Applied Data Science @ Columbia

STAT GR5243/GU4243 Fall 2023

Project 1: What made you happy today?

- Projec title: The effects of demographic factors on happy moments
- This project is conducted by Nashita Rahman - nfr2111
- Project summary: The LDA models shed light on how the nature of happy moments varies across different demographic groups. Young individuals tend to find happiness in a diverse range of activities such as gaming, music, and work-related achievements, while older individuals focus more on everyday life, work, celebrations, and outdoor activities. Men express happiness through a wide array of topics including work, gaming, and fixing things, whereas women emphasize personal and emotional experiences, celebrations, and memories. Married individuals highlight the importance of relationships and interactions in their happy moments, while single individuals place greater emphasis on work-related and financial sources of happiness. Parenthood, regardless of the presence or absence of children, underscores the significance of family, but those without children lean toward happiness derived from various activities and experiences. These findings illuminate the multifaceted nature of happiness, influenced by age, gender, marital status, and parenthood, each contributing unique dimensions to what brings joy to people's lives.

Import necessary packages

In [1]:
```
import gensim
import nltk
import numpy as np
import os
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

from gensim.parsing.preprocessing import remove_stopwords, strip_numeric, strip
from wordcloud import WordCloud
```

Import datasets

In [2]:
```
processed_moments = pd.read_csv('../output/processed_moments.csv')
demographic = pd.read_csv('../data/demographic.csv')
```

In [3]:
```
# PREPROCESSING

# Processed moments preprocessing
# Dropping ground truth category bc it lacks the amount of data the other colum
```

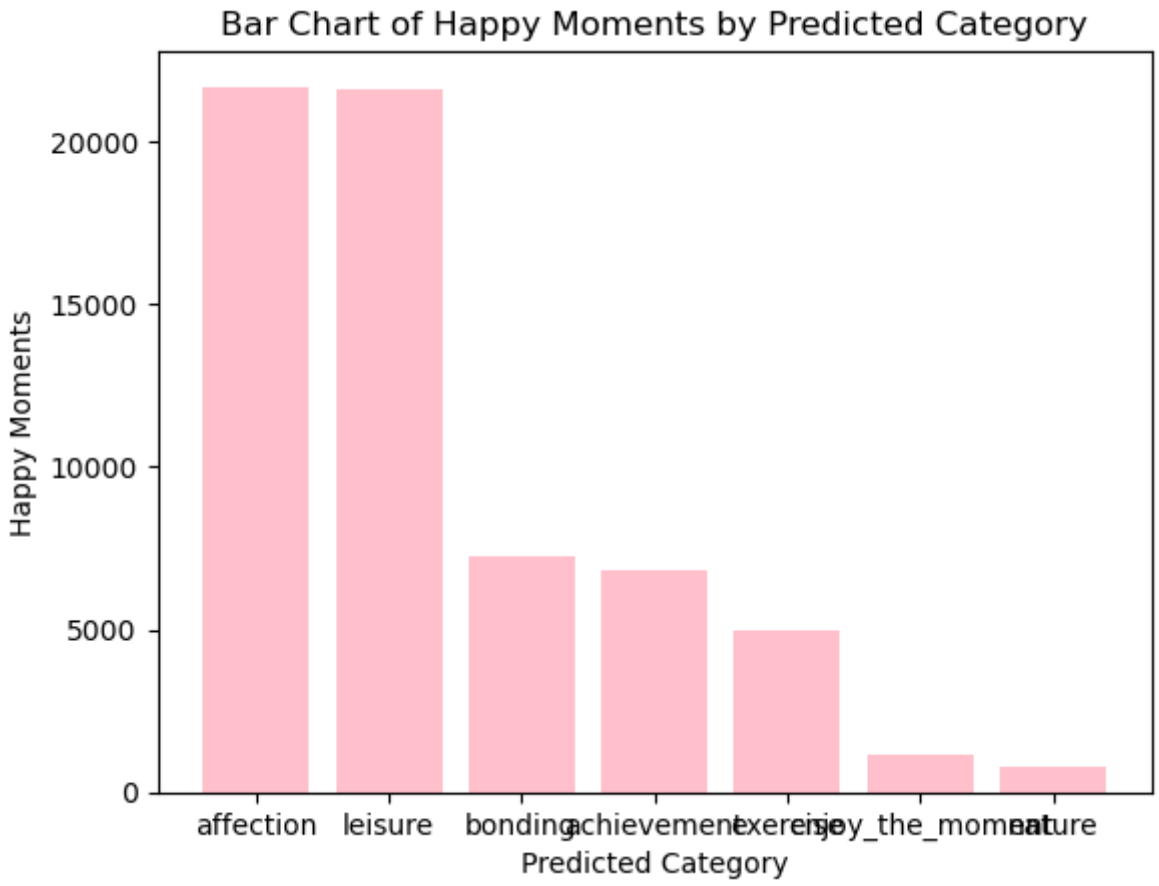
```
processed_moments = processed_moments.drop('ground_truth_category', axis=1)

# Demographic preprocessing
# Age column has some erroneous values that need to be fixed
demographic = demographic.dropna(subset=['age'])
demographic = demographic[demographic['age'].str.isnumeric()]
demographic['age'] = demographic['age'].astype(int)

# Joining the processed moments and demographic datasets by wid
df = pd.merge(processed_moments, demographic, on='wid')
```

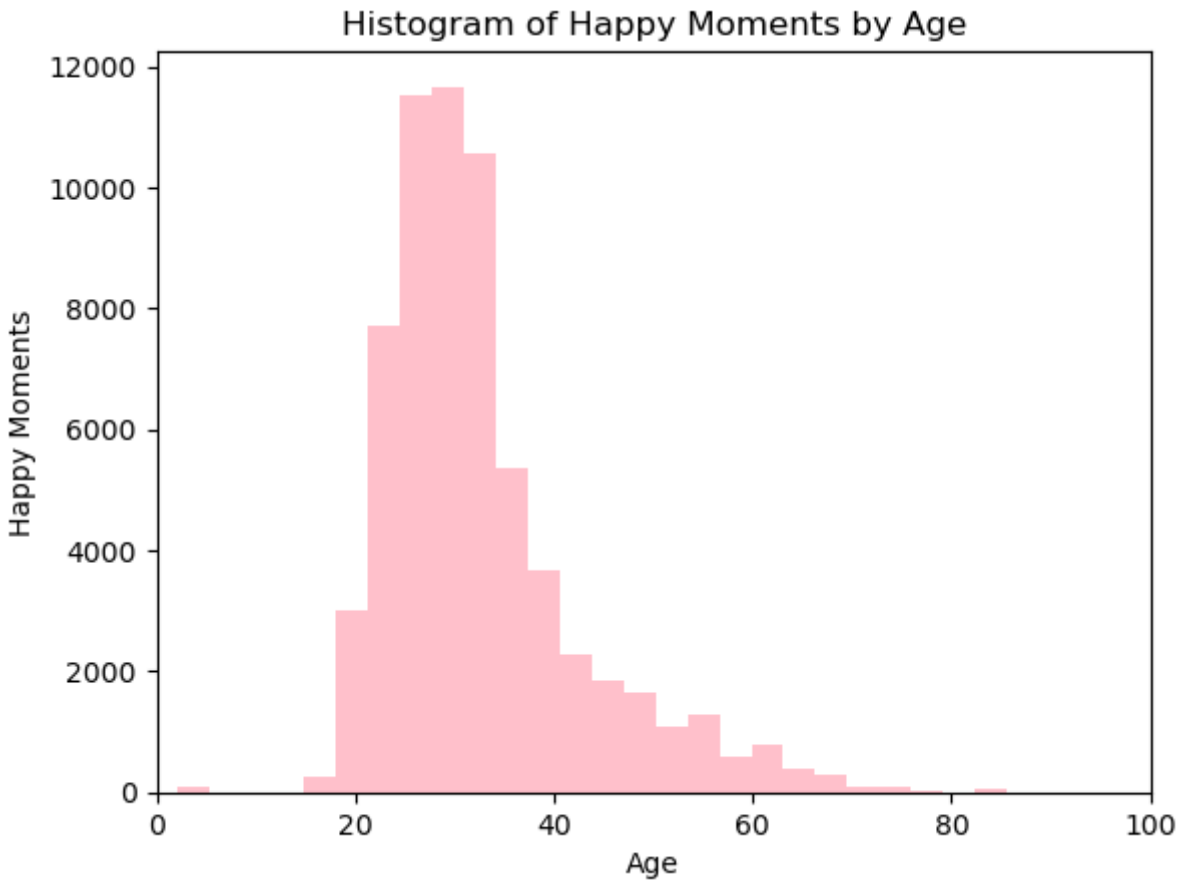
Analyzing happy moments Exploratory Analysis

In [4]:
```
plt.bar(df['predicted_category'].unique(), df['predicted_category'].value_count
plt.xlabel('Predicted Category')
plt.ylabel('Happy Moments')
plt.title('Bar Chart of Happy Moments by Predicted Category')
plt.savefig('../figs/hm_by_predicted_category.png')
plt.show()
```



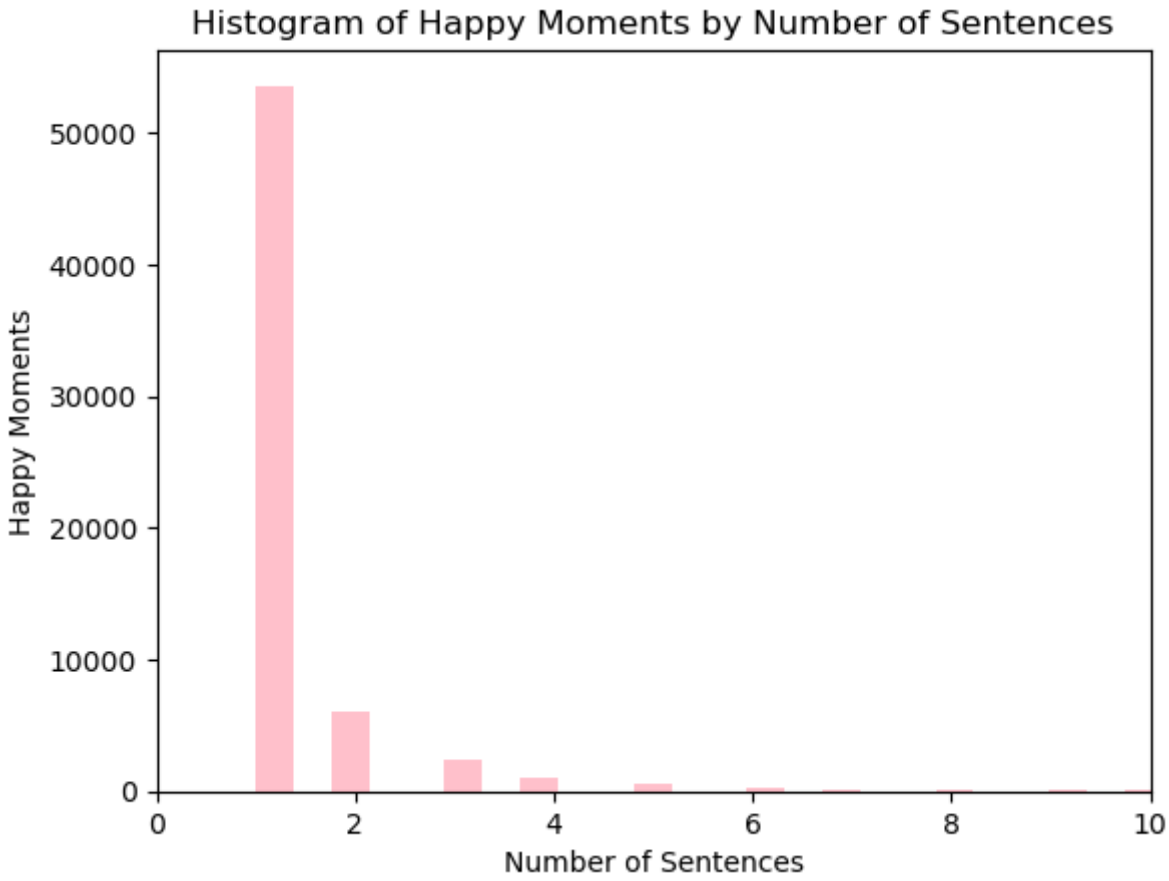
In [5]:
```
print('mean =', np.mean(df['age']))
print('std =', np.std(df['age']))
plt.hist(df['age'], bins=int(70), color='pink')
plt.xlabel('Age')
plt.ylabel('Happy Moments')
plt.title('Histogram of Happy Moments by Age')
plt.xlim(0, 100)
plt.savefig('../figs/hm_by_age.png')
plt.show()
```

```
mean = 32.56561036698247
std = 10.285818727651632
```



In [6]:
```
print('mean =', np.mean(df['num_sentence']))
print('std =', np.std(df['num_sentence']))
plt.hist(df['num_sentence'], bins=int(150), color='pink')
plt.xlabel('Number of Sentences')
plt.ylabel('Happy Moments')
plt.title('Histogram of Happy Moments by Number of Sentences')
plt.xlim(0, 10)
plt.savefig('../figs/hm_by_num_sentences.png')
plt.show()
```

```
mean = 1.3335511270826528
std = 1.1713094162412971
```



Data segmentation

I think it could be interesting to see whether individuals with varying demographic characteristics tend to derive happiness from distinct sources. To analyze this I am deviding the dataset into eight subsets by applying four specific criteria demographic: gender, marital status, parenthood status, and age.

In [7]:
```
# Segment by age
young = df[df['age'] <= np.mean(df['age'])]
old = df[df['age'] > np.mean(df['age'])]
# Segment by gender
male = df[df['gender'] == 'm']
female = df[df['gender'] == 'f']
# Segment by marital status
married = df[df['marital'] == 'married']
single = df[df['marital'] == 'single']
# Segment by parenthood status
parent = df[df['parenthood'] == 'y']
childfree = df[df['parenthood'] == 'n']
```

Generating word clouds and LDA models

Helper functions

In [8]:
```
# Function to generate wordcloud given a dataset
def get_cloud(df, title='', filename=''):
```

```
    plt.imshow(WordCloud(colormap='magma').generate(''.join(df['cleaned_hm'])))
    plt.axis('off')
    plt.title(title)
    plt.savefig(filename)
    plt.show()
```

In [9]:
```
# Define a function to preprocess the document
def preprocess(word):
    word = word.lower()
    word = strip_short(remove_stopwords(strip_numeric(strip_punctuation(word))))
    ss = nltk.SnowballStemmer('english')
    word = [ss.stem(y) for y in word]
    return word

# Define the LDA model function using the input of a different dataset
def LDA_model(data):
    Words = [word for word in data['cleaned_hm']]
    Text = ''.join(Words)
    Text = remove_stopwords(strip_numeric(strip_punctuation(strip_punctuation(T
    # Snowball stemming
    ss = nltk.SnowballStemmer('english')
    Text = [ss.stem(y) for y in Text]
    # Prepare corpus
    dictionary = gensim.corpora.Dictionary([preprocess(tag) for tag in Words])
    tfidf = gensim.models.TfidfModel([dictionary.doc2bow(preprocess(tag)) for t
    dictionary.filter_extremes(no_below=2, no_above=0.99)
    # LDA model
    lda_model = gensim.models.LdaModel(
        corpus=[dictionary.doc2bow(preprocess(tag)) for tag in Words],
        id2word=dictionary,
        random_state=100,
        num_topics=10,
        passes=5,
        chunksize=10000,
        alpha='asymmetric',
        decay=0.5,
        offset=64,
        eta=None,
        eval_every=0,
        iterations=100,
        gamma_threshold=0.001,
        per_word_topics=True
    )
    return lda_model.print_topics(-1)
```

The effect of age on happy moments

In [10]:
```
get_cloud(young, 'Word Cloud for Young People', '../figs/wordcloud_age_young.pr
```

Word Cloud for Young People

In [11]:
```
get_cloud(old, 'Word Cloud for Old People', '../figs/wordcloud_age_old.png')
```

Word Cloud for Old People

In [12]:
```
LDA_model(young)
```

Out[12]:
```
[(0,
  '0.030*"happi" + 0.020*"went" + 0.019*"work" + 0.016*"time" + 0.015*"got" +
0.015*"watch" + 0.015*"today" + 0.012*"day" + 0.012*"favorit" + 0.011*"new"'),
 (1,
  '0.028*"happi" + 0.018*"abl" + 0.017*"final" + 0.011*"work" + 0.010*"walk" +
0.009*"finish" + 0.009*"exam" + 0.009*"day" + 0.009*"outsid" + 0.009*"new"'),
 (2,
  '0.043*"happi" + 0.033*"friend" + 0.020*"day" + 0.020*"moment" + 0.017*"new"
+ 0.016*"time" + 0.014*"old" + 0.014*"birthday" + 0.013*"got" + 0.013*"mont
h"'),
 (3,
  '0.017*"ride" + 0.017*"employe" + 0.017*"great" + 0.015*"work" + 0.015*"hor
s" + 0.012*"big" + 0.012*"time" + 0.012*"hot" + 0.011*"huge" + 0.010*"make"'),
 (4,
  '0.104*"game" + 0.081*"play" + 0.051*"won" + 0.044*"video" + 0.042*"happi" +
0.032*"team" + 0.016*"beat" + 0.013*"state" + 0.012*"pound" + 0.012*"inter
n"'),
 (5,
  '0.090*"got" + 0.057*"job" + 0.054*"work" + 0.024*"promot" + 0.020*"compani"
+ 0.014*"cowork" + 0.013*"took" + 0.013*"receiv" + 0.012*"husband" + 0.011*"bo
yfriend"'),
 (6,
  '0.032*"happi" + 0.027*"song" + 0.021*"listen" + 0.017*"movement" + 0.017*"p
erson" + 0.014*"music" + 0.012*"time" + 0.009*"band" + 0.008*"favorit" + 0.007
*"happiest"'),
 (7,
  '0.028*"water" + 0.027*"puppi" + 0.023*"white" + 0.023*"head" + 0.022*"swam"
+ 0.022*"black" + 0.020*"saw" + 0.018*"eye" + 0.018*"watch" + 0.016*"hand"'),
 (8,
  '0.056*"leagu" + 0.045*"legend" + 0.012*"switch" + 0.009*"wild" + 0.009*"sa
n" + 0.007*"shoot" + 0.006*"nintendo" + 0.006*"zelda" + 0.006*"cherri" + 0.006
*"chicago"'),
 (9,
  '0.078*"ziggi" + 0.078*"sudi" + 0.059*"friend" + 0.039*"said" + 0.037*"fli"
+ 0.036*"great" + 0.031*"play" + 0.024*"gone" + 0.020*"thought" + 0.019*"gav
e"')]
```

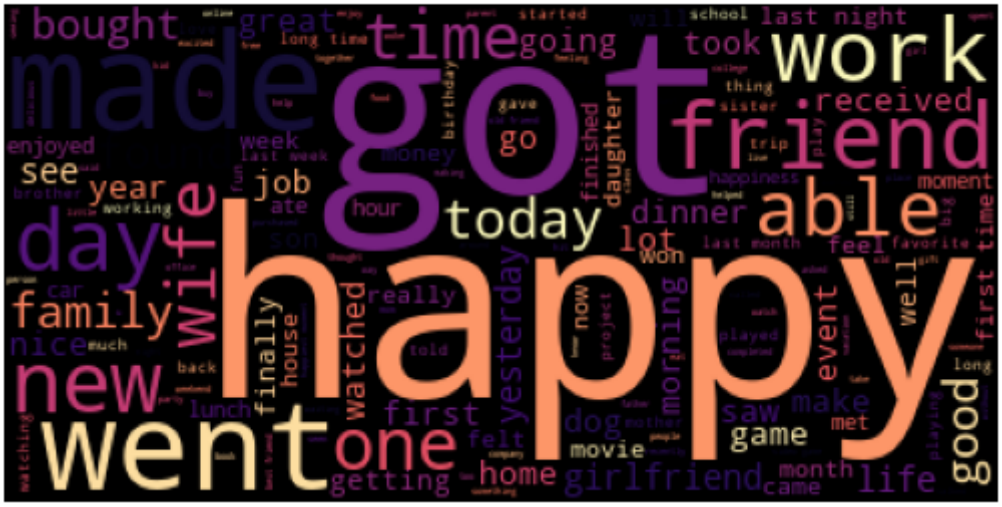
In [13]:
```
LDA_model(old)
```

Out[13]:
```
[(0,
  '0.026*"happi" + 0.020*"got" + 0.017*"day" + 0.016*"time" + 0.016*"work" +
0.016*"went" + 0.011*"friend" + 0.010*"abl" + 0.009*"son" + 0.009*"play"'),
 (1,
  '0.028*"happi" + 0.018*"new" + 0.013*"work" + 0.013*"got" + 0.010*"month" +
0.009*"year" + 0.009*"dinner" + 0.008*"friend" + 0.008*"daughter" + 0.008*"hus
band"'),
 (2,
  '0.038*"birthday" + 0.026*"happi" + 0.026*"friend" + 0.021*"sister" + 0.021
*"talk" + 0.017*"parti" + 0.017*"celebr" + 0.015*"daughter" + 0.014*"phone" +
0.014*"son"'),
 (3,
  '0.014*"master" + 0.009*"qualif" + 0.008*"instead" + 0.007*"time" + 0.007*"t
urk" + 0.006*"smoke" + 0.006*"profit" + 0.005*"mturk" + 0.005*"mechan" + 0.005
*"owe"'),
 (4,
  '0.039*"receiv" + 0.033*"bonus" + 0.014*"mail" + 0.011*"mturk" + 0.011*"got"
+ 0.008*"work" + 0.006*"big" + 0.006*"order" + 0.005*"larg" + 0.005*"jean"'),
 (5,
  '0.016*"cooki" + 0.013*"wash" + 0.009*"chocol" + 0.008*"chip" + 0.006*"iav"
+ 0.005*"ritual" + 0.004*"start" + 0.004*"happi" + 0.004*"dish" + 0.004*"ia
m"'),
 (6,
  '0.084*"life" + 0.039*"moment" + 0.025*"good" + 0.022*"happiest" + 0.014*"se
x" + 0.013*"happi" + 0.006*"work" + 0.005*"day" + 0.005*"bloom" + 0.004*"surge
ri"'),
 (7,
  '0.026*"ride" + 0.025*"bike" + 0.013*"road" + 0.008*"happi" + 0.007*"wait" +
0.007*"went" + 0.006*"bicycl" + 0.006*"new" + 0.006*"stand" + 0.006*"cross"'),
 (8,
  '0.006*"guitar" + 0.004*"vote" + 0.003*"jail" + 0.003*"jelli" + 0.003*"cleve
land" + 0.003*"cavali" + 0.003*"thesi" + 0.002*"stanley" + 0.002*"jute" + 0.00
2*"supervis"'),
 (9,
  '0.032*"film" + 0.013*"releas" + 0.011*"star" + 0.010*"titl" + 0.008*"happi"
+ 0.008*"allu" + 0.008*"version" + 0.006*"kerala" + 0.005*"role" + 0.005*"du
b"')]
```

The effect of gender on happy moments

In [14]:
```
get_cloud(male, 'Word Cloud for Men', '../figs/wordcloud_gender_men.png')
```

Word Cloud for Men



In [15]:
```
get_cloud(female, 'Word Cloud for Women', '../figs/wordcloud_gender_women.png')
```

Word Cloud for Women

In [16]:
```
LDA_model(male)
```

Out[16]:
```
[(0,
  '0.030*"got" + 0.029*"work" + 0.024*"happi" + 0.016*"good" + 0.015*"new" +
0.014*"day" + 0.011*"went" + 0.011*"abl" + 0.010*"final" + 0.010*"job"'),
 (1,
  '0.031*"happi" + 0.020*"new" + 0.016*"day" + 0.014*"wife" + 0.014*"dog" + 0.
012*"love" + 0.012*"got" + 0.011*"watch" + 0.010*"bought" + 0.009*"moment"'),
 (2,
  '0.067*"friend" + 0.032*"time" + 0.029*"went" + 0.023*"happi" + 0.021*"famil
i" + 0.018*"birthday" + 0.016*"month" + 0.013*"day" + 0.013*"long" + 0.013*"ye
ar"'),
 (3,
  '0.036*"happi" + 0.029*"life" + 0.018*"video" + 0.016*"peopl" + 0.013*"game"
+ 0.012*"differ" + 0.010*"beat" + 0.008*"thing" + 0.008*"live" + 0.008*"ne
w"'),
 (4,
  '0.049*"game" + 0.043*"won" + 0.042*"team" + 0.029*"play" + 0.020*"match" +
0.019*"watch" + 0.019*"favorit" + 0.016*"win" + 0.012*"basketbal" + 0.011*"bas
ebal"'),
 (5,
  '0.040*"ate" + 0.031*"delici" + 0.016*"shop" + 0.013*"chicken" + 0.013*"stoc
k" + 0.013*"sandwich" + 0.013*"eat" + 0.012*"cook" + 0.011*"market" + 0.010*"m
eal"'),
 (6,
  '0.069*"happi" + 0.027*"state" + 0.026*"posit" + 0.020*"research" + 0.019*"e
mot" + 0.014*"joy" + 0.013*"psycholog" + 0.013*"defin" + 0.012*"mental" + 0.01
1*"pleasant"'),
 (7,
  '0.054*"fix" + 0.039*"car" + 0.016*"problem" + 0.015*"roti" + 0.015*"happi"
+ 0.014*"said" + 0.011*"repair" + 0.009*"fact" + 0.009*"meant" + 0.008*"par
t"'),
 (8,
  '0.041*"song" + 0.033*"music" + 0.025*"releas" + 0.023*"film" + 0.021*"liste
n" + 0.015*"album" + 0.013*"singl" + 0.010*"star" + 0.009*"radio" + 0.008*"ban
d"'),
 (9,
  '0.013*"contest" + 0.013*"wild" + 0.010*"prize" + 0.009*"sound" + 0.008*"bre
ath" + 0.007*"legend" + 0.007*"apolog" + 0.007*"happi" + 0.006*"zelda" + 0.005
*"holiday"')]
```

In [17]:
```
LDA_model(female)
```

Out[17]:
```
[(0,
  '0.029*"happi" + 0.018*"work" + 0.016*"day" + 0.015*"new" + 0.015*"went" +
0.014*"got" + 0.010*"husband" + 0.010*"abl" + 0.010*"time" + 0.009*"today"'),
 (1,
  '0.035*"happi" + 0.028*"friend" + 0.025*"got" + 0.021*"time" + 0.014*"year"
+ 0.011*"dinner" + 0.011*"daughter" + 0.010*"went" + 0.009*"talk" + 0.008*"da
y"'),
 (2,
  '0.029*"went" + 0.021*"won" + 0.018*"took" + 0.015*"happi" + 0.014*"beauti"
+ 0.013*"lost" + 0.012*"coffe" + 0.011*"saw" + 0.011*"walk" + 0.010*"weigh
t"'),
 (3,
  '0.022*"watch" + 0.017*"happi" + 0.014*"time" + 0.010*"feel" + 0.009*"order"
+ 0.009*"today" + 0.009*"got" + 0.009*"bought" + 0.009*"arriv" + 0.008*"visi
t"'),
 (4,
  '0.024*"friend" + 0.024*"birthday" + 0.022*"parti" + 0.018*"pass" + 0.018*"d
ay" + 0.017*"celebr" + 0.016*"exam" + 0.015*"enjoy" + 0.013*"happi" + 0.013*"s
chool"'),
 (5,
  '0.022*"happi" + 0.021*"cat" + 0.016*"brother" + 0.013*"plant" + 0.013*"pla
y" + 0.013*"time" + 0.012*"bird" + 0.010*"birthday" + 0.010*"garden" + 0.010
*"celebr"'),
 (6,
  '0.054*"moment" + 0.048*"life" + 0.037*"happiest" + 0.028*"day" + 0.022*"sur
pris" + 0.018*"dad" + 0.014*"given" + 0.013*"felt" + 0.013*"present" + 0.012
*"person"'),
 (7,
  '0.020*"rememb" + 0.017*"graduat" + 0.017*"mom" + 0.015*"readi" + 0.014*"deg
re" + 0.014*"happiest" + 0.014*"moment" + 0.014*"everyday" + 0.013*"smile" +
0.011*"bye"'),
 (8,
  '0.004*"nest" + 0.004*"colour" + 0.003*"robin" + 0.003*"key" + 0.003*"fores
t" + 0.003*"harri" + 0.002*"ultrasound" + 0.002*"potter" + 0.002*"pad" + 0.002
*"view"'),
 (9,
  '0.024*"bath" + 0.009*"bubbl" + 0.008*"feedback" + 0.004*"relax" + 0.004*"im
port" + 0.004*"dungeon" + 0.004*"took" + 0.003*"pig" + 0.003*"bomb" + 0.003*"d
ragon"')]
```

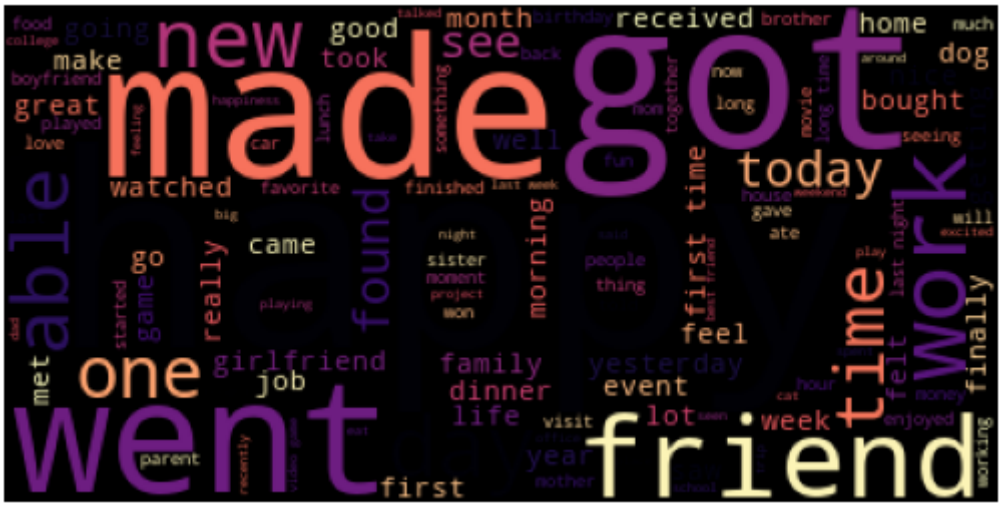
The effect of martial status on happy moments

In [18]:
```
get_cloud(married, 'Word Cloud for Married People', '../figs/wordcloud_marriage
```

Word Cloud for Married People

In [19]:
```
get_cloud(single, 'Word Cloud for Single People', '../figs/wordcloud_marriage_s
```

Word Cloud for Single People



In [20]:
```
LDA_model(married)
```

Out[20]:
```
[(0,
  '0.033*"happi" + 0.019*"got" + 0.018*"work" + 0.016*"time" + 0.015*"day" +
0.013*"went" + 0.011*"wife" + 0.011*"son" + 0.010*"husband" + 0.010*"frien
d"'),
 (1,
  '0.020*"start" + 0.014*"happi" + 0.013*"new" + 0.009*"won" + 0.008*"plant" +
0.008*"week" + 0.008*"daughter" + 0.008*"feel" + 0.007*"garden" + 0.007*"loo
k"'),
 (2,
  '0.029*"friend" + 0.028*"play" + 0.025*"game" + 0.021*"happi" + 0.020*"watc
h" + 0.019*"got" + 0.016*"new" + 0.016*"day" + 0.012*"enjoy" + 0.011*"favori
t"'),
 (3,
  '0.058*"life" + 0.056*"moment" + 0.036*"happiest" + 0.025*"happi" + 0.016*"d
ay" + 0.013*"surpris" + 0.012*"given" + 0.010*"movement" + 0.010*"person" + 0.
009*"date"'),
 (4,
  '0.058*"went" + 0.035*"bought" + 0.033*"new" + 0.025*"shop" + 0.016*"famili"
+ 0.015*"movi" + 0.014*"happi" + 0.013*"templ" + 0.010*"phone" + 0.010*"ic
e"'),
 (5,
  '0.011*"romant" + 0.010*"girlfriend" + 0.010*"went" + 0.007*"grandpa" + 0.00
6*"date" + 0.006*"offic" + 0.005*"societi" + 0.005*"vocat" + 0.005*"happi" +
0.005*"got"'),
 (6,
  '0.055*"happi" + 0.015*"state" + 0.015*"posit" + 0.011*"research" + 0.011*"e
mot" + 0.009*"defin" + 0.008*"mental" + 0.008*"psycholog" + 0.007*"feel" + 0.0
07*"content"'),
 (7,
  '0.020*"colleg" + 0.010*"caught" + 0.008*"function" + 0.006*"life" + 0.006
*"want" + 0.005*"inagur" + 0.005*"outlook" + 0.004*"watch" + 0.004*"happi" +
0.004*"hilari"'),
 (8,
  '0.014*"result" + 0.011*"anxieti" + 0.008*"ate" + 0.007*"key" + 0.005*"exam"
+ 0.005*"think" + 0.005*"music" + 0.004*"publish" + 0.004*"student" + 0.004*"d
irector"'),
 (9,
  '0.007*"happi" + 0.006*"film" + 0.006*"tempel" + 0.005*"titl" + 0.005*"star"
+ 0.005*"allu" + 0.004*"releas" + 0.004*"home" + 0.004*"version" + 0.003*"happ
ier"')]
```

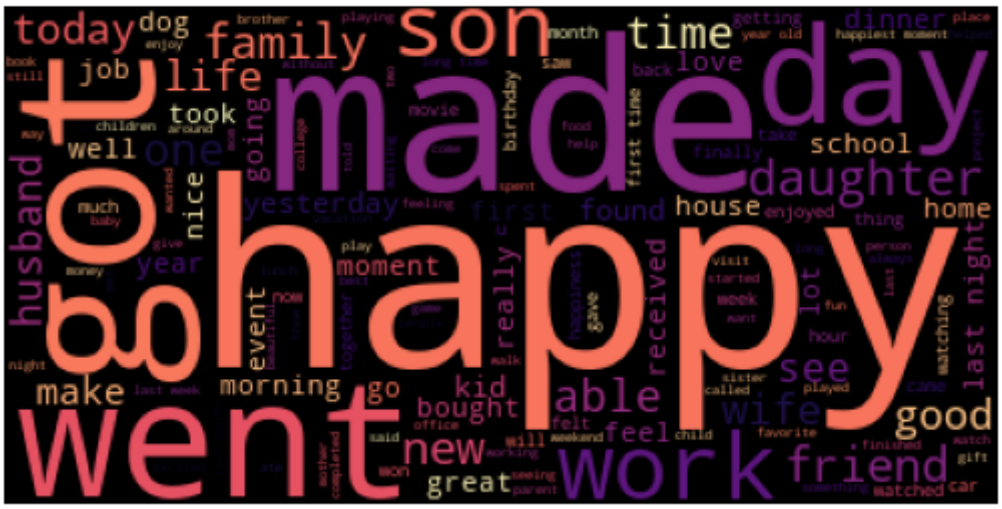
In [21]:
```
LDA_model(single)
```

Out[21]:
```
[(0,
  '0.031*"friend" + 0.024*"happi" + 0.023*"went" + 0.019*"day" + 0.017*"time"
+ 0.016*"got" + 0.011*"play" + 0.009*"famili" + 0.009*"girlfriend" + 0.009*"ho
me"'),
 (1,
  '0.034*"work" + 0.031*"happi" + 0.022*"got" + 0.015*"job" + 0.014*"new" + 0.
013*"month" + 0.012*"abl" + 0.010*"good" + 0.010*"final" + 0.009*"time"'),
 (2,
  '0.049*"new" + 0.032*"money" + 0.026*"got" + 0.020*"happi" + 0.017*"bought"
+ 0.015*"purchas" + 0.013*"car" + 0.011*"today" + 0.011*"dollar" + 0.011*"stor
e"'),
 (3,
  '0.039*"game" + 0.038*"video" + 0.029*"won" + 0.027*"watch" + 0.017*"complim
ent" + 0.015*"funni" + 0.014*"ticket" + 0.013*"test" + 0.011*"youtub" + 0.009
*"scratch"'),
 (4,
  '0.027*"new" + 0.017*"box" + 0.016*"bought" + 0.013*"wash" + 0.013*"fit" +
0.012*"happi" + 0.012*"said" + 0.011*"roti" + 0.010*"car" + 0.009*"clean"'),
 (5,
  '0.040*"summer" + 0.033*"came" + 0.033*"brother" + 0.028*"uncl" + 0.028*"tri
p" + 0.026*"grandma" + 0.017*"bless" + 0.017*"share" + 0.017*"futur" + 0.017
*"cooki"'),
 (6,
  '0.036*"moment" + 0.029*"happiest" + 0.020*"happi" + 0.016*"mom" + 0.015*"li
fe" + 0.011*"readi" + 0.011*"song" + 0.010*"everyday" + 0.010*"feel" + 0.010
*"key"'),
 (7,
  '0.043*"happi" + 0.016*"graduat" + 0.016*"exam" + 0.012*"rememb" + 0.010*"pa
ss" + 0.009*"life" + 0.009*"exercis" + 0.009*"medic" + 0.008*"studi" + 0.008
*"degre"'),
 (8,
  '0.065*"result" + 0.024*"mark" + 0.015*"examin" + 0.015*"govern" + 0.013*"an
xieti" + 0.013*"releas" + 0.013*"subject" + 0.012*"websit" + 0.009*"score" +
0.009*"director"'),
 (9,
  '0.032*"sold" + 0.029*"man" + 0.025*"photo" + 0.024*"listen" + 0.020*"music"
+ 0.016*"ebay" + 0.015*"item" + 0.013*"price" + 0.012*"natur" + 0.010*"blin
d"')]
```

The effect of parenthood on happy moments

In [22]:
```
get_cloud(parent, 'Word Cloud for Parents', '../figs/wordcloud_parenthood_parer
```

Word Cloud for Parents



In [23]:
```
get_cloud(childfree, 'Word Cloud for People Who Are Not Parents', '../figs/word
```

Word Cloud for People Who Are Not Parents

In [24]:
```
LDA_model(parent)
```

Out[24]:
```
[(0,
  '0.025*"got" + 0.025*"happi" + 0.020*"work" + 0.018*"son" + 0.016*"daughter"
+ 0.012*"new" + 0.011*"day" + 0.010*"home" + 0.009*"watch" + 0.009*"today"'),
 (1,
  '0.039*"happi" + 0.023*"friend" + 0.020*"day" + 0.019*"time" + 0.017*"went"
+ 0.014*"famili" + 0.012*"moment" + 0.010*"life" + 0.010*"month" + 0.010*"enjo
y"'),
 (2,
  '0.037*"car" + 0.021*"new" + 0.013*"bought" + 0.011*"happi" + 0.009*"dog" +
0.008*"final" + 0.008*"ice" + 0.008*"pass" + 0.008*"finish" + 0.007*"purcha
s"'),
 (3,
  '0.056*"happi" + 0.020*"state" + 0.019*"posit" + 0.015*"emot" + 0.014*"defi
n" + 0.013*"research" + 0.012*"pleasant" + 0.011*"psycholog" + 0.011*"differ"
+ 0.011*"mental"'),
 (4,
  '0.029*"person" + 0.027*"movement" + 0.022*"happiest" + 0.017*"time" + 0.012
*"result" + 0.012*"chang" + 0.011*"happi" + 0.010*"mark" + 0.010*"far" + 0.010
*"wait"'),
 (5,
  '0.005*"seri" + 0.005*"latest" + 0.004*"wend" + 0.003*"tempel" + 0.003*"epis
od" + 0.002*"harri" + 0.002*"movi" + 0.002*"screen" + 0.002*"new" + 0.002*"bui
lt"'),
 (6,
  '0.019*"film" + 0.013*"happi" + 0.012*"road" + 0.012*"cross" + 0.009*"star"
+ 0.007*"help" + 0.007*"blind" + 0.007*"shop" + 0.006*"music" + 0.006*"perso
n"'),
 (7,
  '0.006*"websit" + 0.005*"mix" + 0.005*"trump" + 0.005*"seri" + 0.004*"fligh
t" + 0.004*"rainbow" + 0.003*"releas" + 0.003*"fan" + 0.003*"special" + 0.003
*"singl"'),
 (8,
  '0.017*"happi" + 0.009*"meant" + 0.008*"fact" + 0.007*"unit" + 0.006*"discus
s" + 0.006*"right" + 0.006*"valu" + 0.006*"wide" + 0.005*"independ" + 0.005*"w
ritten"'),
 (9,
  '0.005*"circus" + 0.005*"art" + 0.004*"paisley" + 0.003*"coupl" + 0.003*"go"
+ 0.003*"william" + 0.003*"wife" + 0.002*"soak" + 0.002*"menu" + 0.002*"ste
w"')]
```

In [25]:
```
LDA_model(childfree)
```

Out[25]:
```
[(0,
  '0.035*"friend" + 0.022*"went" + 0.021*"happi" + 0.020*"time" + 0.019*"day"
+ 0.015*"got" + 0.012*"play" + 0.011*"game" + 0.010*"watch" + 0.009*"famil
i"'),
 (1,
  '0.032*"happi" + 0.026*"got" + 0.025*"work" + 0.018*"new" + 0.012*"month" +
0.012*"job" + 0.009*"day" + 0.009*"week" + 0.009*"abl" + 0.009*"receiv"'),
 (2,
  '0.041*"went" + 0.030*"dinner" + 0.024*"ate" + 0.018*"shop" + 0.018*"eat" +
0.017*"food" + 0.017*"favorit" + 0.016*"delici" + 0.015*"good" + 0.014*"go
t"'),
 (3,
  '0.037*"happi" + 0.020*"dollar" + 0.015*"won" + 0.013*"life" + 0.012*"plant"
+ 0.009*"garden" + 0.009*"man" + 0.008*"song" + 0.008*"ticket" + 0.007*"peop
l"'),
 (4,
  '0.067*"new" + 0.067*"bought" + 0.034*"phone" + 0.021*"card" + 0.014*"paid"
+ 0.012*"credit" + 0.011*"pair" + 0.010*"shoe" + 0.009*"stock" + 0.008*"bu
y"'),
 (5,
  '0.019*"loan" + 0.014*"yoga" + 0.010*"pokemon" + 0.009*"year" + 0.009*"clea
r" + 0.008*"money" + 0.007*"caught" + 0.007*"final" + 0.006*"paid" + 0.006*"tr
uck"'),
 (6,
  '0.049*"fix" + 0.021*"guitar" + 0.020*"learn" + 0.017*"fell" + 0.015*"bike"
+ 0.015*"ride" + 0.014*"asleep" + 0.011*"blood" + 0.010*"insur" + 0.008*"reci
p"'),
 (7,
  '0.073*"mother" + 0.049*"laugh" + 0.025*"sister" + 0.022*"lie" + 0.022*"bor
n" + 0.017*"birth" + 0.016*"sofa" + 0.016*"child" + 0.015*"tickl" + 0.013*"poc
ket"'),
 (8,
  '0.019*"paint" + 0.019*"gas" + 0.015*"parti" + 0.012*"intern" + 0.011*"thre
w" + 0.009*"fair" + 0.009*"station" + 0.008*"friend" + 0.007*"birthday" + 0.00
6*"prank"'),
 (9,
  '0.023*"roti" + 0.022*"said" + 0.015*"youtub" + 0.010*"plate" + 0.010*"cook"
+ 0.009*"lunch" + 0.009*"part" + 0.009*"mother" + 0.009*"seat" + 0.009*"watc
h"')]
```

Conclusions

The effect of age on happy moments

- LDA model outputs: Young
  - Topic 0: This topic seems to be related to daily activities and routines, possibly reflecting a sense of happiness in everyday life.
  - Topic 1: This topic might be related to accomplishments or achievements, including words like 'able,' 'final,' and 'exam.'
  - Topic 2: This topic appears to be about memorable moments, such as birthdays and cherished moments with friends.
  - Topic 3: This topic seems to contain words related to work, perhaps describing a positive working experience or great opportunities.
  - Topic 4: This topic is heavily related to gaming and sports, with words like 'game,' 'play,' 'video,' and 'team.' It seems to be about the joy of winning in these activities.

    - Topic 5: This topic is likely related to career achievements and successes, with words like 'job,' 'work,' and 'promotion.'
    - Topic 6: This topic might be about music and the joy of listening to favorite songs or bands.
    - Topic 7: This topic appears to be related to visual experiences, possibly describing scenes with water, puppies, and other imagery.
    - Topic 8: This topic is somewhat unclear but seems to mention 'league' and 'legend,' which could be related to sports or gaming.
    - Topic 9: This topic includes words like 'friend,' 'great,' and 'play,' possibly describing positive interactions with friends and enjoyable activities.
- LDA model outputs: Old
    - Topic 0: This topic seems to be related to daily routines and work-life balance. It includes words like 'happi,' 'got,' 'day,' 'time,' and 'work,' which might indicate happiness in managing daily activities.
    - Topic 1: This topic appears to be related to new experiences and events, including words like 'new,' 'work,' 'got,' 'month,' and 'year.'
    - Topic 2: This topic is likely about celebrating occasions like birthdays and talking to friends and family members, with words like 'birthday,' 'friend,' 'sister,' and 'talk.'
    - Topic 3: This topic contains somewhat unrelated words like 'master,' 'qualif,' and 'smoke,' which do not seem to form a coherent theme. It might be a less interpretable topic.
    - Topic 4: This topic seems to be related to work-related rewards and bonuses, with words like 'receiv,' 'bonus,' 'mail,' and 'work.'
    - Topic 5: This topic mentions words related to cooking and food, such as 'cooki,' 'wash,' 'chocol,' and 'chip.'
    - Topic 6: This topic appears to be about different moments in life, including happy moments and significant life events. It includes words like 'life,' 'moment,' 'good,' and 'happiest.'
    - Topic 7: This topic seems to be related to outdoor activities, particularly riding bikes and being on the road.
    - Topic 8: This topic contains words that are difficult to interpret and do not seem to form a coherent theme. It might be another less interpretable topic.
    - Topic 9: This topic could be related to film or entertainment, mentioning words like 'film,' 'releas,' 'star,' and 'version.'
- The LDA model for young people captures a wider diversity of topics, including gaming, music, visual experiences, and work-related achievements. The LDA model for old people focuses more on daily life, work, celebrations, and outdoor activities but also includes less coherent topics.

The effect of gender on happy moments

- LDA model outputs: Men
    - Topic 0: Work-related happiness, including words like 'work,' 'day,' and 'job.'

    - Topic 1: Personal and emotional happiness, including words like 'love,' 'wife,' and 'dog.'
    - Topic 2: Happiness related to family and social events, such as birthdays and spending time with friends and family.
    - Topic 3: General life happiness, potentially related to video games and social interactions.
    - Topic 4: Happiness associated with sports and games, particularly winning and watching favorite teams.
    - Topic 5: Food-related happiness, including words like 'ate,' 'delici,' and 'eat.'
    - Topic 6: Happiness related to psychological and emotional well-being, including words like 'posit,' 'research,' and 'emot.'
    - Topic 7: Fixing and repairing things, possibly car-related, with words like 'fix,' 'car,' and 'repair.'
    - Topic 8: Happiness related to music and entertainment, including words like 'song,' 'music,' and 'releas.'
    - Topic 9: Miscellaneous topics, possibly related to contests, games, and legends.
- LDA model outputs: Women
    - Topic 0: General daily life and work-related happiness.
    - Topic 1: Happiness related to spending time with friends, family, and having dinner.
    - Topic 2: Activities and experiences like winning, losing, and coffee-related happiness.
    - Topic 3: Experiences like watching, ordering, and visiting places.
    - Topic 4: Celebrations, birthdays, parties, and school-related happiness.
    - Topic 5: Happiness associated with pets, gardening, and playing.
    - Topic 6: Emotional moments and surprises in life.
    - Topic 7: Memories and achievements like graduation and degrees.
    - Topic 8: Unrelated words, possibly due to noise or less coherent topics.
    - Topic 9: Miscellaneous topics, including bath time and relaxation.
- The LDA model for men covers a broader range of topics, including work, gaming, music, and fixing things. The LDA model for women focuses more on personal and emotional experiences, celebrations, and memories.

The effect of martial status on happy moments

- LDA model outputs: Married
    - Topic 0: Happiness related to family and friends, including words like 'wife,' 'son,' and 'friend.'
    - Topic 1: Starting new things, potentially related to hobbies and gardening.
    - Topic 2: Enjoyment of games, watching, and new experiences.
    - Topic 3: Reflecting on life moments and happiness.
    - Topic 4: Shopping and buying new items.
    - Topic 5: Romantic and social experiences, including relationships and dating.
    - Topic 6: Psychological aspects of happiness, including research and emotions.
    - Topic 7: Possibly college or educational experiences.

    - Topic 8: Unrelated words, possibly due to noise or less coherent topics.
    - Topic 9: Miscellaneous topics, including films and titles.
- LDA model outputs: Single
    - Topic 0: Happiness related to friends, family, and time spent together.
    - Topic 1: Work-related happiness, including jobs and months.
    - Topic 2: Financial happiness, including money, purchases, and stores.
    - Topic 3: Entertainment and fun, including video games, compliments, and tickets.
    - Topic 4: Shopping and purchases of new items.
    - Topic 5: Family-related happiness and activities, including trips and cooking.
    - Topic 6: Reflecting on the happiest moments in life.
    - Topic 7: Academic and exercise-related happiness.
    - Topic 8: Educational results and achievements.
    - Topic 9: Selling items, photography, and music.
- The LDA model for single people has a more prominent focus on work-related and financial happiness, as well as activities. The LDA model for married people has a more coherent focus on relationships and interactions.

The effect of parenthood on happy moments

- LDA model outputs: Parents
    - Topic 0: Happiness related to work and family.
    - Topic 1: General happiness related to time, day, and family.
    - Topic 2: Purchasing items, particularly a car.
    - Topic 3: Psychological aspects of happiness, including emotions and mental well-being.
    - Topic 4: Happiness related to personal achievements, such as results and changes.
    - Topic 5: Unrelated words, possibly due to noise or less coherent topics.
    - Topic 6: Film and entertainment-related topics.
    - Topic 7: Miscellaneous topics, including website, politics, and facts.
    - Topic 8: Unrelated words, possibly due to noise or less coherent topics.
    - Topic 9: Miscellaneous topics, including art, food, and shopping.
- LDA model outputs: People who are childfree
    - Topic 0: Social interactions and activities with friends and family.
    - Topic 1: Work-related happiness, including jobs and receiving something new.
    - Topic 2: Dining and food-related happiness, including eating out and favorite foods.
    - Topic 3: Various aspects of life and happiness, including winning, gardening, and music.
    - Topic 4: Purchases of new items, particularly electronics and accessories.
    - Topic 5: Financial and transaction-related topics, including loans and payments.
    - Topic 6: Learning and activities, including fixing things and playing musical instruments.
    - Topic 7: Family-related happiness, including laughter and childhood memories.
    - Topic 8: Miscellaneous topics, including painting and parties.
    - Topic 9: Food-related topics, including cooking and meals.

- Both of these LDA models heavily mention family as a source of happiness. However, the LDA model for people who are childfree has a heavier focus on activities and experiences.

In [ ]: