

```
In [1]: # Importing libraries
import warnings
warnings.filterwarnings("ignore")

import sqlite3
import pandas as pd
import numpy as np
import nltk
import string
import matplotlib.pyplot as plt
%matplotlib inline
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 nltk.stem.porter import PorterStemmer

import re

import string
from nltk.corpus import stopwords
from nltk.stem import PorterStemmer
from nltk.stem.wordnet import WordNetLemmatizer

from gensim.models import Word2Vec
from gensim.models import KeyedVectors
import pickle
```

```
C:\Users\Sai charan\Anaconda3\lib\site-packages\gensim\utils.py:1197: UserWarning: detected Windows; aliasing chunkize to chunkize_serial
  warnings.warn("detected Windows; aliasing chunkize to chunkize_serial")
```

```
In [2]: con1 = sqlite3.connect('final.sqlite')
        filtered_data = pd.read_sql_query(" SELECT * FROM Reviews ", con1)

        print(filtered_data.shape)
        filtered_data.head()

(364171, 12)
```

Out[2]:

	index	Id	ProductId	UserId	ProfileName	HelpfulnessNumerator	HelpfulnessDenominator
0	138706	150524	0006641040	ACITT7DI6IDDL	shari zychinski	0	
1	138688	150506	0006641040	A2IW4PEEKO2R0U	Tracy	1	
2	138689	150507	0006641040	A1S4A3IQ2MU7V4	sally sue "sally sue"	1	
3	138690	150508	0006641040	AZGXZ2UUK6X	Catherine Hallberg "(Kate)"	1	
4	138691	150509	0006641040	A3CMRKGE0P909G	Teresa	3	

## Data Cleaning: Deduplication

```
In [3]: sorted_data=filtered_data.sort_values('ProductId', axis=0, ascending=True, inplace=False, kind='quicksort', na_position='last')

#Deduplication of entries
final=sorted_data.drop_duplicates(subset={"UserId","ProfileName","Time","Text"}, keep='first', inplace=False)
print(final.shape)

#Checking to see how much % of data still remains
((final.shape[0]*1.0)/(filtered_data.shape[0]*1.0)*100)

(364171, 12)
```

```
Out[3]: 100.0
```

```
In [4]: final = final[final.HelpfulnessNumerator <= final.HelpfulnessDenominator]

print(final.shape)
final[30:50]
```

(364171, 12)

Out[4]:

	index	Id	ProductId	UserId	ProfileName	HelpfulnessNumerator	HelpfulnessDenominator
12	138700	150518	0006641040	AK1L4EJBA23JF	L. M. Kraus	0	
13	138701	150519	0006641040	A12HY5OZ2QNK4N	Elizabeth H. Roessner	0	
14	138702	150520	0006641040	ADBFS9KTQANE	James L. Hammock "Pucks Buddy"	0	
8	138696	150514	0006641040	A2ONB6ZA292PA	Rosalind Matzner	0	
34	476617	515426	141278509X	AB1A5EGHHVA9M	CHelmic	1	
36	22620	24750	2734888454	A13ISQV0U9GZIC	Sandikaye	1	
35	22621	24751	2734888454	A1C298ITT645B6	Hugh G. Pritchard	0	
37	284375	308077	2841233731	A3QD68O22M2XHQ	LABRNTH	0	
142	157910	171225	7310172001	A314APAWYQFKBJ	Diana Hersholt "dog lover"	1	
143	157909	171224	7310172001	AK0CENM3LUM28	Ana Mardoll	1	
144	157924	171240	7310172001	A2JCG7KT8HRSUJ	B. Le	0	
145	157925	171241	7310172001	ACBK6OXAMOHE	Andrew Fox	0	
146	157926	171242	7310172001	A12L1NY994GXSF	Dawn Rene	0	
147	157907	171222	7310172001	A1CV6GLAPUIP80	Reader Rabbit	1	

## Stemming, stop-word removal and Lemmatization

```
In [5]: from nltk.corpus import stopwords
stop = set(stopwords.words('english'))
words_to_keep = set(('not'))
stop -= words_to_keep
#initialising the snowball stemmer
sno = nltk.stem.SnowballStemmer('english')

#function to clean the word of any html-tags
def cleanhtml(sentence):
    cleanr = re.compile('<.*?>')
    cleantext = re.sub(cleanr, ' ', sentence)
    return cleantext

#function to clean the word of any punctuation or special characters
def cleanpunc(sentence):
    cleaned = re.sub(r'[?|!|\'|\"|#]', r'', sentence)
    cleaned = re.sub(r'[.,|)|(|\|/]', r' ', cleaned)
    return cleaned
```

```
In [6]: i=0
str1=' '
final_string=[]
all_positive_words=[] # store words from +ve reviews here
all_negative_words=[] # store words from -ve reviews here.
s=''
for sent in final['Text'].values:
    filtered_sentence=[]
    #print(sent);
    sent=cleanhtml(sent) # remove HTML tags
    for w in sent.split():
        for cleaned_words in cleanpunc(w).split():
            if((cleaned_words.isalpha()) & (len(cleaned_words)>2)):
                if(cleaned_words.lower() not in stop):
                    s=(sno.stem(cleaned_words.lower())).encode('utf8')
                    filtered_sentence.append(s)
                    if (final['Score'].values)[i] == 'positive':
                        all_positive_words.append(s) #list of all words used to describe positive reviews
                    if (final['Score'].values)[i] == 'negative':
                        all_negative_words.append(s) #list of all words used to describe negative reviews
                    else:
                        continue
                else:
                    continue

    str1 = b" ".join(filtered_sentence) #final string of cleaned words

    final_string.append(str1)
    i+=1
```

```
In [7]: final['CleanedText']=final_string
final['CleanedText']=final['CleanedText'].str.decode("utf-8")
#below the processed review can be seen in the CleanedText Column
print('Shape of final',final.shape)
final.head()
```

Shape of final (364171, 12)

Out[7]:

	index	Id	ProductId	UserId	ProfileName	HelpfulnessNumerator	HelpfulnessDenominator
	0	138706	150524	0006641040	ACITT7DI6IDDL	shari zychinski	0
	19	138708	150526	0006641040	A3E9QZFE9KXH8J	R. Mitchell	11
	20	138709	150529	0006641040	A25ACLV5KPB4W	Matt Hetling "Matt"	0
	21	138699	150517	0006641040	ABW4IC5G5G8B5	kevin clark	0
	22	138686	150504	0006641040	AQEYF1AXARWJZ	Les Sinclair "book maven"	1

```
In [8]: # We will collect different 30K rows without repetition from time_sorted_data dataframe
my_final = final[:30000]
my_final.sort_values('Time',inplace=True)

x = my_final['CleanedText'].values
```

## Implementing K-Means++ and K-medoids

### Bag of Words (BoW)

```
In [84]: count_vect = CountVectorizer(min_df = 10)
data = count_vect.fit_transform(x)
print("the type of count vectorizer :",type(data))
print("the shape of out text BOW vectorizer : ",data.get_shape())
print("the number of unique words :", data.get_shape()[1])

the type of count vectorizer : <class 'scipy.sparse.csr.csr_matrix'>
the shape of out text BOW vectorizer : (30000, 5322)
the number of unique words : 5322
```

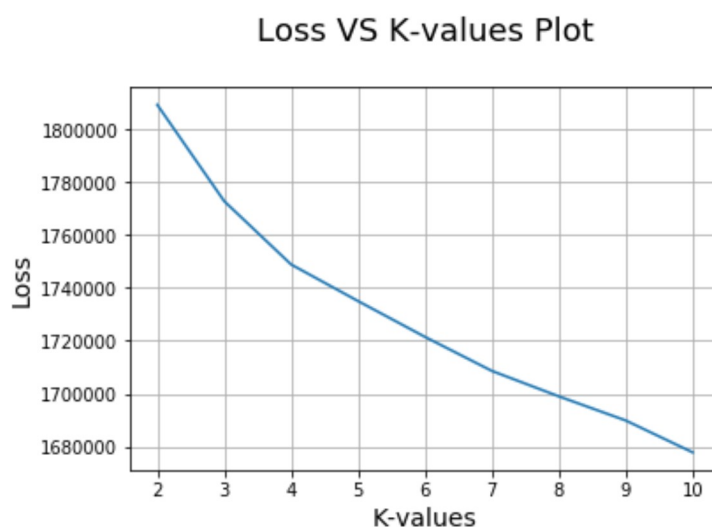


## K-Means++ Implementation

```
In [10]: from sklearn.cluster import KMeans

k_values = [2,3,4,5,6,7,8,9,10]
loss = []
for i in k_values:
    kmeans = KMeans(n_clusters=i, n_jobs=-1).fit(data)
    loss.append(kmeans.inertia_)
```

```
In [11]: plt.plot(k_values, loss)
plt.xlabel('K-values',size=14)
plt.ylabel('Loss',size=14)
plt.title('Loss VS K-values Plot\n',size=18)
plt.grid()
plt.show()
```



OBSERVATION :- From above we can see that there is inflection at  $K = 4$ . Before it loss was decreasing faster as compared to the loss decreasing after it. So, the best value of  $K$  is 4

```
In [12]: optimal_k = 4
# Variable that will be used in the conclusion
bow_means_k = optimal_k

# Implementing K-Means++ using optimal value of K
kmeans = KMeans(n_clusters=optimal_k, n_jobs=-1).fit(data)
```

```
In [13]: reviews = my_final['Text'].values
# Getting all the reviews in different clusters
cluster1 = []
cluster2 = []
cluster3 = []
cluster4 = []
for i in range(kmeans.labels_.shape[0]):
    if kmeans.labels_[i] == 0:
        cluster1.append(reviews[i])
    elif kmeans.labels_[i] == 1:
        cluster2.append(reviews[i])
    elif kmeans.labels_[i] == 2:
        cluster3.append(reviews[i])
    else :
        cluster4.append(reviews[i])

# Number of reviews in different clusters
print("No. of reviews in Cluster-1 : ",len(cluster1))
print("\nNo. of reviews in Cluster-2 : ",len(cluster2))
print("\nNo. of reviews in Cluster-3 : ",len(cluster3))
print("\nNo. of reviews in Cluster-4 : ",len(cluster4))
```

No. of reviews in Cluster-1 : 3271

No. of reviews in Cluster-2 : 998

No. of reviews in Cluster-3 : 931

No. of reviews in Cluster-4 : 24800

## READING REVIEWS MANUALLY

```
In [14]: count=1
         for i in range(3):
             print('Review-%d : \n %s\n'%(count,cluster1[i]))
             count +=1
```

Review-1 :

The wine saver is great in so many ways. Obviously it's wonderful to be able to open a bottle of wine and have it stay good for a day or two, especially if you just like one glass with dinner. This lets you open up and enjoy more expensive wine since you won't waste the whole rest of the bottle. If you are interested in learning about wines, you'll probably want to open up 2 or more wines at a time to compare them, and this way you won't waste anything. The wine saver will keep the undrunk portion until the next time. I do recommend putting the bottle in the fridge if you don't drink it the next day. Just pull the bottle out an hour before you want to drink it if it's red.

Review-2 :

Using the Briel Chamonix machine I purchased from Amazon, I have produced delicious lattes using these pods. The smell of the Espressione pods (regular and decaf) when opened is incredible - dark, rich roasts. FYI - the pods are produced in Italy, which is a good sign. Also, each pod has a handy paper "tab" so that you can remove it from your machine without burning your fingers. I compared these pods to Starbucks espresso pods (regular and decaf), and the Espressione are my favorite. Only one warning: 150 pods (individually wrapped) take up a lot of space, so be sure you know where you're going to store them. More good news: each box of 150 is stamped with a "best if used by" date, which in my case was 16 months after the date of purchase.

Review-3 :

I was very skeptical when I bought this item, so imagine my amazement when I discovered that it actually works! Prior to buying VacuVin, my red wine bottles would last a maximum of 1-2 days if immediately re-corked, and white wine a few days longer if refrigerated. Now, my red wine is good for up to a week, and the white wine lasts even longer. The device is very simple - it pumps air through a plastic cork to re-create the original pressure prior to opening. Simple to use, and the corks are easy to rinse. Perfect for single people like me who don't want to drink a full bottle in one sitting, or who may switch from red one night to white the next.

```
In [15]: count=1
         for i in range(3):
             print('Review-%d : \n %s\n'%(count,cluster2[i]))
             count +=1
```

Review-1 :

I have two Burmese cats, and both of them are being or were raised on this kitten formula (one is 13 months, the other is 8 months). They have been on it since they were 4 months old. I did a lot of research into food recommended by my breeder since we spent a lot of money on the cats. I ultimately decided on this brand because of the research they have done. I plan on putting both of them on the Persian formula (that is for Himalayans and Burmese too) because these types of cats are prone to getting a certain type of UTI and need their urine a different pH.<br />Their fur has never been softer, they are so active and always chasing each other, they are very playful with us. We always comment on how soft the older one is, just like a pillow, and the younger one is getting closer every day. They eat the food I leave out for them every day, so I'm assuming it is satisfying. I just love this brand and plan on sticking with them, even when we get dogs.

Review-2 :

I searched for a food that wasn't full of preservatives and fillers, and didn't cost \$5 a pound. I did a lot of research and read a lot of labels. I have a Siberian Husky pup, and Canidae has really proven itself to me. As a former consumer of cheap discount store dog foods, I have finally realized the advantages of feeding a natural premium food. My dogs poop doesn't really stink! True! Poop is still poop, but the smell is minimal. I have noticed it also quickly disappears into the ground when it rains. With the cheap food, it smelled like rank dog poop! Would make you retch and gag! And there was so much of it! Less poop, firmer stools with barely any odor, less food consumed per day. AND NO BAD BREATH from your dog! A food that swells up when wet is also undesirable. Flint River Ranch brand and Canidae both stay intact, avoiding the bloat factor. I learned, you can learn too! It's also an excellent value - Try Canidae, you'll see!

Review-3 :

If your cat loves poultry and ground food, this is a great one. My previous cat loved this stuff, but my current cat just doesn't like the ground variety as well. He prefers the food with chunks and gobbles it right up. I've bought tons of this food over the years and many of my cats just loved it. It just depends on your cat - if you have one that likes ground, give this one a try!

```
In [16]: count=1
for i in range(3):
    print('Review-%d : \n %s\n'%(count,cluster3[i]))
    count +=1
```

Review-1 :

Blend a little apple cider mix into your Tazo Passion Tea and see what happens. This creates a magical apple cider experience with hints of cranberry and hibiscus flowers gathered in an ancient land.<br /><br />Snuggle Up Spiced Cider is a good choice because you can then spoon in the exact amount you desire, but packets of cider work equally well. This is the type of tea you would expect to enjoy after skiing in some distant forest where you are also munching on homemade brownies or eggnog cookies. Apple cider also works very well with the Gypsy Teas, especially the Lemon Jasmine.<br /><br />The deep purple/red color and hibiscus fragrance of Tazo Tea is enough to make anyone passionate about tea. Yet, is this really tea? It looks like tea, but it is really an herbal infusion made by steeping hibiscus flowers, orange peel, licorice, cinnamon, rose hips, lemongrass and red poppy in a swirling of hot water.<br /><br />The original flavor is also slightly reminiscent of hot apple cider or hot cranberry juice which is why it is extra delicious with a cider mix. Tazo Teas are a mysterious experience into which you can lose yourself with a sense of abandon. So far I've tried more flavors than I can count and have always been impressed.<br /><br />~The Rebecca Review

Review-2 :

"There is a garden overlooking the Yangtze River gorge where an elderly man contemplates his life while sipping a cup of Green Ginger. Perhaps you would like to join him." ~on each tea packet<br /><br />A friend sent me this tea and I've been in love with the slightly bitter flavor. It is quite exotic and fragrant. The ingredients are green teas, natural flavors, ginger and lemongrass.<br /><br />This tea begs to be blended with orange blossom honey.<br /><br />I've tasted many teas and this is by far the most exotic tea I've tried yet. This is pear ginger bliss.<br /><br />For more information on the myseries of Tazo Tea, read the review for the Tazo Assorted Teas.<br /><br />~The Rebecca Review

Review-3 :

"There are few hours in life more agreeable than the hour dedicated to the ceremony known as afternoon tea." -Henry James, Portrait of a Lady<br /><br />This "English Breakfast" style tea is quite satisfying and has a rich flavor that is quite delicious.<br /><br />I'd like to say that their teas are best steeped for an extra 5-10 minutes to draw out all the delicious flavors. You can use a teapot or a coffee maker to make tea. Simply put the tea in a coffee filter or the tea bags in the coffee pot, then allow the boiling water to drip over the tea leaves.<br /><br />The sensual blend of exotic spices, teas and herbs used in their teas include a mixture of the following: Black tea, green tea, hibiscus flowers, chamomile flowers, lemongrass, spearmint leaves, blackberry leaves, cinnamon, Roibos, Gingerroot, rose hips, orange peel, rose petals, licorice root, safflower seeds, black pepper, lemon verbena leaves, cardamom, cloves, peppermint leaves, chicory, sarsaparilla root, lemon balm leaf, star anise and red poppy flowers.<br /><br />This is a blend of black teas, but all their teas are worth trying. Passion and Chai are also favorites.<br /><br />~The Rebecca Review

```
In [17]: count=1
         for i in range(3):
             print('Review-%d : \n %s\n'%(count,cluster4[i]))
             count +=1
```

Review-1 :

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 :

I can remember seeing the show when it aired on television years ago, when I was a child. My sister later bought me the LP (which I have to this day, I'm thirty something).I used this series of books & songs when I did my student teaching for preschoolers & turned the whole school on to it. I am now purchasing it on CD, along with the books for my children 5 & 2. The tradition lives on!

Review-3 :

Beetlejuice is a well written movie ..... everything about it is excellent! From the acting to the special effects you will be delighted you chose to view this movie.

## K-Medoids Implementation

```
In [85]: def print_cluster_reviews_k_medoids(cluster_label_reviews,cluster_number = 0,review_count=3):
         reviews = cluster_label_reviews[cluster_number]
         reviews = reviews[0:review_count]
         print("----: Cluster {0} :----".format(cluster_number))
         index = 1
         for review in reviews:
             print("Review {0} : ".format(index))
             print(CLEANED_TEXT[review])
             print()
             index += 1
```

```
In [86]: def get_cluster_N_medoids(k_medoids):
         return k_medoids.get_clusters(),k_medoids.get_medoids()
```

```
In [88]: svd = TruncatedSVD(n_components=2)
         bow_dtm_2_dim = svd.fit_transform(data)
         bow_dtm_2_dim.shape
```

Out[88]: (30000, 2)

```
In [89]: k_medoids = kmedoids.kmedoids(bow_dtm_2_dim,[2, 9, 12],0.001,ccore=True)
         # Run cluster analysis and obtain results
         k_medoids.process()
```

Out[89]: <pyclustering.cluster.kmedoids.kmedoids at 0x1dd93492a90>

```
In [90]: clusters,medoids = get_cluster_N_medoids(k_medoids)
```

```
In [91]: for i in range(0, len(clusters)):
          list_if_indices = clusters[i]
          # Number of reviews in different clusters
          print("Number of Reviews in Cluster {0} : {1} ".format(i, len(list_if_indices)))
```

Number of Reviews in Cluster 0 : 25940

Number of Reviews in Cluster 1 : 3116

Number of Reviews in Cluster 2 : 944

```
In [92]: print_cluster_reviews_k_medoids(clusters, cluster_number = 0, review_count = 3)
print("\n\n")
print_cluster_reviews_k_medoids(clusters, cluster_number = 1, review_count = 3)
print("\n\n")
print_cluster_reviews_k_medoids(clusters, cluster_number = 2, review_count = 3)
print("\n\n")
```



---: Cluster 0 :---

Review 1 :

I mix this with sugar and sea salt, toss some potatoes in olive oil, and make great baked fries with this stuff. Such a unique flavor, and everyone in my house likes it- even the picky eaters. I love this stuff.

Review 2 :

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 3 :

I can remember seeing the show when it aired on television years ago, when I was a child. My sister later bought me the LP (which I have to this day, I'm thirty something). I used this series of books & songs when I did my student teaching for preschoolers & turned the whole school on to it. I am now purchasing it on CD, along with the books for my children 5 & 2. The tradition lives on!

---: Cluster 1 :---

Review 1 :

This was an instant success for our food- and treat-fussy Siberian, and our "normal" voracious Malamute.  
Pros: 1) Most importantly: If our siberian will accept (as in devour!) these treats, we predict most any other fussy pups will too.  
2) Someone intelligently designed the treat holders and shape of the treats themselves so that they can't be immediately removed and consumed. .. our dogs spend hours working on their treat ball. True to their breeds, our pups get in an occasional gnawing moment, but for the most part they have to lick and nibble their way through their treat. Which means when we leave for an extended time, they are honorably occupied, not bored, and not getting into mischief.  
3) The ball itself is made of a soft-ish, somewhat pliable material, and yet appears to be indestructible. Time will tell.  
4) A variety of flavors. We purchased each of the available flavors to learn which would "win", and immediately discovered that all flavors send our pups into excited expectation when they see their ball coming.  
5) You can put your own treats inside the ball itself, but it looks like a tough nut for any dog to get to.  
Con: and this could be a good thing: it's hard getting a treat into the ball unless you have strong hands/wrists, or a smooth ended spoon handle to help peel the retaining edge of the treat ball opening over the edge of the treat. This means your pup will not likely remove the treat in one lump (unless he uses that smooth spoon handle) and immediately consume it.

Review 2 :

The spectra paste food colors are simply fantastic. They provide a depth and clarity of color that is impossible to achieve with liquid food colors. And the hues are beautiful and distinctive. The black and red, two colors which are very difficult to get when dying frostings, etc. are especially nice - probably the best of any manufacturer I have tried.  
For those who had questions, I am not 100% certain about the sizes, but if the set is the same as mine, the bottles are 3/4 ounce. While that might not seem like much, all you need is a dot of this thick gel paste at the end of a toothpick to color large amounts of material. So there is plenty of coloring ability in each bottle - many, many times more than a same sized liquid food color bottle.  
Highly recommended and at a really good price at the time of this review.

Review 3 :

Long live Camper Van Beethoven!! This is a blessing for long time CVB fans who didn't have a chance to upgrade their collection to CD before most of this went out of print. Wow, where to begin? CVB was a band that matured on record, so to

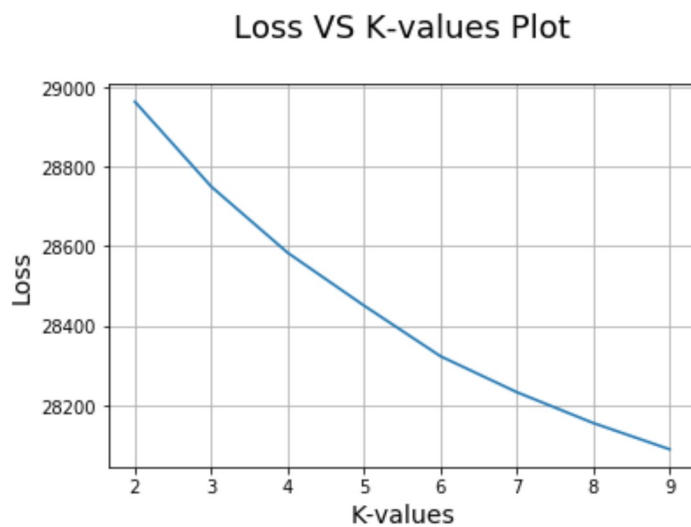
## TFIDF

```
In [30]: tf_idf_vect = TfidfVectorizer(min_df=10)
data = tf_idf_vect.fit_transform(x)
print("the type of count vectorizer :", type(data))
print("the shape of out text TFIDF vectorizer : ", data.get_shape())
print("the number of unique words :", data.get_shape()[1])

the type of count vectorizer : <class 'scipy.sparse.csr.csr_matrix'>
the shape of out text TFIDF vectorizer : (30000, 5322)
the number of unique words : 5322
```

```
In [32]: from sklearn.cluster import KMeans
k_values = [2,3,4,5,6,7,8,9]
loss = []
for i in k_values:
    kmeans = KMeans(n_clusters=i, n_jobs=-1).fit(data)
    loss.append(kmeans.inertia_)
```

```
In [33]: plt.plot(k_values, loss)
plt.xlabel('K-values', size=14)
plt.ylabel('Loss', size=14)
plt.title('Loss VS K-values Plot\n', size=18)
plt.grid()
plt.show()
```



From above we can see that there is inflection at K = 6 . Befor it loss was decreasing faster as compared to the loss decreasing after it . So , the best value of K is 6.

```
In [37]: reviews = my_final['Text'].values
optimal_k = 6
# Variable that will be used in the conclusion
tfidf_means_k = optimal_k

# Implementing K-Means++ using optimal value of K
kmeans = KMeans(n_clusters=optimal_k, n_jobs=-1).fit(data)

# Getting all the reviews in different clusters
cluster1 = []
cluster2 = []
cluster3 = []
cluster4 = []
cluster5 = []
cluster6 = []

for i in range(kmeans.labels_.shape[0]):
    if kmeans.labels_[i] == 0:
        cluster1.append(reviews[i])
    elif kmeans.labels_[i] == 1:
        cluster2.append(reviews[i])
    elif kmeans.labels_[i] == 2:
        cluster3.append(reviews[i])
    elif kmeans.labels_[i] == 3:
        cluster4.append(reviews[i])
    elif kmeans.labels_[i] == 4:
        cluster5.append(reviews[i])
    else :
        cluster6.append(reviews[i])

# Number of reviews in different clusters
print("No. of reviews in Cluster-1 : ",len(cluster1))
print("\nNo. of reviews in Cluster-2 : ",len(cluster2))
print("\nNo. of reviews in Cluster-3 : ",len(cluster3))
print("\nNo. of reviews in Cluster-4 : ",len(cluster4))
print("\nNo. of reviews in Cluster-5 : ",len(cluster5))
print("\nNo. of reviews in Cluster-6 : ",len(cluster6))
```

No. of reviews in Cluster-1 : 19844

No. of reviews in Cluster-2 : 1283

No. of reviews in Cluster-3 : 3252

No. of reviews in Cluster-4 : 1357

No. of reviews in Cluster-5 : 2706

No. of reviews in Cluster-6 : 1558

```
In [38]: count=1
for i in range(3):
    print('Review-%d : \n %s\n'%(count,cluster1[i]))
    count +=1
```

Review-1 :

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 :

I can remember seeing the show when it aired on television years ago, when I was a child. My sister later bought me the LP (which I have to this day, I'm thirty something).I used this series of books & songs when I did my student teaching for preschoolers & turned the whole school on to it. I am now purchasing it on CD, along with the books for my children 5 & 2. The tradition lives on!

Review-3 :

Beetlejuice is a well written movie ..... everything about it is excellent! From the acting to the special effects you will be delighted you chose to view this movie.

```
In [39]: count=1
for i in range(3):
    print('Review-%d : \n %s\n'%(count,cluster2[i]))
    count +=1
```

Review-1 :

Chipotle makes everything taste better. I first encountered this sauce at a small restaurant in New Mexico and it instantly made everything spicy and smoky. We made sure to buy a few bottles in a Mexican grocery store, and I've used this on burgers, chicken, etc. It's awesome. And it's less than \$2. You can't beat that.

Review-2 :

I just love this sauce. I put in on everything.

Review-3 :

This sauce is excellent, and indeed spicy. The brand also makes a mild version, in addition to other wonderful products like Rib Rub. Made in North Carolina. My husband doesn't like the sweeter type BBQ sauces - this brand he loves.

```
In [40]: count=1
         for i in range(3):
             print('Review-%d : \n %s\n'%(count,cluster3[i]))
             count +=1
```

Review-1 :

I've bought a couple of these for my Airedale puppy who constantly wants to chew. These things are amazing in that they have bone, tendon, meat, and knuckles still attached. She dives in and tears it apart - she would work on it non-stop if I didn't periodically take it away from her. The meat and tendons are harmless enough - obviously they have fat and protein so they have to be tied into diet considerations, and the bone appears to hold onto the flavor long after it's stripped (she still licks and gnaws on a bone that's a few weeks bare). The knuckles, once they get torn off, should probably be thrown away after a few minutes because they are small enough to be swallowed and/or choked on. The only downside to this product is when she comes up to me and the whiskers around her mouth are a little discolored and she smells like roasted meat. But, well worth it if you want to give your dog a treat that they can't get enough of.

Review-2 :

I searched for a food that wasn't full of preservatives and fillers, and didn't cost \$5 a pound. I did a lot of research and read a lot of labels. I have a Siberian Husky pup, and Canidae has really proven itself to me. As a former connoisseur of cheap discount store dog foods, I have finally realized the advantages of feeding a natural premium food. My dogs poop doesn't really stink! True! Poop is still poop, but the smell is minimal. I have noticed it also quickly disappears into the ground when it rains. With the cheap food, it smelled like rank dog poop! Would make you retch and gag! And there was so much of it! Less poop, firmer stools with barely any odor, less food consumed per day. AND NO BAD BREATH from your dog! A food that swells up when wet is also undesirable. Flint River Ranch brand and Canidae both stay intact, avoiding the bloat factor. I learned, you can learn too! It's also an excellent value - Try Canidae, you'll see!

Review-3 :

They look and smell like human cookies so keep away from children (or your dog will be upset!).

```
In [41]: count=1
         for i in range(3):
             print('Review-%d : \n %s\n'%(count,cluster4[i]))
             count +=1
```

Review-1 :

Boots (the most handsome red tom cat you'll ever see!) loves these treats. We keep them on a shelf in the pantry - only problem is everytime we open the pantry door he thinks he gets a treat.<p> We've tried other cat treats which he always eats (truth is, Boots eats everything!) but these seem to be his favorite by far.

Review-2 :

My cat Foutchie is heading toward 5 now and no one told her the trouble with a kitten is it eventually becomes a cat (so take that Ogden Nash!). She still vroom-vroom-vooms around at a pace that makes me exhausted. She loves toys, especially stuff she chases. So the bubbles are great full. Especially since they are catnip bubbles. You don't have to worry about them getting soap on them and making them sick. She loves jumping high to pop them.<p>So if you are looking for something to awaken the kitten in your kitty, you might try these.

Review-3 :

I have two Burmese cats, and both of them are being or were raised on this kitten formula (one is 13 months, the other is 8 months). They have been on it since they were 4 months old. I did a lot of research into food recommended by my breeder since we spent a lot of money on the cats. I ultimately decided on this brand because of the research they have done. I plan on putting both of them on the Persian formula (that is for Himalayans and Burmese too) because these types of cats are prone to getting a certain type of UTI and need their urine a different pH.<br />Their fur has never been softer, they are so active and always chasing each other, they are very playful with us. We always comment on how soft the older one is, just like a pillow, and the younger one is getting closer every day. They eat the food I leave out for them every day, so I'm assuming it is satisfying. I just love this brand and plan on sticking with them, even when we get dogs.

```
In [42]: count=1
         for i in range(3):
             print('Review-%d : \n %s\n'%(count,cluster5[i]))
             count +=1
```

Review-1 :

Blend a little apple cider mix into your Tazo Passion Tea and see what happens. This creates a magical apple cider experience with hints of cranberry and hibiscus flowers gathered in an ancient land.<br /><br />Snuggle Up Spiced Cider is a good choice because you can then spoon in the exact amount you desire, but packets of cider work equally well. This is the type of tea you would expect to enjoy after skiing in some distant forest where you are also munching on homemade brownies or eggnog cookies. Apple cider also works very well with the Gypsy Teas, especially the Lemon Jasmine.<br /><br />The deep purple/red color and hibiscus fragrance of Tazo Tea is enough to make anyone passionate about tea. Yet, is this really tea? It looks like tea, but it is really an herbal infusion made by steeping hibiscus flowers, orange peel, licorice, cinnamon, rose hips, lemongrass and red poppy in a swirling of hot water.<br /><br />The original flavor is also slightly reminiscent of hot apple cider or hot cranberry juice which is why it is extra delicious with a cider mix. Tazo Teas are a mysterious experience into which you can lose yourself with a sense of abandon. So far I've tried more flavors than I can count and have always been impressed.<br /><br />~The Rebecca Review

Review-2 :

I think that this is a great Chai! Far from sweet.. I found it very unsweet (and that is exactly why I like it). The Tiger Spice is a great blend of Black Tea, Cinnamon, Ginger, herbs, and more. I was not a big tea drinker (until now) but this is a very smooth and light Chai. All you do is add water and it comes out creamy and perfect (some people add milk but this Chai mix already has a creamer in it). This is my new healthy breakfast wake me up. It is great. I have many Indian friends with whom I drink freshly made Chai with and this is as close as you can come with an instant blend. I highly recommend this.

Review-3 :

For a bag tea--which generally are not as good quality, this brews up like a top quality loose tea. It's a delicious snappy black blend, with a lovely scent and taste. One of my favorites.

```
In [43]: count=1
         for i in range(3):
             print('Review-%d : \n %s\n'%(count,cluster6[i]))
             count +=1
```



## Review-1 :

Using the Briel Chamonix machine I purchased from Amazon, I have produced delicious lattes using these pods. The smell of the Espressione pods (regular and decaf) when opened is incredible - dark, rich roasts. FYI - the pods are produced in Italy, which is a good sign. Also, each pod has a handy paper "tab" so that you can remove it from your machine without burning your fingers. I compared these pods to Starbucks espresso pods (regular and decaf), and the Espressione are my favorite. Only one warning: 150 pods (individually wrapped) take up a lot of space, so be sure you know where you're going to store them. More good news: each box of 150 is stamped with a "best if used by" date, which in my case was 16 months after the date of purchase.

## Review-2 :

After years of using Starbucks' decaf and regular single serving espresso pods, I finally found a substitute that's more economical and MUCH better tasting. I've used three different brands of pods in my Starbucks' Barista machine, and have found that the quality of the pod and the flavor of the espresso inside is, by far, the best in the Espressione pods. The flavor is more full bodied and stronger than in the Starbucks' pods. I actually use one pod to make two shots of espresso, and the flavor is more intense than when I buy my specialty drinks from Starbucks or other coffee houses. Also, the pod itself is of very high quality. Each pod is individually wrapped in an airtight, hard foil wrapper, and I've never opened a package yet that contained a torn pod with coffee spilling out. Such is not the case with other lesser brands. The tab on each pod is also handy, making it easier to switch out a hot spent pod with a new one when making a second round of espresso.

## Review-3 :

Few things I'd like to point out: 1. "Espresso" is not a roast. It's a drink. In northern Italy they make espresso from Vienna roast, in southern Italy they use French. Starbucks appears to be using French roast, so they should just label it French roast like any normal person would. 2. The product description says "Taste Espresso Roast as both a shot of espresso and brewed in a coffee press. This is one versatile coffee!" This is patently absurd. It's pre-ground coffee, and it's either for use in espresso machines or presses, but not both. Here's why: Making a shot of espresso requires a very fine grind. Making coffee press coffee requires a very coarse grind. If you have ground coffee that is appropriate for espresso machines, it will pass through the filter in a coffee press and you will have a cup of sludge. If you use ground coffee that does not pass through a press's filter in an espresso machine, you do not get espresso. You get cafe crema, which is a good, but entirely different, beverage. Therefore, a "Universal Grind" is an impossibility. They should tell you on the label what kind of coffee maker it's ground for, like normal people. 3. It's over-roasted. Starbucks has an excellent bean-sourcing operation that buys better coffee than any other company of its size. Then they roast all their beans Vienna or French. Most beans taste best at a particular level of roasting that varies with the bean, and for most beans that level is not Vienna or French. So why roast all your beans too dark? I guess it's just easier that way. Also, it lets them rename Vienna Roast as "Starbucks Roast" because the Austrians don't seem to care. That said, this is still better than what you get in the supermarket - Starbucks specializes in coffee and sources much better beans than any of the companies that also make cheese, hot dogs, shampoo or cigarettes. Anyway, if you really want some good coffee, do this: Buy a grinder. Nothing makes coffee lose flavor like keeping it around after grinding. If you have a drip coffee machine, get a grinder with spinning blades; if you have something else like a press or vacuum pot, get a grinder with burrs. If you have an espresso machine, get a very nice grinder with burrs. Amazon has good deals on all three kinds, just do a search for 'coffee grinder'. It's not that much money, and if you're somebody who's willing to pay for Starbucks you should invest in one. Find a local coffee shop that roasts their own coffee and gives you the date it's roasted on. Freshness matters! Good beans don't taste nearly as good after they've been sitting around. Ask the people at your local shop to tell you about the correct amount of beans to use. Most people don't use the right amount. If you don't use enough, your coffee comes out bitter, espe

## k-medoids

```
In [44]: from sklearn.decomposition import TruncatedSVD
from pyclustering.utils import calculate_distance_matrix
from pyclustering.cluster import kmedoids
```

```
In [46]: def print_cluster_reviews_k_medoids(cluster_label_reviews, cluster_number = 0, review
_count=3):
    reviews = cluster_label_reviews[cluster_number]
    reviews = reviews[0:review_count]
    print("---: Cluster {0} :---".format(cluster_number))
    index = 1
    for review in reviews:
        print("Review {0} : ".format(index))
        print(CLEANED_TEXT[review])
        print()
        index += 1
```

```
In [47]: def get_cluster_N_medoids(k_medoids):
    return k_medoids.get_clusters(), k_medoids.get_medoids()
```

```
In [48]: svd = TruncatedSVD(n_components=2)
tfidf_dtm_2_dim = svd.fit_transform(data)
tfidf_dtm_2_dim.shape
```

```
Out[48]: (30000, 2)
```

```
In [49]: k_medoids = kmedoids.kmedoids(tfidf_dtm_2_dim, [2, 9, 12], 0.001, ccore=True)
# Run cluster analysis and obtain results
k_medoids.process()
```

```
Out[49]: <pyclustering.cluster.kmedoids.kmedoids at 0x1dd933b5240>
```

```
In [50]: clusters, medoids = get_cluster_N_medoids(k_medoids)
```

```
In [52]: for i in range(0, len(clusters)):
    list_if_indices = clusters[i]
    # Number of reviews in different clusters
    print("Number of Reviews in Cluster {0} : {1} ".format(i, len(list_if_indices)))
```

```
Number of Reviews in Cluster 0 : 14235
Number of Reviews in Cluster 1 : 3136
Number of Reviews in Cluster 2 : 12629
```

```
In [54]: CLEANED_TEXT = my_final['Text'].values
```

```
In [55]: print_cluster_reviews_k_medoids(clusters, cluster_number = 0, review_count = 3)
print("\n\n")
print_cluster_reviews_k_medoids(clusters, cluster_number = 1, review_count = 3)
print("\n\n")
print_cluster_reviews_k_medoids(clusters, cluster_number = 2, review_count = 3)
print("\n\n")
```

---: Cluster 0 :---

Review 1 :

I am not sure what they fed these guys but it is absolutely the best hare I ever eaten. I stuffed it with escargots and a creamy champagne cream sauce with truff les from Luxembourg and and eatable Thai Lilly flowers. Delightful. I added the leftovers to a Flan recipe that went over very well with the children. They all wanted more and skipped the ice cream. I'm ordering more. PS keep dental floss h andy.

Review 2 :

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 s illiness of it all. this is a classic book i am willing to bet my son will STI LL be able to recite from memory when he is in college

Review 3 :

I can remember seeing the show when it aired on television years ago, when I was a child. My sister later bought me the LP (which I have to this day, I'm thirt y something).I used this series of books & songs when I did my student teac hing for preschoolers & turned the whole school on to it. I am now purchasi ng it on CD, along with the books for my children 5 & 2. The tradition liv es on!

---: Cluster 1 :---

Review 1 :

soothing. The aroma & flavor are very good. It does in fact have a "sweet minty flavor". I like! It is a kinda expensive though. I can't find it anywhere for less than \$3.50 a box.<br /><br />(in general i am a huge Yogi tea fan. but i d on't i don't think their stress relief tea is as good as traditional medicinals)

Review 2 :

The grocery store has all kinds of &quot;gourmet&quot; coffees laid out. Each on e I tried tasted like dust. Maybe they are ubiquitous, but Starbucks puts out a decent product and I can get a good cup in the a.m. and that's important.<p>The Breakfast Blend is a bit mild for my tastes (I like really STRONG coffee) but fo r guest and general consumption, this is a mild, aromatic coffee that is better than most.

Review 3 :

This blend is one of Starbucks' gentler blends, but you have to like the taste o f Starbucks to appreciate it. Often described as having a burnt taste, Starbuck s beans are known for their strongly roasted taste. Maxwell House aficionados should steer clear.<p>Before you invest in Starbucks coffee, make sure you like the flavor by tasting the brew at a local franchise. If you don't have one near you, make sure you like a strong coffee flavor. Although I enjoy Starbucks coff ee myself, I am leery of recommending it to the uninitiated. If you like the ta ste, stock up now. The coffee stores well for months in the freezer.

---: Cluster 2 :---

Review 1 :

I almost bought this for my cat i have been looking for a item like this at loca l stores so i placed the order and was reviewing everything i also bought severa l other items so did not think anything of the total until i looked at the invoi ce for this particular item the toy cost a 1.99 but they get you with shipping i t was 6.55 bringing the total to 8.54. i thought it was overpriced and cancelled my order i guess it is buyer beware!!!!

In [ ]:

## Word2Vec

```
In [56]: sent_x = []
for sent in x :
    sent_x.append(sent.split())

# Train your own Word2Vec model using your own train text corpus
# min_count = 5 considers only words that occurred at least 5 times
w2v_model=Word2Vec(sent_x,min_count=5,size=50, workers=4)

w2v_words = list(w2v_model.wv.vocab)
print("number of words that occurred minimum 5 times ",len(w2v_words))

number of words that occurred minimum 5 times 8130
```

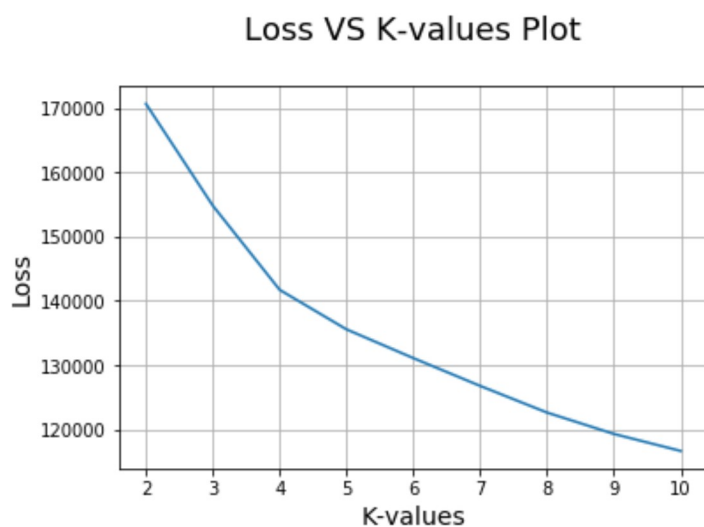
```
In [57]: train_vectors = [];
for sent in sent_x:
    sent_vec = np.zeros(50)
    cnt_words =0;
    for word in sent: #
        if word in w2v_words:
            vec = w2v_model.wv[word]
            sent_vec += vec
            cnt_words += 1
    if cnt_words != 0:
        sent_vec /= cnt_words
    train_vectors.append(sent_vec)

data = train_vectors
```

## K-Means++ Implementation

```
In [58]: k_values = [2,3,4,5,6,7,8,9,10]
loss = []
for i in k_values:
    kmeans = KMeans(n_clusters=i, n_jobs=-1).fit(data)
    loss.append(kmeans.inertia_)
```

```
In [59]: plt.plot(k_values, loss)
plt.xlabel('K-values',size=14)
plt.ylabel('Loss',size=14)
plt.title('Loss VS K-values Plot\n',size=18)
plt.grid()
plt.show()
```



From above we can see that there is inflection at  $K = 4$ . Before it loss was decreasing faster as compared to the loss decreasing after it. So, the best value of  $K$  is 4.

```
In [60]: optimal_k = 4
# Variable that will be used in the conclusion
avg_w2v_means_k = optimal_k

# Implementing K-Means++ using optimal value of K
kmeans = KMeans(n_clusters=optimal_k, n_jobs=-1).fit(data)

# Getting all the reviews in different clusters
cluster1 = []
cluster2 = []
cluster3 = []
cluster4 = []

for i in range(kmeans.labels_.shape[0]):
    if kmeans.labels_[i] == 0:
        cluster1.append(reviews[i])
    elif kmeans.labels_[i] == 1:
        cluster2.append(reviews[i])
    elif kmeans.labels_[i] == 2:
        cluster3.append(reviews[i])
    else :
        cluster4.append(reviews[i])

# Number of reviews in different clusters
print("No. of reviews in Cluster-1 : ",len(cluster1))
print("\nNo. of reviews in Cluster-2 : ",len(cluster2))
print("\nNo. of reviews in Cluster-3 : ",len(cluster3))
print("\nNo. of reviews in Cluster-4 : ",len(cluster4))
```

No. of reviews in Cluster-1 : 4700

No. of reviews in Cluster-2 : 10441

No. of reviews in Cluster-3 : 6160

No. of reviews in Cluster-4 : 8699

```
In [61]: count=1
for i in range(3):
    print('Review-%d : \n %s\n'%(count,cluster1[i]))
    count +=1
```

Review-1 :

For years I have been trying to simulate a truly Italian espresso or cappuccino without success. Several packages arrived [...] just before Christmas. My son had sent a Briel Cadiz Espresso machine, two lovely Russian china cups and a case of Espressione Classic Espresso Pods. It was the kindest, most wonderful Christmas present I can remember. The espresso machine excels in quality and ease of use. The espresso is exquisite. Each time I prepare an espresso or cappuccino I experience my son's loving kindness and a small part of the many gifts of Italy. Heaven on Earth.

Review-2 :

Using the Briel Chamonix machine I purchased from Amazon, I have produced delicious lattes using these pods. The smell of the Espressione pods (regular and decaf) when opened is incredible - dark, rich roasts. FYI - the pods are produced in Italy, which is a good sign. Also, each pod has a handy paper "tab" so that you can remove it from your machine without burning your fingers. I compared these pods to Starbucks espresso pods (regular and decaf), and the Espressione are my favorite. Only one warning: 150 pods (individually wrapped) take up a lot of space, so be sure you know where you're going to store them. More good news: each box of 150 is stamped with a "best if used by" date, which in my case was 16 months after the date of purchase.

Review-3 :

I've purchased both the Espressione Espresso (classic) and the 100% Arabica. My vote is definitely with the 100% Arabica. The flavor has more bite and flavor (much more like European coffee than American).

```
In [62]: count=1
for i in range(3):
    print('Review-%d : \n %s\n'%(count,cluster2[i]))
    count +=1
```

Review-1 :

THIS IS ONE MOVIE THAT SHOULD BE IN YOUR MOVIE COLLECTION. IT IS FILLED WITH COMEDY, ACTION, AND WHATEVER ELSE YOU WANT TO CALL IT.

Review-2 :

These are easy to use, they do not make a mess, and offer vibrant colors. They do not taint what you are decorating as some colors can. I would highly recommend these to anyone who likes to decorate.

Review-3 :

These are so much easier to use than the Wilton paste colors. Colors are vibrant, and do not taint the frosting like some colors can. These are simple to use, and do not make a mess. My only complaint is that I did not find these years ago. This is a must have if you decorate often!



```
In [63]: count=1
for i in range(3):
    print('Review-%d : \n %s\n'%(count,cluster3[i]))
    count +=1
```

Review-1 :

I bought a few of these after my apartment was infested with fruit flies. After only a few hours, the trap had "attracted" many flies and within a few days they were practically gone. This may not be a long term solution, but if flies are driving you crazy, consider buying this. One caution- the surface is very sticky, so try to avoid touching it.

Review-2 :

I THOUGHT THIS MOVIE WAS SO FUNNY, MICHAEL KEATON AS BEETLEJUICE WAS RIGHT ON. AND ALEC BALDWIN PLAYING A GEEKY HUSBAND IS PRICELESS. I LOVE THE DA-OO SCENCE. THE KIDS AND I REALLY ENJOY THIS ONE.

Review-3 :

It was on the other night, and, having been a big fan of the cartoon when it was being shown, decided to watch it (I'm also trying to see all the Winona Ryder movies I know.). <p> It's a well known fact i like seeing her [Winona Ryder] dance at the end of a movie (hey, it's funny!), but apart from that, it was just a hilarious film and i liked it very much. The only thing wrong was that it was too short. Maybe the special effects were a bit "off", but hey, it's an older movie!

```
In [64]: count=1
for i in range(3):
    print('Review-%d : \n %s\n'%(count,cluster4[i]))
    count +=1
```

Review-1 :

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 :

I can remember seeing the show when it aired on television years ago, when I was a child. My sister later bought me the LP (which I have to this day, I'm thirty something). I used this series of books & songs when I did my student teaching for preschoolers & turned the whole school on to it. I am now purchasing it on CD, along with the books for my children 5 & 2. The tradition lives on!

Review-3 :

Beetlejuice is a well written movie ..... everything about it is excellent! From the acting to the special effects you will be delighted you chose to view this movie.

## k-mediods

```
In [65]: svd = TruncatedSVD(n_components=2)
avg_w2v_arr_2_dim = svd.fit_transform(data)
avg_w2v_arr_2_dim.shape
```

Out[65]: (30000, 2)

```
In [66]: k_medoids = kmedoids.kmedoids(avg_w2v_arr_2_dim,[2, 9, 12],0.001,ccore=True)
         # Run cluster analysis and obtain results
         k_medoids.process()
```

```
Out[66]: <pyclustering.cluster.kmedoids.kmedoids at 0x1dd95f2f518>
```

```
In [67]: clusters,medoids = get_cluster_N_medoids(k_medoids)
```

```
In [68]: for i in range(0,len(clusters)):
         list_if_indices = clusters[i]
         # Number of reviews in different clusters
         print("Number of Reviews in Cluster {0} : {1} ".format(i,len(list_if_indices)))
```

```
Number of Reviews in Cluster 0 : 13243
Number of Reviews in Cluster 1 : 10106
Number of Reviews in Cluster 2 : 6651
```

```
In [69]: print_cluster_reviews_k_medoids(clusters, cluster_number = 0, review_count = 3)
print("\n\n")
print_cluster_reviews_k_medoids(clusters, cluster_number = 1, review_count = 3)
print("\n\n")
print_cluster_reviews_k_medoids(clusters, cluster_number = 2, review_count = 3)
print("\n\n")
```

---: Cluster 0 :---

Review 1 :

I toss this on most chicken items I make from whole chickens to wings.<br />Just to spice up things I put a little in prior to grilling and still keep whatever sauce I want.

Review 2 :

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 3 :

I can remember seeing the show when it aired on television years ago, when I was a child. My sister later bought me the LP (which I have to this day, I'm thirty something). I used this series of books & songs when I did my student teaching for preschoolers & turned the whole school on to it. I am now purchasing it on CD, along with the books for my children 5 & 2. The tradition lives on!

---: Cluster 1 :---

Review 1 :

This is definitely a tea that stands alone. This tea is from a bitter, cooling-type herb. Good for problems with digestion, this tea does whatever is necessary to get your GI system back in order. Do not steep too long and remove the tea bag. Sip slowly and enjoy the roasted flavor of this calming and detoxifying drink.

Review 2 :

I'm sick of scads of nasty toothpicks all over my counter when I'm tinting icings; the squeeze-bottle concept is one that is long overdue and exceptionally welcome. The colors are vibrant, do not have an offensive taste, and blend, in my opinion, more smoothly than the Wilton pastes. You'll need to experiment a bit to get hues right at first, but it's well worth it.

Review 3 :

I don't know why anyone would ever use those little liquid food colors again after trying these pastes. The colors are more intense, and you can easily blend lovely shades for your icings. <p>Martha Stewart shows wonderful seasonal cookies from time to time on her TV show. A simple eggwhite-sugar icing on cookies in exciting colors makes such a wonderful table decoration as well as great dessert for parties, teas and other events. These paste colors are a must if you like to make frosted cookies or party cakes. <p>And these are especially fun if you have kids. A cookie-icing contest is a fun event for a birthday party, Christmas party or just a rainy afternoon.

---: Cluster 2 :---

Review 1 :

Wow. I just got a yorkie puppy, 10 weeks at the time of coming home. Breeder sent me home with mighty dog canned food. Worst smelling product I have encountered, my house and puppy stunk. I know dogs are not human, but a product that has that scent cannot possibly be good for the dog. I went to petco and got some supplies including wellness puppy treats. He loved them so much I decided to quickly switch to this dog food. No transition just fed it to him, he loves it, eats more, and has absolutely no problems. I am happy with the ingredient list. Highly recommend. Also when I first opened this I was pleasantly surprised to see actual pieces of vegetables in the can, plus it smelled kind of like tuna, much much better

## TFIDF-Word2Vec

```
In [70]: tf_idf_vect = TfidfVectorizer()

# final_tf_idf1 is the sparse matrix with row= sentence, col=word and cell_val = tf
idf
final_tf_idf1 = tf_idf_vect.fit_transform(x)

# tfidf words/col-names
tfidf_feat = tf_idf_vect.get_feature_names()

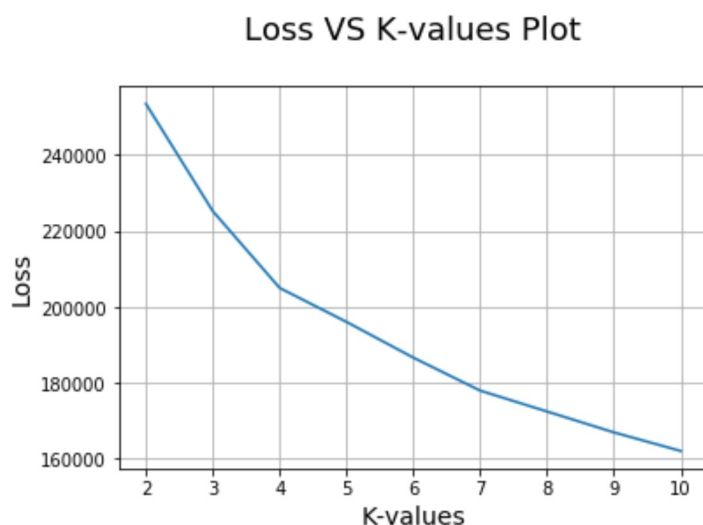
# compute TFIDF Weighted Word2Vec for each review for sent_x .
tfidf_vectors = [];
row=0;
for sent in sent_x:
    sent_vec = np.zeros(50)
    weight_sum =0;
    for word in sent:
        if word in w2v_words:
            vec = w2v_model.wv[word]
            # obtain the tf_idfidf of a word in a sentence/review
            tf_idf = final_tf_idf1[row, tfidf_feat.index(word)]
            sent_vec += (vec * tf_idf)
            weight_sum += tf_idf
    if weight_sum != 0:
        sent_vec /= weight_sum
    tfidf_vectors.append(sent_vec)
    row += 1

data = tfidf_vectors
```

## K-Means++ Implementation

```
In [71]: k_values = [2,3,4,5,6,7,8,9,10]
loss = []
for i in k_values:
    kmeans = KMeans(n_clusters=i, n_jobs=-1).fit(data)
    loss.append(kmeans.inertia_)
```

```
In [72]: plt.plot(k_values, loss)
plt.xlabel('K-values',size=14)
plt.ylabel('Loss',size=14)
plt.title('Loss VS K-values Plot\n',size=18)
plt.grid()
plt.show()
```



**From above we can see that there is inflection at K = 4 . Befor it loss was decreasing faster as compared to the loss decreasing after it . So , the best value of K is 4**

```
In [73]: optimal_k = 4
# Variable that will be used in the conclusion
tfidf_w2v_means_k = optimal_k

# Implementing K-Means++ using optimal value of K
kmeans = KMeans(n_clusters=optimal_k, n_jobs=-1).fit(data)

# Getting all the reviews in different clusters
cluster1 = []
cluster2 = []
cluster3 = []
cluster4 = []

for i in range(kmeans.labels_.shape[0]):
    if kmeans.labels_[i] == 0:
        cluster1.append(reviews[i])
    elif kmeans.labels_[i] == 1:
        cluster2.append(reviews[i])
    elif kmeans.labels_[i] == 2:
        cluster3.append(reviews[i])
    else :
        cluster4.append(reviews[i])

# Number of reviews in different clusters
print("No. of reviews in Cluster-1 : ",len(cluster1))
print("\nNo. of reviews in Cluster-2 : ",len(cluster2))
print("\nNo. of reviews in Cluster-3 : ",len(cluster3))
print("\nNo. of reviews in Cluster-4 : ",len(cluster4))
```

No. of reviews in Cluster-1 : 3719

No. of reviews in Cluster-2 : 5283

No. of reviews in Cluster-3 : 10593

No. of reviews in Cluster-4 : 10405

```
In [74]: count=1
         for i in range(3):
             print('Review-%d : \n %s\n'%(count,cluster1[i]))
             count +=1
```

Review-1 :

For years I have been trying to simulate a truly Italian espresso or cappuccino without success. Several packages arrived [...] just before Christmas. My son had sent a Briel Cadiz Espresso machine, two lovely Russian china cups and a case of Espressione Classic Espresso Pods. It was the kindest, most wonderful Christmas present I can remember. The espresso machine excels in quality and ease of use. The espresso is exquisite. Each time I prepare an espresso or cappuccino I experience my son's loving kindness and a small part of the many gifts of Italy. Heaven on Earth.

Review-2 :

Using the Briel Chamonix machine I purchased from Amazon, I have produced delicious lattes using these pods. The smell of the Espressione pods (regular and decaf) when opened is incredible - dark, rich roasts. FYI - the pods are produced in Italy, which is a good sign. Also, each pod has a handy paper "tab" so that you can remove it from your machine without burning your fingers. I compared these pods to Starbucks espresso pods (regular and decaf), and the Espressione are my favorite. Only one warning: 150 pods (individually wrapped) take up a lot of space, so be sure you know where you're going to store them. More good news: each box of 150 is stamped with a "best if used by" date, which in my case was 16 months after the date of purchase.

Review-3 :

After years of using Starbucks' decaf and regular single serving espresso pods, I finally found a substitute that's more economical and MUCH better tasting. I've used three different brands of pods in my Starbucks' Barista machine, and have found that the quality of the pod and the flavor of the espresso inside is, by far, the best in the Espressione pods. The flavor is more full bodied and stronger than in the Starbucks' pods. I actually use one pod to make two shots of espresso, and the flavor is more intense than when I buy my specialty drinks from Starbucks or other coffee houses. Also, the pod itself is of very high quality. Each pod is individually wrapped in an airtight, hard foil wrapper, and I've never opened a package yet that contained a torn pod with coffee spilling out. Such is not the case with other lesser brands. The tab on each pod is also handy, making it easier to switch out a hot spent pod with a new one when making a second round of espresso.



```
In [75]: count=1
for i in range(3):
    print('Review-%d : \n %s\n'%(count,cluster2[i]))
    count +=1
```

Review-1 :

Boots (the most handsome red tom cat you'll ever see!) loves these treats. We keep them on a shelf in the pantry - only problem is everytime we open the pantry door he thinks he gets a treat.<p> We've tried other cat treats which he always eats (truth is, Boots eats everything!) but these seem to be his favorite by far.

Review-2 :

My cat Foutchie is heading toward 5 now and no one told her the trouble with a kitten is it eventually becomes a cat (so take that Ogden Nash!). She still vroom-vooms around at a pace that makes me exhausted. She loves toys, especially stuff she chases. So the bubbles are great full. Especially since they are catnip bubbles. You don't have to worry about them getting soap on them and making them sick. She loves jumping high to pop them.<p>So if you are looking for something to awaken the kitten in your kitty, you might try these.

Review-3 :

Are you supposed to avoid sugars if you have a urinary tract infection?

```
In [76]: count=1
for i in range(3):
    print('Review-%d : \n %s\n'%(count,cluster3[i]))
    count +=1
```

Review-1 :

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 :

I can remember seeing the show when it aired on television years ago, when I was a child. My sister later bought me the LP (which I have to this day, I'm thirty something). I used this series of books & songs when I did my student teaching for preschoolers & turned the whole school on to it. I am now purchasing it on CD, along with the books for my children 5 & 2. The tradition lives on!

Review-3 :

Beetlejuice is a well written movie ..... everything about it is excellent! From the acting to the special effects you will be delighted you chose to view this movie.

```
In [77]: count=1
        for i in range(3):
            print('Review-%d : \n %s\n'%(count,cluster4[i]))
            count +=1
```

Review-1 :

I'm getting crazy.I'm looking for Beatlejuice french version video.<p>Is it really impossible today not to find the French VHS version of this film ?<p>Could U please tell me something about it ? Tks

Review-2 :

These are easy to use, they do not make a mess, and offer vibrant colors. They do not taint what you are decorting as some colors can. I would highly recomme nd these to anyone to likes to decorate.

Review-3 :

This are so much easier to use than the Wilson paste colors. Colors are vibra nt, and do not taint the frosting like some colors can. These are simple to us e, and do not make a mess. My only complaint is that I did not find these year s ago. This is a must have if you decorate often!

## k-mediods

```
In [79]: svd = TruncatedSVD(n_components=2)
        tfidf_w2v_arr_2_dim = svd.fit_transform(data)
        tfidf_w2v_arr_2_dim.shape
```

Out[79]: (30000, 2)

```
In [80]: k_medoids = kmedoids.kmedoids(tfidf_w2v_arr_2_dim,[2, 9, 12],0.001,ccore=True)
        # Run cluster analysis and obtain results
        k_medoids.process()
```

Out[80]: <pyclustering.cluster.kmedoids.kmedoids at 0x1dd9e820710>

```
In [81]: clusters,medoids = get_cluster_N_medoids(k_medoids)
```

```
In [82]: for i in range(0,len(clusters)):
        list_if_indices = clusters[i]
        # Number of reviews in different clusters
        print("Number of Reviews in Cluster {0} : {1} ".format(i,len(list_if_indices)))
```

Number of Reviews in Cluster 0 : 5676

Number of Reviews in Cluster 1 : 7068

Number of Reviews in Cluster 2 : 17256

```
In [83]: print_cluster_reviews_k_medoids(clusters, cluster_number = 0, review_count = 3)
print("\n\n")
print_cluster_reviews_k_medoids(clusters, cluster_number = 1, review_count = 3)
print("\n\n")
print_cluster_reviews_k_medoids(clusters, cluster_number = 2, review_count = 3)
print("\n\n")
```

---: Cluster 0 :---

Review 1 :

My cat recently freaked out over some catnip that came with a bed I bought him so I thought I'd order some more. I couldn't find the same brand but this "Xtreme" catnip sounded like it might do the trick. Not even close. My husband swears he saw our cat lift up his paw and yawn after smelling it. If your cat is like mine you'll probably get a better reaction out of any other catnip out there.

Review 2 :

I bought a few of these after my apartment was infested with fruit flies. After only a few hours, the trap had "attracted" many flies and within a few days they were practically gone. This may not be a long term solution, but if flies are driving you crazy, consider buying this. One caution- the surface is very sticky, so try to avoid touching it.

Review 3 :

I have just recently purchased the Woodstream Corp 0610 gopher trap, and within 10 minutes of laying the traps i was catching gophers. This product is the best i have ever used. The 0610 gopher trap is easy to set and works with great success. Also, remember to use a wire attached to the trap and tie it to a stake to prevent gophers from dragging the trap further into their hole once they are caught. I hope you find this product as easy to use as i did, Good luck.

---: Cluster 1 :---

Review 1 :

Seriously? It's flavored sugar with candy that you lick and dip in the sugar so as to eat sugar with your candy that is mostly pressed sugar.<br /><br />That said. It's sooo good. And if your parents let you have it they obviously loved you because no one ACTUALLY wants to hand a kid a bunch of sugar and say, "GO FOR IT!"<br /><br />At least that's my take on it.<br /><br />-Mike

Review 2 :

I bought these to decorate some dia de los muertos skulls with icing. The Ateco food coloring gave nice bright, vibrant colors and mixed with the icing very well. I was totally satisfied with the purchase.

Review 3 :

For years I have been trying to simulate a truly Italian espresso or cappuccino without success. Several packages arrived [...] just before Christmas. My son had sent a Briel Cadiz Espresso machine, two lovely Russian china cups and a case of Espressione Classic Espresso Pods. It was the kindest, most wonderful Christmas present I can remember. The espresso machine excels in quality and ease of use. The espresso is exquisite. Each time I prepare an espresso or cappuccino I experience my son's loving kindness and a small part of the many gifts of Italy. Heaven on Earth.

---: Cluster 2 :---

Review 1 :

Stay away from these people! They are passing off kosher salt as pretzel salt. Save your time and money!! They are ripoffs.

Review 2 :

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

## Conclusion

```
In [93]: from prettytable import PrettyTable

names = ['K-means++ for BoW', 'K-medoids for BoW', 'K-means++ for TFIDF', 'K-medoids f
or TFIDF', 'K-means++ for Avg_Word2Vec', \
        'K-medoids for Avg_Word2Vec', 'K-means++ for tfidf_Word2Vec', 'K-medoids for
tfidf_Word2Vec']

optimal_k = [4,3,6,3,4,3,4,3]

numbering = [1,2,3,4,5,6,7,8]

# Initializing prettytable
table = PrettyTable()

# Adding columns
table.add_column("S.NO.", numbering)
table.add_column("MODEL", names)
table.add_column("Number of Clusters ", optimal_k)

# Printing the Table
print(table)
```

S.NO.	MODEL	Number of Clusters
1	K-means++ for BoW	4
2	K-medoids for BoW	3
3	K-means++ for TFIDF	6
4	K-medoids for TFIDF	3
5	K-means++ for Avg_Word2Vec	4
6	K-medoids for Avg_Word2Vec	3
7	K-means++ for tfidf_Word2Vec	4
8	K-medoids for tfidf_Word2Vec	3

In [ ]:

In [ ]: