

Project_1_Yiyang_Zeng

February 6, 2019

Exploration on Different Categories of Happy Moments

Yiyang Zeng

For this project, I would like to explore different categories of happy moments, what might be the main drivers for these categories and how are categories different from each other.

The source of the dataset comes from a corpus of 100,000 crowd-sourced happy moments via Amazon's Mechanical Turk. Detailed information about the dataset is on <https://rit-public.github.io/HappyDB/>.

Step 0: Load All the Packages

Brief descriptions are as following:

Pandas: **pandas** is a Python package providing fast, flexible, and expressive data structures designed to make working with "relational" or "labeled" data both easy and intuitive.

Numpy: **numpy** is a package providing an array object of arbitrary homogeneous items, fast mathematical operations over arrays and linear algebra, fourier transforms, random number generation.

PIL: **PIL** is a package for working with images.

Matplotlib.pyplot: **matplotlib.pyplot** is a state-based interface to matplotlib. It provides a MATLAB-like way of plotting. Pyplot is mainly intended for interactive plots and simple cases of programmatic plot generation.

wordcloud: **wordcloud** is a package for generating wordclouds.

gensim: **gensim** contains interfaces and functionality to compute pair-wise document similarities within a corpus of documents.

seaborn: **seaborn** is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.

```
In [61]: import pandas as pd
         from pandas import DataFrame
         pd.set_option('display.max_columns', 50)
         import numpy as np
         from PIL import Image
         import PIL
         import matplotlib.pyplot as plt
         %matplotlib inline
         from wordcloud import WordCloud, STOPWORDS
         import gensim
         from gensim import corpora, models
         from gensim.models import LdaModel
         from gensim.parsing.preprocessing import remove_stopwords, strip_punctuation, strip_num
         import seaborn as sns
```

Step 1: Load the Dataset

The dataset I will be using is called `processed_moments.csv`. It is a preprocessed version of the HappyDB dataset, created by Arpita Shah and Tian Zheng. Some of the more important columns for this project are `cleaned_hm`, `predicted_category` and `text`.

`cleaned_hm`: This is the cleaned version of the happy moments written by people.

`predicted_category`: This is the predicted category of the happy moments and it is the most important thing that I care about in this project.

`text`: This is the lists of informative words extracted from `cleaned_hm`.

```
In [12]: # processed_moments is under data folder
df = pd.read_csv("processed_moments.csv")
df
```

```
Out[12]:
```

	hmid	wid	reflection_period	\
0	27673	2053		24h
1	27674	2		24h
2	27675	1936		24h
3	27676	206		24h
4	27677	6227		24h
5	27678	45		24h
6	27679	195		24h
7	27680	740		24h
8	27681	3		24h
9	27682	4833		24h
10	27683	7334		24h
11	27684	78		24h
12	27685	21		24h
13	27686	8		24h
14	27687	586		24h
15	27688	489		24h
16	27689	976		24h
17	27690	3972		24h
18	27691	14		24h
19	27692	1230		24h
20	27693	9950		24h
21	27694	4862		24h
22	27695	4891		24h
23	27696	737		24h
24	27697	498		24h
25	27698	2124		24h
26	27699	156		24h
27	27700	1783		24h
28	27701	3382		24h
29	27702	54		24h
...
100362	128737	1161		24h
100363	128738	2780		24h
100364	128739	2545		24h

100365	128740	10983	24h
100366	128741	3948	24h
100367	128742	525	24h
100368	128743	13605	24h
100369	128744	5180	24h
100370	128745	1234	24h
100371	128746	248	24h
100372	128747	2974	24h
100373	128748	8880	24h
100374	128749	2616	24h
100375	128750	323	24h
100376	128751	11402	24h
100377	128752	301	24h
100378	128753	597	24h
100379	128754	105	24h
100380	128755	22	24h
100381	128756	412	24h
100382	128757	1767	24h
100383	128758	4428	24h
100384	128759	11050	24h
100385	128760	5	24h
100386	128761	4298	24h
100387	128762	112	24h
100388	128763	714	24h
100389	128764	3934	24h
100390	128765	1629	24h
100391	128766	141	24h

original_hm \

```

0      I went on a successful date with someone I fel...
1      I was happy when my son got 90% marks in his e...
2          I went to the gym this morning and did yoga.
3      We had a serious talk with some friends of our...
4      I went with grandchildren to butterfly display...
5          I meditated last night.
6      I made a new recipe for peasant bread, and it ...
7      I got gift from my elder brother which was rea...
8          YESTERDAY MY MOMS BIRTHDAY SO I ENJOYED
9      Watching cupcake wars with my three teen children
10     I came in 3rd place in my Call of Duty video g...
11     I completed my 5 miles run without break. It m...
12         went to movies with my friends it was fun
13     I was shorting Gold and made $200 from the trade.
14     Hearing Songs It can be nearly impossible to g...
15     My son performed very well for a test preparat...
16         I helped my neighbour to fix their car damages.
17     Managed to get the final trophy in a game I wa...
18     A hot kiss with my girl friend last night made...

```

19 My new BCAAs came in the mail. Yay! Strawberry...
 20 Got A in class.
 21 My sister called me from abroad this morning a...
 22 The cake I made today came out amazing. It tas...
 23 There are two types of people in the world: th...
 24 My grandmother start to walk from the bed afte...
 25 i was able to hit a top spin serve in tennis
 26 I napped with my husband on the bed this after...
 27 My co-woker started playing a Carley Rae Jepse...
 28 My son woke me up to a fantastic breakfast of ...
 29 After a long time my brother gave a suprise vi...
 ...
 100362 DRY DAY CELEBRATION
 100363 My bills were paid on time
 100364 I watched my favorite gamer on Twitch.tv and h...
 100365 I was able to accomplish one of my goals at wo...
 100366 I heard a song I really like on the radio this...
 100367 A friend called just to see how i'd been doing.
 100368 Last night while attending my grandmother's wa...
 100369 My friend came to my home and he give some gif...
 100370 I went shopping for Father's Day dinner on Sun...
 100371 I woke up in the middle of the night and reali...
 100372 I had pizza for dinner and it was great
 100373 My husband called me just to tell me he loved me.
 100374 I was able to weed my entire garden and plant ...
 100375 A coworker praised the job I did on a recent p...
 100376 Fixed my car.
 100377 I had a little time to myself in the afternoon.
 100378 I worked out, which always makes me feel good.
 100379 Finally got to watch the new Resident Evil movie.
 100380 I made plans to go to a new restaurant with th...
 100381 Those rare, perfectly timed naps that leave yo...
 100382 I got to talk to an old friend and reminisce o...
 100383 Yesterday my relations came to my house. That ...
 100384 I am fond of seeing cricket match. Yesterday t...
 100385 My daughter made a very funny sound when I was...
 100386 I spent time with my daughter.
 100387 My husband announced he is getting a decent bo...
 100388 Had a can of Pepsi to drink.
 100389 Cuddling with my girlfriend last night.
 100390 I had a great meeting yesterday at work with m...
 100391 I had a great workout last night.

		cleaned_hm	modified \
0	I went on a successful date with someone I fel...		True
1	I was happy when my son got 90% marks in his e...		True
2	I went to the gym this morning and did yoga.		True
3	We had a serious talk with some friends of our...		True

4	I went with grandchildren to butterfly display...	True
5	I meditated last night.	True
6	I made a new recipe for peasant bread, and it ...	True
7	I got gift from my elder brother which was rea...	True
8	YESTERDAY MY MOMS BIRTHDAY SO I ENJOYED	True
9	Watching cupcake wars with my three teen children	True
10	I came in 3rd place in my Call of Duty video g...	True
11	I completed my 5 miles run without break. It m...	True
12	went to movies with my friends it was fun	True
13	I was shorting Gold and made \$200 from the trade.	True
14	Hearing Songs It can be nearly impossible to g...	False
15	My son performed very well for a test preparat...	True
16	I helped my neighbour to fix their car damages.	True
17	Managed to get the final trophy in a game I wa...	True
18	A hot kiss with my girl friend last night made...	True
19	My new BCAAs came in the mail. Yay! Strawberry...	True
20	Got A in class.	True
21	My sister called me from abroad this morning a...	True
22	The cake I made today came out amazing. It tas...	True
23	There are two types of people in the world: th...	False
24	My grandmother start to walk from the bed afte...	True
25	i was able to hit a top spin serve in tennis	True
26	I napped with my husband on the bed this after...	True
27	My co-woker started playing a Carley Rae Jepse...	True
28	My son woke me up to a fantastic breakfast of ...	True
29	After a long time my brother gave a suprise vi...	True
...
100362	DRY DAY CELEBRATION	True
100363	My bills were paid on time	True
100364	I watched my favorite gamer on Twitch.tv and h...	True
100365	I was able to accomplish one of my goals at wo...	True
100366	I heard a song I really like on the radio this...	True
100367	A friend called just to see how i'd been doing.	True
100368	Last night while attending my grandmother's wa...	True
100369	My friend came to my home and he give some gif...	True
100370	I went shopping for Father's Day dinner on Sun...	True
100371	I woke up in the middle of the night and reali...	True
100372	I had pizza for dinner and it was great	True
100373	My husband called me just to tell me he loved me.	True
100374	I was able to weed my entire garden and plant ...	True
100375	A coworker praised the job I did on a recent p...	True
100376	Fixed my car.	True
100377	I had a little time to myself in the afternoon.	True
100378	I worked out, which always makes me feel good.	True
100379	Finally got to watch the new Resident Evil movie.	True
100380	I made plans to go to a new restaurant with th...	True
100381	Those rare, perfectly timed naps that leave yo...	True
100382	I got to talk to an old friend and reminisce o...	True

100383	Yesterday my relations came to my house. That ...	True
100384	I am fond of seeing cricket match. Yesterday t...	True
100385	My daughter made a very funny sound when I was...	True
100386	I spent time with my daughter.	True
100387	My husband announced he is getting a decent bo...	True
100388	Had a can of Pepsi to drink.	True
100389	Cuddling with my girlfriend last night.	True
100390	I had a great meeting yesterday at work with m...	True
100391	I had a great workout last night.	True

	num_sentence	ground_truth_category	predicted_category	id \
0	1	NaN	affection	1
1	1	NaN	affection	2
2	1	NaN	exercise	3
3	2	bonding	bonding	4
4	1	NaN	affection	5
5	1	leisure	leisure	6
6	1	NaN	achievement	7
7	1	NaN	affection	8
8	1	NaN	enjoy_the_moment	9
9	1	NaN	affection	10
10	1	NaN	leisure	11
11	2	NaN	exercise	12
12	1	NaN	bonding	13
13	1	NaN	achievement	14
14	2	NaN	enjoy_the_moment	15
15	1	NaN	affection	16
16	1	NaN	bonding	17
17	1	NaN	achievement	18
18	1	NaN	bonding	19
19	3	NaN	affection	20
20	1	NaN	achievement	21
21	2	NaN	affection	22
22	2	NaN	achievement	23
23	2	NaN	enjoy_the_moment	24
24	1	affection	affection	25
25	1	NaN	achievement	26
26	1	NaN	affection	27
27	1	NaN	leisure	28
28	1	NaN	affection	29
29	1	NaN	affection	30
...
100362	1	NaN	achievement	100506
100363	1	NaN	achievement	100507
100364	1	NaN	leisure	100508
100365	1	NaN	achievement	100509
100366	1	NaN	enjoy_the_moment	100510
100367	1	NaN	bonding	100511

100368	1	NaN	bonding	100512
100369	1	NaN	bonding	100513
100370	1	NaN	affection	100514
100371	1	enjoy_the_moment	enjoy_the_moment	100515
100372	1	NaN	enjoy_the_moment	100516
100373	1	NaN	affection	100517
100374	1	NaN	nature	100518
100375	1	NaN	achievement	100519
100376	1	NaN	achievement	100520
100377	1	NaN	enjoy_the_moment	100521
100378	1	NaN	enjoy_the_moment	100522
100379	1	NaN	leisure	100523
100380	1	NaN	bonding	100524
100381	3	NaN	achievement	100525
100382	1	NaN	bonding	100526
100383	2	affection	affection	100527
100384	6	NaN	enjoy_the_moment	100528
100385	1	NaN	affection	100529
100386	1	NaN	affection	100530
100387	1	NaN	affection	100531
100388	1	NaN	enjoy_the_moment	100532
100389	1	affection	affection	100533
100390	1	NaN	bonding	100534
100391	1	NaN	exercise	100535

	text
0	successfully date sympathy connected
1	son marks examination
2	gym morning yoga
3	talked friend flaky understood evening hang
4	grandchildren butterfly display crohn conserva...
5	meditation night
6	recipe peasant bread spectacular
7	gift elderly brother surprise
8	mom birthday enjoyed
9	watched cupcakes wars teen children
10	called duty video game
11	completed miles run break makes feel strong
12	movie friend fun
13	short gold trade
14	hear song impossible angry youre ease angry fe...
15	son performance test prepared
16	helped neighbours fixed car damage
17	managed finally trophy game played
18	hot kiss girl friend night day
19	bcaa mail yay strawberry lemonade flavor amino...
20	class
21	sister called abroad morning occassion family

```

22             cake amazing tasted amazing
23     type people world choose choose unhappy contra...
24             grandmother started walk bed time
25             hit top spin served tennis
26     nap husband bed afternoon sweet cuddled close
27     coworker started played carley rae jepsen song ...
28     son woke fantastic breakfast eggs special hamb...
29             time brother suprise visit house
...
100362             dry day celebrated
100363             bill paid time
100364     watched favorite gamer twitchtv hilarious
100365     accomplished goal customer thanked profusely
100366             heard song radio morning coffee
100367             friend called id
100368     night attended grandmother wake enjoyed friend...
100369             friend home gift energy project
100370             shopping father day dinner sunday
100371     woke middle night realized hours sleep left
100372             pizza dinner
100373             husband called love
100374     weed entire garden planted tomato basil rain
100375     coworker praised job recently project supervisor
100376             fixed car
100377             time afternoon
100378             makes feel
100379             finally watched residence evil movie
100380             planned restaurant friend havent
100381     rare perfect time nap leave feel refreshed gro...
100382             talked friend reminisce time
100383             relatives house time
100384     fond cricket match semi finally champions trop...
100385     daughter funny sound change diaper night laugh
100386             spent time daughter
100387             husband announced decent bonus quarter
100388             pepsi drink
100389             cuddled girlfriend night
100390     meet boss colleagues lunch afterward excited p...
100391             workout night

```

```
[100392 rows x 11 columns]
```

Step 2: Preliminary Data Exploration

`df.info()` enables us to take a closer look at the columns in the dataset as well as their data types.

```
In [3]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```



```

RangeIndex: 100392 entries, 0 to 100391
Data columns (total 11 columns):
hmid          100392 non-null int64
wid           100392 non-null int64
reflection_period  100392 non-null object
original_hm    100392 non-null object
cleaned_hm     100392 non-null object
modified       100392 non-null bool
num_sentence   100392 non-null int64
ground_truth_category  14107 non-null object
predicted_category  100392 non-null object
id            100392 non-null int64
text          100391 non-null object
dtypes: bool(1), int64(4), object(6)
memory usage: 7.8+ MB

```

By the following function, we see there are seven unique categories of happy moments in the dataset. And they are: affection, exercise, bonding, leisure, achievement, enjoy_the_moment and nature.

```
In [7]: df["predicted_category"].unique()
```

```
Out[7]: array(['affection', 'exercise', 'bonding', 'leisure', 'achievement',
              'enjoy_the_moment', 'nature'], dtype=object)
```

Step 3: A Look at the Number of Sentences by Categories

First, I would like to see the relationship between number of sentences and different categories of happy moments. Thus, I draw seven boxplots on one graph as following, and my findings are:

There are many outliers in terms of number of sentences in the dataset, which makes the boxplots hard to see.

Relatively speaking, the category exercise and leisure seem to be stable, in terms of number of sentences.

Enjoy_the_moment and affections have large ranges of number of sentences.

```

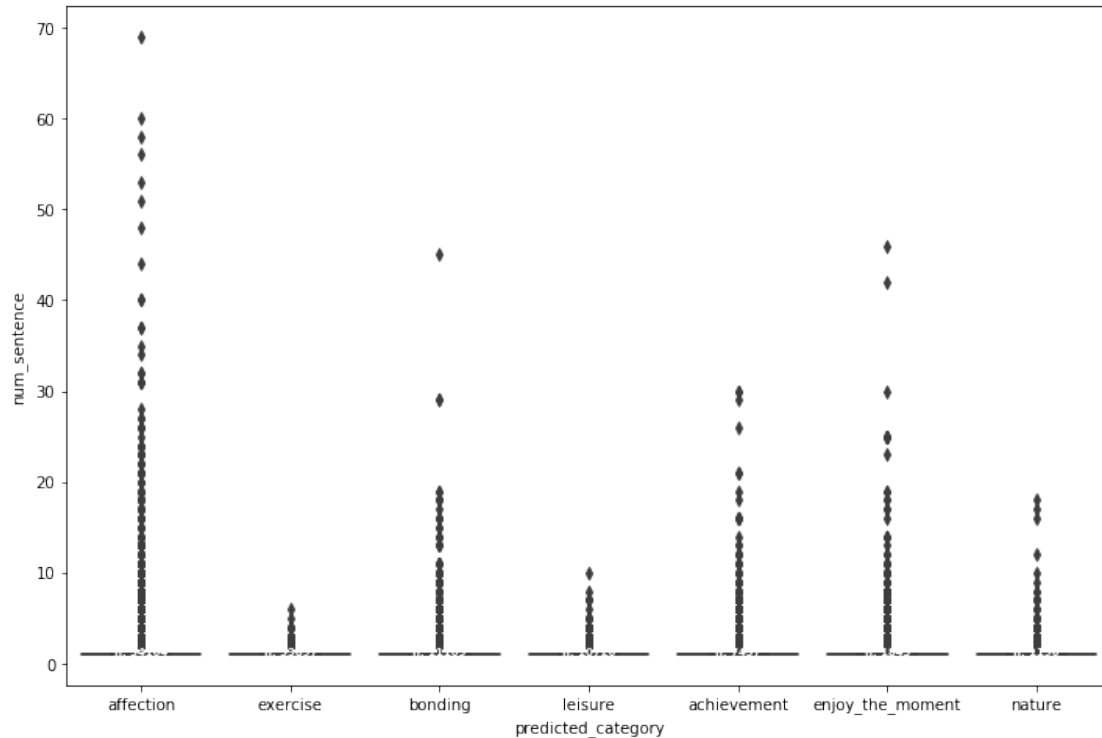
In [22]: plt.figure(figsize = (12,8))
         ax = sns.boxplot(x="predicted_category", y="num_sentence", data=df)

         # Calculate number of obs per group & median to position labels
         medians = df.groupby(['predicted_category'])['num_sentence'].median().values
         nobs = df['predicted_category'].value_counts().values
         nobs = [str(x) for x in nobs.tolist()]
         nobs = ["n: " + i for i in nobs]

         pos = range(len(nobs))
         for tick,label in zip(pos,ax.get_xticklabels()):
             ax.text(pos[tick], medians[tick] + 0.03, nobs[tick],
                     horizontalalignment='center', size='x-small', color='w', weight='semibold')

         plt.show()

```



To be able to better visualize the number of sentences by categories. I also calculate their means. From the following chart, we can see that average speaking:

Enjoy_the_moment has a mean of 1.48, the highest one out of seven. People like to talk more about enjoying certain moments.

Leisure only has a mean of 1.1, the lowest one out of seven. It might be hard for people to elaborate more on leisure, as most leisure activities can be easily described in less amounts of sentences.

```
In [86]: sen = df.groupby('predicted_category', as_index=False)['num_sentence'].mean()
sen
```

```
Out[86]: predicted_category  num_sentence
0      achievement         1.227483
1      affection          1.460455
2      bonding            1.366959
3  enjoy_the_moment       1.483032
4      exercise           1.144649
5      leisure            1.100979
6      nature             1.319045
```

Using the means of number of sentences calculated above, I also draw its corresponding barplot below. The plot helps visualize the difference in means.

```
In [106]: bars1 = list(sen["num_sentence"])
plt.figure(figsize = (9,7))
```

```

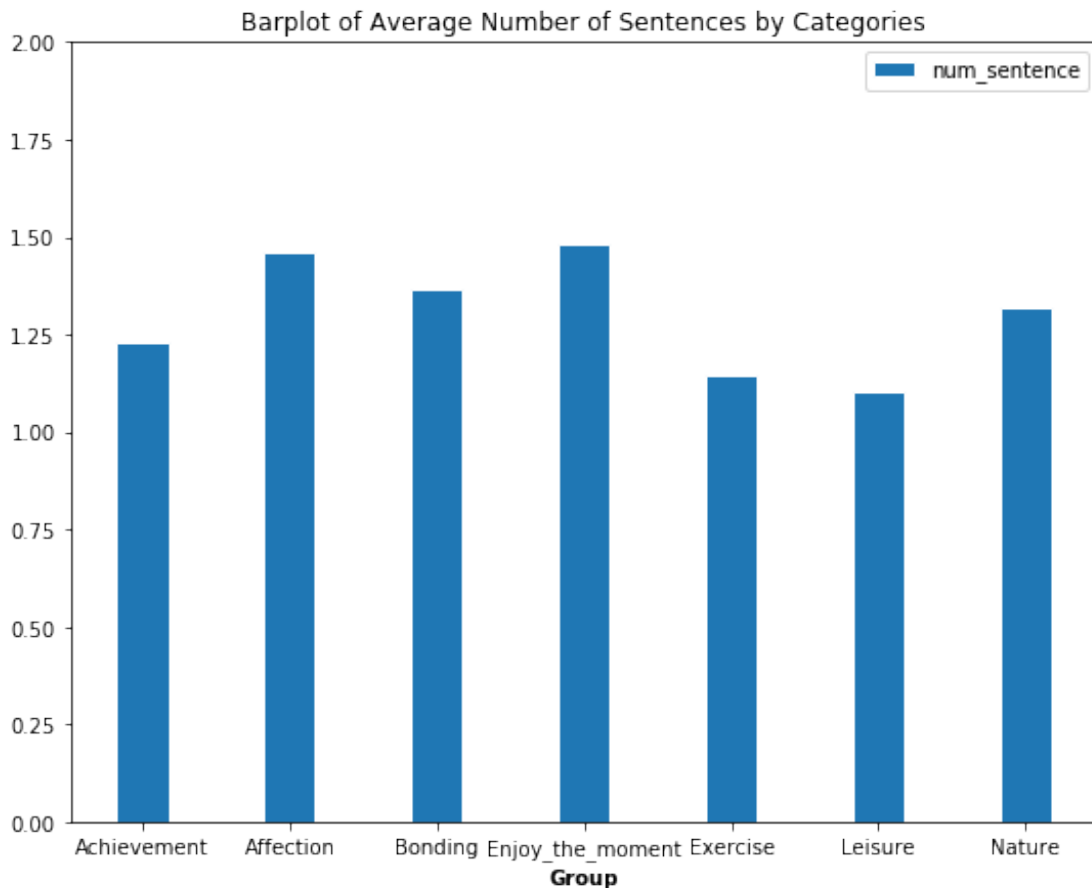
# Set position of bar on X axis
r1 = np.arange(len(bars1))

# Make the plot
plt.bar(r1, bars1, width=0.35, edgecolor='white', label='num_sentence')

# Add xticks on the middle of the group bars
plt.xlabel('Group', fontweight='bold')
plt.xticks([r for r in range(len(bars1))], ['Achievement', 'Affection', 'Bonding', 'Enjoy_the_moment', 'Exercise', 'Leisure', 'Nature'])

# Create legend & Show graphic
plt.title("Barplot of Average Number of Sentences by Categories")
plt.ylim(0,2)
plt.legend()
plt.show()

```



Step 4: Word Cloud

By common sense, I think the seven categories of happy moments should all contain different sets of frequent words, as they are describing different happy moments. Therefore, I decide to

draw word clouds by seven categories to find some of the frequent words.

I start by drawing a word cloud for the whole sample based on the text column of the data set. I feel the bigger words in the word cloud are either people ("friend", "family", "wife", "husband") or terms related to time("time", "finally", "first"):

When thinking about happy moments, people tend to think of some memorable events with people close to them, such as family, husband, wife, friend and so on.

Time is also important. Happy moments could be something people have done for the very first time or something they have been thinking about for long time.

*** I used a heart-shaped mask for all my word clouds, as I think a red heart can be a symbol of happiness. This mask image is under data folder.

```
In [28]: #draw a word cloud based on the text column of the whole dataset
plt.figure(figsize = (8,6))
#happy.png is under data folder
heart = np.array(Image.open("C:/Users/zengy/Desktop/2019 Spring Semester/Applied Data
stopwords = set(STOPWORDS)
stopwords.update(["made", "one", "day", "found"])
txt = " ".join(str(i) for i in df.text)
txt = txt.lower()
txt_wordcloud = WordCloud(stopwords=stopwords, background_color=(240,100,100), contour
                        mask = heart, contour_color='white', max_words=100).genera
plt.imshow(txt_wordcloud, interpolation="bilinear")
plt.title("Word Cloud Based on Text of the Whole Dataset")
plt.axis('off')
plt.show()
```


Some of the findings for the seven categories:

affection: things with family members or significant other are associated with affection happy moments.

enjoy_the_moment: enjoy_the_moment can be a lot of different things. People however have a tendency to associate the moment with enjoying food.

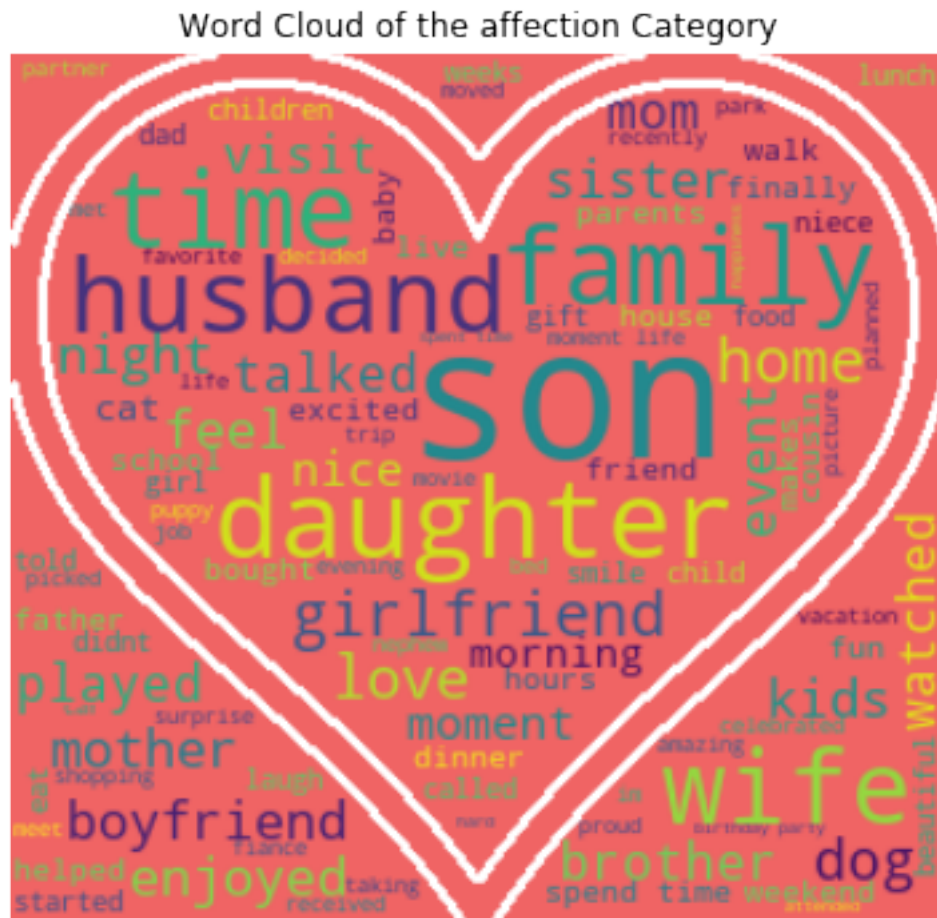
leisure: people like to mention movies, video games or vacation for leisure. These are possibly the most popular ways of leisure.

```
In [32]: word_cloud('achievement')
```



<Figure size 864x648 with 0 Axes>

```
In [33]: word_cloud('affection')
```



<Figure size 864x648 with 0 Axes>

```
In [34]: word_cloud('bonding')
```

Word Cloud of the bonding Category



<Figure size 864x648 with 0 Axes>

```
In [35]: word_cloud('enjoy_the_moment')
```


Word Cloud of the enjoy_the_moment Category



<Figure size 864x648 with 0 Axes>

```
In [36]: word_cloud('exercise')
```

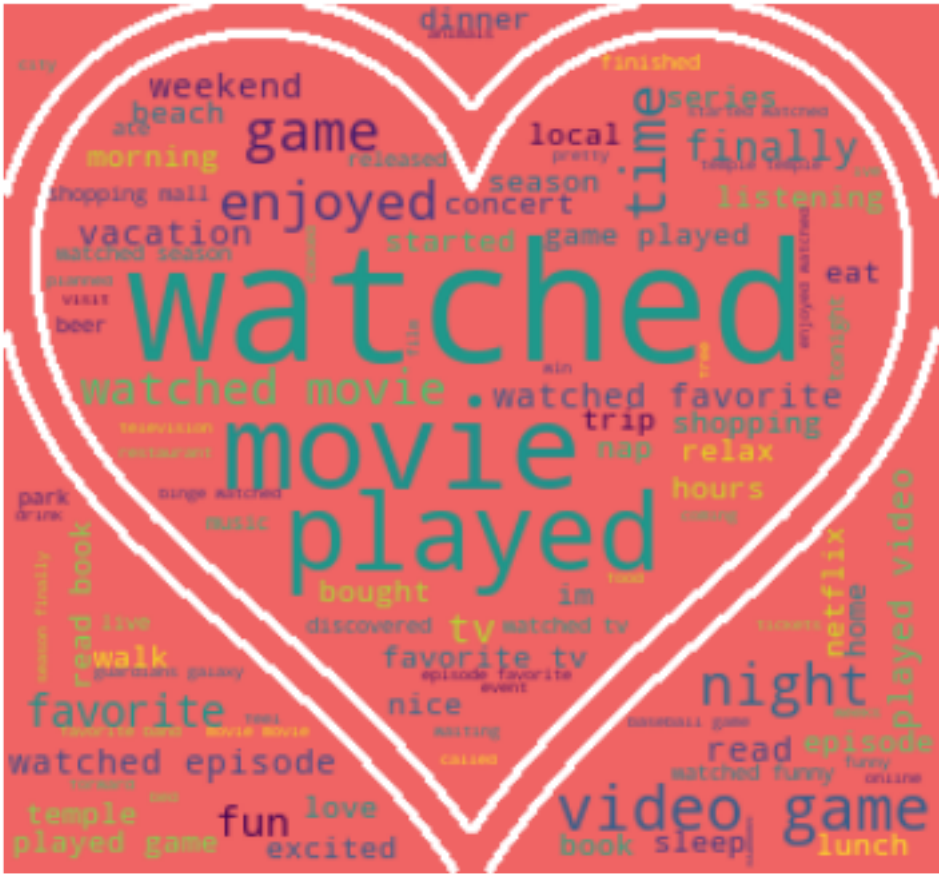
Word Cloud of the exercise Category



<Figure size 864x648 with 0 Axes>

```
In [37]: word_cloud('leisure')
```

Word Cloud of the leisure Category



<Figure size 864x648 with 0 Axes>

```
In [38]: word_cloud('nature')
```



```

with open(nrc, 'r') as f:
    all_lines = list()
    for line in f:
        if count < 46:
            count += 1
            continue
        line = line.strip().split('\t')
        if int(line[2]) == 1:
            if emotion_dict.get(line[0]):
                emotion_dict[line[0]].append(line[1])
            else:
                emotion_dict[line[0]] = [line[1]]
    return emotion_dict

emotion_dict = get_nrc_data()
emotions = {x for y in emotion_dict.values() for x in y}

#measure emotions associated with specific text
def emotion_analyzer(text, emotions=emotions):
    #Set up the result dictionary
    emotion_count = dict()
    for emotion in emotions:
        emotion_count[emotion] = 0

    #Analyze the text and normalize by total number of words
    total_words = len(text.split())
    for word in text.split():
        if emotion_dict.get(word):
            for emotion in emotion_dict.get(word):
                emotion_count[emotion] += 1/len(text.split())
    return emotion_count

```

```

In [21]: df["Disgust"] = np.nan
df["Surprise"] = np.nan
df["Fear"] = np.nan
df["Trust"] = np.nan
df["Anticipation"] = np.nan
df["Sadness"] = np.nan
df["Anger"] = np.nan
df["Joy"] = np.nan
df["Positive"] = np.nan
df["Negative"] = np.nan
for i in range(0, df.shape[0]):
    senti_results = emotion_analyzer(df.iloc[i, 4])
    df.iloc[i, 11] = senti_results["disgust"]
    df.iloc[i, 12] = senti_results["surprise"]
    df.iloc[i, 13] = senti_results["fear"]
    df.iloc[i, 14] = senti_results["trust"]

```

```

df.iloc[i,15] = senti_results["anticipation"]
df.iloc[i,16] = senti_results["sadness"]
df.iloc[i,17] = senti_results["anger"]
df.iloc[i,18] = senti_results["joy"]
df.iloc[i,19] = senti_results["positive"]
df.iloc[i,20] = senti_results["negative"]
df

```

```

Out[21]:
      hmid  wid reflection_period \
0    27673  2053                24h
1    27674    2                24h
2    27675  1936                24h
3    27676   206                24h
4    27677  6227                24h
5    27678   45                24h
6    27679   195                24h
7    27680   740                24h
8    27681    3                24h
9    27682  4833                24h
10   27683  7334                24h
11   27684   78                24h
12   27685   21                24h
13   27686    8                24h
14   27687   586                24h
15   27688   489                24h
16   27689   976                24h
17   27690  3972                24h
18   27691   14                24h
19   27692  1230                24h
20   27693  9950                24h
21   27694  4862                24h
22   27695  4891                24h
23   27696   737                24h
24   27697   498                24h
25   27698  2124                24h
26   27699   156                24h
27   27700  1783                24h
28   27701  3382                24h
29   27702    54                24h
...     ...     ...     ...
100362 128737  1161                24h
100363 128738  2780                24h
100364 128739  2545                24h
100365 128740 10983                24h
100366 128741  3948                24h
100367 128742   525                24h
100368 128743 13605                24h
100369 128744  5180                24h

```

100370	128745	1234	24h
100371	128746	248	24h
100372	128747	2974	24h
100373	128748	8880	24h
100374	128749	2616	24h
100375	128750	323	24h
100376	128751	11402	24h
100377	128752	301	24h
100378	128753	597	24h
100379	128754	105	24h
100380	128755	22	24h
100381	128756	412	24h
100382	128757	1767	24h
100383	128758	4428	24h
100384	128759	11050	24h
100385	128760	5	24h
100386	128761	4298	24h
100387	128762	112	24h
100388	128763	714	24h
100389	128764	3934	24h
100390	128765	1629	24h
100391	128766	141	24h

original_hm \

```

0      I went on a successful date with someone I fel...
1      I was happy when my son got 90% marks in his e...
2          I went to the gym this morning and did yoga.
3      We had a serious talk with some friends of our...
4      I went with grandchildren to butterfly display...
5          I meditated last night.
6      I made a new recipe for peasant bread, and it ...
7      I got gift from my elder brother which was rea...
8          YESTERDAY MY MOMS BIRTHDAY SO I ENJOYED
9      Watching cupcake wars with my three teen children
10     I came in 3rd place in my Call of Duty video g...
11     I completed my 5 miles run without break. It m...
12         went to movies with my friends it was fun
13     I was shorting Gold and made $200 from the trade.
14     Hearing Songs It can be nearly impossible to g...
15     My son performed very well for a test preparat...
16         I helped my neighbour to fix their car damages.
17     Managed to get the final trophy in a game I wa...
18     A hot kiss with my girl friend last night made...
19     My new BCAAs came in the mail. Yay! Strawberry...
20         Got A in class.
21     My sister called me from abroad this morning a...
22     The cake I made today came out amazing. It tas...
23     There are two types of people in the world: th...

```

24 My grandmother start to walk from the bed afte...
 25 i was able to hit a top spin serve in tennis
 26 I napped with my husband on the bed this after...
 27 My co-woker started playing a Carley Rae Jypse...
 28 My son woke me up to a fantastic breakfast of ...
 29 After a long time my brother gave a suprise vi...
 ...
 100362 DRY DAY CELEBRATION
 100363 My bills were paid on time
 100364 I watched my favorite gamer on Twitch.tv and h...
 100365 I was able to accomplish one of my goals at wo...
 100366 I heard a song I really like on the radio this...
 100367 A friend called just to see how i'd been doing.
 100368 Last night while attending my grandmother's wa...
 100369 My friend came to my home and he give some gif...
 100370 I went shopping for Father's Day dinner on Sun...
 100371 I woke up in the middle of the night and reali...
 100372 I had pizza for dinner and it was great
 100373 My husband called me just to tell me he loved me.
 100374 I was able to weed my entire garden and plant ...
 100375 A coworker praised the job I did on a recent p...
 100376 Fixed my car.
 100377 I had a little time to myself in the afternoon.
 100378 I worked out, which always makes me feel good.
 100379 Finally got to watch the new Resident Evil movie.
 100380 I made plans to go to a new restaurant with th...
 100381 Those rare, perfectly timed naps that leave yo...
 100382 I got to talk to an old friend and reminisce o...
 100383 Yesterday my relations came to my house. That ...
 100384 I am fond of seeing cricket match. Yesterday t...
 100385 My daughter made a very funny sound when I was...
 100386 I spent time with my daughter.
 100387 My husband announced he is getting a decent bo...
 100388 Had a can of Pepsi to drink.
 100389 Cuddling with my girlfriend last night.
 100390 I had a great meeting yesterday at work with m...
 100391 I had a great workout last night.

		cleaned_hm	modified	\
0	I went on a successful date with someone I fel...		True	
1	I was happy when my son got 90% marks in his e...		True	
2	I went to the gym this morning and did yoga.		True	
3	We had a serious talk with some friends of our...		True	
4	I went with grandchildren to butterfly display...		True	
5	I meditated last night.		True	
6	I made a new recipe for peasant bread, and it ...		True	
7	I got gift from my elder brother which was rea...		True	
8	YESTERDAY MY MOMS BIRTHDAY SO I ENJOYED		True	

9	Watching cupcake wars with my three teen children	True
10	I came in 3rd place in my Call of Duty video g...	True
11	I completed my 5 miles run without break. It m...	True
12	went to movies with my friends it was fun	True
13	I was shorting Gold and made \$200 from the trade.	True
14	Hearing Songs It can be nearly impossible to g...	False
15	My son performed very well for a test preparat...	True
16	I helped my neighbour to fix their car damages.	True
17	Managed to get the final trophy in a game I wa...	True
18	A hot kiss with my girl friend last night made...	True
19	My new BCAAs came in the mail. Yay! Strawberry...	True
20	Got A in class.	True
21	My sister called me from abroad this morning a...	True
22	The cake I made today came out amazing. It tas...	True
23	There are two types of people in the world: th...	False
24	My grandmother start to walk from the bed afte...	True
25	i was able to hit a top spin serve in tennis	True
26	I napped with my husband on the bed this after...	True
27	My co-woker started playing a Carley Rae Jepse...	True
28	My son woke me up to a fantastic breakfast of ...	True
29	After a long time my brother gave a suprise vi...	True
...
100362	DRY DAY CELEBRATION	True
100363	My bills were paid on time	True
100364	I watched my favorite gamer on Twitch.tv and h...	True
100365	I was able to accomplish one of my goals at wo...	True
100366	I heard a song I really like on the radio this...	True
100367	A friend called just to see how i'd been doing.	True
100368	Last night while attending my grandmother's wa...	True
100369	My friend came to my home and he give some gif...	True
100370	I went shopping for Father's Day dinner on Sun...	True
100371	I woke up in the middle of the night and reali...	True
100372	I had pizza for dinner and it was great	True
100373	My husband called me just to tell me he loved me.	True
100374	I was able to weed my entire garden and plant ...	True
100375	A coworker praised the job I did on a recent p...	True
100376	Fixed my car.	True
100377	I had a little time to myself in the afternoon.	True
100378	I worked out, which always makes me feel good.	True
100379	Finally got to watch the new Resident Evil movie.	True
100380	I made plans to go to a new restaurant with th...	True
100381	Those rare, perfectly timed naps that leave yo...	True
100382	I got to talk to an old friend and reminisce o...	True
100383	Yesterday my relations came to my house. That ...	True
100384	I am fond of seeing cricket match. Yesterday t...	True
100385	My daughter made a very funny sound when I was...	True
100386	I spent time with my daughter.	True
100387	My husband announced he is getting a decent bo...	True

100388	Had a can of Pepsi to drink.	True
100389	Cuddling with my girlfriend last night.	True
100390	I had a great meeting yesterday at work with m...	True
100391	I had a great workout last night.	True

	num_sentence	ground_truth_category	predicted_category	id \
0	1	NaN	affection	1
1	1	NaN	affection	2
2	1	NaN	exercise	3
3	2	bonding	bonding	4
4	1	NaN	affection	5
5	1	leisure	leisure	6
6	1	NaN	achievement	7
7	1	NaN	affection	8
8	1	NaN	enjoy_the_moment	9
9	1	NaN	affection	10
10	1	NaN	leisure	11
11	2	NaN	exercise	12
12	1	NaN	bonding	13
13	1	NaN	achievement	14
14	2	NaN	enjoy_the_moment	15
15	1	NaN	affection	16
16	1	NaN	bonding	17
17	1	NaN	achievement	18
18	1	NaN	bonding	19
19	3	NaN	affection	20
20	1	NaN	achievement	21
21	2	NaN	affection	22
22	2	NaN	achievement	23
23	2	NaN	enjoy_the_moment	24
24	1	affection	affection	25
25	1	NaN	achievement	26
26	1	NaN	affection	27
27	1	NaN	leisure	28
28	1	NaN	affection	29
29	1	NaN	affection	30
...
100362	1	NaN	achievement	100506
100363	1	NaN	achievement	100507
100364	1	NaN	leisure	100508
100365	1	NaN	achievement	100509
100366	1	NaN	enjoy_the_moment	100510
100367	1	NaN	bonding	100511
100368	1	NaN	bonding	100512
100369	1	NaN	bonding	100513
100370	1	NaN	affection	100514
100371	1	enjoy_the_moment	enjoy_the_moment	100515
100372	1	NaN	enjoy_the_moment	100516

100373	1	NaN	affection	100517
100374	1	NaN	nature	100518
100375	1	NaN	achievement	100519
100376	1	NaN	achievement	100520
100377	1	NaN	enjoy_the_moment	100521
100378	1	NaN	enjoy_the_moment	100522
100379	1	NaN	leisure	100523
100380	1	NaN	bonding	100524
100381	3	NaN	achievement	100525
100382	1	NaN	bonding	100526
100383	2	affection	affection	100527
100384	6	NaN	enjoy_the_moment	100528
100385	1	NaN	affection	100529
100386	1	NaN	affection	100530
100387	1	NaN	affection	100531
100388	1	NaN	enjoy_the_moment	100532
100389	1	affection	affection	100533
100390	1	NaN	bonding	100534
100391	1	NaN	exercise	100535

		text	Disgust	Surprise	\
0		successfully date sympathy connected	0.000000	0.000000	
1		son marks examination	0.000000	0.083333	
2		gym morning yoga	0.000000	0.000000	
3		talked friend flaky understood evening hang	0.040000	0.040000	
4		grandchildren butterfly display crohn conserva...	0.000000	0.000000	
5		meditation night	0.000000	0.000000	
6		recipe peasant bread spectacular	0.000000	0.000000	
7		gift elderly brother surprise	0.000000	0.166667	
8		mom birthday enjoyed	0.000000	0.000000	
9		watched cupcakes wars teen children	0.000000	0.000000	
10		called duty video game	0.000000	0.000000	
11		completed miles run break makes feel strong	0.000000	0.000000	
12		movie friend fun	0.000000	0.000000	
13		short gold trade	0.000000	0.000000	
14		hear song impossible angry youre ease angry fe...	0.050000	0.016667	
15		son performance test prepared	0.000000	0.000000	
16		helped neighbours fixed car damage	0.000000	0.000000	
17		managed finally trophy game played	0.000000	0.083333	
18		hot kiss girl friend night day	0.000000	0.083333	
19		bcaa mail yay strawberry lemonade flavor amino...	0.000000	0.000000	
20		class	0.000000	0.000000	
21		sister called abroad morning occassion family	0.000000	0.000000	
22		cake amazing tasted amazing	0.000000	0.000000	
23		type people world choose choose unhappy contra...	0.000000	0.000000	
24		grandmother started walk bed time	0.000000	0.000000	
25		hit top spin served tennis	0.000000	0.000000	
26		nap husband bed afternoon sweet cuddled close	0.000000	0.050000	

27	cowoker started played carley rae jepsen song ...	0.000000	0.000000
28	son woke fantastic breakfast eggs special hamb...	0.000000	0.000000
29	time brother suprise visit house	0.000000	0.000000
...
100362	dry day celebrated	0.000000	0.000000
100363	bill paid time	0.000000	0.000000
100364	watched favorite gamer twitchtv hilarious	0.000000	0.000000
100365	accomplished goal customer thanked profusely	0.000000	0.000000
100366	heard song radio morning coffee	0.000000	0.000000
100367	friend called id	0.000000	0.000000
100368	night attended grandmother wake enjoyed friend...	0.000000	0.000000
100369	friend home gift energy project	0.000000	0.052632
100370	shopping father day dinner sunday	0.000000	0.111111
100371	woke middle night realized hours sleep left	0.000000	0.000000
100372	pizza dinner	0.000000	0.000000
100373	husband called love	0.000000	0.000000
100374	weed entire garden planted tomato basil rain	0.000000	0.000000
100375	coworker praised job recently project supervisor	0.000000	0.000000
100376	fixed car	0.000000	0.000000
100377	time afternoon	0.000000	0.000000
100378	makes feel	0.000000	0.000000
100379	finally watched residence evil movie	0.000000	0.000000
100380	planned restaurant friend havent	0.000000	0.000000
100381	rare perfect time nap leave feel refreshed gro...	0.033333	0.100000
100382	talked friend reminisce time	0.000000	0.071429
100383	relatives house time	0.000000	0.000000
100384	fond cricket match semi finally champions trop...	0.000000	0.020408
100385	daughter funny sound change diaper night laugh	0.043478	0.043478
100386	spent time daughter	0.000000	0.000000
100387	husband announced decent bonus quarter	0.000000	0.083333
100388	pepsi drink	0.000000	0.000000
100389	cuddled girlfriend night	0.000000	0.000000
100390	meet boss colleagues lunch afterward excited p...	0.000000	0.027027
100391	workout night	0.000000	0.000000

	Fear	Trust	Anticipation	Sadness	Anger	Joy \
0	0.000000	0.071429	0.071429	0.071429	0.000000	0.071429
1	0.083333	0.083333	0.083333	0.000000	0.000000	0.083333
2	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	0.040000	0.040000	0.040000	0.040000	0.040000	0.040000
4	0.000000	0.100000	0.100000	0.000000	0.000000	0.100000
5	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
6	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
7	0.000000	0.166667	0.083333	0.000000	0.000000	0.083333
8	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
9	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
10	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
11	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

12	0.000000	0.000000	0.111111	0.000000	0.000000	0.111111
13	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
14	0.016667	0.016667	0.050000	0.033333	0.050000	0.016667
15	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
16	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
17	0.000000	0.083333	0.083333	0.000000	0.000000	0.083333
18	0.000000	0.083333	0.083333	0.000000	0.083333	0.166667
19	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
20	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
21	0.000000	0.050000	0.100000	0.000000	0.000000	0.050000
22	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
23	0.000000	0.000000	0.027027	0.000000	0.000000	0.027027
24	0.000000	0.000000	0.166667	0.000000	0.000000	0.000000
25	0.000000	0.181818	0.090909	0.000000	0.090909	0.000000
26	0.000000	0.100000	0.050000	0.000000	0.000000	0.100000
27	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
28	0.000000	0.000000	0.000000	0.000000	0.000000	0.058824
29	0.000000	0.071429	0.142857	0.000000	0.000000	0.000000
...
100362	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
100363	0.000000	0.000000	0.166667	0.000000	0.000000	0.000000
100364	0.000000	0.090909	0.000000	0.000000	0.000000	0.090909
100365	0.000000	0.000000	0.000000	0.000000	0.000000	0.052632
100366	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
100367	0.000000	0.100000	0.000000	0.000000	0.000000	0.100000
100368	0.000000	0.035714	0.000000	0.000000	0.000000	0.000000
100369	0.000000	0.052632	0.052632	0.000000	0.000000	0.105263
100370	0.000000	0.111111	0.111111	0.000000	0.000000	0.111111
100371	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
100372	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
100373	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
100374	0.000000	0.000000	0.000000	0.000000	0.000000	0.062500
100375	0.000000	0.000000	0.000000	0.000000	0.000000	0.071429
100376	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
100377	0.000000	0.000000	0.100000	0.000000	0.000000	0.000000
100378	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
100379	0.111111	0.000000	0.111111	0.000000	0.000000	0.000000
100380	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
100381	0.066667	0.033333	0.033333	0.066667	0.033333	0.033333
100382	0.000000	0.142857	0.071429	0.000000	0.000000	0.142857
100383	0.000000	0.066667	0.133333	0.000000	0.000000	0.066667
100384	0.000000	0.020408	0.020408	0.020408	0.000000	0.020408
100385	0.000000	0.000000	0.000000	0.000000	0.000000	0.086957
100386	0.000000	0.000000	0.166667	0.000000	0.000000	0.000000
100387	0.000000	0.000000	0.083333	0.000000	0.000000	0.083333
100388	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
100389	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
100390	0.000000	0.081081	0.054054	0.000000	0.000000	0.027027

100391	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
--------	----------	----------	----------	----------	----------	----------

	Positive	Negative
0	0.142857	0.000000
1	0.083333	0.083333
2	0.000000	0.000000
3	0.080000	0.040000
4	0.100000	0.000000
5	0.000000	0.000000
6	0.000000	0.000000
7	0.250000	0.000000
8	0.000000	0.000000
9	0.000000	0.000000
10	0.000000	0.000000
11	0.000000	0.000000
12	0.111111	0.000000
13	0.000000	0.000000
14	0.016667	0.066667
15	0.000000	0.000000
16	0.000000	0.000000
17	0.083333	0.000000
18	0.166667	0.000000
19	0.000000	0.000000
20	0.000000	0.000000
21	0.050000	0.000000
22	0.000000	0.000000
23	0.027027	0.000000
24	0.083333	0.000000
25	0.090909	0.181818
26	0.100000	0.000000
27	0.000000	0.000000
28	0.117647	0.000000
29	0.142857	0.000000
...
100362	0.000000	0.000000
100363	0.000000	0.000000
100364	0.090909	0.000000
100365	0.105263	0.000000
100366	0.062500	0.000000
100367	0.100000	0.000000
100368	0.000000	0.000000
100369	0.105263	0.000000
100370	0.222222	0.000000
100371	0.000000	0.000000
100372	0.111111	0.000000
100373	0.000000	0.000000
100374	0.062500	0.000000
100375	0.142857	0.000000

```

100376  0.000000  0.000000
100377  0.000000  0.000000
100378  0.000000  0.000000
100379  0.000000  0.000000
100380  0.000000  0.000000
100381  0.066667  0.100000
100382  0.214286  0.000000
100383  0.066667  0.000000
100384  0.020408  0.020408
100385  0.086957  0.000000
100386  0.000000  0.166667
100387  0.166667  0.000000
100388  0.000000  0.000000
100389  0.000000  0.000000
100390  0.081081  0.000000
100391  0.000000  0.000000

```

```
[100392 rows x 21 columns]
```

I aggregate the mean positive emotions by categories. As this data set is about happy moments, I want to focus on positive moments here (joy, anticipation, trust, surprise).

```

In [72]: #take a look at the relationship between positive emotions and predicted_category
pos_cate = df.groupby('predicted_category', as_index=False)[('Joy', 'Anticipation', 'Trust', 'Surprise')]
pos_cate

```

```

Out[72]:
predicted_category      Joy  Anticipation      Trust  Surprise
0      achievement  0.035054      0.035926  0.036917  0.017148
1      affection    0.046649      0.038766  0.034132  0.015277
2      bonding      0.058872      0.034243  0.053036  0.016536
3  enjoy_the_moment  0.056271      0.044527  0.046218  0.018740
4      exercise     0.027147      0.033712  0.028498  0.018057
5      leisure      0.039408      0.030086  0.035053  0.013688
6      nature        0.047575      0.042358  0.027719  0.021618

```

To better visualize the chart above, I draw the grouped boxplot as the following. Some of my findings are:

Bonding and enjoy_the_moment seem to bring people more joy than others.

Leisure has the lowest surprise portion. The explanation is that for leisure activities, people already have pretty clear ideas of what will happen. So surprise is kind of low.

Bonding has the highest trust. As explained above, bonding is more about activities with people other than family members. To have happy moments with them, trust is a necessity.

Genrally, joy has portions much higher than any other categories. So, joy is the predominant emotion in happy moments.

```

In [83]: barWidth = 0.20
plt.figure(figsize = (12,9))

bars1 = list(pos_cate["Joy"])

```

```

bars2 = list(pos_cate["Anticipation"])
bars3 = list(pos_cate["Trust"])
bars4 = list(pos_cate["Surprise"])

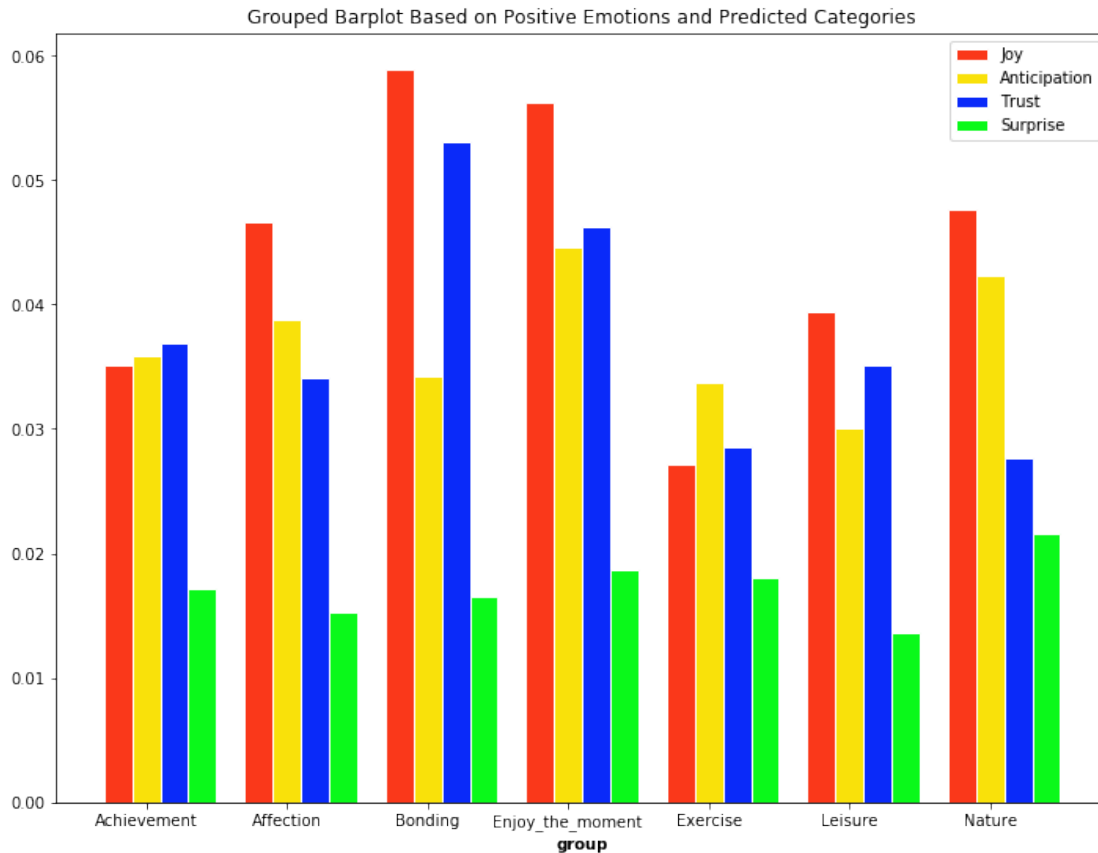
# Set position of bar on X axis
r1 = np.arange(len(bars1))
r2 = [x + barWidth for x in r1]
r3 = [x + barWidth for x in r2]
r4 = [x + barWidth for x in r3]

# Make the plot
plt.bar(r1, bars1, color='#FA381A', width=barWidth, edgecolor='white', label='Joy')
plt.bar(r2, bars2, color='#F9E10A', width=barWidth, edgecolor='white', label='Anticipation')
plt.bar(r3, bars3, color='#0A2AF9', width=barWidth, edgecolor='white', label='Trust')
plt.bar(r4, bars4, color='#0AF91A', width=barWidth, edgecolor='white', label='Surprise')

# Add xticks on the middle of the group bars
plt.xlabel('group', fontweight='bold')
plt.xticks([r + barWidth for r in range(len(bars1))], ['Achievement', 'Affection', 'Boredom'])

# Create legend & Show graphic
plt.title("Grouped Barplot Based on Positive Emotions and Predicted Categories")
plt.legend()
plt.show()

```

Step 6: Topic Analysis

I also want to take a look at the topics talked about in each category and to see if the topics align with some of my findings above. I first want to construct a LDA model with 10 topics based on the text of the whole data set and then take a look at the relationship between topics weights and categories.

```
In [40]: #construct a LDA model on text
def preprocess(doc,n):
    return(strip_short(remove_stopwords(strip_numeric(strip_punctuation(str(doc).lower)))

lda_texts = [i for i in df["text"]]
corp = [preprocess(i,2) for i in lda_texts]
corp_dict = corpora.Dictionary(corp)
corp_dict.filter_extremes(no_below = 2, no_above = 0.97) #tags that appear at least 2
corpus = [corp_dict.doc2bow(preprocess(i,2)) for i in lda_texts]

In [41]: lda_train_model = LdaModel(corpus=corpus,
                                     id2word=corp_dict,
                                     random_state=1,
                                     num_topics=10,
                                     passes=50,
```

```

chunksizes=10000,
alpha='asymmetric',
decay=0.5,
offset=64,
eta=None,
eval_every=0,
iterations=80,
gamma_threshold=0.001,
per_word_topics=True)

```

Some findings about the topics by the LDA model:

The second topic seems to be related with happy moments when going to school.

The fourth topic seems to be related with happy moments working.

The ninth topic seems to be related with natural happy moments.

The tenth topic seems to be related with enjoying foods.

The vast majority topics here can be easily interpreted.

```
In [43]: lda_train_model.show_topics(num_topics = 10, formatted=False, num_words = 10)
```

```

Out[43]: [(0,
  [('time', 0.036702696),
   ('day', 0.029357994),
   ('friend', 0.028094964),
   ('home', 0.02133608),
   ('played', 0.020257771),
   ('night', 0.019810792),
   ('game', 0.015881015),
   ('nice', 0.013022319),
   ('wife', 0.012447071),
   ('birthday', 0.012255518)]),
 (1,
  [('friend', 0.035255561),
   ('day', 0.022346426),
   ('school', 0.021193678),
   ('time', 0.013409452),
   ('received', 0.01320421),
   ('talked', 0.013131115),
   ('college', 0.011014912),
   ('meet', 0.010066762),
   ('class', 0.009790211),
   ('called', 0.009705198)]),
 (2,
  [('moment', 0.064454585),
   ('life', 0.04032126),
   ('movie', 0.035499904),
   ('family', 0.032792542),
   ('time', 0.03230947),
   ('enjoyed', 0.02361735),

```

```

        ('finished', 0.022173386),
        ('project', 0.021142533),
        ('vacation', 0.019067734),
        ('person', 0.018342892)]),
(3,
 [ ('job', 0.07483118),
   ('moved', 0.025881086),
   ('happiness', 0.020560946),
   ('offer', 0.016249256),
   ('positive', 0.014394519),
   ('company', 0.013747236),
   ('business', 0.013183209),
   ('interview', 0.0122508975),
   ('cake', 0.010657622),
   ('live', 0.010332561)]),
(4,
 [ ('bought', 0.053013988),
   ('car', 0.0385913),
   ('money', 0.03379518),
   ('shopping', 0.027191343),
   ('purchased', 0.022756644),
   ('buy', 0.020901704),
   ('cleaned', 0.018912762),
   ('pay', 0.018068505),
   ('surprise', 0.01662841),
   ('gift', 0.015889196)]),
(5,
 [ ('feel', 0.027817167),
   ('trip', 0.026872857),
   ('life', 0.014303698),
   ('song', 0.013016175),
   ('event', 0.012692717),
   ('listening', 0.0119677605),
   ('people', 0.011843975),
   ('music', 0.010344065),
   ('world', 0.010111439),
   ('time', 0.00996193)]),
(6,
 [ ('watched', 0.16418143),
   ('favorite', 0.10868935),
   ('funny', 0.024790004),
   ('enjoyed', 0.021420168),
   ('video', 0.019768467),
   ('season', 0.01877177),
   ('laugh', 0.017711595),
   ('travel', 0.016625457),
   ('friend', 0.016078748),
   ('water', 0.015388882)]),

```

```

(7,
 [('finally', 0.062141474),
  ('started', 0.053088974),
  ('run', 0.04180326),
  ('free', 0.04106716),
  ('team', 0.0355677),
  ('father', 0.033038523),
  ('tv', 0.029079758),
  ('match', 0.021234395),
  ('win', 0.018016241),
  ('helped', 0.015010331)]),
(8,
 [('book', 0.05913682),
  ('read', 0.04147855),
  ('garden', 0.031725336),
  ('planted', 0.023807874),
  ('flowers', 0.019871803),
  ('released', 0.017777622),
  ('film', 0.01643239),
  ('hike', 0.013460284),
  ('rain', 0.011050527),
  ('tree', 0.010846492)]),
(9,
 [('dinner', 0.12095568),
  ('lunch', 0.06310201),
  ('eat', 0.051549677),
  ('food', 0.050624825),
  ('ate', 0.04849463),
  ('delicious', 0.037690237),
  ('restaurant', 0.0373117),
  ('cooked', 0.037169967),
  ('chicken', 0.015001298),
  ('family', 0.014524359)]])

```

Now, I want to get topic weights by each document or observation and add these values in the data set

```

In [57]: df["Topic 1"] = np.nan
df["Topic 2"] = np.nan
df["Topic 3"] = np.nan
df["Topic 4"] = np.nan
df["Topic 5"] = np.nan
df["Topic 6"] = np.nan
df["Topic 7"] = np.nan
df["Topic 8"] = np.nan
df["Topic 9"] = np.nan
df["Topic 10"] = np.nan
for i in range(0,df.shape[0]):

```

```
corp_x = corp_dict.doc2bow(preprocess(df.iloc[i,10],2))
topic_w = lda_train_model.get_document_topics(corp_x, minimum_probability = 0)
for j in range(0,10):
    df.iloc[i,(11+j)] = topic_w[j][1]
```

To visualize the relationship between the topic weights and the categories of happy moments, I obtain the mean value of the topic weights for each topic by categories as following.

```
In [59]: topic_cate = df.groupby('predicted_category', as_index=False)[('Topic 1','Topic 2','Topic 3','Topic 4','Topic 5','Topic 6','Topic 7','Topic 8','Topic 9','Topic 10')]
```

topic_cate

```
Out[59]:
```

	predicted_category	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	\
0	achievement	0.157197	0.244544	0.090674	0.110860	0.168186	
1	affection	0.462094	0.132507	0.082855	0.049085	0.061093	
2	bonding	0.354800	0.283035	0.081772	0.050430	0.045816	
3	enjoy_the_moment	0.285873	0.115807	0.101254	0.067366	0.072057	
4	exercise	0.219898	0.417937	0.087833	0.037309	0.033664	
5	leisure	0.285546	0.084392	0.129010	0.049717	0.069557	
6	nature	0.470660	0.049822	0.059825	0.034386	0.038580	
	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10		
0	0.068702	0.032631	0.052166	0.039322	0.035718		
1	0.067522	0.035264	0.032000	0.025166	0.052413		
2	0.051645	0.040046	0.027025	0.022032	0.043400		
3	0.098419	0.061732	0.040852	0.036367	0.120274		
4	0.053684	0.024259	0.080411	0.025104	0.019902		
5	0.077359	0.139445	0.060264	0.059880	0.044830		
6	0.070087	0.048583	0.041175	0.145929	0.040955		

I further make the table into a heatmap as following. Some of the findings are:

The first three topics seem to be more generalized, as almost all seven categories have high topic weights on them.

Achievement seems to have higher weights on topic 4 and topic 5. It makes perfect sense, as topic 4 is about careers and work and topic 5 is about money and purchasing stuff.

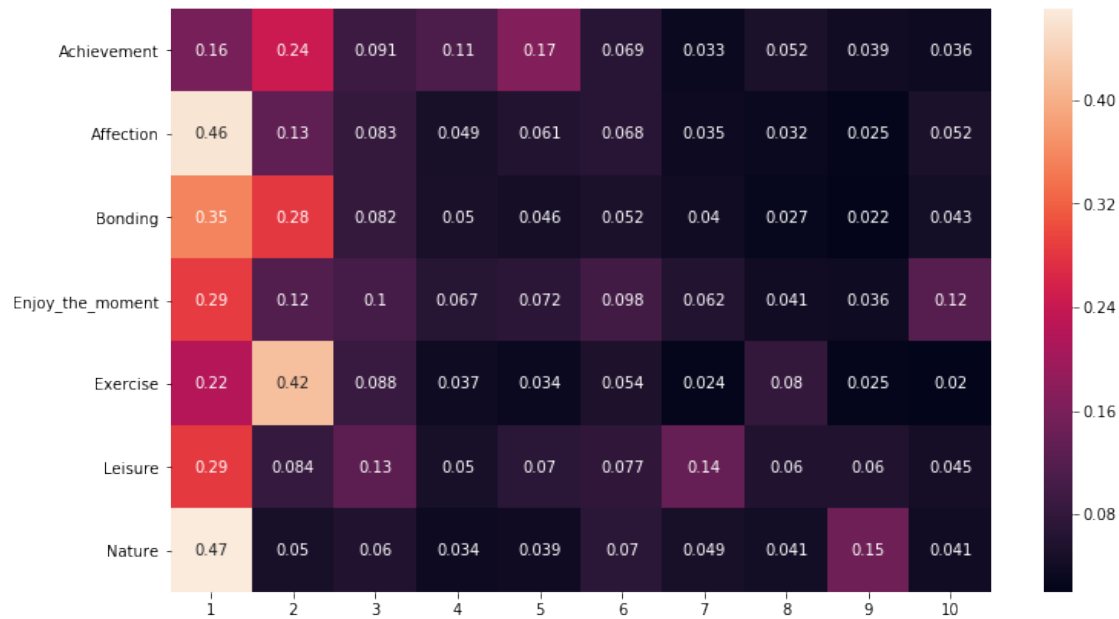
Enjoy_the_moment seems to have higher weights on topic 6 and topic 10. Topic 6 contains a lot of words related to music and topic 10 contains words related to food. People tend to enjoy moments with music and food.

Leisure seems to have higher weights on topic 7 that is about traveling and watching movies.

Nature seems to have higher weights on topic 9 that is about nature and natural objects.

```
In [77]: Index= ['Achievement', 'Affection', 'Bonding', 'Enjoy_the_moment', 'Exercise', 'Leisure']
Cols = ['1', '2', '3', '4', '5', '6', '7', '8', '9', '10']
df = DataFrame(np.array(topic_cate.iloc[:,1:11]), index=Index, columns=Cols)
plt.figure(figsize = (12,7))
sns.heatmap(df, annot=True)
```

```
Out[77]: <matplotlib.axes._subplots.AxesSubplot at 0x181687dbb38>
```



Step 7: Conclusion

From the project, I learn that there is a distinct pattern for each of the category of happy moments. Specifically:

Achievement is about working, education and being successful.

Affection is about doing things with family.

Bonding is about doing things with people other than family.

Enjoy_the_moment can be a lot of things, but people tend to associate music and food with a happy moment.

Exercise is about exercising.

Leisure is about watching movies, playing video games and traveling.

Nature is about enjoying the natural environment.