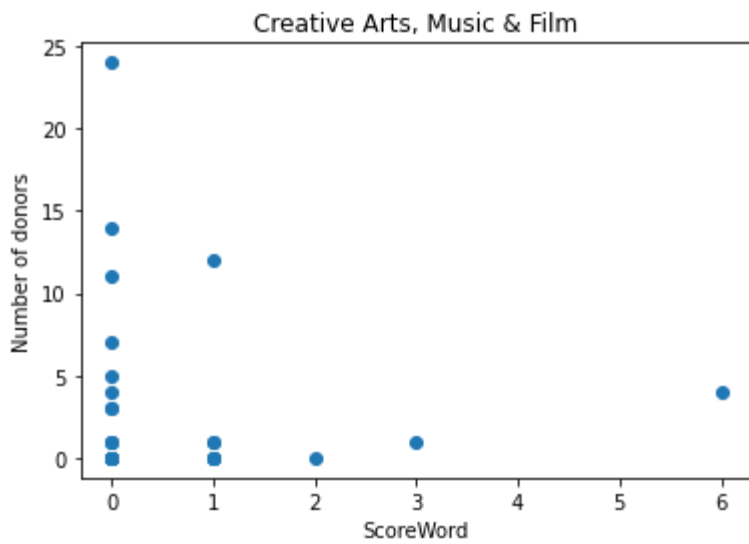
    fig=plt.scatter(x, y )
    plt.title(cat)
    plt.xlabel('ScoreWord')
    plt.ylabel('Number of donors')
```

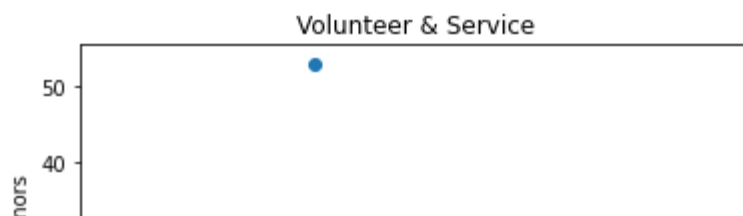
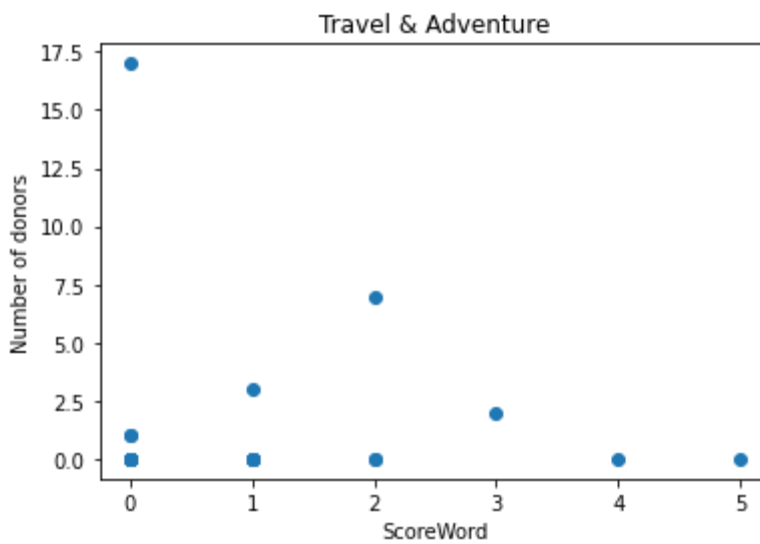
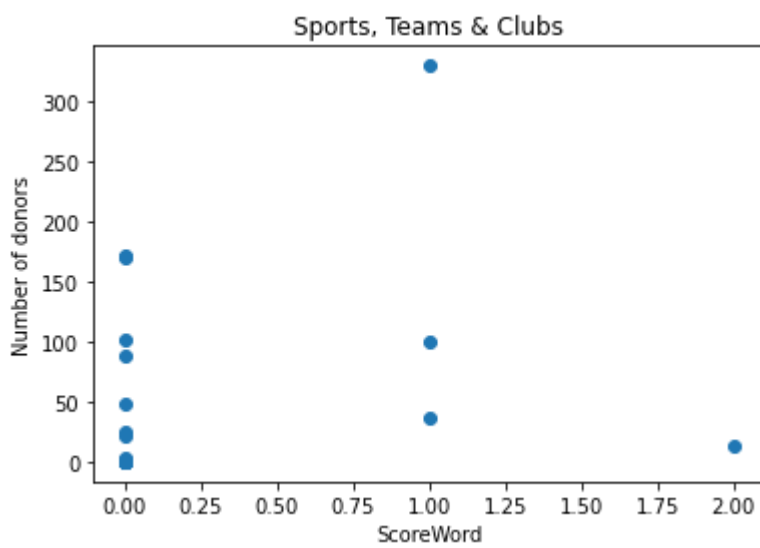
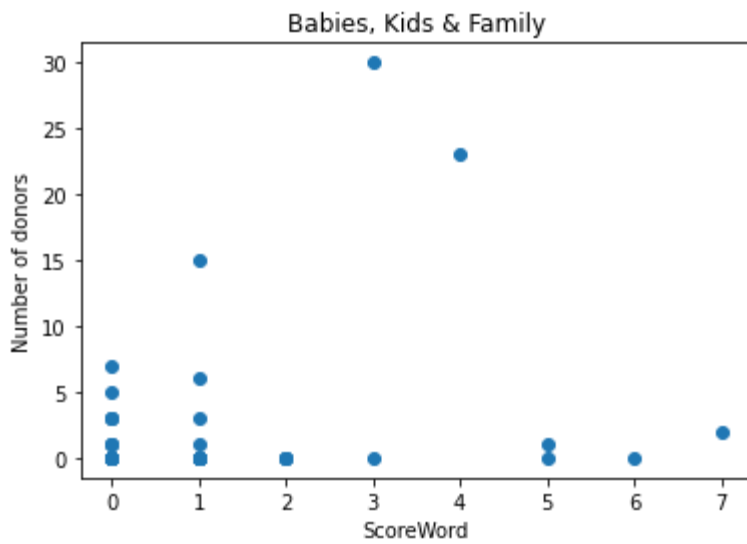












Relation between Duration in days and Number of donors for each category

In [684...

```

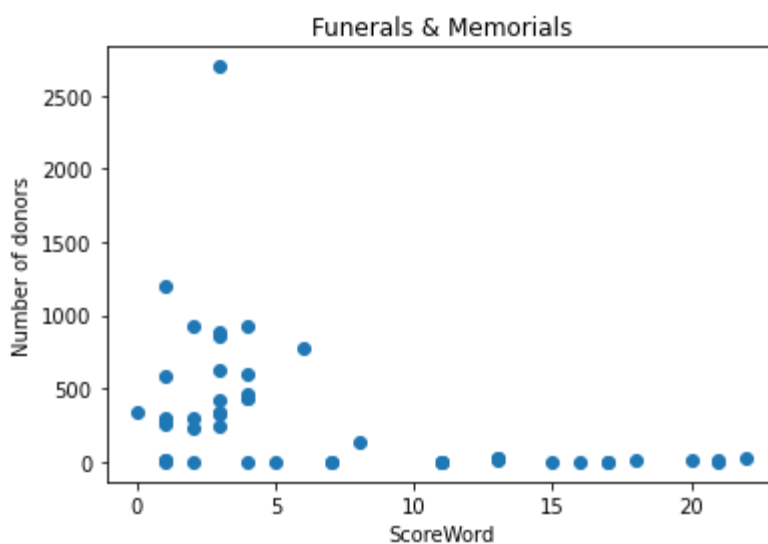
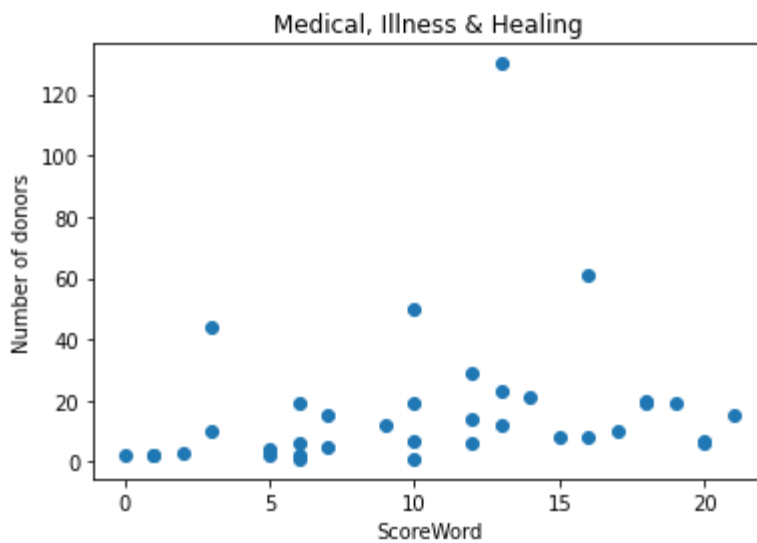
deceember=df.loc[mask_deceember]

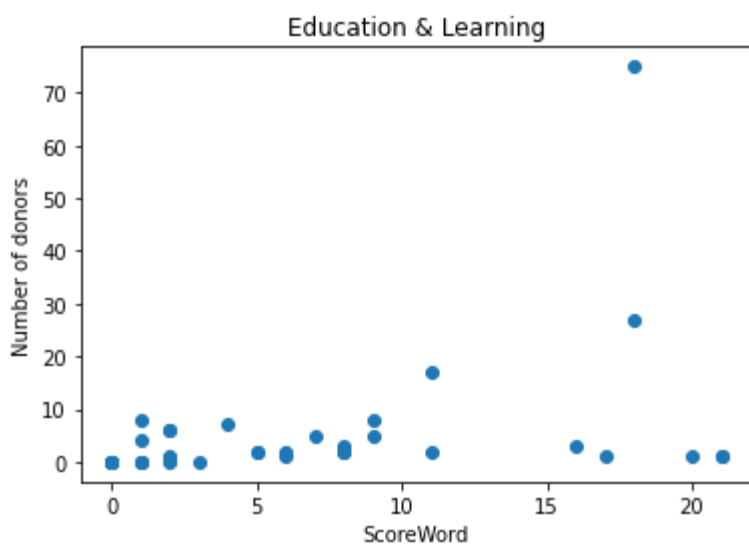
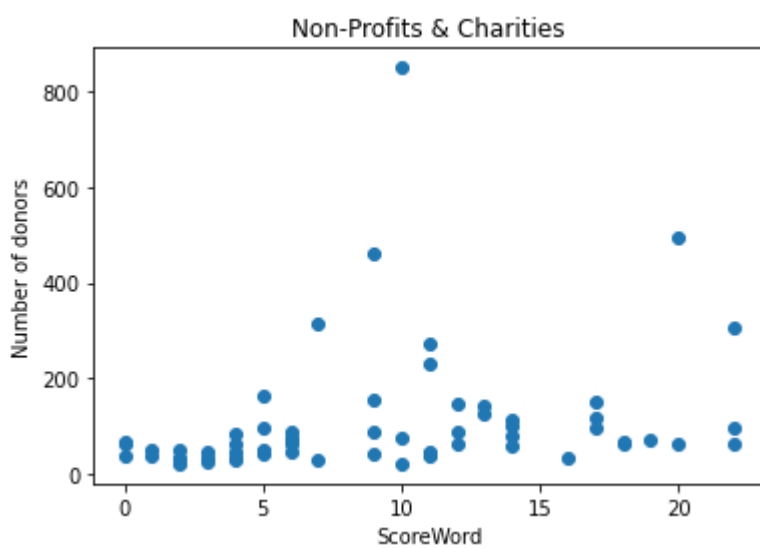
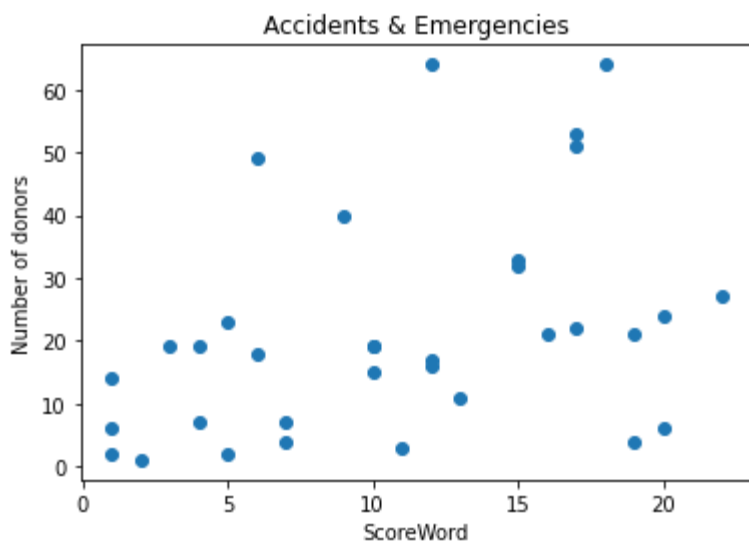
for cat in categorie:
    plt.figure()

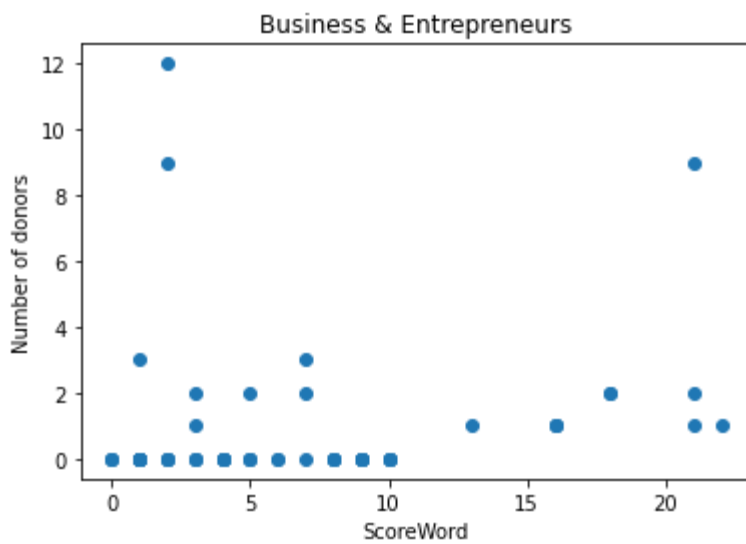
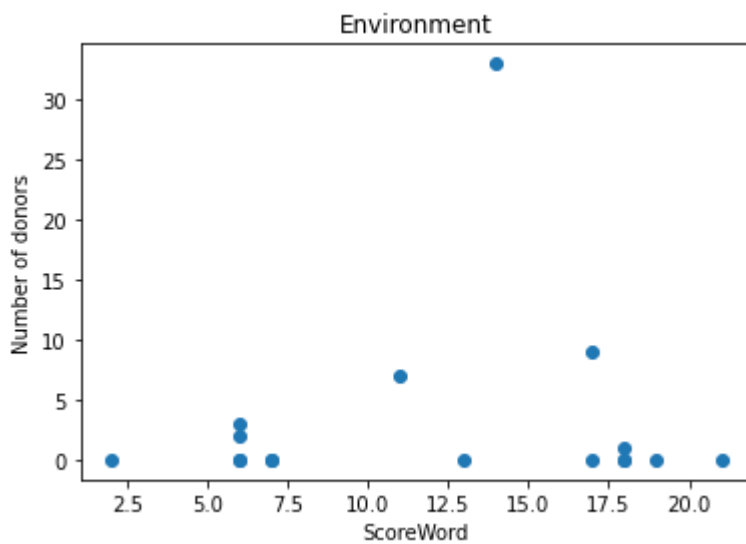
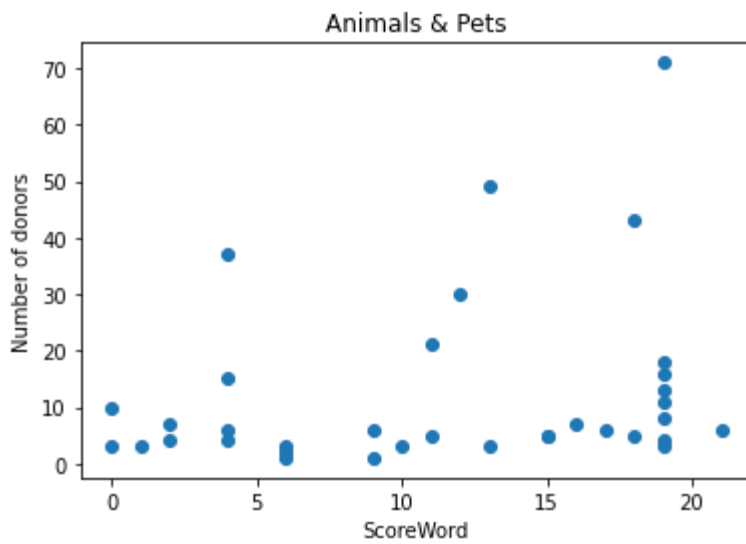
    x=deceember.loc[(deceember['Categorie'] ==cat)][ 'Duration in days']
    y=deceember.loc[(deceember['Categorie'] ==cat)][ 'NumberDonors']

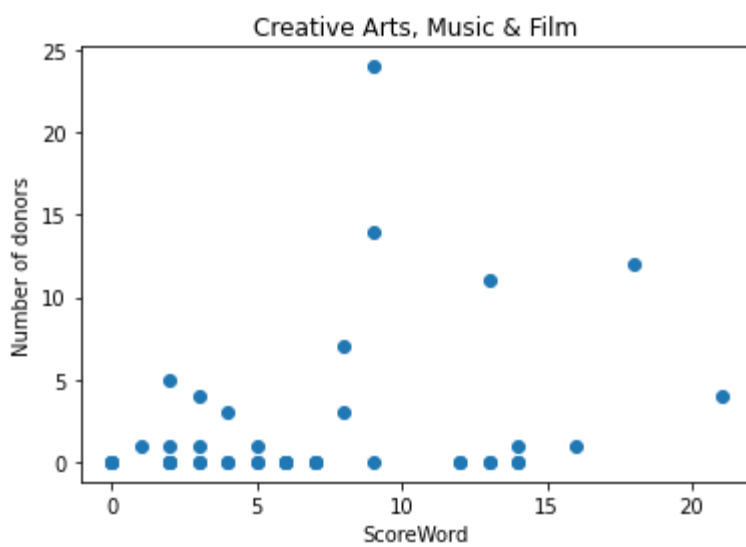
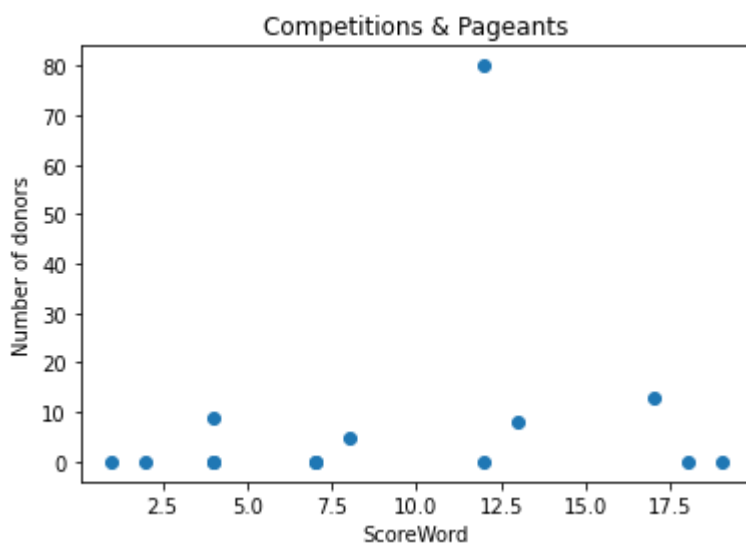
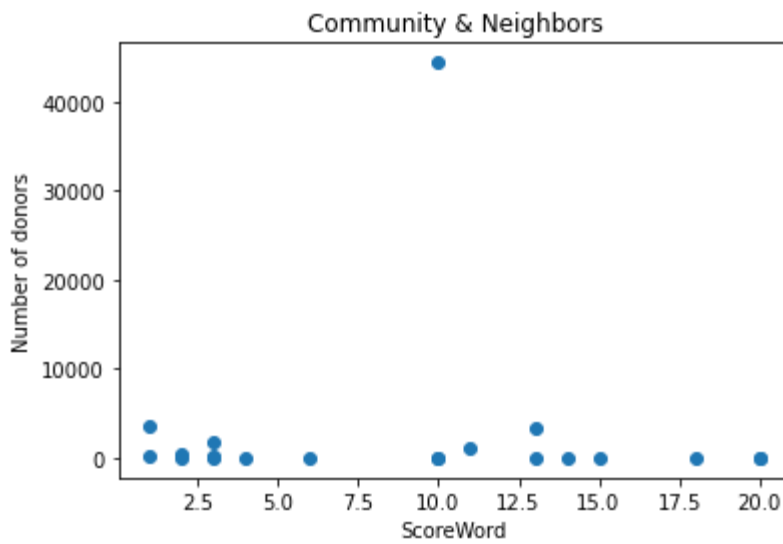
    fig=plt.scatter(x, y )
    plt.title(cat)
    plt.xlabel('ScoreWord')
    plt.ylabel('Number of donors')

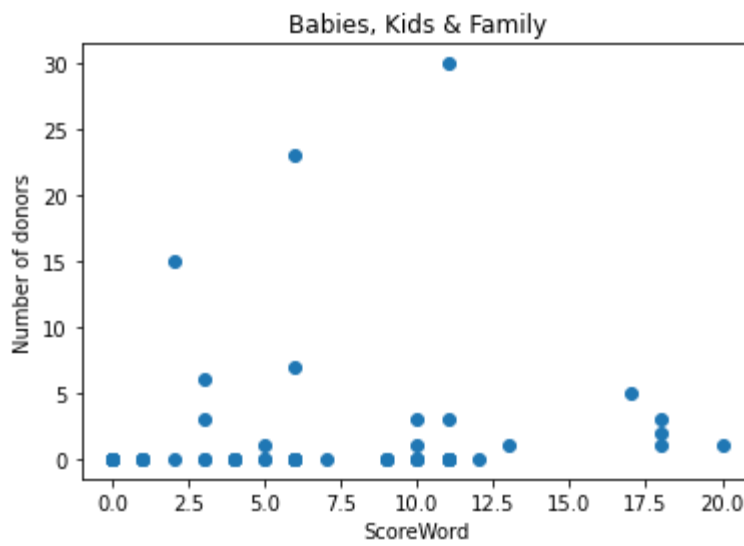
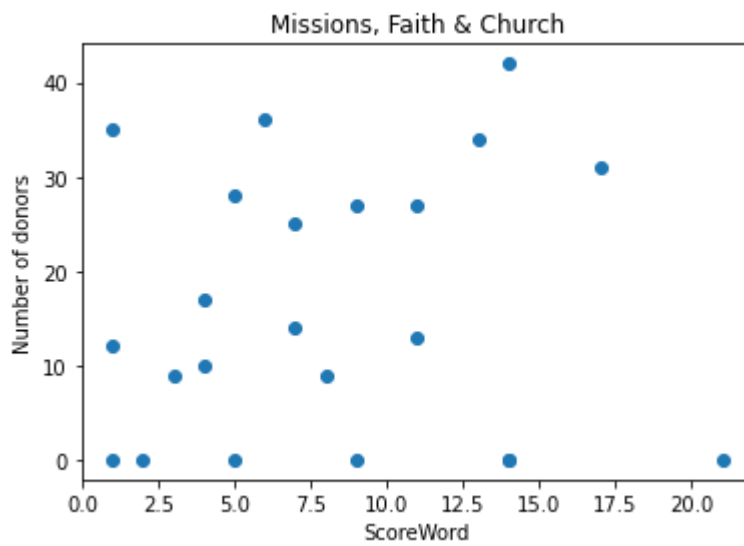
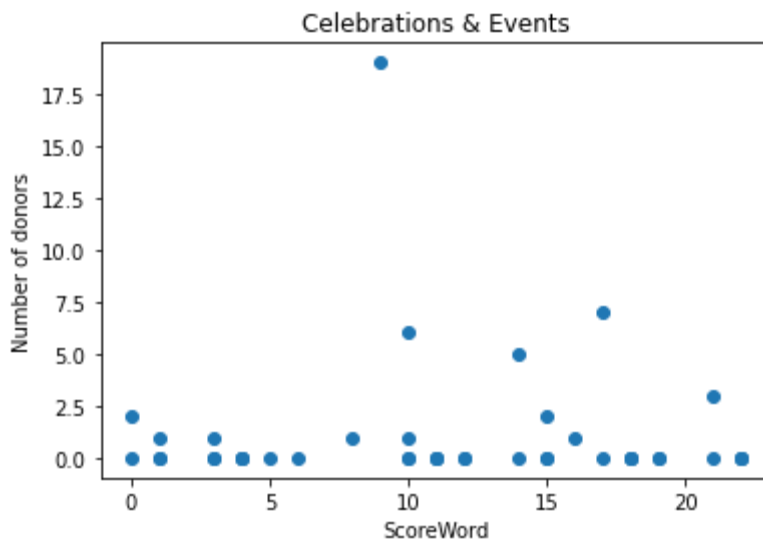
```

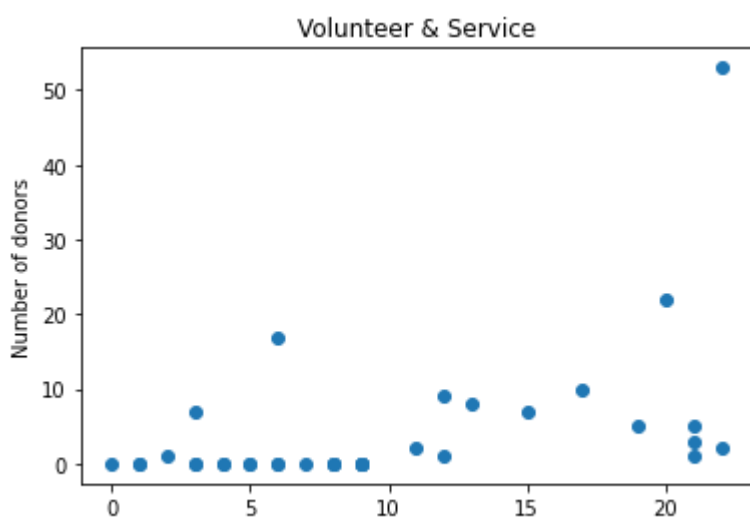
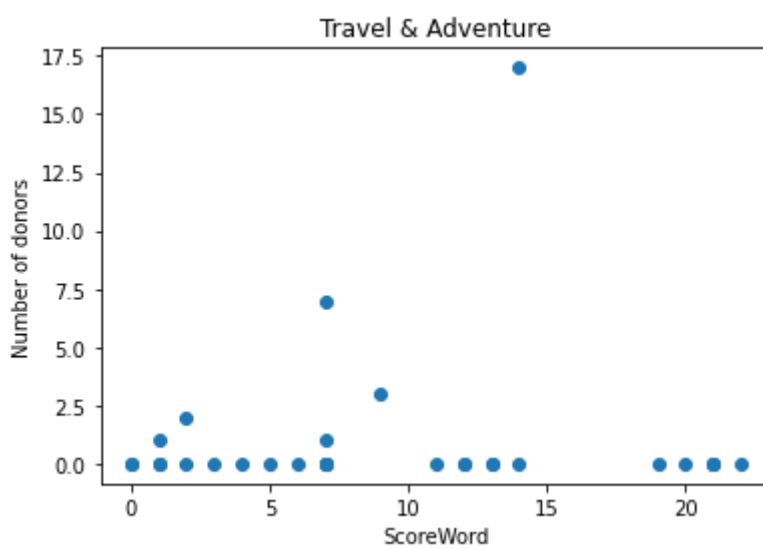
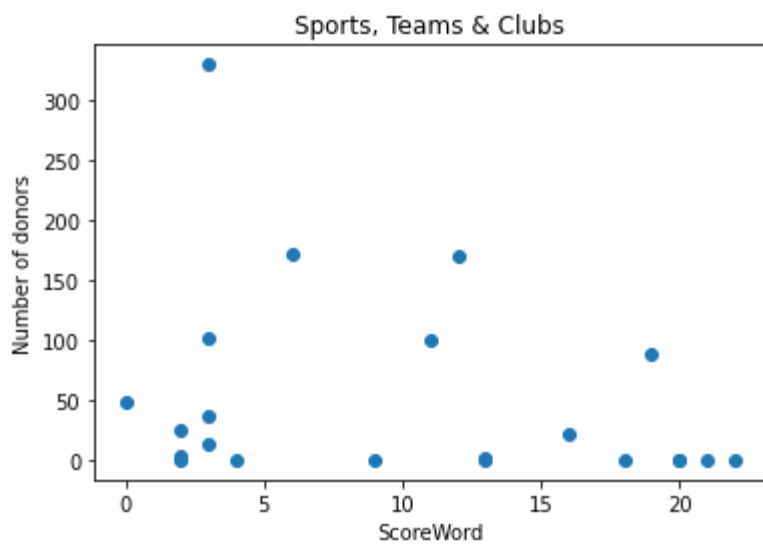












Relation between Number of Donors and Amount collected

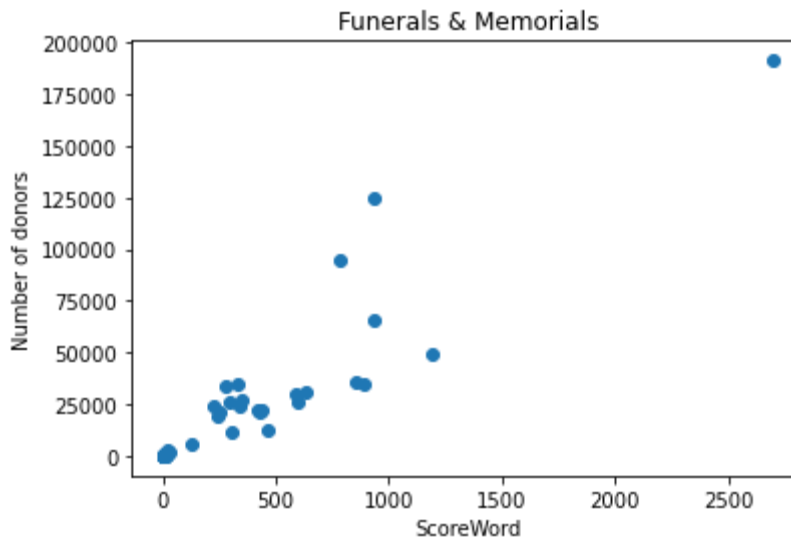
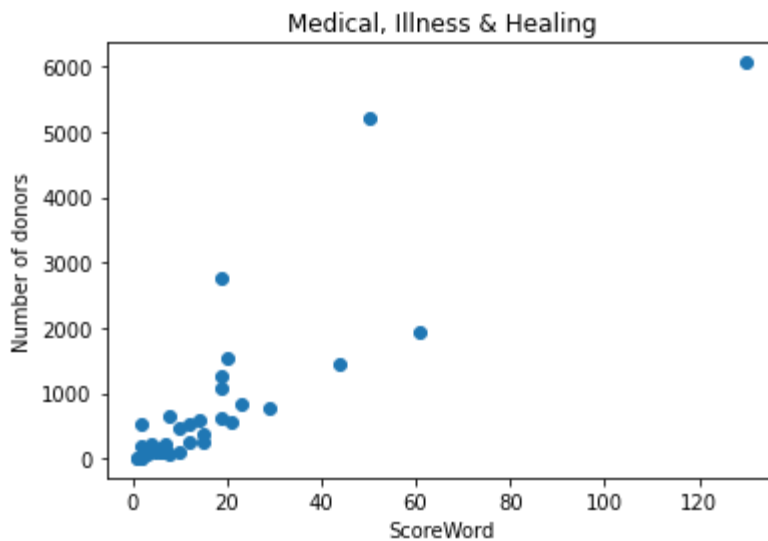
In [691]...

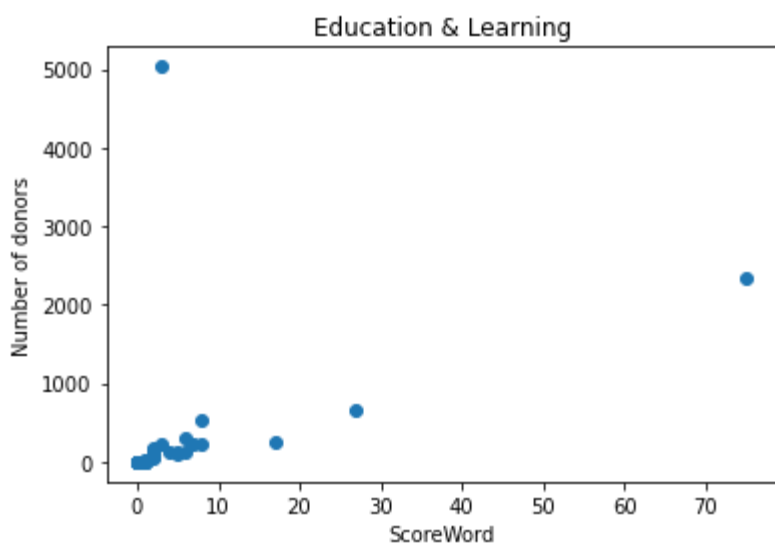
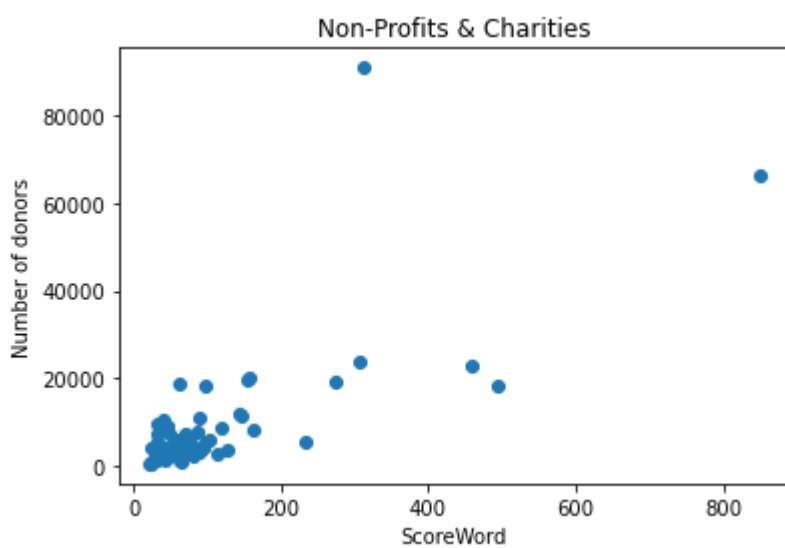
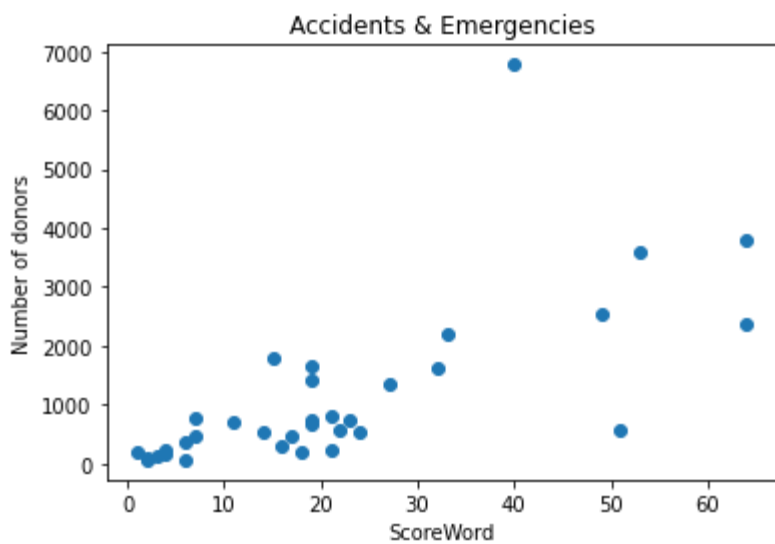
```
december=df.loc[mask_december]

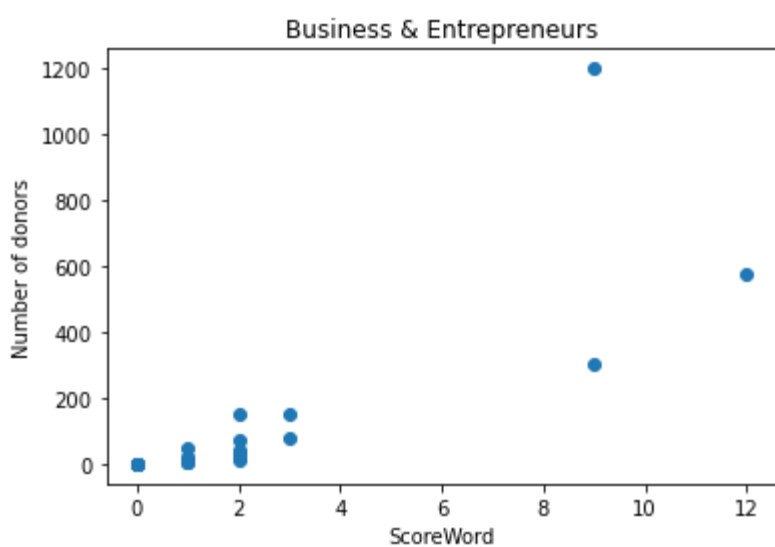
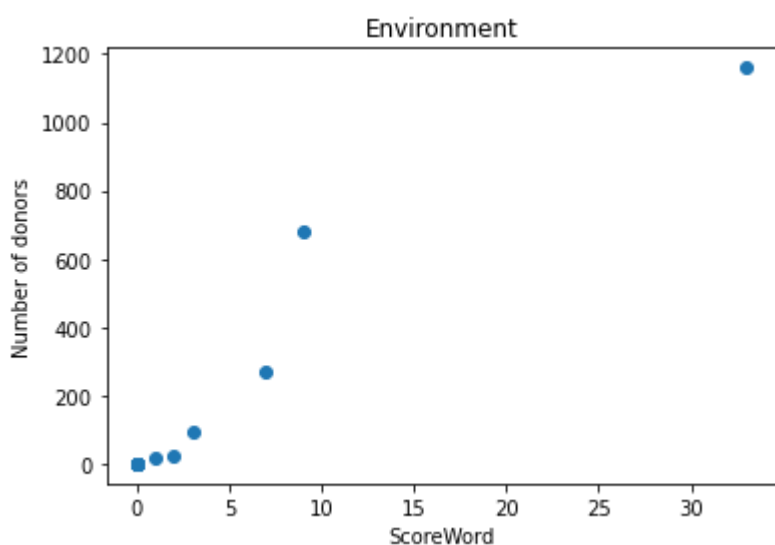
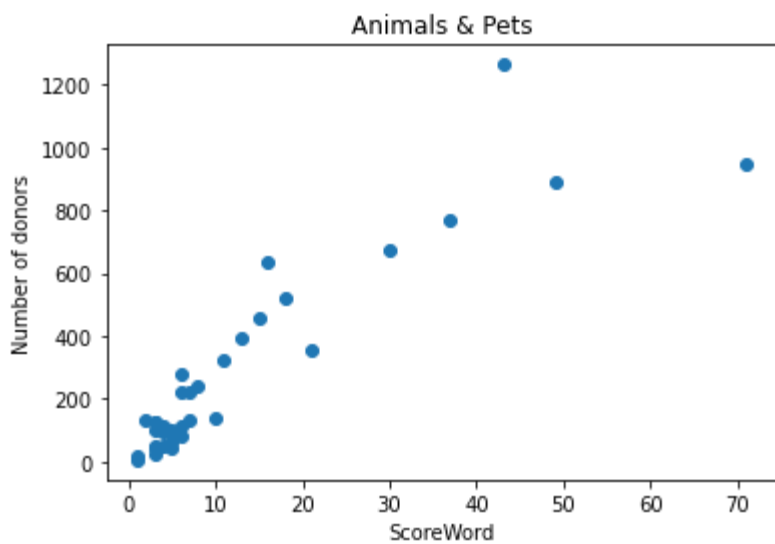
for cat in categorie:
    plt.figure()

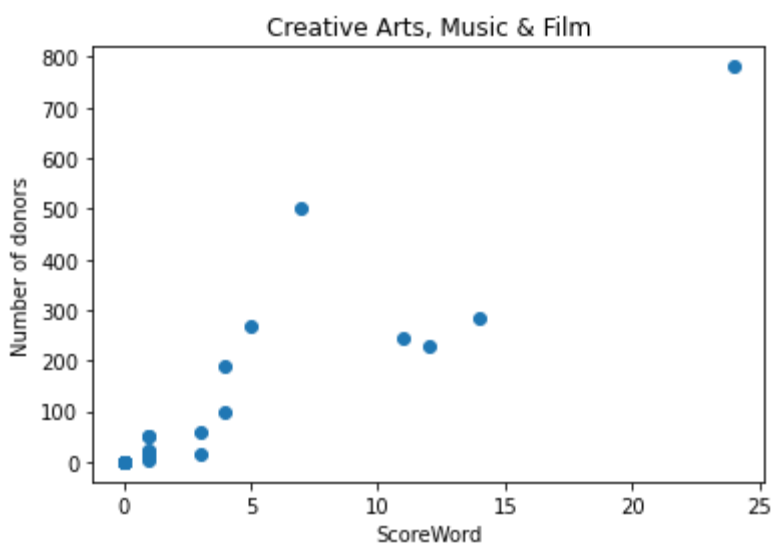
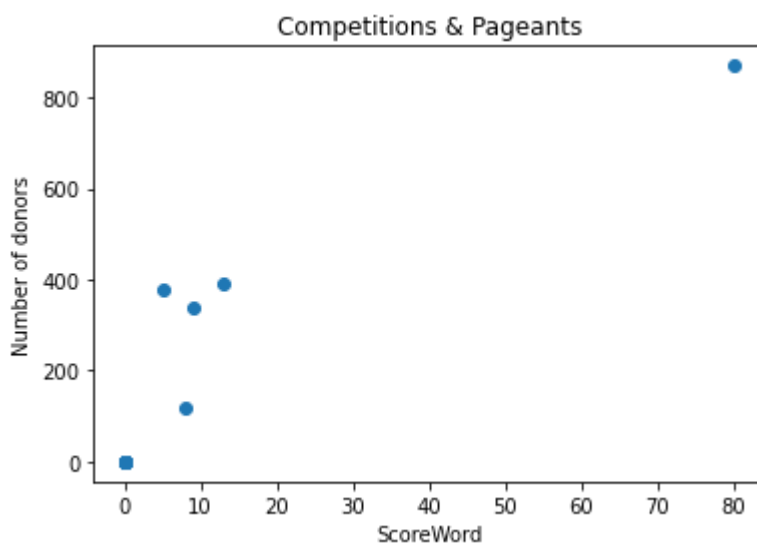
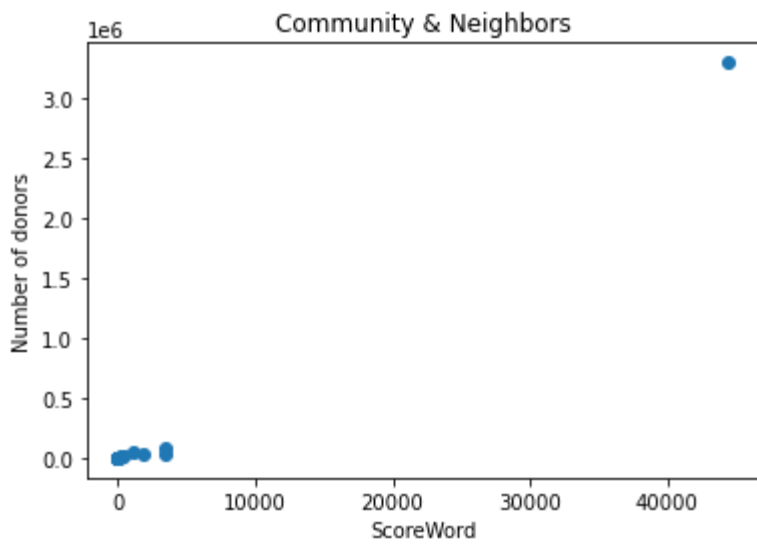
    x=december.loc[(december['Categorie'] ==cat)]['NumberDonors']
    y=december.loc[(december['Categorie'] ==cat)]['Amount collected ' ]

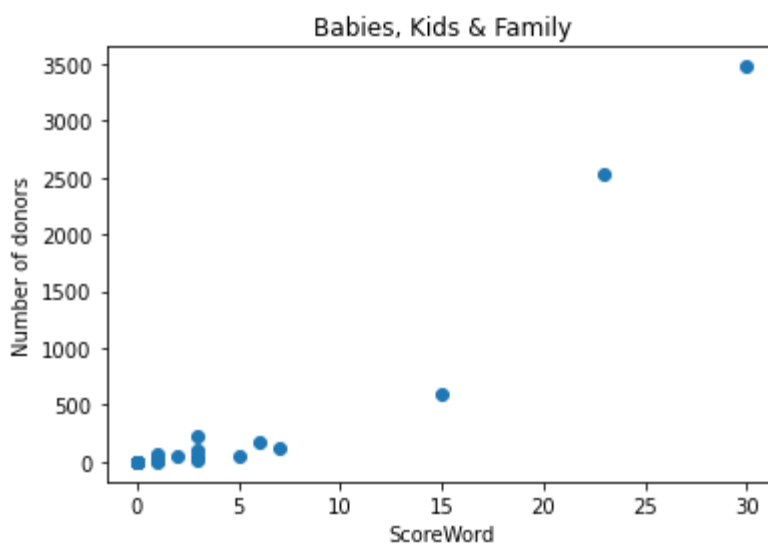
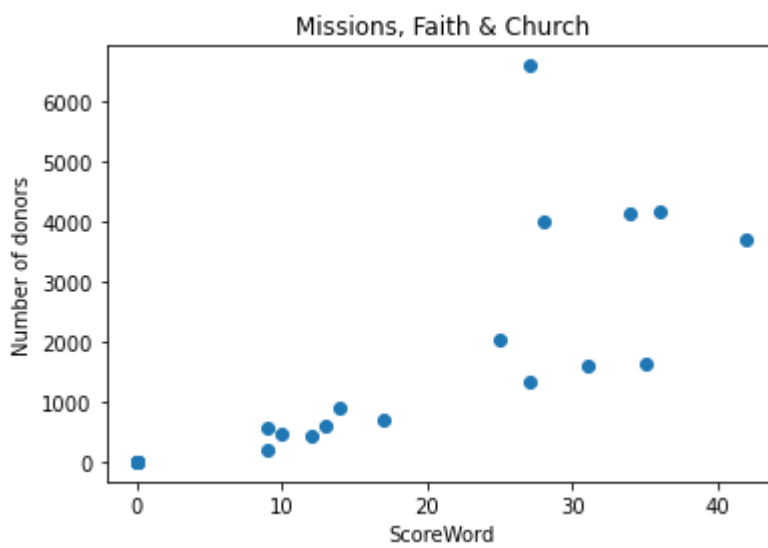
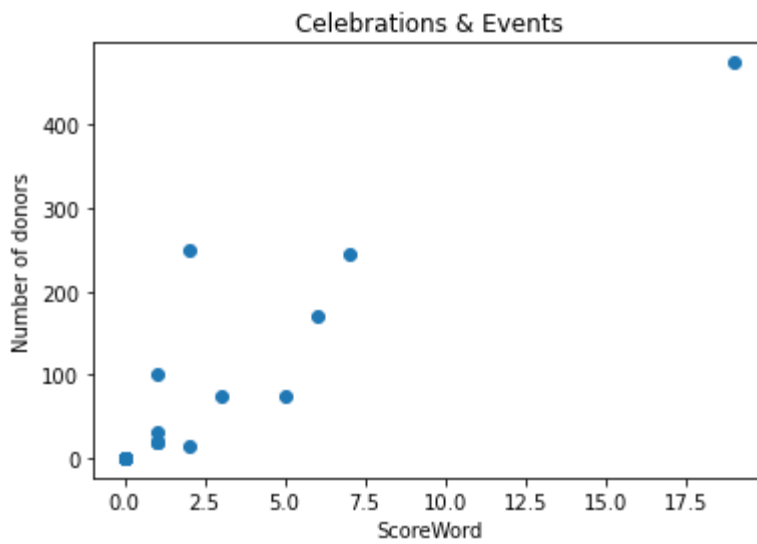
    fig=plt.scatter(x, y )
    plt.title(cat)
    plt.xlabel('ScoreWord')
    plt.ylabel('Number of donors')
```

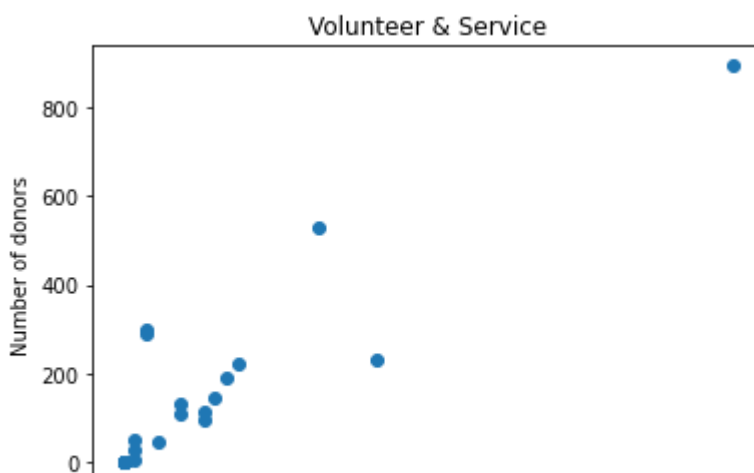
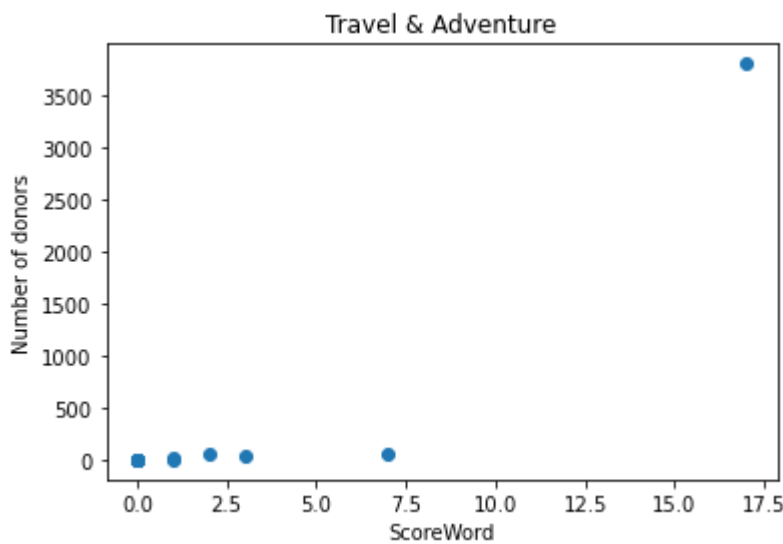
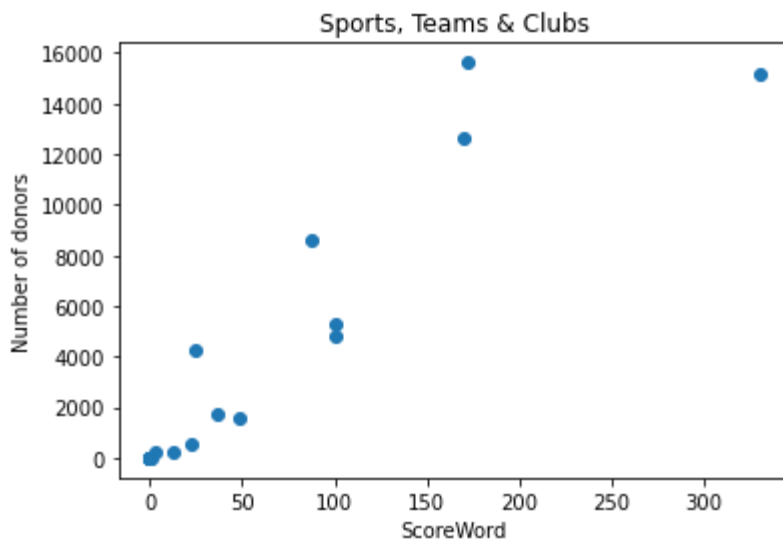












In [740...

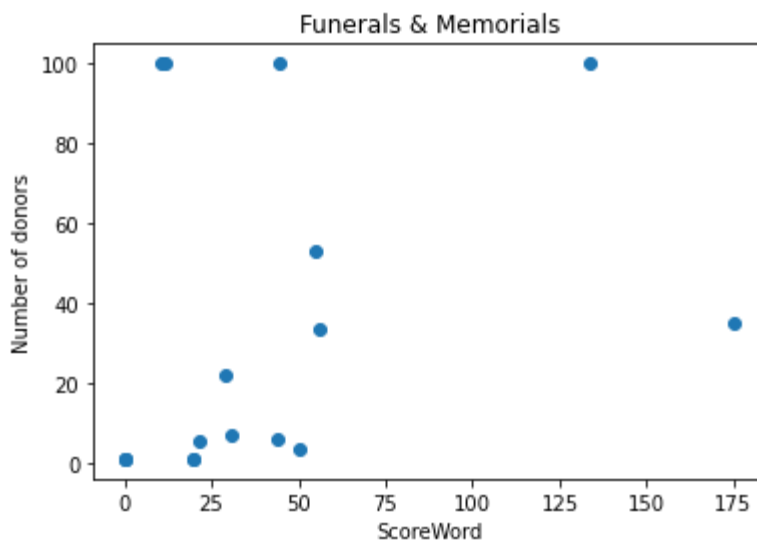
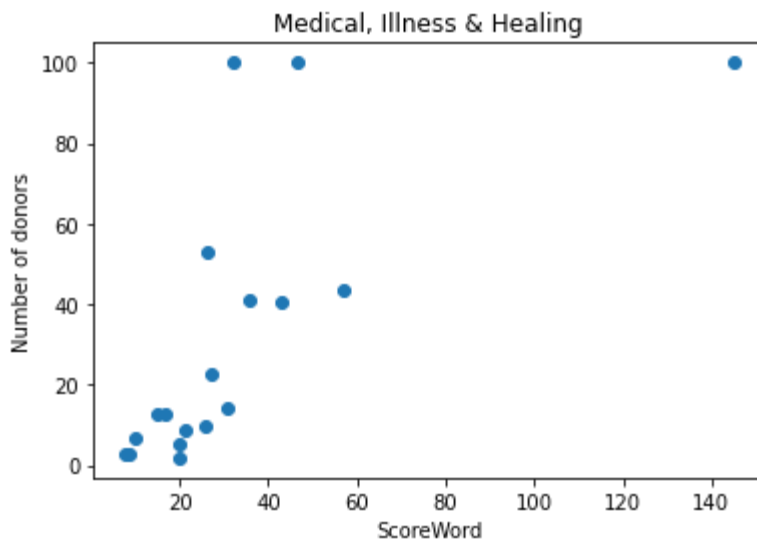
```

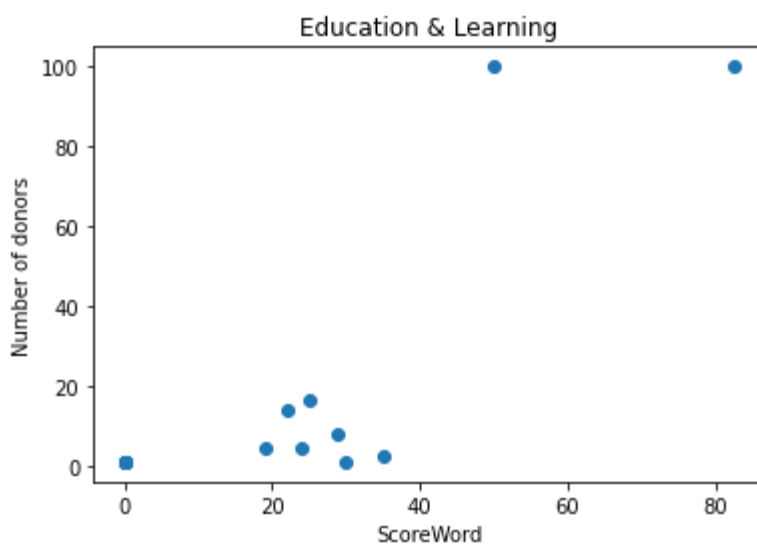
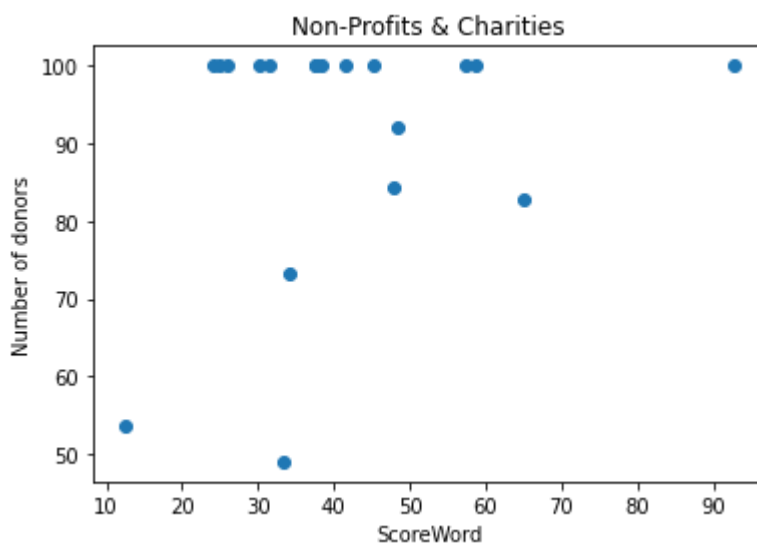
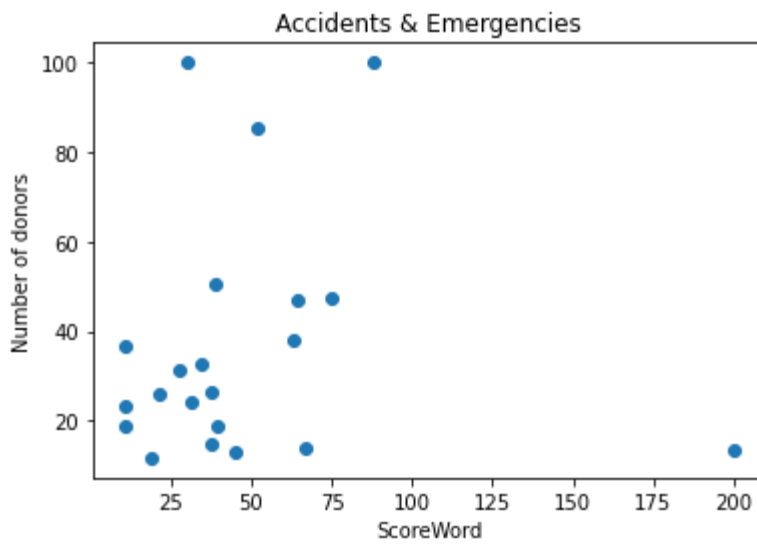
#SmallCap
decemberSmallCap=df.loc[mask_december&mask_SmallCap]
#print(decemberSmallCap.corr())
for cat in catégorie:
    plt.figure()

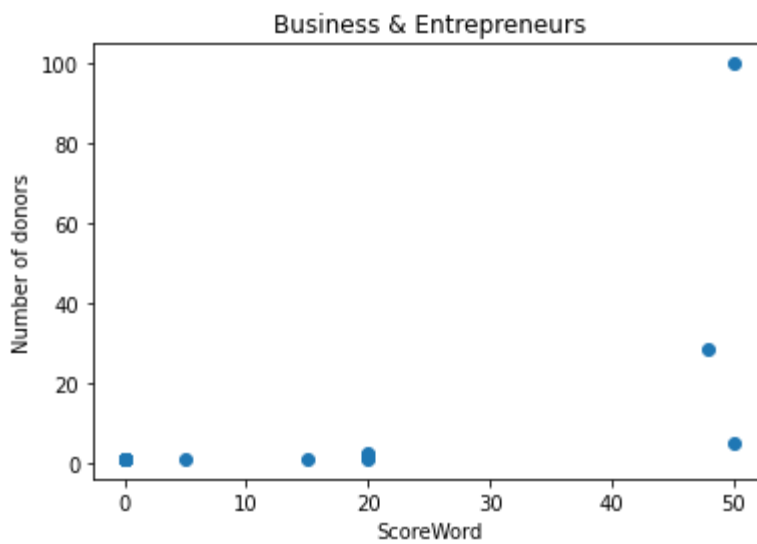
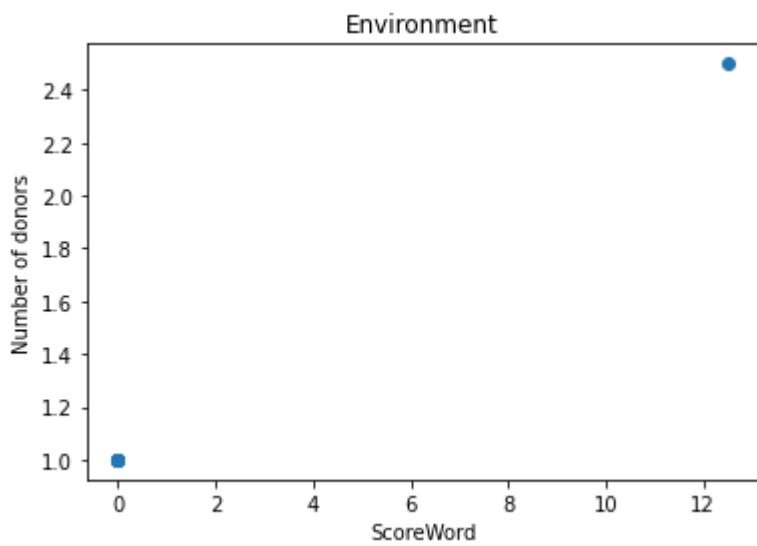
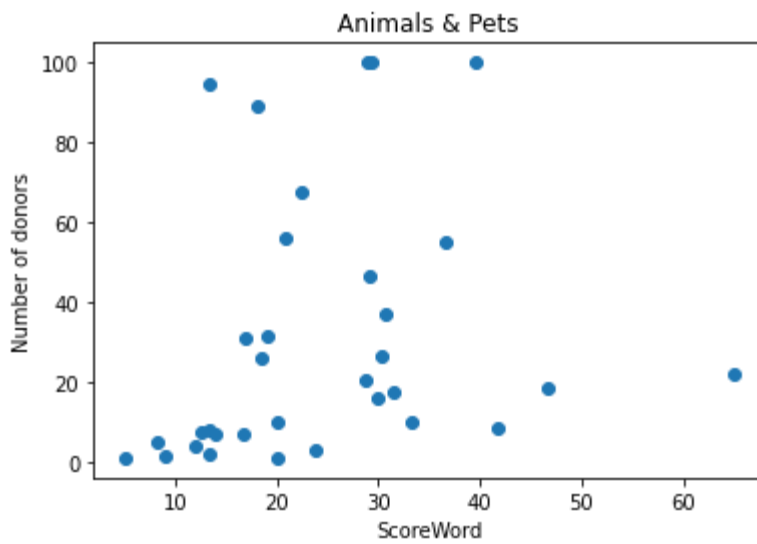
    x=decemberSmallCap.loc[(decemberSmallCap['Catégorie'] ==cat)][['MeanDonati
    y=decemberSmallCap.loc[(decemberSmallCap['Catégorie'] ==cat)][['Pourcentag

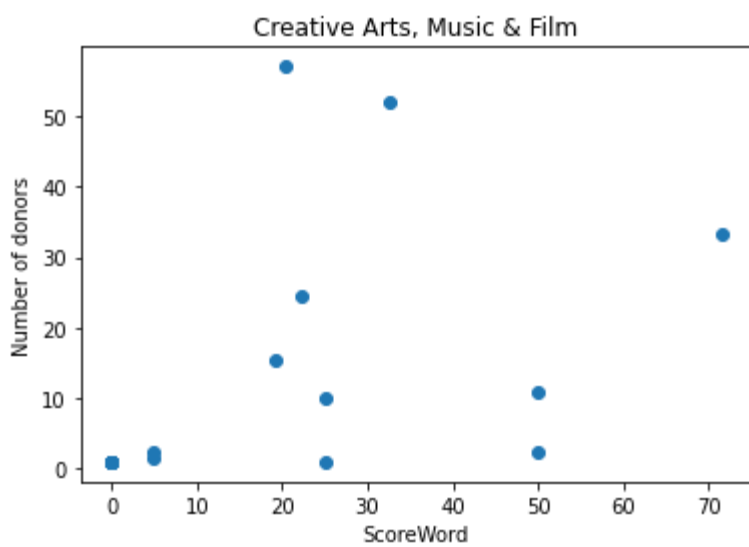
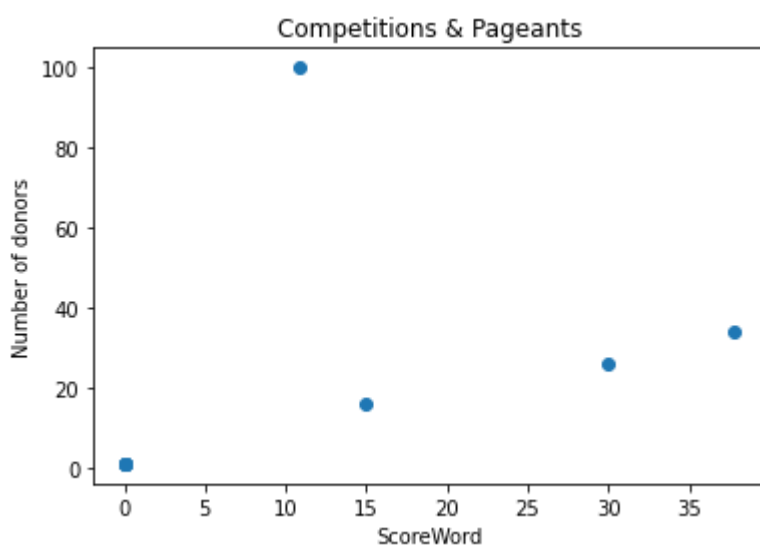
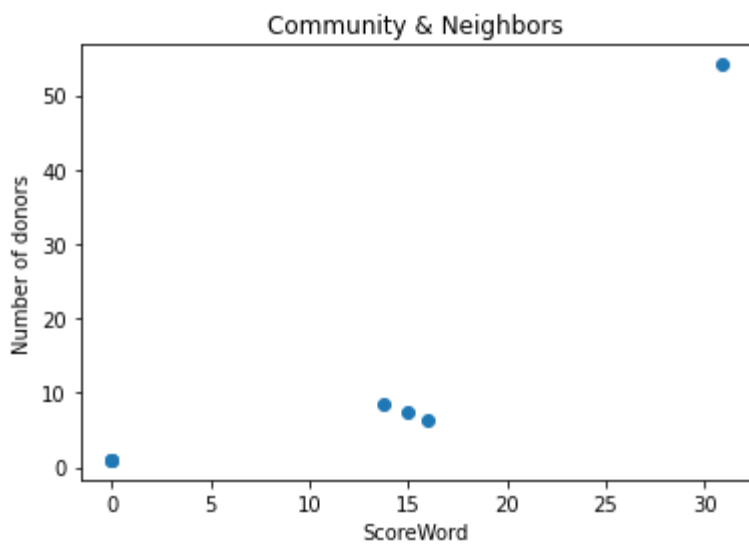
    fig=plt.scatter(x, y )
    plt.title(cat)
    plt.xlabel('ScoreWord')
    plt.ylabel('Number of donors')

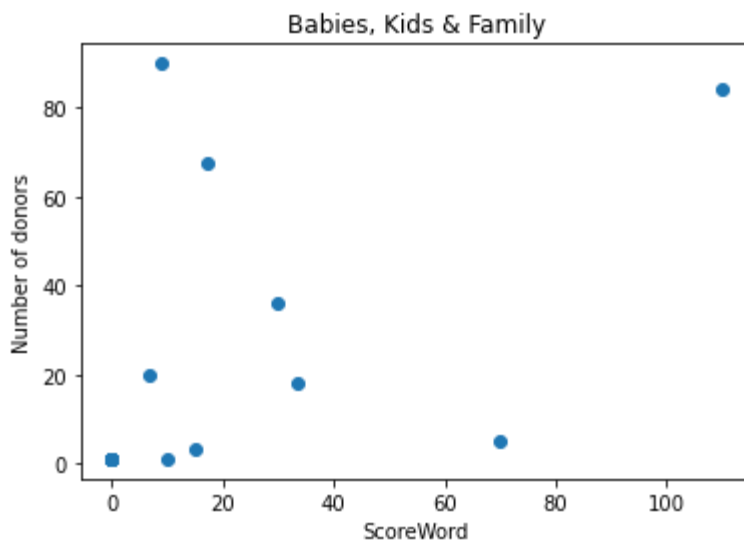
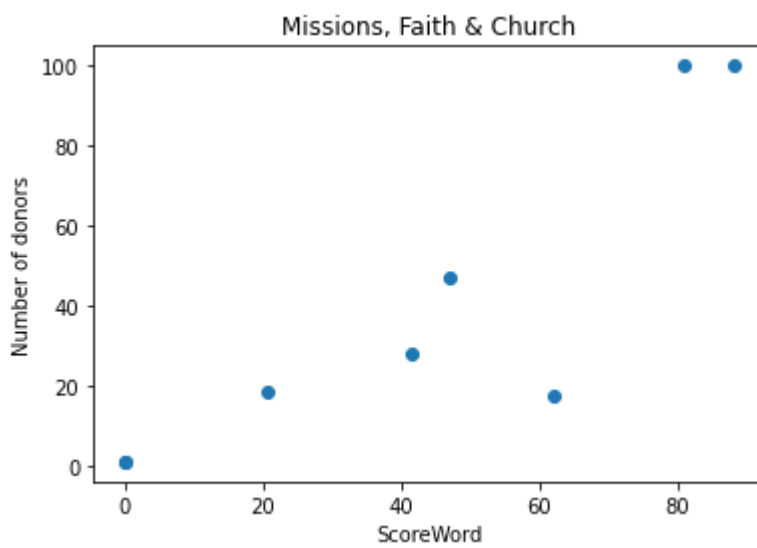
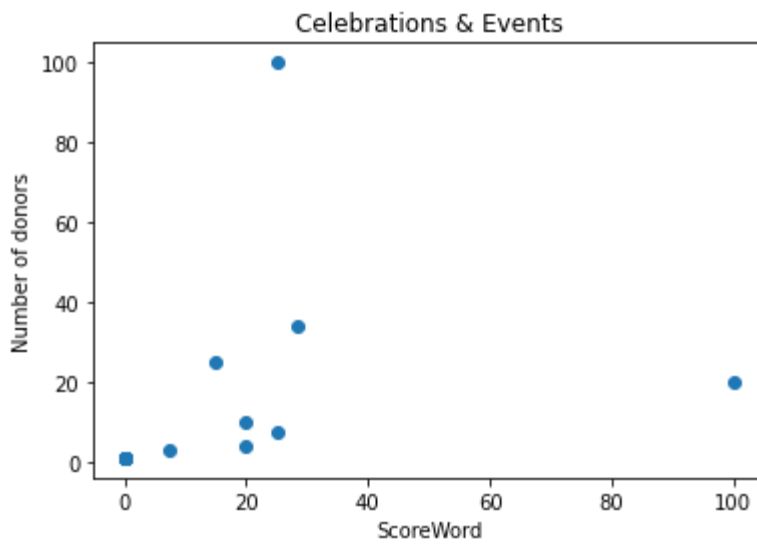
```

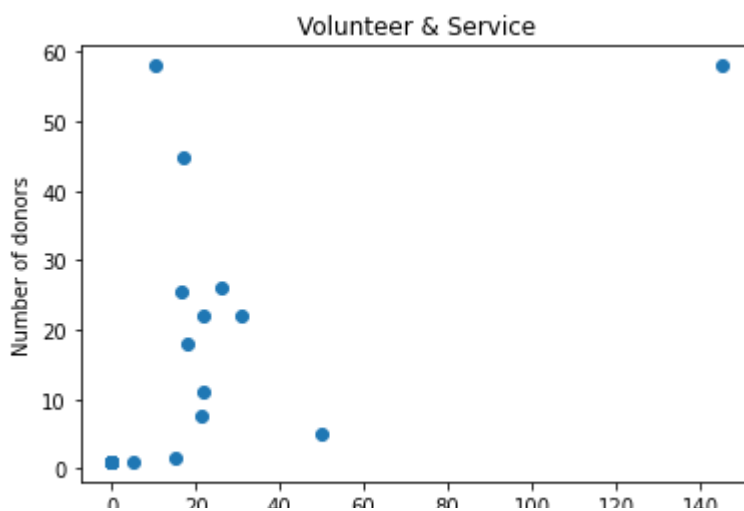
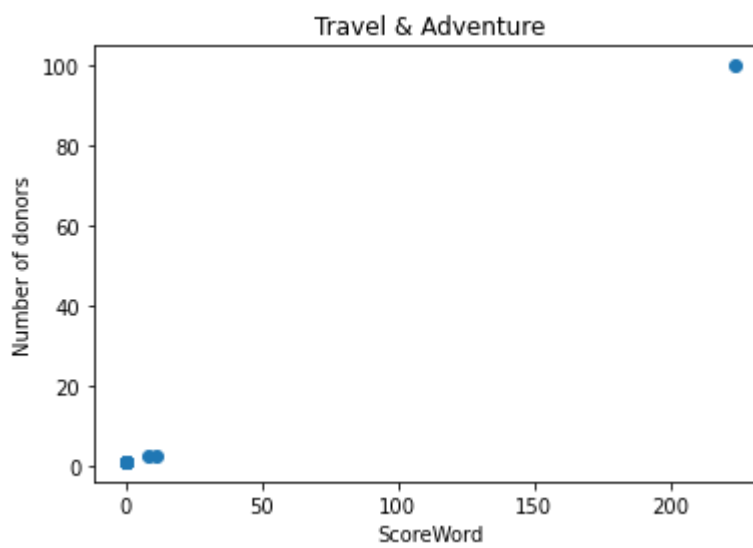
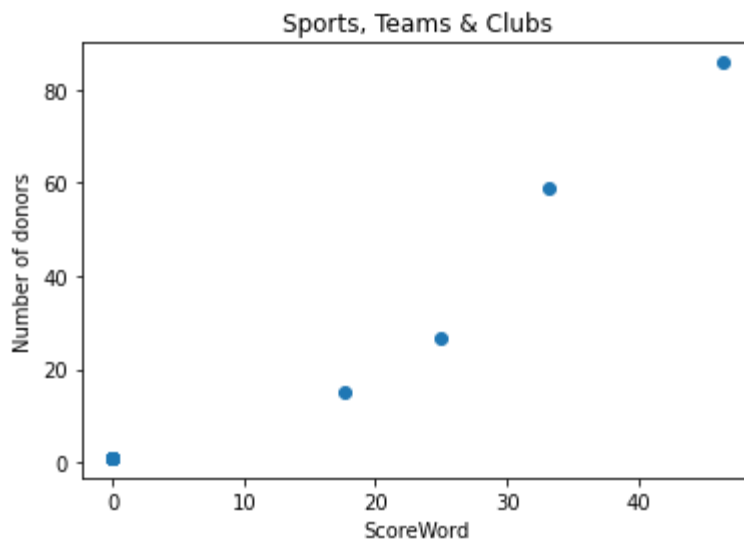












Relation between Pourcentage Raised and MeanDonation for each cap

In [741]...

```
#SmallCap
decemberSmallCap=df.loc[mask_december&mask_SmallCap]
plt.figure()

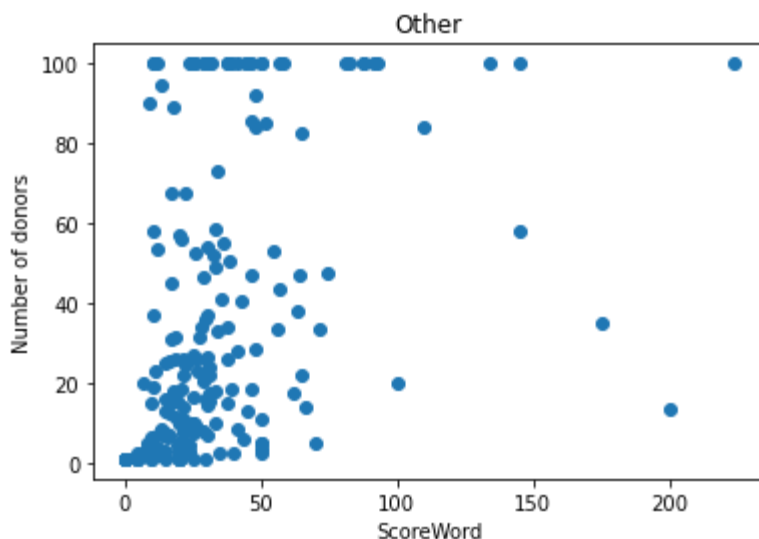
x=decemberSmallCap['MeanDonation']
y=decemberSmallCap['Pourcentage Raised']

fig=plt.scatter(x, y )
plt.title(cat)
plt.xlabel('ScoreWord')
plt.ylabel('Number of donors')

decemberSmallCap.corr()
```

Out[741]...

	Unnamed: 0	Pourcentage Raised	NumberDonors	Duration in days	Amount collected	Amoui targete
Unnamed: 0	1.000000	-0.399011	-0.282834	-0.108535	-0.272957	-0.22954
Pourcentage Raised	-0.399011	1.000000	0.570283	0.102381	0.614059	0.06277
NumberDonors	-0.282834	0.570283	1.000000	0.148741	0.940682	0.14366
Duration in days	-0.108535	0.102381	0.148741	1.000000	0.122893	0.07862
Amount collected	-0.272957	0.614059	0.940682	0.122893	1.000000	0.19828
Amount targeted	-0.229545	0.062775	0.143667	0.078624	0.198289	1.00000
MeanDonation	-0.349844	0.575202	0.208286	0.073083	0.371615	0.20803
categoryNumeric	0.998403	-0.387048	-0.280630	-0.105449	-0.270712	-0.22553
townNumeric	0.528655	-0.145016	-0.096483	-0.119439	-0.089499	-0.08373
ScoreWord	-0.124959	0.154265	0.058714	0.121605	0.081912	0.19501



In [742]...

```
#BigCap
decemberBigCap=df.loc[mask_december&mask_BigCap]
plt.figure()

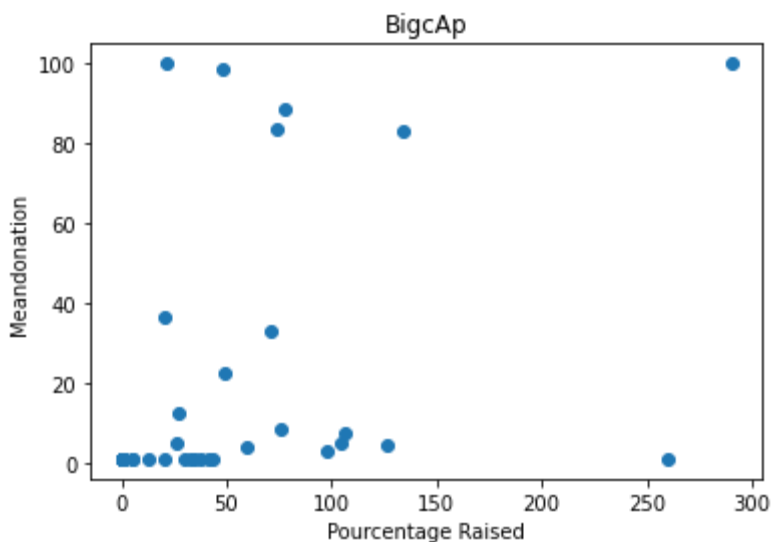
x=decemberBigCap['MeanDonation']
y=decemberBigCap['Pourcentage Raised']

fig=plt.scatter(x, y )
plt.title('BigcAp')
plt.xlabel('Pourcentage Raised')
plt.ylabel('Meandonation')

decemberBigCap.corr()
```

Out[742]...

	Unnamed: 0	Pourcentage Raised	NumberDonors	Duration in days	Amount collected	Amou targeted
Unnamed: 0	1.000000	-0.219787	0.000348	-0.191494	-0.004742	0.325467
Pourcentage Raised	-0.219787	1.000000	0.400304	-0.092646	0.385324	-0.103601
NumberDonors	0.000348	0.400304	1.000000	0.035005	0.997585	-0.038764
Duration in days	-0.191494	-0.092646	0.035005	1.000000	0.045787	-0.125230
Amount collected	-0.004742	0.385324	0.997585	0.045787	1.000000	-0.036612
Amount targeted	0.325467	-0.103601	-0.038764	-0.125230	-0.036612	1.000000
MeanDonation	-0.545189	0.459672	0.101394	0.000220	0.120569	-0.143537
categoryNumeric	0.998363	-0.230210	-0.010118	-0.178616	-0.014327	0.333142
townNumeric	0.519797	0.082650	0.134935	-0.045368	0.121935	-0.088371
ScoreWord	-0.308450	0.195202	-0.082512	-0.022870	-0.088408	-0.183241



In [743...

```
#SmallCap
decemberMidCap=df.loc[mask_december&mask_MidCap]
plt.figure()

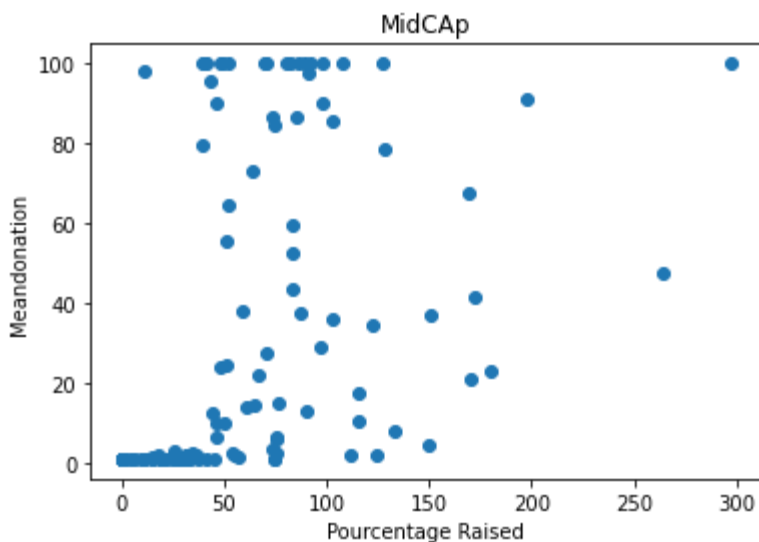
x=decemberMidCap['MeanDonation']
y=decemberMidCap['Pourcentage Raised']

fig=plt.scatter(x, y )
plt.title('MidCap')
plt.xlabel('Pourcentage Raised')
plt.ylabel('Meandonation')

decemberMidCap.corr()
```

Out[743...

	Unnamed: 0	Pourcentage Raised	NumberDonors	Duration in days	Amount collected	Amour targete
Unnamed: 0	1.000000	-0.336653	-0.142380	-0.001731	-0.332635	0.01842
Pourcentage Raised	-0.336653	1.000000	0.511479	-0.050785	0.879013	0.03572
NumberDonors	-0.142380	0.511479	1.000000	-0.037458	0.708378	0.22576
Duration in days	-0.001731	-0.050785	-0.037458	1.000000	-0.129860	-0.13190
Amount collected	-0.332635	0.879013	0.708378	-0.129860	1.000000	0.20021
Amount targeted	0.018428	0.035725	0.225769	-0.131904	0.200214	1.00000
MeanDonation	-0.298328	0.536859	0.059112	0.103832	0.390574	0.05999
categoryNumeric	0.998192	-0.355597	-0.156200	0.012318	-0.352957	0.01216
townNumeric	0.558099	-0.122648	-0.026966	-0.021303	-0.140713	-0.02386
ScoreWord	-0.179021	0.063950	-0.019471	0.082737	0.043328	-0.02733



OLS regression because strong correlation between NumberDonors and Amount

collected

In [737...

```
from sklearn import linear_model

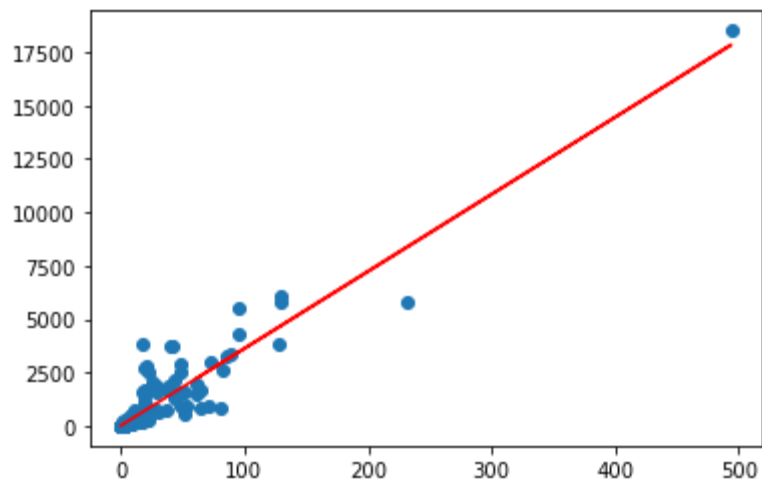
X=decemberSmallCap[['NumberDonors']]
y=decemberSmallCap["Amount collected "]

model=linear_model.LinearRegression()
results=model.fit(X,y)

y_pred = model.predict(X)
plt.scatter(X,y)
plt.plot(X,y_pred,color='r')

print(results.score(X, y, sample_weight=None))
```

0.8848828208290608



In []: