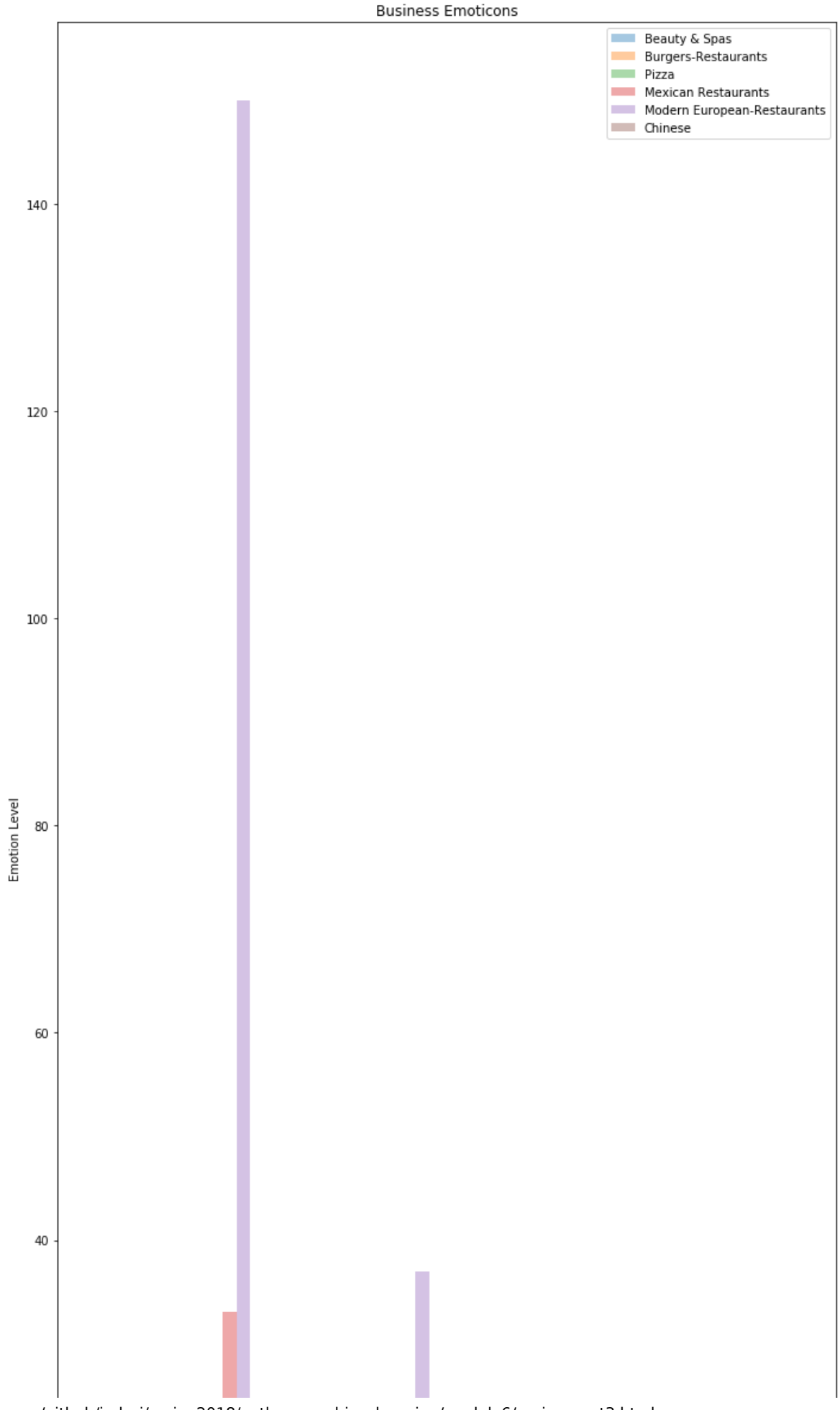
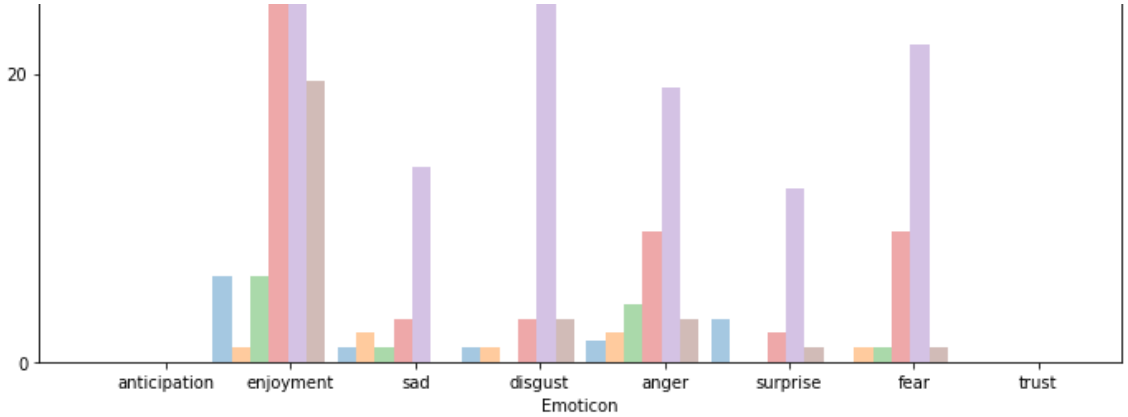
In [71]:
```
# Assignment 3 - Part 1
# You have been given a barGraph.csv file.
# Using the data of this file you have to draw a bar graph showing al
l 8 emotions corresponding to each business.

import os
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

data = pd.read_csv('barGraph.csv')
opacity = .4
bwidth = .15
xlabels = data['Business'].values
i = 1
for l in xlabels:
    vals = data[data['Business'] == l].values
    yvals = vals[0][1:]
    index = np.arange(len(yvals))
    plt.bar(index + (i * bwidth), yvals, bwidth, alpha=opacity, label
=l)
    i = i + 1

plt.xlabel('Emoticon')
plt.ylabel('Emotion Level')
plt.title('Business Emoticons')
plt.xticks(index + bwidth + .55, data.columns.values[1:])
plt.legend()
plt.tight_layout()
plt.rcParams["figure.figsize"] = [10, 20]
plt.show()
plt.clf()
plt.cla()
plt.close()
```


Business Emoticons
Beauty & Spas
Burgers-Restaurants
Pizza
Mexican Restaurants
Modern European-Restaurants
Chinese
Emotion Level
140
120
100
80
60
40



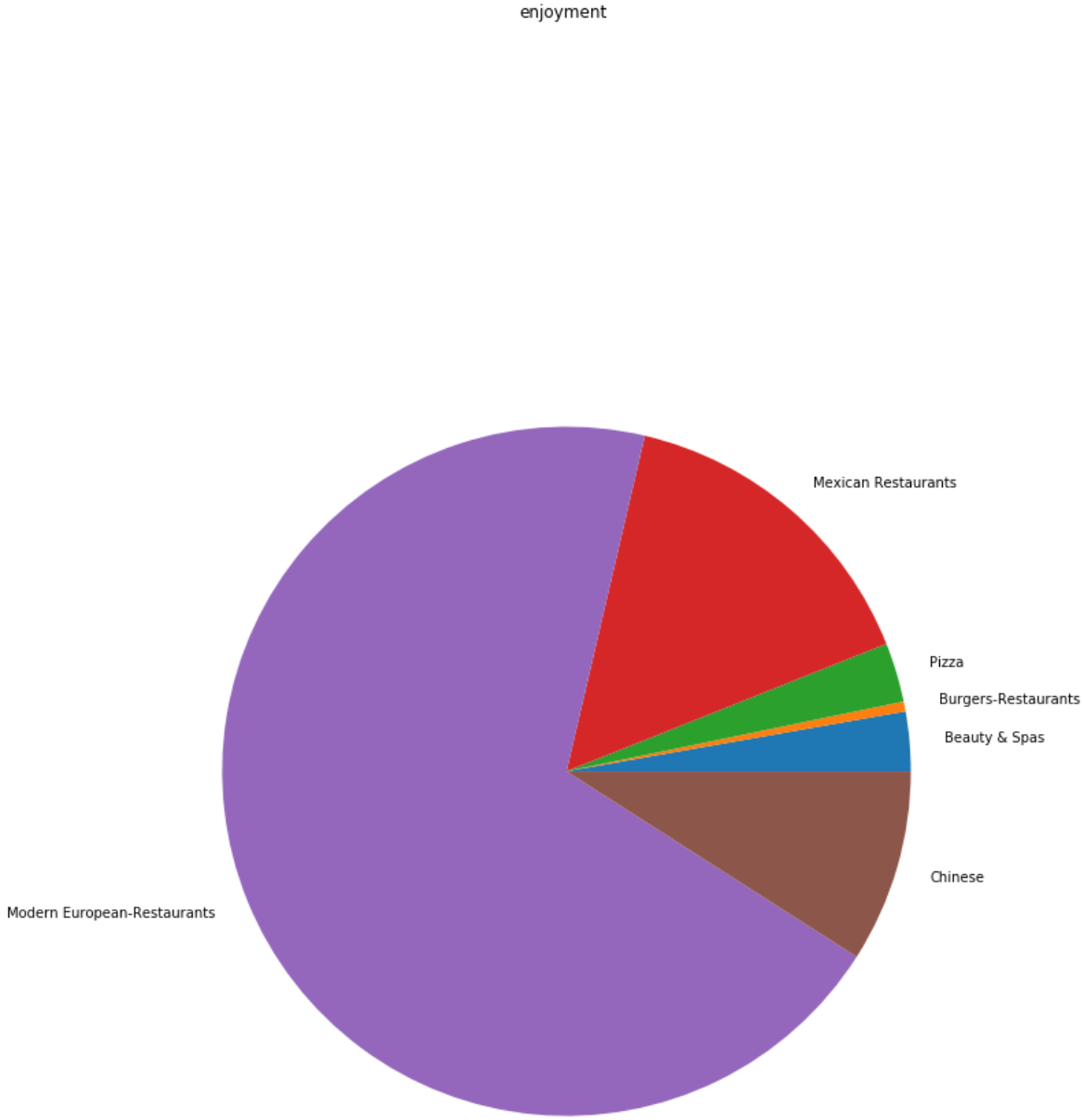
In [72]:
```
data
```

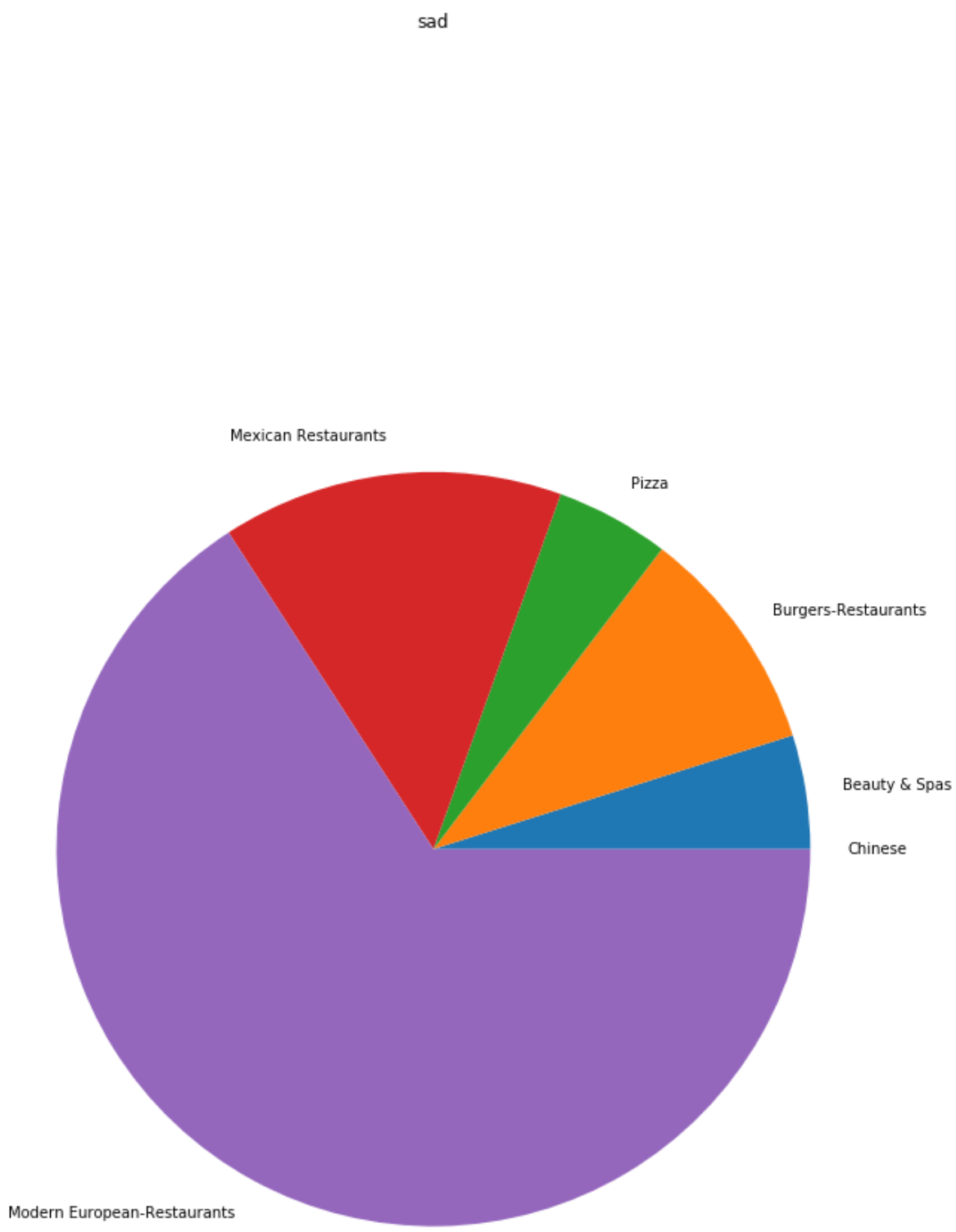
Out[72]:

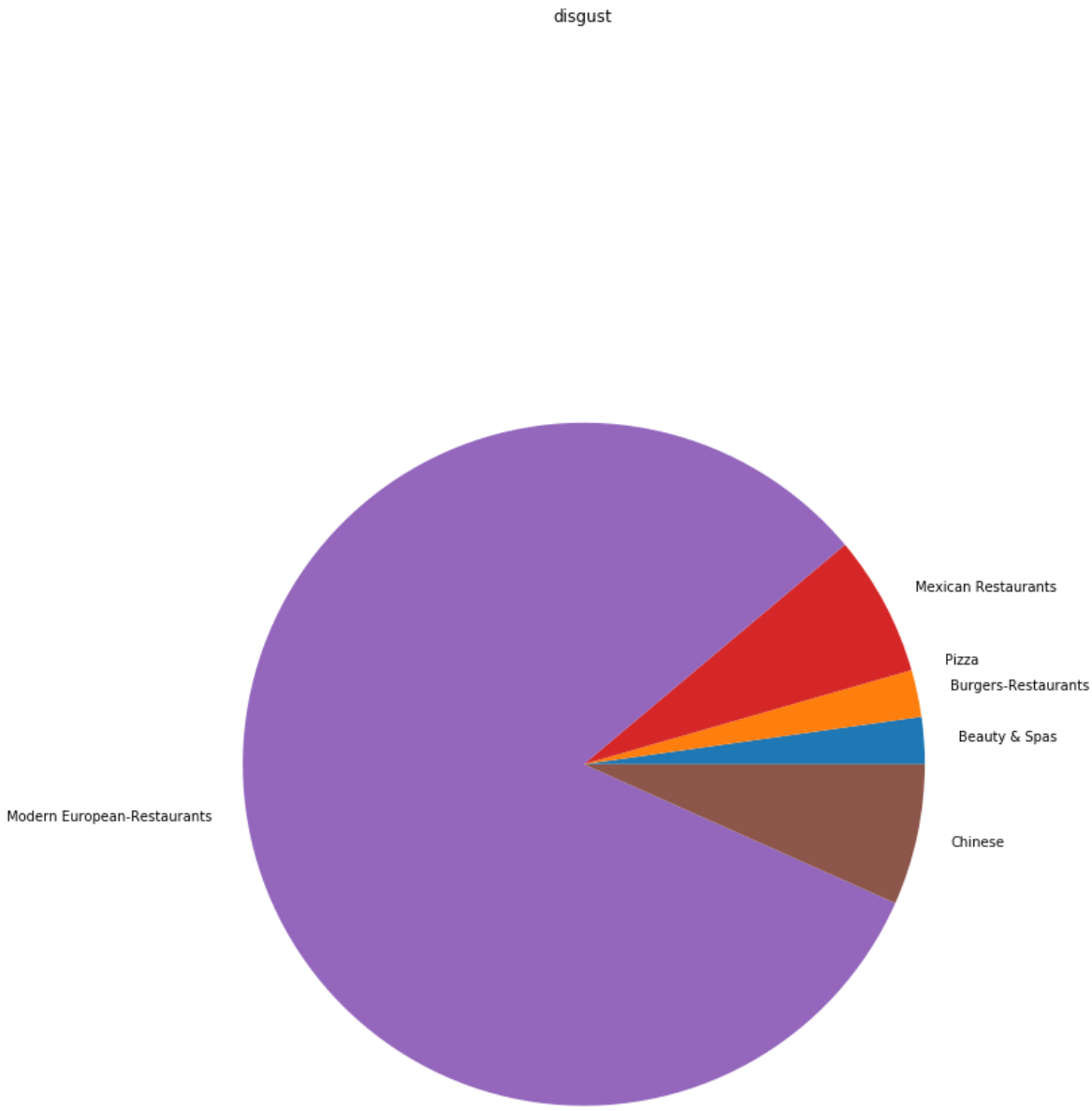
	Business	anticipation	enjoyment	sad	disgust	anger	surprise	fear	trust
0	Beauty & Spas	0	6.0	1.0	1	1.5	3	0	0
1	Burgers-Restaurants	0	1.0	2.0	1	2.0	0	1	0
2	Pizza	0	6.0	1.0	0	4.0	0	1	0
3	Mexican Restaurants	0	33.0	3.0	3	9.0	2	9	0
4	Modern European-Restaurants	0	150.0	13.5	37	19.0	12	22	0
5	Chinese	0	19.5	0.0	3	3.0	1	1	0

In [73]:
```
# Assignment 3 - Part 2
# Using the data present in barGraph.csv file generate pie-chart show
ing percentage of emotions
# for each business.

labels = data['Business'].values
emoticons = data.columns.values[1:]

for e in emoticons:
    vals = data[e].values
    if vals.sum() > 0:
        sizes = vals / vals.sum() * 100
        plt.figure()
        plt.title(e)
        plt.pie(sizes, labels = labels)
        plt.axis('equal')

plt.clf()
plt.cla()
plt.close()
```

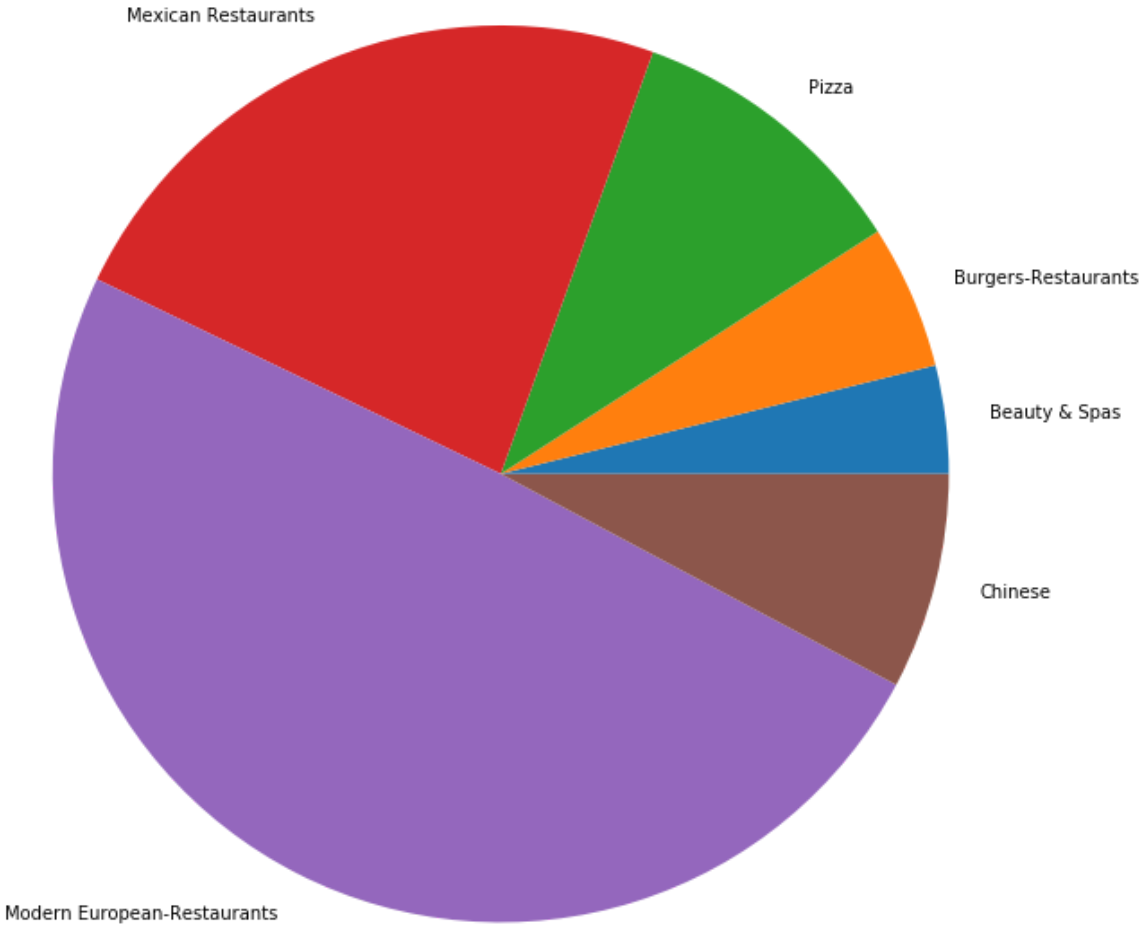
In [74]:
```
# Assignment 3 - Part 3. Generate a word cloud of your favorite news
 article or story or anything.
# This word cloud should contain words having 4 letters or more.
from wordcloud import WordCloud

text = open('words.txt').read()
wordcloud = WordCloud().generate(text)
plt.imshow(wordcloud)
plt.axis("off")
# Code borrowed from course module instructions
wordcloud = WordCloud(background_color="white", max_words=2000,max_fo
nt_size=40, relative_scaling=.4).generate(text)
plt.figure()
plt.imshow(wordcloud)
# plt.axis("off")
plt.show()
plt.clf()
plt.cla()
plt.close()
```


enjoyment
Mexican Restaurants
Pizza
Burgers-Restaurants
Beauty & Spas
Chinese
Modern European-Restaurants


sad
Mexican Restaurants
Pizza
Burgers-Restaurants
Beauty & Spas
Chinese
Modern European-Restaurants

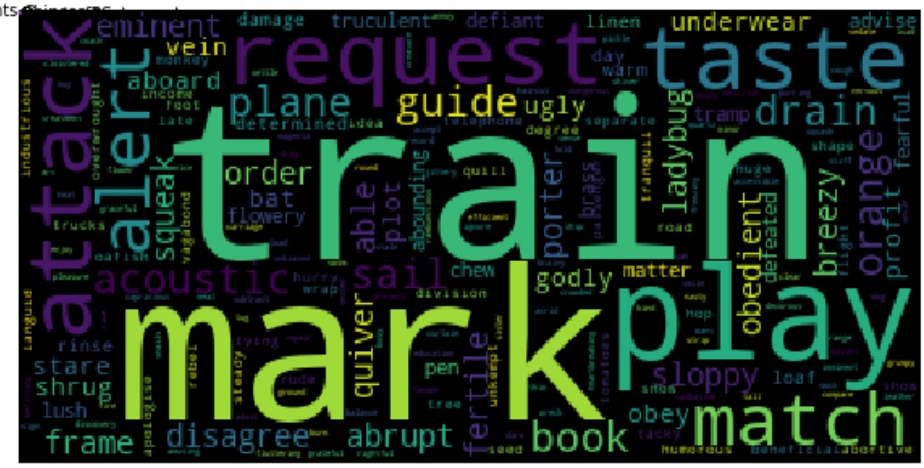

disgust
Mexican Restaurants
Pizza
Burgers-Restaurants
Beauty & Spas
Chinese
Modern European-Restaurants

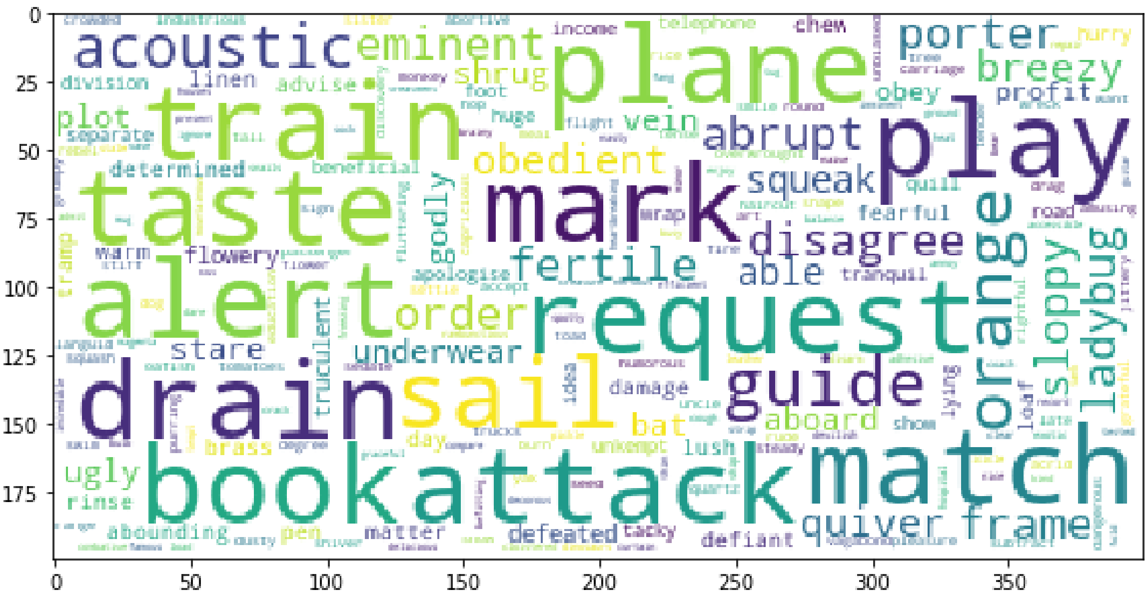
anger


Mexican Restaurants
Pizza
Burgers-Restaurants
Beauty & Spas
Chinese
Modern European-Restaurants

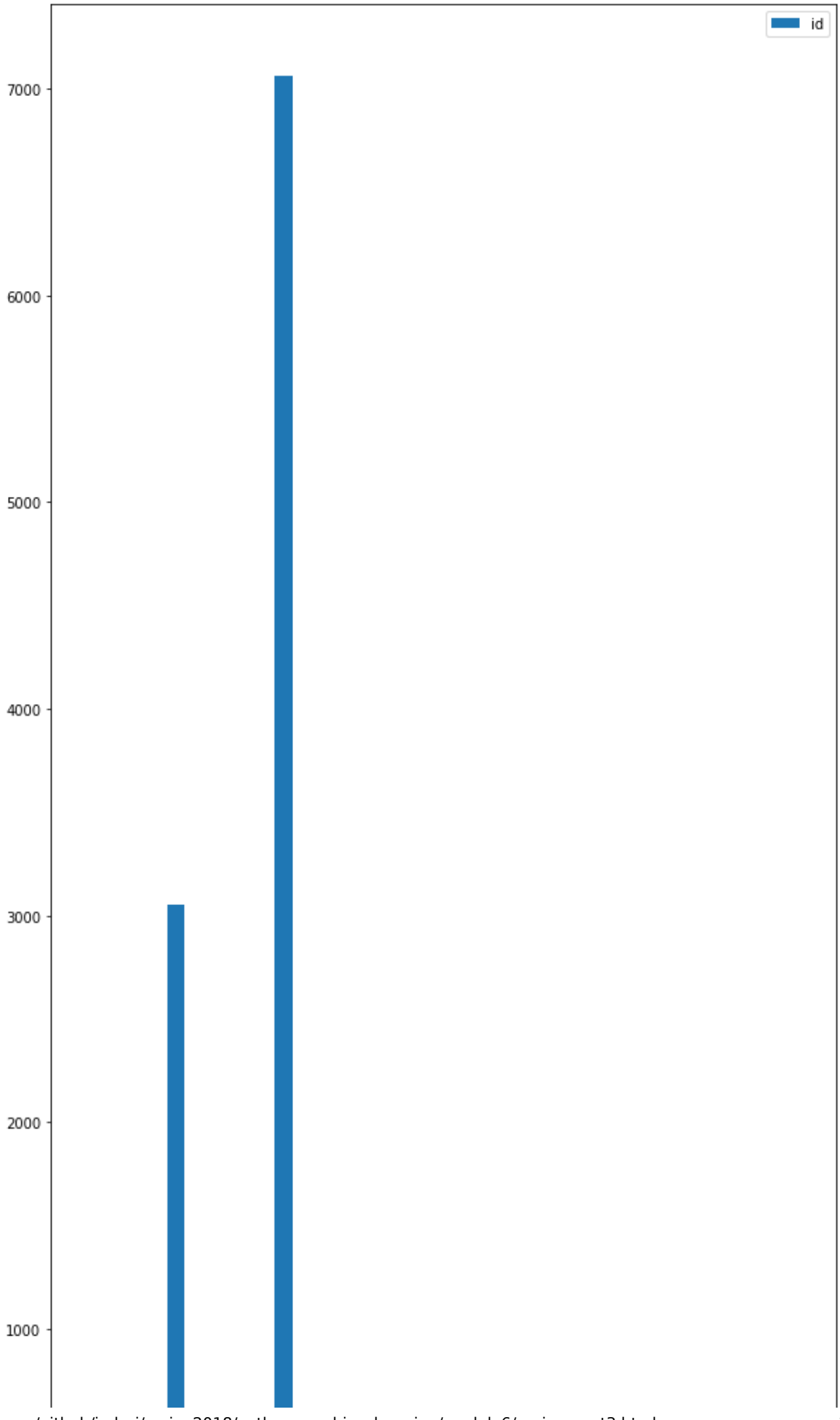
surprise

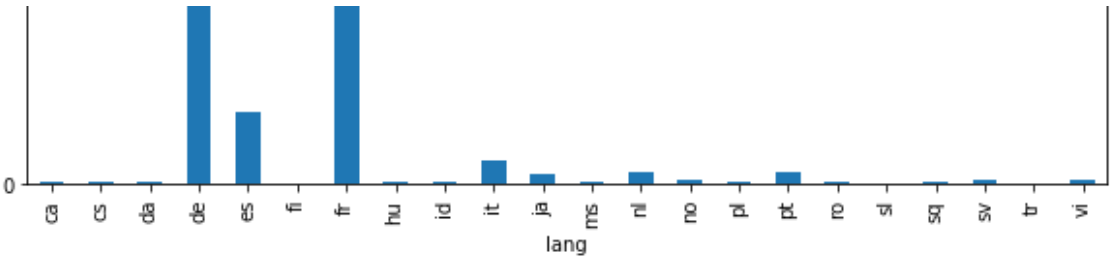
Modern European-Restaurants




0
25
50
75
100
125
150
175
0
50
100
150
200
250
300
350
acoustic
eminent
plane
porter
breezy
train
play
taste
mark
abrupt
obedient
alert
request
orange
disagree
fertile
drain
sail
guide
book
attack
match
quiver
frame
ladybug
sloppy
squeak
order
underwear
stare
truculent
aboard
ugly
rinse
abounding
defeated
defiant
plot
separate
determined
division
linen
advise
shrug
vein
obey
profit
fearful
able
flowery
godly
warm
damage
bat
lush
brass
unkempt
matter
pen
tacky
telephone
income
chew
industrious
crowded
abortive
hurry
beneficial
apologise
tranquil
road
quill
idea
lying
loaf
day
burn
show
acrid

In [75]:
```
"""
Assignment 3 - Part 4
You have been given a file ReviewID.txt. It has 10646 records in it,
 each record is made up of two fields
separated by a colon: like AzSn8aTOyVTUePaIQtTUYA:es . The first fiel
d is review ID and the second field
is language in which reviews has been written. Read this file and cre
ate a bar graph showing the percentage
of the reviews written in a particular language. The aim of this prob
lem is to generate a graph using which
we can do a comparative analysis of the languages used for writing re
views.
"""
review_data = pd.read_csv('ReviewID.txt', delimiter=':', header=None,
 names=['id', 'lang'])
review_data.head(10)

# Group by language
grouped_reviews = review_data.groupby(['lang']).count()
grouped_reviews.plot.bar()
plt.show()
```


id
7000
6000
5000
4000
3000
2000
1000


0
ca
cs
da
de
es
fi
fr
hu
id
it
ja
ms
nl
no
pl
pt
ro
sl
sq
sv
tr
vi
lang