

# K Means Clustering, Agglomerative Clustering, DBSCAN Clustering:

Objective :

1. Perform featurization BoW,TFIDF, Avg Word2Vec, tf-idf-Word2Vec.
2. At the end of the assignment, please elaborate on the step by step procedure you followed to solve the assignment.

## 2. Agglomerative Clustering:

1. Apply agglomerative algorithm and try a different number of clusters like 2,5 etc.
2. Same as that of K-means, plot word clouds for each cluster.
3. You can take around 5000 reviews or so(as this is very computationally expensive one) to perform hierarchical clustering because they do take a considerable amount of time to run.

### Step-By-Step procedure

1. In this, we need to work with all 4-vectorizers (BOW, TFIDF,Avg w2v and TF-IDF weighted w2v) where we will convert our texted review into numerical(vector) form in order to apply any Model on it.
2. After that we will take our cleandtext(i.e cleand text means we have already cleaned our data by removing stops words, other this which are going to affect our model) and then we will take only test data not their respective class label because as we know we are going to apply K-means clustering algo which do not required class labels. what it does is it mainly group/clusters the similar data points
3. As we know Agglomerative is a Clustering algo but it works in different way that the k-means works. As we know Agglomerative clustering starts with each points being a individual clusters and then it starts to group the clusters based on some sense of similarities/distance and so we don't need to give the or tune to getting best k as a hyperparameters and if we want n\_clusters we can get it at that range. So we try to implement Agglomerative clustering with different number of clusters and try to see the reviews in the clusters

In [1]:

```
%matplotlib inline
import warnings
warnings.filterwarnings("ignore")
import warnings
warnings.filterwarnings("ignore", category=DeprecationWarning)

import sqlite3
import pandas as pd
import numpy as np
import nltk
import string
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.feature_extraction.text import TfidfTransformer
from sklearn.feature_extraction.text import TfidfVectorizer

from sklearn.feature_extraction.text import CountVectorizer
from sklearn.metrics import confusion_matrix
from sklearn import metrics
from sklearn.metrics import roc_curve, auc
from nltk.stem.porter import PorterStemmer

import re
# Tutorial about Python regular expressions: https://pymotw.com/2/re/
import string

from gensim.models import Word2Vec
from gensim.models import KeyedVectors
import pickle
#taking cleaned data i.e in Reviews table from final sql database
#making connection with database
conn = sqlite3.connect('final.sqlite')
final = pd.read_sql_query(""" SELECT * FROM Reviews""", conn)
```

```
C:\Users\nisha\Anaconda3\lib\site-packages\gensim\utils.py:1212: UserWarning: detected Windows; aliasing chunkize to chunkize_serial
warnings.warn("detected Windows; aliasing chunkize to chunkize_serial")
```

In [2]:

```
final = final[:5000]
print(len(final))
```

5000

In [3]:

```
CleanedText = final['CleanedText'];
text=final.CleanedText.values
```

In [4]:

```
from tqdm import tqdm
import os

from wordcloud import WordCloud, STOPWORDS
import matplotlib.pyplot as plt
stopwords = set(STOPWORDS)
import seaborn as sns;

def cluster_wordcloud_generated_image_fun(text):
    wordcloud = WordCloud(max_font_size=50, max_words=100, stopwords=stopwords ,background_color="white").generate(text)
    plt.figure()
    plt.imshow(wordcloud, interpolation="bilinear")
    plt.axis("off")
    plt.show()
```

## Bow

Applying Bow vectorizer on data

In [5]:

```
#BOW
from sklearn.feature_extraction.text import CountVectorizer
vectorizer = CountVectorizer(min_df=10)
vocabulary= vectorizer.fit(text)
#print("the shape of out text BOW vectorizer ",vocabulary.get_shape())

bow_data= vectorizer.transform(text)
print("the shape of out text BOW vectorizer ",bow_data.get_shape())
```

the shape of out text BOW vectorizer (5000, 2240)

## Agglomerative Clustering - Bow

In [15]:

```
from sklearn.cluster import AgglomerativeClustering
reviews = final['Text'].values
n_clusters_lists = [2,3,5]

for cluster in n_clusters_lists:
    agclf = AgglomerativeClustering(n_clusters=cluster).fit(bow_data.toarray())
    # agclf.fit(bow_data)
    print("#" * 25, "Agglomerative Clustering for Number of Clusters = ", cluster, " ", "*" * 25)
    print("\nReviews for each of the clusters : \n")

    labels = agclf.labels_
    # Printing two reviews from each cluster
    for i in range(cluster):
        review_count = 1 ;
        print("#" * 80)
```

```

print("Cluster ",i+1);
print(" " * 80)
for j in range(agclf.labels_.shape[0]):
    if agclf.labels_[j] == i:
        if review_count == 3:
            break;
        else:
            print('Review - ',review_count,'And review Index No - ',j)
            print(reviews[j])
            cluster_wordcloud_generated_image_fun(reviews[j])
            print('\n');
            review_count = review_count+1;
# cluster_1.append(reviews[i])

print(" " * 100)
print("\n\n")

```

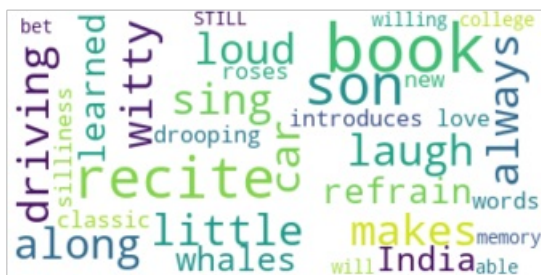
##### Agglomerative Clustering for Number of Clusters = 2  
 \*\*\*\*\*

Reviews for each of the clusters :

\*\*\*\*\*  
 Cluster 1  
 \*\*\*\*\*

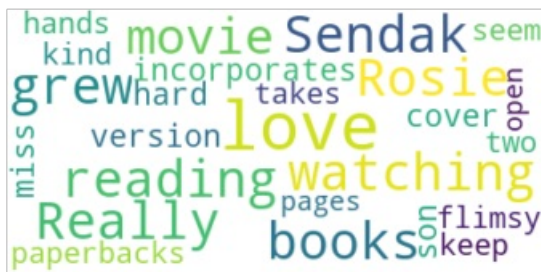
Review - 1 And review Index No - 0

this witty little book makes my son laugh at loud. i recite it in the car as we're driving along and he always can sing the refrain. he's learned about whales, India, drooping roses: i love all the new words this book introduces and the silliness of it all. this is a classic book i am willing to bet my son will STILL be able to recite from memory when he is in college



Review - 2 And review Index No - 1

I grew up reading these Sendak books, and watching the Really Rosie movie that incorporates them, and love them. My son loves them too. I do however, miss the hard cover version. The paperbacks seem kind of flimsy and it takes two hands to keep the pages open.



\*\*\*\*\*  
 Cluster 2  
 \*\*\*\*\*

Review - 1 And review Index No - 898

It was while at a friends home watching her use the grocery store food colour she had bought that I was reminded of why I love these Ateco full spectrum food colours. I am very very French in what I eat and how I bake and the products I need/want/use.<br /><br />Being a gel, you need less and t

he colours can be soft or bold. If you are new to the gel I recommend using a toothpick to get a wee bit to mix in the dough, frosting, candy, cake, cookie you are making to better see if you actually need more.

I love the violet for when I make my B&eacutere&egrave;re Abraham French macarons which calls for purple food colour to match the violet extract flavor. Or green for when I make his Pistachio or Kiwi macarons. And the orange colour is magnificent and is used for one of my favorite orange liqueur frostings.

Fact is these food colours are unlike any run of the mill food colour.

They have the USA-K symbol on the box which I denotes kosher in the USA.



Review - 2 And review Index No - 1134

This kibble has been the staple diet of my 80 lb Great Dane Mix (named Kobe) for the past year. He came to me two years ago from a rescue group who transports black dogs from the southern U.S, where black dogs have a very high kill rate, to the northern U.S. to be adopted. He was about 5 yrs old at the time I adopted him, sight unseen, from Arkansas. Severe allergies were included among the myriad of issues that afflicted him. This manifested itself in runny eyes, full body skin yeast infection resulting in what the vet termed "elephant skin" on his abdomen, throat, elbows, and hunches. Also bald spots over his neck, throat, thighs, and toes from constant scratching and biting at himself. In many places the "elephant skin" was perpetually split/cracked, infected, and bleeding. After treating the bacterial and yeast infections with both topical and oral agents which did clear up the bacteria but left the yeast as a chronic secondary infection of the allergies, I searched for other remedies. Along with seasonal, environmental, and other allergies to dust, fabrics, vegetation, etc. I also found a lot of good information about food allergies. Taking this information into consideration I scouted around for an "allergenic" kibble that was made by a company that has never had a recall, does not do animal testing, provides appropriate nutrition, and is somewhat affordable. Wysong seemed to fit the bill. Anergex has been a key part of Kobe's allergy treatment program for the past year. Along with "therapeutic baths", Ketokonazole lotions, hydrocortisone lotions, diligent laundering of doggie coats/bedding/linens (no wool), and oral antihistamines Kobe's skin is soft and smooth without splits/cracks/infection; his bald spots have shrunk to half the size; and his eyes only run occasionally. He still does scratch and bite at himself intermittently. I notice this usually happens when I deviate from the Anergex diet, giving him rawhide chews, beef bones with marrow, or cookies (organic) with mixed ingredients. I don't really know what else, short of steroids, can be done to help him. I have avoided steroids due to their own detrimental health effects. Anergex seems to be the food solution to his allergies, which is a good thing! I add water to the kibble and let it sit about 10 minutes to form a kind of "gravy". He likes this better than the straight dry kibble and it also helps to keep him hydrated. Note: My other dog, a 9yr old husky-shepherd mix eats Wysong Maintenance and is strong and svelt; and the cats eat Wysong Uretic.



```
##### Agglomerative Clustering for Number of Clusters = 3
*****
```

Reviews for each of the clusters :



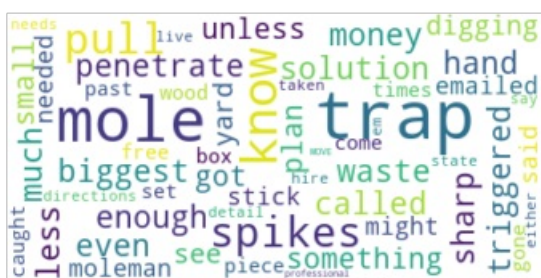
This kibble has been the staple diet of my 80 lb Great Dane Mix (named Kobe) for the past year. He came to me two years ago from a rescue group who transports black dogs from the southern U.S, where black dogs have a very high kill rate, to the northern U.S. to be adopted. He was about 5 yrs old at the time I adopted him, sight unseen, from Arkansas. Severe allergies were included among the myriad of issues that afflicted him. This manifested itself in runny eyes, full body yeast infection resulting in what the vet termed "elephant skin" on his abdomen, throat, elbows, and hunches. Also bald spots over his neck, throat, thighs, and toes from constant scratching and biting at himself. In many places the "elephant skin" was perpetually split/cracked, infected, and bleeding. After treating the bacterial and yeast infections with both topical and oral agents which did clear up the bacteria but left the yeast as a chronic secondary infection of the allergies, I searched for other remedies. Along with seasonal, environmental, and other allergies to dust, fabrics, vegetation, etc. I also found a lot of good information about food allergies. Taking this information into consideration I scouted around for an "allergenic" kibble that was made by a company that has never had a recall, does not do animal testing, provides appropriate nutrition, and is somewhat affordable. Wysong seemed to fit the bill. Anergex has been a key part of Kobe's allergy treatment program for the past year. Along with "therapeutic baths", Ketokonazole lotions, hydrocortisone lotions, diligent laundering of doggie coats/bedding/linens (no wool), and oral antihistamines Kobe's skin is soft and smooth without splits/cracks/infection; his bald spots have shrunk to half the size; and his eyes only run occasionally. He still does scratch and bite at himself intermittently. I notice this usually happens when I deviate from the Anergex diet, giving him rawhide chews, beef bones with marrow, or cookies (organic) with mixed ingredients. I don't really know what else, short of steroids, can be done to help him. I have avoided steroids due to their own detrimental health effects. Anergex seems to be the food solution to his allergies, which is a good thing! I add water to the kibble and let it sit about 10 minutes to form a kind of "gravy". He likes this better than the straight dry kibble and it also helps to keep him hydrated. Note: My other dog, a 9yr old husky-shepherd mix eats Wysong Maintenance and is strong and svelt; and the cats eat Wysong Uretic.



Cluster 3

Review - 1 And review Index No - 439

This was the biggest waste of our money. I don't know what the solution is, but this trap is not it. The so called "spikes" are not sharp enough to even penetrate my hand when triggered, much less something as small as a mole. When you pull the trap up, you don't know if you got the mole, because, unless you plan on digging up your yard to see if the mole might be there, you don't know. We emailed the moleman, and he said that we needed to stick a piece of wood under the trap so that when the "spikes" come down, they do not just trap the mole. So the past 6 times this trap has gone off, we have caught the mole, only to pull the trap out and set him free. No where on the box or in the directions does it state the detail that needs to be taken. I say, either live with em, or hire a professional. Or, MOVE!!!!!!!!!!!!



Review - 2 And review Index No - 441

This gopher trap is designed poorly. We have gophers and bought this trap online. The only problem is that the part that should come down and trap/kill the gopher gets stuck with the dirt when you set the trap up. I'm not sure if we didn't dig down far enough or what the problem was, but the gopher issue was not solved with this product.





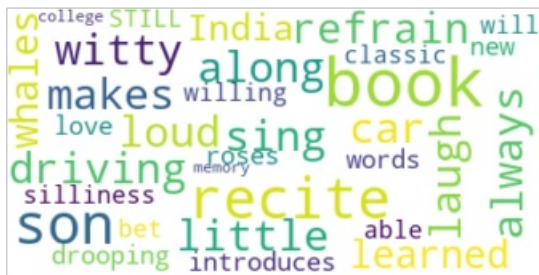
```
##### Agglomerative Clustering for Number of Clusters = 5
*****
```

Reviews for each of the clusters :

```
*****
Cluster 1
*****
```

Review - 1 And review Index No - 0

this witty little book makes my son laugh at loud. i recite it in the car as we're driving along and he always can sing the refrain. he's learned about whales, India, drooping roses: i love all the new words this book introduces and the silliness of it all. this is a classic book i am willing to bet my son will STILL be able to recite from memory when he is in college



Review - 2 And review Index No - 1

I grew up reading these Sendak books, and watching the Really Rosie movie that incorporates them, and love them. My son loves them too. I do however, miss the hard cover version. The paperbacks seem kind of flimsy and it takes two hands to keep the pages open.



```
*****
Cluster 2
*****
```

Review - 1 And review Index No - 898

It was while at a friends home watching her use the grocery store food colour she had bought that I was reminded of why I love these Ateco full spectrum food colours. I am very very French in what I eat and how I bake and the products I need/want/use. Being a gel, you need less and the colours can be soft or bold. If you are new to the gel I recommend using a toothpick to get a

ne colours can be soft or bold. If you are new to the gel I recommend using a toothpick to get a wee bit to mix in the dough, frosting, candy, cake, cookie you are making to better see if you actually need more.<br /><br />I love the violet for when I make my B&eacuteng&egrave;re Abraham French macarons which calls for purple food colour to match the violet extract flavor. Or green for when I make his Pistachio or Kiwi macarons. And the orange colour is magnificent and is used for one of my favorite orange liqueur frostings.<br /><br />Fact is these food colours are unlike any run of the mill food colour.<br /><br />They have the USA-K symbol on the box which I denotes kosher in the USA.



Review - 2 And review Index No - 1134

This kibble has been the staple diet of my 80 lb Great Dane Mix (named Kobe) for the past year. He came to me two years ago from a rescue group who transports black dogs from the southern U.S, where black dogs have a very high kill rate, to the northern U.S. to be adopted. He was about 5 yrs old at the time I adopted him, sight unseen, from Arkansas. Severe allergies were included among the myriad of issues that afflicted him. This manifested itself in runny eyes, full body skin yeast infection resulting in what the vet termed "elephant skin" on his abdomen, throat, elbows, and hunches. Also bald spots over his neck, throat, thighs, and toes from constant scratching and biting at himself. In many places the "elephant skin" was perpetually split/cracked, infected, and bleeding. After treating the bacterial and yeast infections with both topical and oral agents which did clear up the bacteria but left the yeast as a chronic secondary infection of the allergies, I searched for other remedies. Along with seasonal, environmental, and other allergies to dust, fabrics, vegetation, etc. I also found a lot of good information about food allergies.  
Taking this information into consideration I scouted around for an "allergenic" kibble that was made by a company that has never had a recall, does not do animal testing, provides appropriate nutrition, and is somewhat affordable. Wysong seemed to fit the bill. Anergex has been a key part of Kobe's allergy treatment program for the past year. Along with "therapeutic baths", Ketokonazole lotions, hydrocortisone lotions, diligent laundering of doggie coats/bedding/linens (no wool), and oral antihistamines Kobe's skin is soft and smooth without splits/cracks/infection; his bald spots have shrunk to half the size; and his eyes only run occasionally. He still does scratch and bite at himself intermittently. I notice this usually happens when I deviate from the Anergex diet, giving him rawhide chews, beef bones with marrow, or cookies (organic) with mixed ingredients.  
I don't really know what else, short of steroids, can be done to help him. I have avoided steroids due to their own detrimental health effects. Anergex seems to be the food solution to his allergies, which is a good thing! I add water to the kibble and let it sit about 10 minutes to form a kind of "gravy". He likes this better than the straight dry kibble and it also helps to keep him hydrated.  
Note: My other dog, a 9yr old husky-shepherd mix eats Wysong Maintenance and is strong and svelt; and the cats eat Wysong Uretic.



```
*****
Cluster 3
```

Review - 1 And review Index No - 439

This was the biggest waste of our money. I don't know what the solution is, but this trap is not it. The so called "spikes" are not sharp enough to even penetrate my hand when triggered, much less something as small as a mole. When you pull the trap up, you don't know if you got the mole, because, unless you plan on digging up your yard to see if the mole might be there, you don't know. We emailed the moleman, and he said that we needed to stick a piece of wood under the trap so that when the "spikes" came down, they do not just trap the mole. So the past 6 times this trap has co

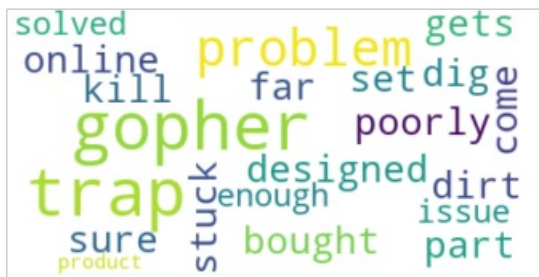


when the spikes come down, they do not just trap the mole. So the past 6 times this trap has gone off, we have caught the mole, only to pull the trap out and set him free. No where on the box or in the directions does it state the detail that needs to be taken. I say, either live with em, or hire a professional. Or, MOVE!!!!!!!!!!!!



Review - 2 And review Index No - 441

This gopher trap is designed poorly. We have gophers and bought this trap online. The only problem is that the part that should come down and trap/kill the gopher gets stuck with the dirt when you set the trap up. I'm not sure if we didn't dig down far enough or what the problem was, but the gopher issue was not solved with this product.



```
*****
Cluster 4
```

Review - 1 And review Index No - 2224

It absolutely amazes me that some people spend over two dollars for a single cup of coffee, yet still buy those low cost tea bags from the grocery store, and then claim to like coffee better. Many people have not tasted a real good cup of tea. Good tea does not come in a 200 count per cardboard box - the tea should have higher quality and shouldn't be in a cheap cardboard box at all. It should be in a sealed tin or packed in an airtight bag in order to preserve the flavor, and the leaves should be specially selected. The cheapest teas are bought in bulk from the lowest quality teas parts of the tea leaf. Yes, cost does matter in tea, just like coffee.

Unless you're really lazy (which, I admit, I am), teas should also never, ever, be served from a bag. Tea should be in loose form, not confined in a bag. Why? Because tea leaves need to expand in the water to release their full flavor. You can buy loose tea leaves buy on amazon, high end grocery store, or in specialty markets (e.g. Cost Plus), or tea shoppes.

Now, unless you want to brew a whole pot with tea leaves in the pot, there are not many ways to make tea right for just one person. This mug solve the single serving dilemma. It's big, sturdy, and made of glass. The most important thing though is that it comes with an infuser. It looks like a cylindrical filter that is placed into the mug. Just spoon a bit of the tea into the filter, dunk in the mug, and the tea will be brewed right then and there. The result is some of the best tea you'll have ever drank. The brewing is really easy, add tea to the filter, and add sub boiling hot water, wait, remove the infuser, and drink. If you need to add sugar or cream, well, that's your choice, but I believe the right tea needs none of that.

To make tea, place one to two spoonfuls in the filter, then pour in 160 degree water for green, steeped for a little less than 2 minutes, 180 degrees for white tea, steeped for over 5 minutes, and about 200 degrees for black or oolong, steeped for about three to four minutes. No - you don't have to measure the temperature each time, only when first starting out. Just bring the water to boiling, then let it cool down until the desired temperature is reached. Count those seconds. Next time you make tea, boil, wait for those amount of seconds and then pour in the mug. Easy.

I suggest trying different teas (without adding anything), until you find your right one. My personal favorite is Green Well Green Tea, but I still have a lot to try.

Tea is like wine, cheese, coffee, or beer. Quality and taste vary tremendously, and if you only try the cheap stuff, you'll never really enjoy it like you should.





ing some other reviews here and I wanted to clear up some misconceptions and misunderstandings.  
First, if it looks like the moles are going around your trap, the most likely cause is that the trap is set too deep. That ties to another reviewer who said they set the trap with the trigger at grass level. That would be too shallow. Before you set the trap, tamp down the run to GROUND level (not grass level). When you cut the notches for the trap blades, and place the trap in the ground, the trigger should lay even with the rest of the ground.  
Second, while true that moles will eat grubs, their primary food source are worms. Worms are usually more plentiful, and they move through the ground, thus making them easier to find for a hungry mole. Grubs are more stationary. A mole will eat them, but it requires more luck to find them.  
Third, if you don't get a mole in the first day, don't panic. Give it a few days. Remember, the best place to put the traps are in the long straight runs. Those are the mole highways. Just because it's a highway though, doesn't mean your mole will use that highway every day. Give it some time. It's not necessary to move your traps too often. Even if your mole seems to be active in another area of your yard, if you picked a good spot to place your trap, he will come back and use his highway again.  
Fourth, moles are territorial. There may be some overlap, but it's minimal. HOWEVER, keep this in mind. Just like you, moles prefer to travel along pre-dug roads. You don't build a new road every time you want to go to the grocery store. Moles don't like to either. So, even if you kill your mole, it is VERY likely that another mole will move into your yard. Keep in mind also, that the runs you see at ground level, are only a fraction of the amount of tunnels that might be running through your yard. Moles don't hibernate. They follow the worms in the winter, and worms dig in deeper during the winter. For a mole to survive the winter, that mole has a whole labyrinth of tunnels deep under ground. Easy pickings for another mole looking to move into your yard.  
EDIT 4-29-10: Taught my boss how to use these traps and he got his first mole this weekend. The moles in St. Louis are having a bad day!  
EDIT 5-28-10: Got another one last night. I've had a mole kicking around the area for the last few months and just couldn't get a good run to trap. Finally, the idiot moved into another area of the yard, and to do that he had to make a couple long straight runs to get there. Well, that was his last mistake. I had the traps in the ground less than 12 hours. I woke up this morning to a dead mole. There really is nothing better with your Corn Flakes than a dead mole.  
EDIT 7-22-10: So about a week after I killed my last mole in May I got another one destroying my yard. Took me 2 months, but that mole is now dead. What a wonderful thing heavy rain is. When it rains, it brings the worms up, and the moles are soon to follow. Right into the waiting arms of the sharp teeth of my mole trap.  
EDIT 9-9-10: Pics are up! If anyone has any suggestions or questions about the pics, just leave a comment and I'll see what I can do.  
EDIT 9-10-10: Less than 12 hours later, I'm the proud parent of a dead mole. I kill more moles before 9am than most people do all day.  
EDIT 9-30-10: Posted 3 pics that show how to prep the trap itself before setting it. I also tried to label some components just so if there are questions in the future, we all have a reference for terminology. I'm sure my verbiage isn't right, but as long as we all understand the different parts, that's all that matters.  
EDIT 10-22-10: There have been a number of commenters and reviewers that have asked about sandy/loose soil. We haven't been able to give a very good answer because I don't have any experience in that kind of condition. Another reviewer does though, and he did a really good job of explaining his technique. The following is a link to his comments. Thanks Bjay25.  
[...]  
So it looks like Amazon won't hyper link in reviews, but just copy that link and paste it into another tab on your browser and that will take you right to Bjay25's comments. That should be real helpful for people in the south that are dealing with sandy soil.  
EDIT 1-25-11: Doesn't everyone hunt moles in the dead of winter? I got my first winter mole last night. I have all but moved in with my girlfriend, and after taking care of her yard last summer, we started noticing some mole activity. Mostly in a flower bed, which is a tough place to trap. I kept waiting for a good run to trap, but never got one. Then winter came, and I kind of forgot about it. I figured I'd get him next spring. So then I was out taking down xmas lights, and the yard was TORE UP! I tried stomping down runs, and was hoping I would get a highway to trap, but couldn't get a good one. I finally got fed up, set a few traps and stomped down all the other runs. About a week ago, it snowed pretty good here. So the traps have been covered. Yesterday was warmish, and the snow started to melt. That allowed the traps to peek above the snow, and sure enough, the trap was sprung. I figured a kid prolly kicked it. When I pulled it up, there he was. In the middle of winter, on a day when we got freezing rain in the morning, I got a winter mole. Remember, moles don't hibernate. Usually they are deeper, to avoid the cold during the winter, but this one must have been desperate. Maybe he thought he was safe since there was snow on the ground. Wrongo buddy!  
EDIT 6-9-11: This year is the hatch of the 13 year cicadas in St. Louis. What does that have to do with moles? Well, the cicada larvae emerge from the ground to mutate and mate. What are they while they are in the ground and preparing to emerge? They are mole food is what they are. I've read a few articles now about the effects of cicadas on the ecosystem, and one thing that is consistent is that up to 2 years before a cicada emergence, mole activity is increased. On the other hand, you can expect to see decreased mole activity after they emerge. So, that might explain why so many in our area have been having mole issues. Hopefully after the stupid cicadas are gone, some of the moles will go with them.  
EDIT 5-3-12: I would love to talk to the 3 people who didn't find this review helpful. I'm not gonna dwell on it, but I'm just sayin...





\* BOW ENDS \*

## TF-IDF

In [16]:

```
#tfidf
# tf_idf_vect = TfidfVectorizer(ngram_range=(1,2))
tf_idf_vect = TfidfVectorizer()
vocabulary = tf_idf_vect.fit(text)
#print("the shape of out text TF-IDF vectorizer ",tf_idf_x_tr.get_shape())
tf_idf_data = tf_idf_vect.transform(text)
print("the shape of out text TF-IDF vectorizer ",tf_idf_data.get_shape())
```

the shape of out text TF-IDF vectorizer (5000, 10429)

## AgglomerativeClustering - TF-IDF

Apply agglomerative algorithm and try a different number of clusters

In [17]:

```
from sklearn.cluster import AgglomerativeClustering
reviews = final['Text'].values
n_clusters_lists = [2,3,5]

for cluster in n_clusters_lists:
    agclf = AgglomerativeClustering(n_clusters=cluster).fit(tf_idf_data.toarray())
    # agclf.fit(tf_idf_data)
    print("#" * 25, "Agglomerative Clustering for Number of Clusters = ", cluster, " ", "*" * 25)
    print("\nReviews for each of the clusters : \n")

    lables = agclf.labels_
    # Printing two reviews from each cluster
    for i in range(cluster):
        review_count = 1 ;
        print("#" * 80)
        print("Cluster ",i+1);
        print("#" * 80)
        for j in range(agclf.labels_.shape[0]):
            if agclf.labels_[j] == i:
                if review_count == 3:
                    break;
                else:
                    print('Review - ',review_count,'And review Index No - ',j)
                    print(reviews[j])
                    cluster_wordcloud_generated_image_fun(reviews[j])
                    print('\n');
                    review_count = review_count+1;
        # cluster_1.append(reviews[i])

    print("#" * 100)
    print("\n\n")
```

##### Agglomerative Clustering for Number of Clusters = 2

— —

—

\*\*\*\*\*

Cluster 1

\*\*\*\*\*

Review - 1 And review Index No - 0

this witty little book makes my son laugh at loud. i recite it in the car as we're driving along and he always can sing the refrain. he's learned about whales, India, drooping roses: i love all the new words this book introduces and the silliness of it all. this is a classic book i am willing to bet my son will STILL be able to recite from memory when he is in college



Review - 2 And review Index No - 1

I grew up reading these Sendak books, and watching the Really Rosie movie that incorporates them, and love them. My son loves them too. I do however, miss the hard cover version. The paperbacks seem kind of flimsy and it takes two hands to keep the pages open.



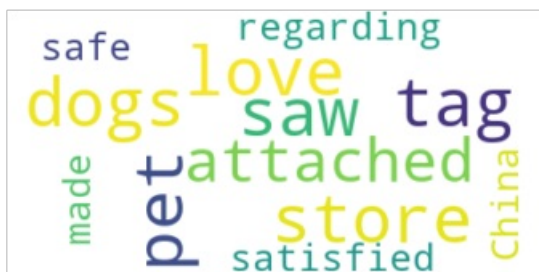
\*\*\*\*\*

Cluster 2

\*\*\*\*\*

Review - 1 And review Index No - 35

Our dogs just love them. I saw them in a pet store and a tag was attached regarding them being made in China and it satisfied me that they were safe.



Review - 2 And review Index No - 36

My dogs loves this chicken but its a product from China, so we wont be buying it anymore. Its very hard to find any chicken products made in the USA but they are out there, but this one isnt. Its too bad too because its a good product but I wont take any chances till they know what is going on with the china imports.





##### Agglomerative Clustering for Number of Clusters = 3  
\*\*\*\*\*

Reviews for each of the clusters :

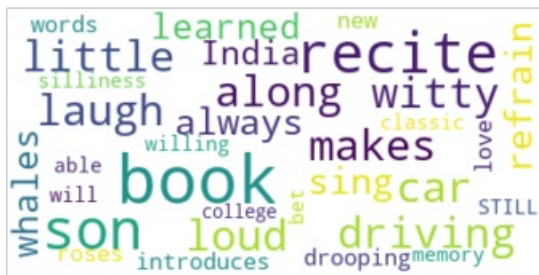
\*\*\*\*\*

Cluster 1

\*\*\*\*\*

Review - 1 And review Index No - 0

this witty little book makes my son laugh at loud. i recite it in the car as we're driving along and he always can sing the refrain. he's learned about whales, India, drooping roses: i love all the new words this book introduces and the silliness of it all. this is a classic book i am willing to bet my son will STILL be able to recite from memory when he is in college



Review - 2 And review Index No - 1

I grew up reading these Sendak books, and watching the Really Rosie movie that incorporates them, and love them. My son loves them too. I do however, miss the hard cover version. The paperbacks seem kind of flimsy and it takes two hands to keep the pages open.



\*\*\*\*\*

Cluster 2

\*\*\*\*\*

Review - 1 And review Index No - 35

Our dogs just love them. I saw them in a pet store and a tag was attached regarding them being made in China and it satisfied me that they were safe.





Review - 2 And review Index No - 36

My dogs loves this chicken but its a product from China, so we wont be buying it anymore. Its ver y hard to find any chicken products made in the USA but they are out there, but this one isnt. It s too bad too because its a good product but I wont take any chances till they know what is going on with the china imports.



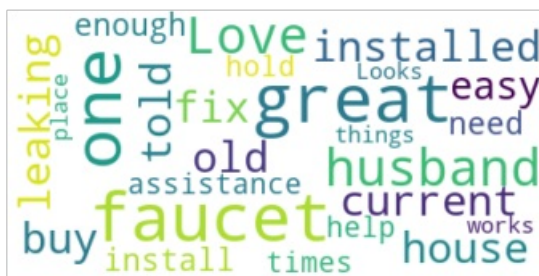
\*\*\*\*\*

Cluster 3

\*\*\*\*\*

Review - 1 And review Index No - 138

Love this faucet. My husband had installed the same one in our old house so when our current faucet was leaking I told him not to fix it and we would buy the same one. It was easy enough for him to install but he did need my assistance a few times to help hold some things in place. Looks great and works great.



Review - 2 And review Index No - 204

What a horrible product. First of all the item does not work. I have had it up for a week with lots of flies around it and they simply avoid it. Infact the will sit on the window next to it but never go to it. Also it does no look like the one pictured You get the idea that there are small boxes for the flies to enter. Not so. Mine came with black and pink strips running from the top to the bottom. WASTE OF MONEY



##### Agglomerative Clustering for Number of Clusters = 5  
\*\*\*\*\*

Reviews for each of the clusters :

\*\*\*\*\*

Cluster 1

\*\*\*\*\*

Review - 1 And review Index No - 35

Our dogs just love them. I saw them in a pet store and a tag was attached regarding them being made in China and it satisfied me that they were safe.



Review - 2 And review Index No - 36

My dogs loves this chicken but its a product from China, so we wont be buying it anymore. Its very hard to find any chicken products made in the USA but they are out there, but this one isnt. Its too bad too because its a good product but I wont take any chances till they know what is going on with the china imports.



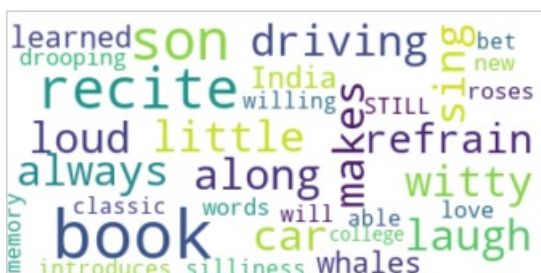
\*\*\*\*\*

Cluster 2

\*\*\*\*\*

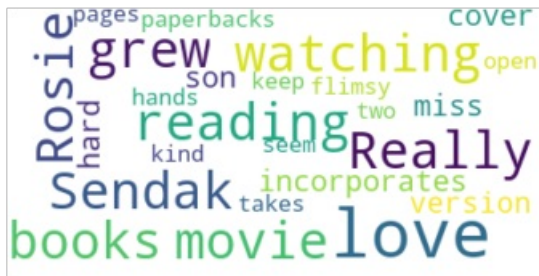
Review - 1 And review Index No - 0

this witty little book makes my son laugh at loud. i recite it in the car as we're driving along and he always can sing the refrain. he's learned about whales, India, drooping roses: i love all the new words this book introduces and the silliness of it all. this is a classic book i am willing to bet my son will STILL be able to recite from memory when he is in college



Review - 2 And review Index No - 1

I grew up reading these Sendak books, and watching the Really Rosie movie that incorporates them, and love them. My son loves them too. I do however, miss the hard cover version. The paperbacks seem kind of flimsy and it takes two hands to keep the pages open.



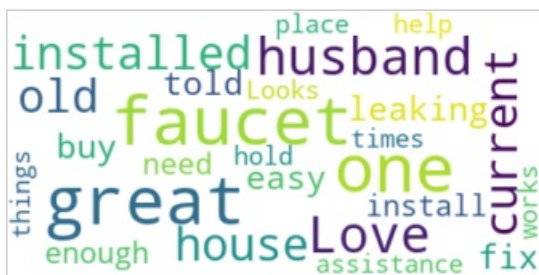
\*\*\*\*\*

Cluster 3

\*\*\*\*\*

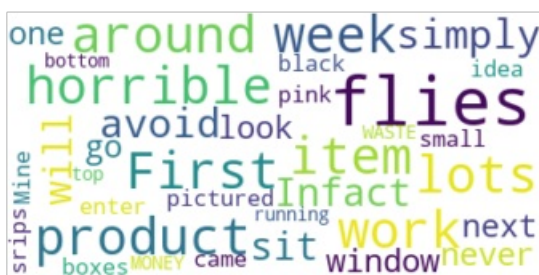
Review - 1 And review Index No - 138

Love this faucet. My husband had installed the same one in our old house so when our current faucet was leaking I told him not to fix it and we would buy the same one. It was easy enough for him to install but he did need my assistance a few times to help hold some things in place. Looks great and works great.



Review - 2 And review Index No - 204

What a horrible product. First of all the item does not work. I have had it up for a week with lots of flies around it and they simply avoid it. Infact the will sit on the window next to it but never go to it. Also it does not look like the one pictured You get the idea that there are small boxes for the flies to enter. Not so. Mine came with black and pink strips running from the top to the bottom. WASTE OF MONEY



\*\*\*\*\*

Cluster 4

\*\*\*\*\*

Review - 1 And review Index No - 904

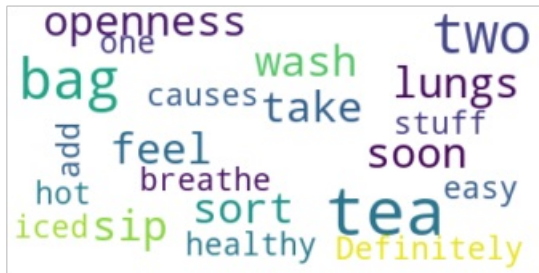
It's just as described. Its minty, helps with my mild-dry mouth and provides refreshment.<br />I'd recommend and/or purchase this again.





Review - 2 And review Index No - 906

As soon as I take a sip or two I can feel a sort of wash and openness in my lungs. Definitely causes one to "breathe easy." Have a hot tea or add a tea bag or two to your iced tea bags. Very healthy stuff in there.



\*\*\*\*\*

Cluster 5

\*\*\*\*\*

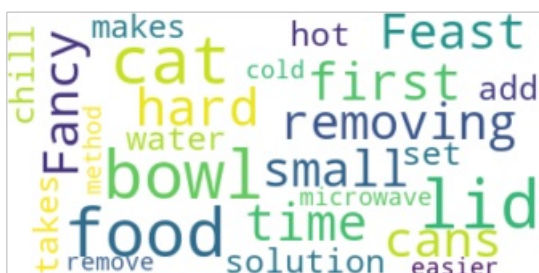
Review - 1 And review Index No - 891

I am an average baker and used these for cupcake icing. So far I am extremely pleased. One single drop colored about a cup of white frosting a pleasing pink. In an attempt to get black, I used 7 or 8 drops of black but ended up with dark grey. A few more drops would have probably done it. Each bottle is 3/4 of an ounce but using them one drop at a time should make them last for years. These do not compare to the store bought mccormick coloring- I will never buy them again!



Review - 2 And review Index No - 1228

At first I had a hard time removing these lids from small cans such as Fancy Feast. The solution is to just add hot water to a bowl and set the can in the bowl. This takes the chill off the cat food and also makes the lid easier to remove. (The cats don't like cold food and they don't like the microwave method!!)





\*\*\* TF-IDF ENDS \*

## Word2Vec

In [18]:

```
#Word2Vec mode
#splitting train sentence in words
# Train your own Word2Vec model using your own text corpus
i=0
list_of_sent=[]
for sent in text:
    list_of_sent.append(sent.split())

print(len(text))
# print("\n-----Splitting each sentence into words-----word list of ie data corpus-----\n")
# print(X_tr_list_of_sent[:2])
#word list of ie data corpus
```

5000

In [19]:

```
#The Word to Vec model produces a vocabulary, with each word being represented by
#an n-dimensional numpy array
data_w2v_model=Word2Vec(list_of_sent,min_count=1,size=50, workers=4)
data_w2v_model.wv['man']
wlist =list(data_w2v_model.wv.vocab)
# wlist is a list of words
len(wlist)
```

Out [19] :

10429

In [20]:

```
#CALCULATE AVG WORD2VEC FOR data
w2v_words = list(data_w2v_model.wv.vocab)
# compute average word2vec for each review.
data_sent_vectors = []; # the avg-w2v for each sentence/review is stored in this list
for sent in tqdm(list_of_sent): # for each review/sentence
    sent_vec = np.zeros(50) # as word vectors are of zero length
    cnt_words = 0; # num of words with a valid vector in the sentence/review
    for word in sent: # for each word in a review/sentence
        if word in w2v_words:
            vec = data_w2v_model.wv[word]
            sent_vec += vec
            cnt_words += 1
    if cnt_words != 0:
        sent_vec /= cnt_words
    data_sent_vectors.append(sent_vec)
print(len(data_sent_vectors))
print(len(data_sent_vectors[0]))
```

[illegible]

5000

50

### AgglomerativeClustering - AVG WORD2VEC

In [22]:

```

from sklearn.cluster import AgglomerativeClustering
reviews = final['Text'].values
n_clusters_lists = [2,3,5]

for cluster in n_clusters_lists:
    agclf = AgglomerativeClustering(n_clusters=cluster).fit(data_sent_vectors)
    # agclf.fit(tf_idf_data)
    print("#" * 25, "Agglomerative Clustering for Number of Clusters = ", cluster, " ", "*" * 25)
    print("\nReviews for each of the clusters : \n")

    labels = agclf.labels_
    # Printing two reviews from each cluster
    for i in range(cluster):
        review_count = 1 ;
        print("#" * 80)
        print("Cluster ", i+1);
        print("#" * 80)
        for j in range(agclf.labels_.shape[0]):
            if agclf.labels_[j] == i:
                if review_count == 3:
                    break;
                else:
                    print('Review - ', review_count, 'And review Index No - ', j)
                    print(reviews[j])
                    cluster_wordcloud_generated_image_fun(reviews[j])
                    print('\n');
                    review_count = review_count+1;
        # cluster_1.append(reviews[i])

    print("#" * 100)
    print("\n\n")

```

```
##### Agglomerative Clustering for Number of Clusters = 2
*****
```

Reviews for each of the clusters :

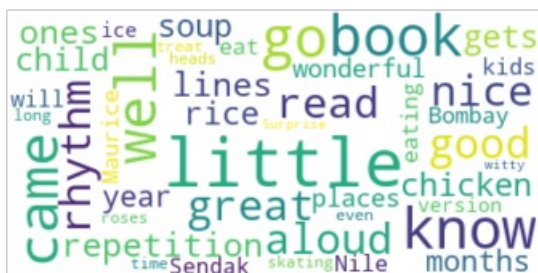
```

Cluster 1

```

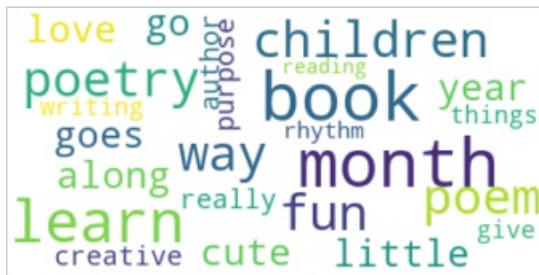
Review - 1 And review Index No - 3

This is a great little book to read aloud- it has a nice rhythm as well as good repetition that little ones like, in the lines about "chicken soup with rice". The child gets to go through the months of the year and go to wonderful places like Bombay and down the Nile, all the while, eating, well, you know what they get to eat. Some kids will have Maurice Sendak's version of ice skating or how to treat roses in their heads for a long time and they won't even know where it came from. Surprise! It came from this little witty book. :-)



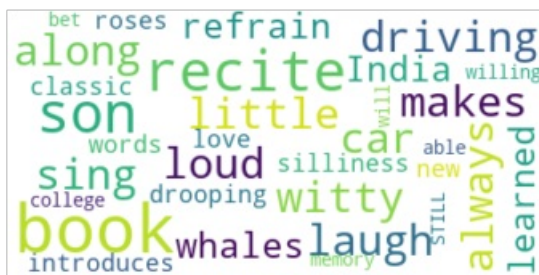
Review - 2 And review Index No - 4

This is a book of poetry about the months of the year. It goes through each month and has a cute little poem to go along with it. I love this book because it is a really fun way to learn the months and the poems are very creative. The author's purpose for writing this book was to give children a fun way to learn the months. The children can also learn things about poetry and rhythm through reading this book.



```
*****
Cluster 2
*****
```

```
Review - 1 And review Index No - 0
this witty little book makes my son laugh at loud. i recite it in the car as we're driving along a
nd he always can sing the refrain. he's learned about whales, India, drooping roses: i love all t
he new words this book introduces and the silliness of it all. this is a classic book i am
willing to bet my son will STILL be able to recite from memory when he is in college
```



```
Review - 2 And review Index No - 1
I grew up reading these Sendak books, and watching the Really Rosie movie that incorporates them,
and love them. My son loves them too. I do however, miss the hard cover version. The paperbacks se
em kind of flimsy and it takes two hands to keep the pages open.
```



```
##### Agglomerative Clustering for Number of Clusters = 3
*****
```

Reviews for each of the clusters :

```
*****
Cluster 1
*****
```

```
Review - 1 And review Index No - 3
This is a great little book to read aloud- it has a nice rhythm as well as good repetition that li
ttle ones like, in the lines about "chicken soup with rice". The child gets to go through the mon
ths of the year and go to wonderful places like Bombay and down the Nile, all the while, eating, w
```

◀ ▶

This is a book of poetry about the months of the year. It goes through each month and has a cute

Cluster 2

\*\*\*\*\*

this witty little book makes my son laugh at loud. i recite it in the car as we're driving along a

I grew up reading these Sendak books, and watching the Really Rosie movie that incorporates them.





\*\*\*\*\*  
Cluster 3  
\*\*\*\*\*

Review - 1 And review Index No - 34

This product by Archer Farms is the best drink mix ever. Just mix a flavored packet with your 16 oz. water bottle. Contains the all natural sweetner Stevia, real fruit flavoring and no food coloring. Just colored with fruit or vegetable colors. Pure and natural and tastes great. There are eight packets in a box and only contains 10 calories per packet. Thank you Archer Farms!



Review - 2 And review Index No - 107

My pitbull, Sammy, loves these treats! they are useful for training as well as do not make her breath smell overly bad.



##### Agglomerative Clustering for Number of Clusters = 5  
\*\*\*\*\*

Reviews for each of the clusters :

\*\*\*\*\*  
Cluster 1  
\*\*\*\*\*

Review - 1 And review Index No - 34

This product by Archer Farms is the best drink mix ever. Just mix a flavored packet with your 16 oz. water bottle. Contains the all natural sweetner Stevia, real fruit flavoring and no food coloring. Just colored with fruit or vegetable colors. Pure and natural and tastes great. There are eight packets in a box and only contains 10 calories per packet. Thank you Archer Farms!







Review - 2 And review Index No - 107

My pitbull, Sammy, loves these treats! they are useful for training as well as do not make her breath smell overly bad.



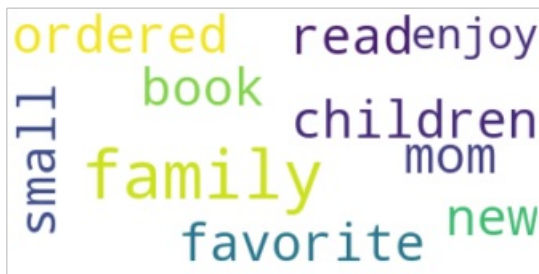
\*\*\*\*\*

Cluster 2

\*\*\*\*\*

Review - 1 And review Index No - 9

This book is a family favorite and was read to my children over and over again when they were small. We ordered it for a new mom-to-be so that she and her family can enjoy it too!



Review - 2 And review Index No - 12

Great book, perfect condition arrived in a short amount of time, long before the expected delivery date



\*\*\*\*\*

Cluster 3

\*\*\*\*\*

Review - 1 And review Index No - 3

This is a great little book to read aloud- it has a nice rhythm as well as good repetition that little ones like, in the lines about "chicken soup with rice". The child gets to go through the months of the year and go to wonderful places like Bombay and down the Nile, all the while, eating, well, you know what they get to eat. Some kids will have Maurice Sendak's version of ice skating or how to treat roses in their heads for a long time and they won't even know where it came from.

Surprise! It came from this little witty book. :-)



Review - 2 And review Index No - 4

This is a book of poetry about the months of the year. It goes through each month and has a cute little poem to go along with it. I love this book because it is a really fun way to learn the months and the poems are very creative. The author's purpose for writing this book was to give children a fun way to learn the months. The children can also learn things about poetry and rhythm through reading this book.



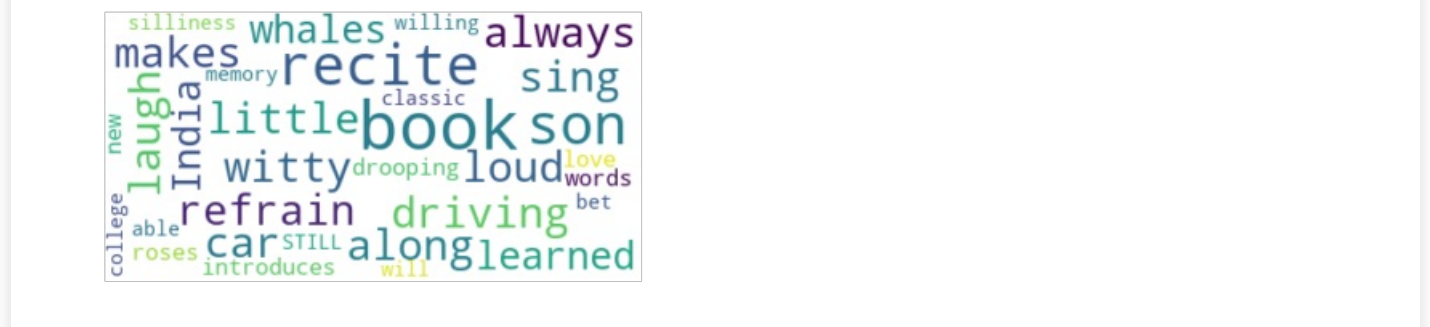
\*\*\*\*\*

Cluster 4

\*\*\*\*\*

Review - 1 And review Index No - 0

this witty little book makes my son laugh at loud. i recite it in the car as we're driving along and he always can sing the refrain. he's learned about whales, India, drooping roses: i love all the new words this book introduces and the silliness of it all. this is a classic book i am willing to bet my son will STILL be able to recite from memory when he is in college



Review - 2 And review Index No - 1

I grew up reading these Sendak books, and watching the Really Rosie movie that incorporates them, and love them. My son loves them too. I do however, miss the hard cover version. The paperbacks seem kind of flimsy and it takes two hands to keep the pages open.



\*\*\*\*\*  
Cluster 5  
\*\*\*\*\*

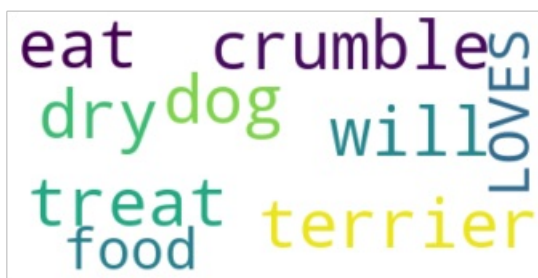
Review - 1 And review Index No - 37

This book is easy to read and the ingredients are available in stores unlike some recipe books where you have to go to special stores to find ingredients. I use it to make food not only for my son but for myself! The only thing I wish was that there were more recipes like these. It is great for any one who likes a healthy and yummy snack.



Review - 2 And review Index No - 46

This is the only dog treat that my terrier will eat. We crumble over his dry food and he LOVES it.



\*\*AVG W2V ENDS\*

## TF-IDF weighted Word2Vec

In [23]:

```
# we are converting a dictionary with word as a key, and the idf as a value
dictionary = dict(zip(tf_idf_vect.get_feature_names(), list(tf_idf_vect.idf_)))
```

In [24]:

```
# TF-IDF weighted Word2Vec
tfidf_feat = tf_idf_vect.get_feature_names() # tfidf words/col-names
# final_tf_idf is the sparse matrix with row= sentence, col=word and cell_val = tfidf

data_tfidf_sent_vectors = []; # the tfidf-w2v for each sentence/review is stored in this list
row=0;
for sent in tqdm(list_of_sent): # for each review/sentence
    sent_vec = np.zeros(50) # as word vectors are of zero length
```



ths of the year and go to wonderful places like Bombay and down the Nile, all the while, eating, well, you know what they get to eat. Some kids will have Maurice Sendak's version of ice skating or how to treat roses in their heads for a long time and they won't even know where it came from. Surprise! It came from this little witty book. :-)



Review - 2 And review Index No - 4

This is a book of poetry about the months of the year. It goes through each month and has a cute little poem to go along with it. I love this book because it is a really fun way to learn the months and the poems are very creative. The author's purpose for writing this book was to give children a fun way to learn the months. The children can also learn things about poetry and rhythm through reading this book.



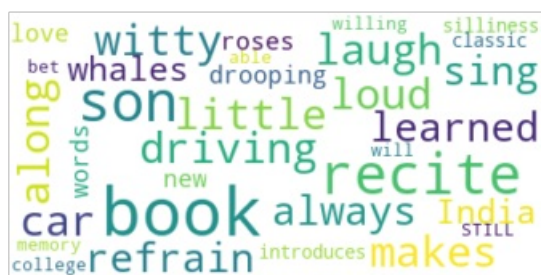
\*\*\*\*\*

Cluster 2

\*\*\*\*\*

Review - 1 And review Index No - 0

this witty little book makes my son laugh at loud. i recite it in the car as we're driving along and he always can sing the refrain. he's learned about whales, India, drooping roses: i love all the new words this book introduces and the silliness of it all. this is a classic book i am willing to bet my son will STILL be able to recite from memory when he is in college



Review - 2 And review Index No - 1

I grew up reading these Sendak books, and watching the Really Rosie movie that incorporates them, and love them. My son loves them too. I do however, miss the hard cover version. The paperbacks seem kind of flimsy and it takes two hands to keep the pages open.







##### Agglomerative Clustering for Number of Clusters = 3  
 \*\*\*\*\*

Reviews for each of the clusters :

\*\*\*\*\*

Cluster 1

\*\*\*\*\*

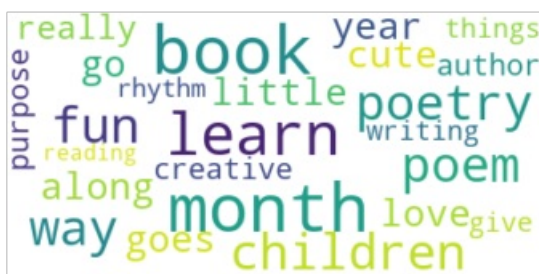
Review - 1 And review Index No - 3

This is a great little book to read aloud- it has a nice rhythm as well as good repetition that little ones like, in the lines about "chicken soup with rice". The child gets to go through the months of the year and go to wonderful places like Bombay and down the Nile, all the while, eating, well, you know what they get to eat. Some kids will have Maurice Sendak's version of ice skating or how to treat roses in their heads for a long time and they won't even know where it came from. Surprise! It came from this little witty book. :-)



Review - 2 And review Index No - 4

This is a book of poetry about the months of the year. It goes through each month and has a cute little poem to go along with it. I love this book because it is a really fun way to learn the months and the poems are very creative. The author's purpose for writing this book was to give children a fun way to learn the months. The children can also learn things about poetry and rhythm through reading this book.



\*\*\*\*\*

Cluster 2

\*\*\*\*\*

Review - 1 And review Index No - 0

this witty little book makes my son laugh at loud. i recite it in the car as we're driving along and he always can sing the refrain. he's learned about whales, India, drooping roses: i love all the new words this book introduces and the silliness of it all. this is a classic book i am willing to bet my son will STILL be able to recite from memory when he is in college



Review - 2 And review Index No - 1

I grew up reading these Sendak books, and watching the Really Rosie movie that incorporates them, and love them. My son loves them too. I do however, miss the hard cover version. The paperbacks seem kind of flimsy and it takes two hands to keep the pages open.



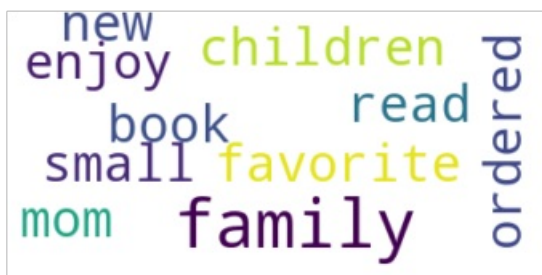
\*\*\*\*\*

Cluster 3

\*\*\*\*\*

Review - 1 And review Index No - 9

This book is a family favorite and was read to my children over and over again when they were small. We ordered it for a new mom-to-be so that she and her family can enjoy it too!



Review - 2 And review Index No - 12

Great book, perfect condition arrived in a short amount of time, long before the expected delivery date



##### Agglomerative Clustering for Number of Clusters = 5  
\*\*\*\*\*

Reviews for each of the clusters :

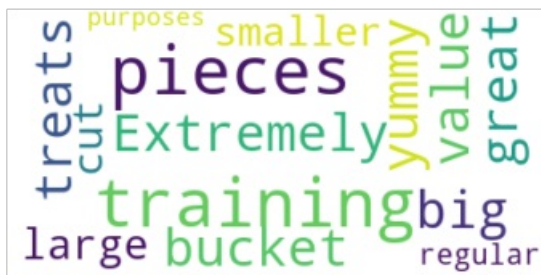
\*\*\*\*\*

Cluster 1

\*\*\*\*\*

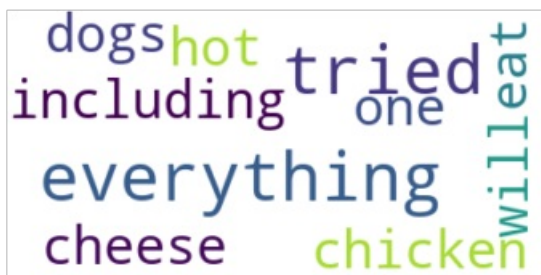
Review - 1 And review Index No - 77

Extremely big bucket of yummy treats for training! This is a great value. The pieces are large, but I cut them up into smaller pieces for regular training purposes.



Review - 2 And review Index No - 87

We tried everything including cheese, chicken & hot dogs and this is the only one he will eat.



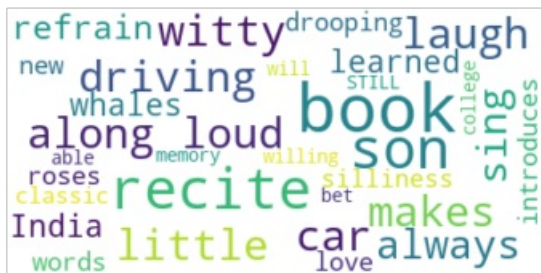
\*\*\*\*\*

Cluster 2

\*\*\*\*\*

Review - 1 And review Index No - 0

this witty little book makes my son laugh at loud. i recite it in the car as we're driving along and he always can sing the refrain. he's learned about whales, India, drooping roses: i love all the new words this book introduces and the silliness of it all. this is a classic book i am willing to bet my son will STILL be able to recite from memory when he is in college



Review - 2 And review Index No - 1

I grew up reading these Sendak books, and watching the Really Rosie movie that incorporates them, and love them. My son loves them too. I do however, miss the hard cover version. The paperbacks seem kind of flimsy and it takes two hands to keep the pages open.



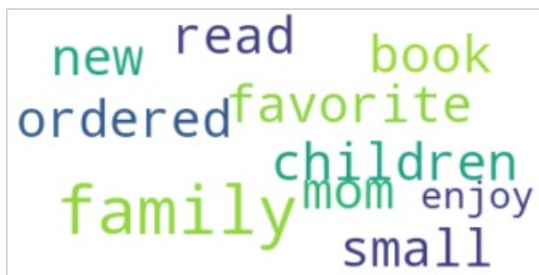
\*\*\*\*\*

Cluster 3

\*\*\*\*\*

Review - 1 And review Index No - 9

This book is a family favorite and was read to my children over and over again when they were small. We ordered it for a new mom-to-be so that she and her family can enjoy it too!



Review - 2 And review Index No - 12

Great book, perfect condition arrived in a short amount of time, long before the expected delivery date



\*\*\*\*\*

Cluster 4

\*\*\*\*\*

Review - 1 And review Index No - 3

This is a great little book to read aloud- it has a nice rhythm as well as good repetition that little ones like, in the lines about "chicken soup with rice". The child gets to go through the months of the year and go to wonderful places like Bombay and down the Nile, all the while, eating, well, you know what they get to eat. Some kids will have Maurice Sendak's version of ice skating or how to treat roses in their heads for a long time and they won't even know where it came from. Surprise! It came from this little witty book. :-)



Review - 2 And review Index No - 4

This is a book of poetry about the months of the year. It goes through each month and has a cute little poem to go along with it. I love this book because it is a really fun way to learn the months and the poems are very creative. The author's purpose for writing this book was to give children a fun way to learn the months. The children can also learn things about poetry and rhythm through reading this book.



\*\*\*\*\*

Cluster 5

\*\*\*\*\*

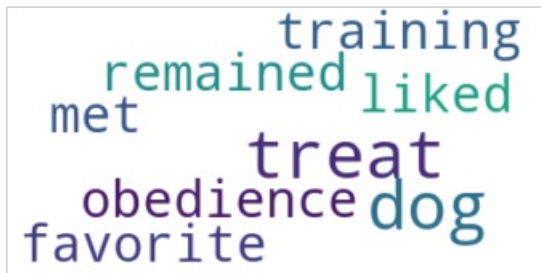
Review - 1 And review Index No - 38

I have been feeding my greyhounds these treats for years. My hounds can be a little finicky, but they love these treats. They are expensive, relative to other biscuits, but I find they are a good addition to their diet. The treats are easy on the teeth. Since they are protein treat, be careful not to overindulge your pet on a regular basis. They are great treats to take on walks as you can stuff them in your pocket easily. I have found prices can vary from vendor to vendor, so shop around.



Review - 2 And review Index No - 42

This was the only treat my dog liked during obedience training and it has remained her favorite ever since. I haven't met a dog that didn't like this treat!





# Conclusion