

# DonorsChoose

DonorsChoose.org receives hundreds of thousands of project proposals each year for classroom projects in need of funding. Right now, a large number of volunteers is needed to manually screen each submission before it's approved to be posted on the DonorsChoose.org website.

Next year, DonorsChoose.org expects to receive close to 500,000 project proposals. As a result, there are three main problems they need to solve:

- How to scale current manual processes and resources to screen 500,000 projects so that they can be posted as quickly and as efficiently as possible
- How to increase the consistency of project vetting across different volunteers to improve the experience for teachers
- How to focus volunteer time on the applications that need the most assistance

The goal of the competition is to predict whether or not a DonorsChoose.org project proposal submitted by a teacher will be approved, using the text of project descriptions as well as additional metadata about the project, teacher, and school. DonorsChoose.org can then use this information to identify projects most likely to need further review before approval.

## About the DonorsChoose Data Set

The `train.csv` data set provided by DonorsChoose contains the following features:

Feature	Description
<code>project_id</code>	A unique identifier for the proposed project. <b>Example</b>
<code>project_title</code>	Title of the project. <b>Examples:</b> <ul style="list-style-type: none"> <li>• Art Will Make You Happy!</li> <li>• First Grade Fun</li> </ul>
<code>project_grade_category</code>	Grade level of students for which the project is targeted. Enumerated values: <ul style="list-style-type: none"> <li>• Grades PreK-2</li> <li>• Grades 3-5</li> <li>• Grades 6-8</li> <li>• Grades 9-12</li> </ul>
<code>project_subject_categories</code>	One or more (comma-separated) subject categories from the following enumerated list of values: <ul style="list-style-type: none"> <li>• Applied Learning</li> <li>• Care &amp; Hunger</li> <li>• Health &amp; Sports</li> <li>• History &amp; Civics</li> <li>• Literacy &amp; Language</li> <li>• Math &amp; Science</li> <li>• Music &amp; The Arts</li> <li>• Special Needs</li> <li>• Warmth</li> </ul> <b>Examples:</b> <ul style="list-style-type: none"> <li>• Music &amp; The Arts</li> <li>• Literacy &amp; Language, Math &amp; Science</li> </ul>
<code>school_state</code>	State where school is located ( <u>Two-letter U.S. postal code</u> ( <a href="https://en.wikipedia.org/wiki/List_of_U.S._state_abbreviations">https://en.wikipedia.org/wiki/List_of_U.S._state_abbreviations</a> )). <b>Example:</b> WY
<code>project_subject_subcategories</code>	One or more (comma-separated) subject subcategories. <b>Examples:</b> <ul style="list-style-type: none"> <li>• Literacy</li> <li>• Literature &amp; Writing, Social Sciences</li> </ul>
<code>project_resource_summary</code>	An explanation of the resources needed for the project. <ul style="list-style-type: none"> <li>• My students need hands on literacy materials to meet sensory needs!</li> </ul>

Feature	Description
project_essay_1	First application essay*
project_essay_2	Second application essay*
project_essay_3	Third application essay*
project_essay_4	Fourth application essay*
project_submitted_datetime	Datetime when project application was submitted. <b>Example:</b> 12:43:56.245
teacher_id	A unique identifier for the teacher of the proposed project. <b>Example:</b> bdf8baa8fedef6bfeec7ae4ff1c15c56
teacher_prefix	Teacher's title. One of the following enumerated values: <ul style="list-style-type: none"> <li>• nan</li> <li>• Dr.</li> <li>• Mr.</li> <li>• Mrs.</li> <li>• Ms.</li> <li>• Teacher.</li> </ul>
teacher_number_of_previously_posted_projects	Number of project applications previously submitted by the teacher. <b>Example:</b> 2

\* See the section **Notes on the Essay Data** for more details about these features.

Additionally, the `resources.csv` data set provides more data about the resources required for each project. Each line in this file represents a resource required by a project:

Feature	Description
id	A project_id value from the <code>train.csv</code> file. <b>Example:</b> p036502
description	Description of the resource. <b>Example:</b> Tenor Saxophone Reeds, Box of 25
quantity	Quantity of the resource required. <b>Example:</b> 3
price	Price of the resource required. <b>Example:</b> 9.95

**Note:** Many projects require multiple resources. The `id` value corresponds to a `project_id` in `train.csv`, so you use it as a key to retrieve all resources needed for a project:

The data set contains the following label (the value you will attempt to predict):

Label	Description
project_is_approved	A binary flag indicating whether DonorsChoose approved the project. A value of 0 indicates the project was not approved, and a value of 1 indicates the project was approved.



## Notes on the Essay Data

Prior to May 17, 2016, the prompts for the essays were as follows:

- `__project_essay_1:__` "Introduce us to your classroom"
- `__project_essay_2:__` "Tell us more about your students"
- `__project_essay_3:__` "Describe how your students will use the materials you're requesting"
- `__project_essay_3:__` "Close by sharing why your project will make a difference"

Starting on May 17, 2016, the number of essays was reduced from 4 to 2, and the prompts for the first 2 essays were changed to the following:

- `__project_essay_1:__` "Describe your students: What makes your students special? Specific details about their background, your neighborhood, and your school are all helpful."
- `__project_essay_2:__` "About your project: How will these materials make a difference in your students' learning and improve their school lives?"

For all projects with `project_submitted_datetime` of 2016-05-17 and later, the values of `project_essay_3` and `project_essay_4` will be NaN.

In [1]:

```
%matplotlib inline
import warnings
warnings.filterwarnings("ignore")

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 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

from tqdm import tqdm
import os

from chart_studio.plotly import plotly
import plotly.offline as offline
import plotly.graph_objs as go
offline.init_notebook_mode()
from collections import Counter
```

## 1.1 Reading Data

In [2]:

```
project_data = pd.read_csv('train_data.csv')
resource_data = pd.read_csv('resources.csv')
```

In [3]:

```
print("Number of data points in train data", project_data.shape)
print('-'*50)
print("The attributes of data :", project_data.columns.values)
```

Number of data points in train data (109248, 17)

-----

The attributes of data : ['Unnamed: 0' 'id' 'teacher\_id' 'teacher\_prefix'  
'school\_state'  
'project\_submitted\_datetime' 'project\_grade\_category'  
'project\_subject\_categories' 'project\_subject\_subcategories'  
'project\_title' 'project\_essay\_1' 'project\_essay\_2' 'project\_essay\_3'  
'project\_essay\_4' 'project\_resource\_summary'  
'teacher\_number\_of\_previously\_posted\_projects' 'project\_is\_approved']

In [4]:

```
print("Number of data points in train data", resource_data.shape)
print(resource_data.columns.values)
resource_data.head(2)
```

Number of data points in train data (1541272, 4)

['id' 'description' 'quantity' 'price']

Out[4]:

	id	description	quantity	price
0	p233245	LC652 - Lakeshore Double-Space Mobile Drying Rack	1	149.00
1	p069063	Bouncy Bands for Desks (Blue support pipes)	3	14.95

## 1.2 preprocessing of project\_subject\_categories

In [5]:

```
categories = list(project_data['project_subject_categories'].values)
# remove special characters from list of strings python: https://stackoverflow.com/a/47301924/4084039

# https://www.geeksforgeeks.org/removing-stop-words-nltk-python/
# https://stackoverflow.com/questions/23669024/how-to-strip-a-specific-word-from-a-string
# https://stackoverflow.com/questions/8270092/remove-all-whitespace-in-a-string-in-python
cat_list = []
for i in categories:
    temp = ""
    # consider we have text like this "Math & Science, Warmth, Care & Hunger"
    for j in i.split(','): # it will split it in three parts ["Math & Science", "Warmth", "Care & Hunger"]
        if 'The' in j.split(): # this will split each of the category based on space "Math & Science"=> "Math", "&", "Science"
            j=j.replace('The','') # if we have the words "The" we are going to replace it with '' (i.e removing 'The')
            j = j.replace(' ', '') # we are replacing all the ' ' (space) with '' (empty) ex: "Math & Science"=> "Math&Science"
            temp+=j.strip()+" " # " abc ".strip() will return "abc", remove the trailing spaces
        temp = temp.replace('&','_') # we are replacing the & value into
    cat_list.append(temp.strip())

project_data['clean_categories'] = cat_list
project_data.drop(['project_subject_categories'], axis=1, inplace=True)

from collections import Counter
my_counter = Counter()
for word in project_data['clean_categories'].values:
    my_counter.update(word.split())

cat_dict = dict(my_counter)
sorted_cat_dict = dict(sorted(cat_dict.items(), key=lambda kv: kv[1]))
```

## 1.3 preprocessing of project\_subject\_subcategories

In [6]:

```
sub_categories = list(project_data['project_subject_subcategories'].values)
# remove special characters from list of strings python: https://stackoverflow.com/a/47301924/4084039

# https://www.geeksforgeeks.org/removing-stop-words-nltk-python/
# https://stackoverflow.com/questions/23669024/how-to-strip-a-specific-word-from-a-string
# https://stackoverflow.com/questions/8270092/remove-all-whitespace-in-a-string-in-python

sub_cat_list = []
for i in sub_categories:
    temp = ""
    # consider we have text like this "Math & Science, Warmth, Care & Hunger"
    for j in i.split(','): # it will split it in three parts ["Math & Science", "Warmth", "Care & Hunger"]
        if 'The' in j.split(): # this will split each of the category based on space "Math & Science" => "Math", "&", "Science"
            j = j.replace('The', '') # if we have the words "The" we are going to replace it with '' (i.e removing 'The')
            j = j.replace(' ', '') # we are replacing all the ' ' (space) with '' (empty) ex: "Math & Science" => "Math&Science"
            temp += j.strip() + " #"
    temp = temp.replace('&', '_')
    sub_cat_list.append(temp.strip())

project_data['clean_subcategories'] = sub_cat_list
project_data.drop(['project_subject_subcategories'], axis=1, inplace=True)

# count of all the words in corpus python: https://stackoverflow.com/a/22898595/4084039
my_counter = Counter()
for word in project_data['clean_subcategories'].values:
    my_counter.update(word.split())

sub_cat_dict = dict(my_counter)
sorted_sub_cat_dict = dict(sorted(sub_cat_dict.items(), key=lambda kv: kv[1]))
```

## PREPROCESSING OF PROJECT GRADE CATEGORY

In [7]:

```
grade_categories = list(project_data['project_grade_category'].values)
clean_grades = []
for i in grade_categories:
    temp = ""
    for j in i.split(','):
        j = j.replace(' ', '_')
        j = j.replace('-', '_')
        temp += j
    clean_grades.append(temp)
project_data['clean_grades'] = clean_grades
project_data.drop(['project_grade_category'], axis=1, inplace=True)
```



## 1.3 Text preprocessing

In [8]:

```
# merge two column text dataframe:
project_data["essay"] = project_data["project_essay_1"].map(str) + \
    project_data["project_essay_2"].map(str) + \
    project_data["project_essay_3"].map(str) + \
    project_data["project_essay_4"].map(str)
```

In [9]:

```
project_data.head(2)
```

Out[9]:

	Unnamed: 0	id	teacher_id	teacher_prefix	school_s
0	160221	p253737	c90749f5d961ff158d4b4d1e7dc665fc	Mrs.	IN
1	140945	p258326	897464ce9ddc600bced1151f324dd63a	Mr.	FL

In [10]:

```
#### 1.4.2.3 Using Pretrained Models: TFIDF weighted W2V
```

In [11]:

```
# printing some random reviews
print(project_data['essay'].values[0])
print("="*50)
print(project_data['essay'].values[150])
print("="*50)
print(project_data['essay'].values[1000])
print("="*50)
print(project_data['essay'].values[20000])
print("="*50)
print(project_data['essay'].values[99999])
print("="*50)
```

My students are English learners that are working on English as their second or third languages. We are a melting pot of refugees, immigrants, and native-born Americans bringing the gift of language to our school. \r\n\r\n We have over 24 languages represented in our English Learner program with students at every level of mastery. We also have over 40 countries represented with the families within our school. Each student brings a wealth of knowledge and experiences to us that open our eyes to new cultures, beliefs, and respect.\r\n\r\nThe limits of your language are the limits of your world.\r\n\r\n-Ludwig Wittgenstein Our English learner's have a strong support system at home that begs for more resources. Many times our parents are learning to read and speak English along side of their children. Sometimes this creates barriers for parents to be able to help their child learn phonetics, letter recognition, and other reading skills.\r\n\r\n\r\nBy providing these dvd's and players, students are able to continue their mastery of the English language even if no one at home is able to assist. All families with students within the Level 1 proficiency status, will be offered to be a part of this program. These educational videos will be specially chosen by the English Learner Teacher and will be sent home regularly to watch. The videos are to help the child develop early reading skills.\r\n\r\n\r\nParents that do not have access to a dvd player will have the opportunity to check out a dvd player to use for the year. The plan is to use these videos and educational dvd's for the years to come for other EL students.\r\n\r\nnannan

=====  
The 51 fifth grade students that will cycle through my classroom this year all love learning, at least most of the time. At our school, 97.3% of the students receive free or reduced price lunch. Of the 560 students, 97.3% are minority students. \r\n\r\nThe school has a vibrant community that loves to get together and celebrate. Around Halloween there is a whole school parade to show off the beautiful costumes that students wear. On Cinco de Mayo we put on a big festival with crafts made by the students, dances, and games. At the end of the year the school hosts a carnival to celebrate the hard work put in during the school year, with a dunk tank being the most popular activity. My students will use these five brightly colored Hokki stools in place of regular, stationary, 4-legged chairs. As I will only have a total of ten in the classroom and not enough for each student to have an individual one, they will be used in a variety of ways. During independent reading time they will be used as special chairs students will each use on occasion. I will utilize them in place of chairs at my small group tables during math and reading times. The rest of the day they will be used by the students who need the highest amount of movement in their life in order to stay focused on school.\r\n\r\n\r\nWhenever asked what the classroom is missing, my students always say more Hokki Stools. They can't get their fill of the 5 stools we already have. When the students are sitting in group with me on the Hokki Stools, they are always moving, but at the same time doing their work. Anytime the students get to pick where they can sit, the Hokki Stools are the first to be taken. There are always students who head over to the kidney table to get one of the stools who are disappointed as there are not enough of them. \r\n\r\n\r\nWe ask a lot of students to sit for 7 hours a day. The Hokki stools will be a compromise that allow my students to do desk work and move at the same time. These stools will help students to meet their 60 minutes a day of movement by allowing them to activate their core muscles for balance while they sit. For many of my students, these chairs will take away the barrier that exists in schools for a child who can't sit still.nannan

=====  
How do you remember your days of school? Was it in a sterile environment with plain walls, rows of desks, and a teacher in front of the room? A typical day in our room is nothing like that. I work hard to create a warm inviting themed room for my students look forward to coming to each day.\r\n\r\n\r\nMy class is made up of 28 wonderfully unique boys and girls of mixed r

aces in Arkansas.\r\nThey attend a Title I school, which means there is a high enough percentage of free and reduced-price lunch to qualify. Our school is an "open classroom" concept, which is very unique as there are no walls separating the classrooms. These 9 and 10 year-old students are very eager learners; they are like sponges, absorbing all the information and experiences and keep on wanting more. With these resources such as the comfy red throw pillows and the whimsical nautical hanging decor and the blue fish nets, I will be able to help create the mood in our classroom setting to be one of a themed nautical environment. Creating a classroom environment is very important in the success in each and every child's education. The nautical photo props will be used with each child as they step foot into our classroom for the first time on Meet the Teacher evening. I'll take pictures of each child with them, have them developed, and then hung in our classroom ready for their first day of 4th grade. This kind gesture will set the tone before even the first day of school! The nautical thank you cards will be used throughout the year by the students as they create thank you cards to their team groups.\r\n\r\nYour generous donations will help me to help make our classroom a fun, inviting, learning environment from day one.\r\n\r\nIt costs lost of money out of my own pocket on resources to get our classroom ready. Please consider helping with this project to make our new school year a very successful one. Thank you!nannan

=====

My kindergarten students have varied disabilities ranging from speech and language delays, cognitive delays, gross/fine motor delays, to autism. They are eager beavers and always strive to work their hardest working past their limitations. \r\n\r\nThe materials we have are the ones I seek out for my students. I teach in a Title I school where most of the students receive free or reduced price lunch. Despite their disabilities and limitations, my students love coming to school and come eager to learn and explore. Have you ever felt like you had ants in your pants and you needed to groove and move as you were in a meeting? This is how my kids feel all the time. They want to be able to move as they learn or so they say. Wobble chairs are the answer and I love them because they develop their core, which enhances gross motor and in turn fine motor skills. \r\n\r\nThey also want to learn through games, my kids don't want to sit and do worksheets. They want to learn to count by jumping and playing. Physical engagement is the key to our success. The number toss and color and shape mats can make that happen. My students will forget they are doing work and just have the fun a 6 year old deserves.nannan

=====

The mediocre teacher tells. The good teacher explains. The superior teacher demonstrates. The great teacher inspires. -William A. Ward\r\n\r\nMy school has 803 students which is makeup is 97.6% African-American, making up the largest segment of the student body. A typical school in Dallas is made up of 23.2% African-American students. Most of the students are on free or reduced lunch. We aren't receiving doctors, lawyers, or engineers children from rich backgrounds or neighborhoods. As an educator I am inspiring minds of young children and we focus not only on academics but one smart, effective, efficient, and disciplined students with good character. In our classroom we can utilize the Bluetooth for swift transitions during class. I use a speaker which doesn't amplify the sound enough to receive the message. Due to the volume of my speaker my students can't hear videos or books clearly and it isn't making the lessons as meaningful. But with the Bluetooth speaker my students will be able to hear and I can stop, pause and replay it at any time.\r\n\r\nThe cart will allow me to have more room for storage of things that are needed for the day and has an extra part to it I can use. The table top chart has all of the letter, words and pictures for students to learn about different letters and it is more accessible.nannan

=====

In [12]:

```
# https://stackoverflow.com/a/47091490/4084039
import re

def decontracted(phrase):
    # specific
    phrase = re.sub(r"won't", "will not", phrase)
    phrase = re.sub(r"can't", "can not", phrase)

    # general
    phrase = re.sub(r"n't", " not", phrase)
    phrase = re.sub(r"\ 're", " are", phrase)
    phrase = re.sub(r"\ 's", " is", phrase)
    phrase = re.sub(r"\ 'd", " would", phrase)
    phrase = re.sub(r"\ 'll", " will", phrase)
    phrase = re.sub(r"\ 't", " not", phrase)
    phrase = re.sub(r"\ 've", " have", phrase)
    phrase = re.sub(r"\ 'm", " am", phrase)
    return phrase
```

In [13]:

```
sent = decontracted(project_data['essay'].values[20000])
print(sent)
print("="*50)
```

My kindergarten students have varied disabilities ranging from speech and language delays, cognitive delays, gross/fine motor delays, to autism. They are eager beavers and always strive to work their hardest working past their limitations. \r\n\r\nThe materials we have are the ones I seek out for my students. I teach in a Title I school where most of the students receive free or reduced price lunch. Despite their disabilities and limitations, my students love coming to school and come eager to learn and explore. Have you ever felt like you had ants in your pants and you needed to groove and move as you were in a meeting? This is how my kids feel all the time. They want to be able to move as they learn or so they say. Wobble chairs are the answer and I love them because they develop their core, which enhances gross motor and in turn fine motor skills. \r\nThey also want to learn through games, my kids do not want to sit and do worksheets. They want to learn to count by jumping and playing. Physical engagement is the key to our success. The number toss and color and shape mats can make that happen. My students will forget they are doing work and just have the fun a 6 year old deserves. \n\n\n

=====

In [14]:

```
# \r \n \t remove from string python: http://texthandler.com/info/remove-line-breaks-python/
sent = sent.replace('\r', ' ')
sent = sent.replace('\n', ' ')
sent = sent.replace('\t', ' ')
print(sent)
```

My kindergarten students have varied disabilities ranging from speech and language delays, cognitive delays, gross/fine motor delays, to autism. They are eager beavers and always strive to work their hardest working past their limitations. The materials we have are the ones I seek out for my students. I teach in a Title I school where most of the students receive free or reduced price lunch. Despite their disabilities and limitations, my students love coming to school and come eager to learn and explore. Have you ever felt like you had ants in your pants and you needed to groove and move as you were in a meeting? This is how my kids feel all the time. They want to be able to move as they learn or so they say. Wobble chairs are the answer and I love them because they develop their core, which enhances gross motor and in turn fine motor skills. They also want to learn through games, my kids do not want to sit and do worksheets. They want to learn to count by jumping and playing. Physical engagement is the key to our success. The number toss and color and shape mats can make that happen. My students will forget they are doing work and just have the fun a 6 year old deserves. nannan

In [15]:

```
#remove special character: https://stackoverflow.com/a/5843547/4084039
sent = re.sub('[^A-Za-z0-9]+', '', sent)
print(sent)
```

My kindergarten students have varied disabilities ranging from speech and language delays cognitive delays gross fine motor delays to autism They are eager beavers and always strive to work their hardest working past their limitations The materials we have are the ones I seek out for my students I teach in a Title I school where most of the students receive free or reduced price lunch Despite their disabilities and limitations my students love coming to school and come eager to learn and explore Have you ever felt like you had ants in your pants and you needed to groove and move as you were in a meeting This is how my kids feel all the time They want to be able to move as they learn or so they say Wobble chairs are the answer and I love them because they develop their core which enhances gross motor and in turn fine motor skills They also want to learn through games my kids do not want to sit and do worksheets They want to learn to count by jumping and playing Physical engagement is the key to our success The number toss and color and shape mats can make that happen My students will forget they are doing work and just have the fun a 6 year old deserves nannan

In [16]:

```
# https://gist.github.com/sebleier/554280
# we are removing the words from the stop words list: 'no', 'nor', 'not'
stopwords= ['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', "you'r
e", "you've",\
            "you'll", "you'd", 'your', 'yours', 'yourself', 'yourselves', 'he', 'him',
'his', 'himself', \
            'she', "she's", 'her', 'hers', 'herself', 'it', "it's", 'its', 'itself', 't
hey', 'them', 'their',\
            'theirs', 'themselves', 'what', 'which', 'who', 'whom', 'this', 'that', "th
at'll", 'these', 'those', \
            'am', 'is', 'are', 'was', 'were', 'be', 'been', 'being', 'have', 'has', 'ha
d', 'having', 'do', 'does', \
            'did', 'doing', 'a', 'an', 'the', 'and', 'but', 'if', 'or', 'because', 'as'
, 'until', 'while', 'of', \
            'at', 'by', 'for', 'with', 'about', 'against', 'between', 'into', 'through'
, 'during', 'before', 'after',\
            'above', 'below', 'to', 'from', 'up', 'down', 'in', 'out', 'on', 'off', 'ov
er', 'under', 'again', 'further',\
            'then', 'once', 'here', 'there', 'when', 'where', 'why', 'how', 'all', 'an
y', 'both', 'each', 'few', 'more',\
            'most', 'other', 'some', 'such', 'only', 'own', 'same', 'so', 'than', 'too'
, 'very', \
            's', 't', 'can', 'will', 'just', 'don', "don't", 'should', "should've", 'no
w', 'd', 'll', 'm', 'o', 're', \
            've', 'y', 'ain', 'aren', "aren't", 'couldn', "couldn't", 'didn', "didn't",
'doesn', "doesn't", 'hadn',\
            "hadn't", 'hasn', "hasn't", 'haven', "haven't", 'isn', "isn't", 'ma', 'migh
tn', "mighntn't", 'mustn',\
            "mustn't", 'needn', "needn't", 'shan', "shan't", 'shouldn', "shouldn't", 'w
asn', "wasn't", 'weren', "weren't", \
            'won', "won't", 'wouldn', "wouldn't"]
```

In [17]:

```
# Combining all the above stundents
from tqdm import tqdm
preprocessed_essays = []
# tqdm is for printing the status bar
for sentence in tqdm(project_data['essay'].values):
    sent = decontracted(sentence)
    sent = sent.replace('\r', ' ')
    sent = sent.replace('\n', ' ')
    sent = sent.replace('\n', ' ')
    sent = re.sub('[^A-Za-z0-9]+', ' ', sent)
    # https://gist.github.com/sebleier/554280
    sent = ' '.join(e for e in sent.split() if e not in stopwords)
    preprocessed_essays.append(sent.lower().strip())
```

```
100%|████████████████████████████████████████████████████████████████████████████████|
██████| 109248/109248 [01:42<00:00, 1069.63it/s]
```





In [21]:

```
project_data.columns
```

Out[21]:

```
Index(['Unnamed: 0', 'id', 'teacher_id', 'teacher_prefix', 'school_state',
      'project_submitted_datetime', 'project_title', 'project_essay_1',
      'project_essay_2', 'project_essay_3', 'project_essay_4',
      'project_resource_summary',
      'teacher_number_of_previously_posted_projects', 'project_is_approved',
      'clean_categories', 'clean_subcategories', 'clean_grades', 'essay'],
      dtype='object')
```

we are going to consider

- school\_state : categorical data
- clean\_categories : categorical data
- clean\_subcategories : categorical data
- project\_grade\_category : categorical data
- teacher\_prefix : categorical data
- project\_title : text data
- text : text data
- project\_resource\_summary: text data (optional)
- quantity : numerical (optional)
- teacher\_number\_of\_previously\_posted\_projects : numerical
- price : numerical

### 1.5.1 Vectorizing Categorical data

- <https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/handling-categorical-and-numerical-features/> (<https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/handling-categorical-and-numerical-features/>)

In [22]:

```
# we use count vectorizer to convert the values into one
from sklearn.feature_extraction.text import CountVectorizer
vectorizer = CountVectorizer(vocabulary=list(sorted_cat_dict.keys()), lowercase=False,
                             binary=True)
categories_one_hot = vectorizer.fit_transform(project_data['clean_categories'].values)
print(vectorizer.get_feature_names())
print("Shape of matrix after one hot encoding ", categories_one_hot.shape)
```

```
['Warmth', 'Care_Hunger', 'History_Civics', 'Music_Arts', 'AppliedLearning',
 'SpecialNeeds', 'Health_Sports', 'Math_Science', 'Literacy_Language']
Shape of matrix after one hot encoding (109248, 9)
```

In [23]:

```
# we use count vectorizer to convert the values into one
vectorizer = CountVectorizer(vocabulary=list(sorted_sub_cat_dict.keys()), lowercase=False, binary=True)
sub_categories_one_hot = vectorizer.fit_transform(project_data['clean_subcategories'].values)
print(vectorizer.get_feature_names())
print("Shape of matrix after one hot encoding ", sub_categories_one_hot.shape)
```

```
['Economics', 'CommunityService', 'FinancialLiteracy', 'ParentInvolvement', 'Extracurricular', 'Civics_Government', 'ForeignLanguages', 'NutritionEducation', 'Warmth', 'Care_Hunger', 'SocialSciences', 'PerformingArts', 'CharacterEducation', 'TeamSports', 'Other', 'College_CareerPrep', 'Music', 'History_Geography', 'Health_LifeScience', 'EarlyDevelopment', 'ESL', 'Gym_Fitness', 'EnvironmentalScience', 'VisualArts', 'Health_Wellness', 'AppliedSciences', 'SpecialNeeds', 'Literature_Writing', 'Mathematics', 'Literacy']
Shape of matrix after one hot encoding (109248, 30)
```

In [0]:

```
# you can do the similar thing with state, teacher_prefix and project_grade_category also
```

In [24]:

```
#vectorizing student state
vectorizer=CountVectorizer()
school_state_one_hot=vectorizer.fit_transform(project_data['school_state'].values)
print(vectorizer.get_feature_names())
print('shape of matrix after one hot encoding', school_state_one_hot.shape)
```

```
['ak', 'al', 'ar', 'az', 'ca', 'co', 'ct', 'dc', 'de', 'fl', 'ga', 'hi', 'ia', 'id', 'il', 'in', 'ks', 'ky', 'la', 'ma', 'md', 'me', 'mi', 'mn', 'mo', 'ms', 'mt', 'nc', 'nd', 'ne', 'nh', 'nj', 'nm', 'nv', 'ny', 'oh', 'ok', 'or', 'pa', 'ri', 'sc', 'sd', 'tn', 'tx', 'ut', 'va', 'vt', 'wa', 'wi', 'wv', 'wy']
shape of matrix after one hot encoding (109248, 51)
```

In [25]:

```
#vectorizing project grade category
vectorizer=CountVectorizer()
project_grade_one_hot=vectorizer.fit_transform(project_data['clean_grades'].values)
print(vectorizer.get_feature_names())
print('shape of matrix after one hot encoding', project_grade_one_hot.shape)
```

```
['grades_3_5', 'grades_6_8', 'grades_9_12', 'grades_prek_2']
shape of matrix after one hot encoding (109248, 4)
```

In [26]:

```
#vectorizing teacher prefix
x=project_data['teacher_prefix'].fillna('')
vectorizer = CountVectorizer()

teacher_prefix_one_hot = vectorizer.fit_transform(x.values)
print(vectorizer.get_feature_names())
print("Shape of matrix after one hot encodig ",teacher_prefix_one_hot.shape)
```

```
['dr', 'mr', 'mrs', 'ms', 'teacher']
Shape of matrix after one hot encodig (109248, 5)
```

## 1.5.2 Vectorizing Text data

### 1.5.2.1 Bag of words

In [27]:

```
# We are considering only the words which appeared in at least 10 documents(rows or projects).
vectorizer = CountVectorizer(min_df=10)
text_bow = vectorizer.fit_transform(preprocessed_essays)
print("Shape of matrix after one hot encodig ",text_bow.shape)
```

```
Shape of matrix after one hot encodig (109248, 16623)
```

In [28]:

```
# you can vectorize the title also
# before you vectorize the title make sure you preprocess it
vectorizer=CountVectorizer(min_df=10)
title_bow=vectorizer.fit_transform(preprocessed_title)
print('Shape of matrix after vectorizing',title_bow.shape)
```

```
Shape of matrix after vectorizing (109248, 3222)
```

### 1.5.2.2 TFIDF vectorizer

In [29]:

```
from sklearn.feature_extraction.text import TfidfVectorizer
vectorizer = TfidfVectorizer(min_df=10)
text_tfidf = vectorizer.fit_transform(preprocessed_essays)
print("Shape of matrix after one hot encodig ",text_tfidf.shape)
```

```
Shape of matrix after one hot encodig (109248, 16623)
```

In [30]:

```
vectorizer=TfidfVectorizer(min_df=10)
title_tfidf=vectorizer.fit_transform(preprocessed_title)
print('Shape of matrix after vectorizing',title_tfidf.shape)
```

```
Shape of matrix after vectorizing (109248, 3222)
```

### 1.5.2.3 Using Pretrained Models: Avg W2V

In [0]:

```

...
# Reading glove vectors in python: https://stackoverflow.com/a/38230349/4084039
def loadGloveModel(gloveFile):
    print ("Loading Glove Model")
    f = open(gloveFile,'r', encoding="utf8")
    model = {}
    for line in tqdm(f):
        splitLine = line.split()
        word = splitLine[0]
        embedding = np.array([float(val) for val in splitLine[1:]])
        model[word] = embedding
    print ("Done.",len(model)," words loaded!")
    return model
model = loadGloveModel('glove.42B.300d.txt')

# =====
Output:

Loading Glove Model
1917495it [06:32, 4879.69it/s]
Done. 1917495 words loaded!

# =====

words = []
for i in preproced_texts:
    words.extend(i.split(' '))

for i in preproced_titles:
    words.extend(i.split(' '))
print("all the words in the coupus", len(words))
words = set(words)
print("the unique words in the coupus", len(words))

inter_words = set(model.keys()).intersection(words)
print("The number of words that are present in both glove vectors and our coupus", \
      len(inter_words), "(", np.round(len(inter_words)/len(words)*100,3), "%)")

words_courpus = {}
words_glove = set(model.keys())
for i in words:
    if i in words_glove:
        words_courpus[i] = model[i]
print("word 2 vec length", len(words_courpus))

# stronging variables into pickle files python: http://www.jessicayung.com/how-to-use-pickle-to-save-and-load-variables-in-python/

import pickle
with open('glove_vectors', 'wb') as f:
    pickle.dump(words_courpus, f)

...

```

```
'\n# Reading glove vectors in python: https://stackoverflow.com/a/3823034
9/4084039\ndef loadGloveModel(gloveFile):\n    print ("Loading Glove Mode
l")\n    f = open(gloveFile,\'r\', encoding="utf8")\n    model = {}\n    f
or line in tqdm(f):\n        splitLine = line.split()\n        word = spli
tline[0]\n        embedding = np.array([float(val) for val in splitLine
[1:]])\n        model[word] = embedding\n        print ("Done.",len(model)," w
ords loaded!")\n    return model\nmodel = loadGloveModel(\'glove.42B.300d.
txt\')\n\n# =====\n\nOutput:\n    \nLoading Glove Mod
el\n1917495it [06:32, 4879.69it/s]\nDone. 1917495 words loaded!\n\n# ====
=====
\n\nwords = []\nfor i in preproced_texts:\n    wor
ds.extend(i.split(\' \'))\n\nfor i in preproced_titles:\n    words.extend
(i.split(\' \'))\n\nprint("all the words in the coupus", len(words))\nwords
= set(words)\n\nprint("the unique words in the coupus", len(words))\n\ninter
_words = set(model.keys()).intersection(words)\n\nprint("The number of words
that are present in both glove vectors and our coupus", len(inter_wo
rds), "(" ,np.round(len(inter_words)/len(words)*100,3), "%")\n\nwords_courpu
s = {}\nwords_glove = set(model.keys())\nfor i in words:\n    if i in word
s_glove:\n        words_courpus[i] = model[i]\n\nprint("word 2 vec length",
len(words_courpus))\n\n\n# stronging variables into pickle files python: h
ttp://www.jessicayung.com/how-to-use-pickle-to-save-and-load-variables-in-
python/\n\nimport pickle\nwith open(\'glove_vectors\', \'wb\') as f:\n
pickle.dump(words courpus, f)\n\n\n'
```

```
# stronging variables into pickle files python: http://www.jessicayung.com/how-to-use-pickle-to-save-and-load-variables-in-python/
# make sure you have the glove_vectors file
with open('glove_vectors', 'rb') as f:
    model = pickle.load(f)
    glove words = set(model.keys())
```

```
# Average Word2Vec
# compute average word2vec for each review.
avg_w2v_vectors = []; # the avg-w2v for each sentence/review is stored in this list
for sentence in tqdm(preprocessed_essays): # for each review/sentence
    vector = np.zeros(300) # as word vectors are of zero length
    cnt_words = 0; # num of words with a valid vector in the sentence/review
    for word in sentence.split(): # for each word in a review/sentence
        if word in glove_words:
            vector += model[word]
            cnt_words += 1
    if cnt_words != 0:
        vector /= cnt_words
    avg_w2v_vectors.append(vector)

print(len(avg_w2v_vectors))
print(len(avg_w2v_vectors[0]))
```

```
100%|██████████| 109248/109248 [00:59<00:00, 1823.16it/s]
```



In [23]:

```
data1=project_data.drop(['id','teacher_id','project_essay_1','project_essay_2','project_essay_3','project_essay_4','project_is_approved'],axis=1)
data1.head(2)
data=data1[0:100000]
data[0:2]
```

Out[23]:

	Unnamed: 0	teacher_prefix	school_state	project_submitted_datetime	project_title
0	160221	Mrs.	IN	2016-12-05 13:43:57	Educational Support for English Learners at Home
1	140945	Mr.	FL	2016-10-25 09:22:10	Wanted: Projector for Hungry Learners

In [24]:

```
# check this one: https://www.youtube.com/watch?v=0H0qOcln3Z4&t=530s
# standardization sklearn: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html
from sklearn.preprocessing import StandardScaler

# price_standardized = standardScalar.fit(project_data['price'].values)
# this will rise the error
# ValueError: Expected 2D array, got 1D array instead: array=[725.05 213.03 329. ... 399. 287.73 5.5 ].
# Reshape your data either using array.reshape(-1, 1)

price_scalar = StandardScaler()
price_scalar.fit(project_data['price'].values.reshape(-1,1)) # finding the mean and standard deviation of this data
print(f"Mean : {price_scalar.mean_[0]}, Standard deviation : {np.sqrt(price_scalar.var_[0])}")

# Now standardize the data with above mean and variance.
price_standardized = price_scalar.transform(project_data['price'].values.reshape(-1, 1))
```

Mean : 298.1193425966608, Standard deviation : 367.49634838483496



In [25]:

```
price_standardized
```

Out[25]:

```
array([[ -0.3905327 ],
       [  0.00239637],
       [  0.59519138],
       ...,
       [-0.15825829],
       [-0.61243967],
       [-0.51216657]])
```

In [26]:

```
projects_scalar = StandardScaler()
projects_scalar.fit(project_data['teacher_number_of_previously_posted_projects'].values
.reshape(-1,1)) # finding the mean and standard deviation of this data
print(f"Mean : {price_scalar.mean_[0]}, Standard deviation : {np.sqrt(price_scalar.var_
[0])}")

# Now standardize the data with above maen and variance.
projects_standardized = projects_scalar.transform(project_data['teacher_number_of_previ
ously_posted_projects'].values.reshape(-1,1))
projects_standardized
```

C:\Users\HP\AppData\Local\Continuum\anaconda3\lib\site-packages\sklearn\utils\validation.py:475: DataConversionWarning:

Data with input dtype int64 was converted to float64 by StandardScaler.

Mean : 298.1193425966608, Standard deviation : 367.49634838483496

C:\Users\HP\AppData\Local\Continuum\anaconda3\lib\site-packages\sklearn\utils\validation.py:475: DataConversionWarning:

Data with input dtype int64 was converted to float64 by StandardScaler.

Out[26]:

```
array([[ -0.40152481],
       [-0.14951799],
       [-0.36552384],
       ...,
       [-0.29352189],
       [-0.40152481],
       [-0.40152481]])
```

In [27]:

```
projects_scalar = StandardScaler()
projects_scalar.fit(project_data['quantity'].values.reshape(-1,1)) # finding the mean and standard deviation of this data
print(f"Mean : {price_scalar.mean_[0]}, Standard deviation : {np.sqrt(price_scalar.var_[0])}")

# Now standardize the data with above mean and variance.
quantity_standardized = projects_scalar.transform(project_data['quantity'].values.reshape(-1,1))
quantity_standardized
```

C:\Users\HP\AppData\Local\Continuum\anaconda3\lib\site-packages\sklearn\utils\validation.py:475: DataConversionWarning:

Data with input dtype int64 was converted to float64 by StandardScaler.

Mean : 298.1193425966608, Standard deviation : 367.49634838483496

C:\Users\HP\AppData\Local\Continuum\anaconda3\lib\site-packages\sklearn\utils\validation.py:475: DataConversionWarning:

Data with input dtype int64 was converted to float64 by StandardScaler.

Out[27]:

```
array([[ 0.23047132],
       [-0.60977424],
       [ 0.19227834],
       ...,
       [-0.4951953 ],
       [-0.03687954],
       [-0.45700232]])
```

## 1.5.4 Merging all the above features

- we need to merge all the numerical vectors i.e categorical, text, numerical vectors

In [36]:

```
print(categories_one_hot.shape)
print(sub_categories_one_hot.shape)
print(teacher_prefix_one_hot.shape)
print(school_state_one_hot.shape)
print(project_grade_one_hot.shape)
print(title_bow.shape)
print(text_bow.shape)
print(price_standardized.shape)
print(projects_standardized.shape)
```

```
(109248, 9)
(109248, 30)
(109248, 5)
(109248, 51)
(109248, 4)
(109248, 3222)
(109248, 16623)
(109248, 1)
(109248, 1)
```

In [37]:

```
# merge two sparse matrices: https://stackoverflow.com/a/19710648/4084039
from scipy.sparse import hstack
# with the same hstack function we are concatenating a sparse matrix and a dense matrix
:
X = hstack((categories_one_hot, sub_categories_one_hot, text_bow, price_standardized))
X.shape
```

Out[37]:

```
(109248, 16663)
```

In [28]:

```
y1=project_data['project_is_approved']
print(y1.shape)
y=y1[0:100000]
```

```
(109248,)
```

In [0]:

```
# please write all the code with proper documentation, and proper titles for each subsection
# when you plot any graph make sure you use
    # a. Title, that describes your plot, this will be very helpful to the reader
    # b. Legends if needed
    # c. X-axis label
    # d. Y-axis label
```

## Computing Sentiment Scores

In [0]:

```
import nltk
from nltk.sentiment.vader import SentimentIntensityAnalyzer

# import nltk
# nltk.download('vader_lexicon')

sid = SentimentIntensityAnalyzer()

for_sentiment = 'a person is a person no matter how small dr seuss i teach the smallest
students with the biggest enthusiasm \
for learning my students learn in many different ways using all of our senses and multi
ple intelligences i use a wide range\
of techniques to help all my students succeed students in my class come from a variety
of different backgrounds which makes\
for wonderful sharing of experiences and cultures including native americans our school
is a caring community of successful \
learners which can be seen through collaborative student project based learning in and
out of the classroom kindergarteners \
in my class love to work with hands on materials and have many different opportunities
to practice a skill before it is\
mastered having the social skills to work cooperatively with friends is a crucial aspec
t of the kindergarten curriculum\
montana is the perfect place to learn about agriculture and nutrition my students love
to role play in our pretend kitchen\
in the early childhood classroom i have had several kids ask me can we try cooking with
real food i will take their idea \
and create common core cooking lessons where we learn important math and writing concep
ts while cooking delicious healthy \
food for snack time my students will have a grounded appreciation for the work that wen
t into making the food and knowledge \
of where the ingredients came from as well as how it is healthy for their bodies this p
roject would expand our learning of \
nutrition and agricultural cooking recipes by having us peel our own apples to make hom
emade applesauce make our own bread \
and mix up healthy plants from our classroom garden in the spring we will also create o
ur own cookbooks to be printed and \
shared with families students will gain math and literature skills as well as a life lo
ng enjoyment for healthy cooking \
nannan'
ss = sid.polarity_scores(for_sentiment)

for k in ss:
    print('{0}: {1}, '.format(k, ss[k]), end='')

# we can use these 4 things as features/attributes (neg, neu, pos, compound)
# neg: 0.0, neu: 0.753, pos: 0.247, compound: 0.93
```

D:\installed\Anaconda3\lib\site-packages\nltk\twitter\\_\_init\_\_.py:20: User Warning:

The twython library has not been installed. Some functionality from the tw  
itter package will not be available.

neg: 0.01, neu: 0.745, pos: 0.245, compound: 0.9975,

# Assignment 10: Clustering

- step 1: Choose any vectorizer (data matrix) that you have worked in any of the assignments, and got the best AUC value.
- step 2: Choose any of the [feature selection](https://scikit-learn.org/stable/modules/feature_selection.html) ([https://scikit-learn.org/stable/modules/feature\\_selection.html](https://scikit-learn.org/stable/modules/feature_selection.html))/[reduction algorithms](https://scikit-learn.org/stable/modules/decomposition.html) (<https://scikit-learn.org/stable/modules/decomposition.html>) ex: selectkbest features, pretrained word vectors, model based feature selection etc and reduce the number of features to 5k features.
- step 3: Apply all three kmeans, Agglomerative clustering, DBSCAN
  - **K-Means Clustering:**
    - Find the best 'k' using the elbow-knee method (plot k vs inertia\_)
  - **Agglomerative Clustering:**
    - Apply [agglomerative algorithm](https://stackabuse.com/hierarchical-clustering-with-python-and-scikit-learn/) (<https://stackabuse.com/hierarchical-clustering-with-python-and-scikit-learn/>) and try a different number of clusters like 2,5 etc.
    - As this is very computationally expensive, take **5k** datapoints only to perform hierarchical clustering because they do take a considerable amount of time to run.
  - **DBSCAN Clustering:**
    - Find the best 'eps' using the [elbow-knee method](https://stackoverflow.com/a/48558030/4084039) (<https://stackoverflow.com/a/48558030/4084039>).
    - Take **5k** datapoints only.
- step 4: Summarize each cluster by manually observing few points from each cluster.
- step 5: You need to plot the word cloud with essay text for each cluster for each of algorithms mentioned in step 3.

In [31]:

```
from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test=train_test_split(data,y,test_size=0.30,stratify=y)
X_train,X_cv,y_train,y_cv=train_test_split(X_train,y_train,test_size=0.30,stratify=y_train)
```

## 2. Clustering

### 2.1 Choose the best data matrix on which you got the best AUC

In [0]:

```
# please write all the code with proper documentation, and proper titles for each subsection
# go through documentations and blogs before you start coding
# first figure out what to do, and then think about how to do.
# reading and understanding error messages will be very much helpfull in debugging your code
# when you plot any graph make sure you use
    # a. Title, that describes your plot, this will be very helpful to the reader
    # b. Legends if needed
    # c. X-axis label
    # d. Y-axis label
```

## 2.2 Make Data Model Ready: encoding numerical, categorical features

In [0]:

```
# please write all the code with proper documentation, and proper titles for each subsection
# go through documentations and blogs before you start coding
# first figure out what to do, and then think about how to do.
# reading and understanding error messages will be very much helpfull in debugging your code
# make sure you featurize train and test data separatly

# when you plot any graph make sure you use
    # a. Title, that describes your plot, this will be very helpful to the reader
    # b. Legends if needed
    # c. X-axis label
    # d. Y-axis label
```

In [32]:

```
statevectorizer=CountVectorizer()
statevectorizer.fit(X_train['school_state'].values)

X_train_state_encoded=statevectorizer.transform(X_train['school_state'].values)
X_cv_state_encoded=statevectorizer.transform(X_cv['school_state'].values)
X_test_state_encoded=statevectorizer.transform(X_test['school_state'].values)

print("AFTER VECTORIZATION")
print('='*50)
print(X_train_state_encoded.shape,y_train.shape)
print(X_cv_state_encoded.shape,y_cv.shape)
print(X_test_state_encoded.shape,y_test.shape)
```

AFTER VECTORIZATION

```
=====
(49000, 51) (49000,)
(21000, 51) (21000,)
(30000, 51) (30000,)
```

In [33]:

```
X_train['teacher_prefix'].unique()
```

Out[33]:

```
array(['Ms.', 'Mrs.', 'Mr.', 'Teacher', 'Dr.', nan], dtype=object)
```

In [34]:

```
X_train['teacher_prefix'].fillna('',inplace=True)
X_cv['teacher_prefix'].fillna('',inplace=True)
X_test['teacher_prefix'].fillna('',inplace=True)
```

In [35]:

```
prefixvectorizer=CountVectorizer()
prefixvectorizer.fit(X_train['teacher_prefix'].values)

X_train_prefix_encoded=prefixvectorizer.transform(X_train['teacher_prefix'].values)
X_cv_prefix_encoded=prefixvectorizer.transform(X_cv['teacher_prefix'].values)
X_test_prefix_encoded=prefixvectorizer.transform(X_test['teacher_prefix'].values)
print('AFTER VECTORIZATION')
print('='*50)
print(prefixvectorizer.get_feature_names())

print(X_train_prefix_encoded.shape,y_train.shape)
print(X_cv_prefix_encoded.shape,y_cv.shape)
print(X_test_prefix_encoded.shape,y_test.shape)
```

AFTER VECTORIZATION

```
=====
['dr', 'mr', 'mrs', 'ms', 'teacher']
(49000, 5) (49000,)
(21000, 5) (21000,)
(30000, 5) (30000,)
```

In [36]:

```
#encoding grade category
gradevectorizer=CountVectorizer()
gradevectorizer.fit(X_train['clean_grades'].values)

X_train_grade_encoded=gradevectorizer.transform(X_train['clean_grades'].values)
X_cv_grade_encoded=gradevectorizer.transform(X_cv['clean_grades'].values)
X_test_grade_encoded=gradevectorizer.transform(X_test['clean_grades'].values)

print("AFTER VECTORIZATION")
print('='*50)
print(gradevectorizer.get_feature_names())
print(X_train_grade_encoded.shape,y_train.shape)
print(X_cv_grade_encoded.shape,y_cv.shape)
print(X_test_grade_encoded.shape,y_test.shape)
```

AFTER VECTORIZATION

```
=====
['grades_3_5', 'grades_6_8', 'grades_9_12', 'grades_prek_2']
(49000, 4) (49000,)
(21000, 4) (21000,)
(30000, 4) (30000,)
```

In [37]:

```
#encoding clean category
categoryvectorizer=CountVectorizer()
categoryvectorizer.fit(X_train['clean_categories'].values)

X_train_category_encoded=categoryvectorizer.transform(X_train['clean_categories'].values)
X_cv_category_encoded=categoryvectorizer.transform(X_cv['clean_categories'].values)
X_test_category_encoded=categoryvectorizer.transform(X_test['clean_categories'].values)

print('AFTER VECTORIZATION')
print('='*50)
print(categoryvectorizer.get_feature_names())
print(X_train_category_encoded.shape,y_train.shape)
print(X_cv_category_encoded.shape,y_cv.shape)
print(X_test_category_encoded.shape,y_test.shape)
```

AFTER VECTORIZATION

```
=====
['appliedlearning', 'care_hunger', 'health_sports', 'history_civics', 'literacy_language', 'math_science', 'music_arts', 'specialneeds', 'warmth']
(49000, 9) (49000,)
(21000, 9) (21000,)
(30000, 9) (30000,)
```

In [38]:

```
#encoding subcategories
subcategoryvectorizer=CountVectorizer()
subcategoryvectorizer.fit(X_train['clean_subcategories'].values)
X_train_subcategories_encoded=subcategoryvectorizer.transform(X_train['clean_subcategories'].values)
X_cv_subcategories_encoded=subcategoryvectorizer.transform(X_cv['clean_subcategories'].values)
X_test_subcategories_encoded=subcategoryvectorizer.transform(X_test['clean_subcategories'].values)
print("AFTER VECTORIZATION")
print('='*50)
print(subcategoryvectorizer.get_feature_names())
print(X_train_subcategories_encoded.shape,y_train.shape)
print(X_cv_subcategories_encoded.shape,y_cv.shape)
print(X_test_subcategories_encoded.shape,y_test.shape)
```

AFTER VECTORIZATION

```
=====
['appliedsciences', 'care_hunger', 'charactereducation', 'civics_government', 'college_careerprep', 'communityservice', 'earlydevelopment', 'economics', 'environmentalscience', 'esl', 'extracurricular', 'financialliteracy', 'foreignlanguages', 'gym_fitness', 'health_lifescience', 'health_wellness', 'history_geography', 'literacy', 'literature_writing', 'mathematics', 'music', 'nutritioneducation', 'other', 'parentinvolvement', 'performingarts', 'socialsciences', 'specialneeds', 'teamsports', 'visualarts', 'warmth']
(49000, 30) (49000,)
(21000, 30) (21000,)
(30000, 30) (30000,)
```



In [39]:

```
#encoding numerical categories---price
from sklearn.preprocessing import Normalizer
normalizer=Normalizer()
normalizer.fit(X_train['price'].values.reshape(1,-1))

X_train_price_norm=normalizer.transform(X_train['price'].values.reshape(1,-1))
X_cv_price_norm=normalizer.transform(X_cv['price'].values.reshape(1,-1))
X_test_price_norm=normalizer.transform(X_test['price'].values.reshape(1,-1))

print("after vectorization")
print(X_train_price_norm.shape,y_train.shape)
print(X_cv_price_norm.shape,y_cv.shape)
print(X_test_price_norm.shape,y_test.shape)
```

```
after vectorization
(1, 49000) (49000,)
(1, 21000) (21000,)
(1, 30000) (30000,)
```

In [40]:

```
price_train_norm=X_train_price_norm.reshape(49000,1)
price_cv_norm=X_cv_price_norm.reshape(21000,1)
price_test_norm=X_test_price_norm.reshape(30000,1)
print(price_train_norm.shape)
print(price_cv_norm.shape)
print(price_test_norm.shape)
```

```
(49000, 1)
(21000, 1)
(30000, 1)
```

In [41]:

```
#encoding numerical category quantity
normalizer=Normalizer()
normalizer.fit(X_train['quantity'].values.reshape(1,-1))

X_train_quantity_norm=normalizer.transform(X_train['quantity'].values.reshape(1,-1))
X_cv_quantity_norm=normalizer.transform(X_cv['quantity'].values.reshape(1,-1))
X_test_quantity_norm=normalizer.transform(X_test['quantity'].values.reshape(1,-1))

print('after vectorization')
print(X_train_quantity_norm.shape,y_train.shape)
print(X_cv_quantity_norm.shape,y_cv.shape)
print(X_test_quantity_norm.shape,y_test.shape)
```

```
after vectorization
(1, 49000) (49000,)
(1, 21000) (21000,)
(1, 30000) (30000,)
```

In [42]:

```
train_quantity_norm=X_train_quantity_norm.reshape(49000,1)
cv_quantity_norm=X_cv_quantity_norm.reshape(21000,1)
test_quantity_norm=X_test_quantity_norm.reshape(30000,1)
print(train_quantity_norm.shape)
print(cv_quantity_norm.shape)
print(test_quantity_norm.shape)
```

```
(49000, 1)
(21000, 1)
(30000, 1)
```

In [43]:

```
#encoding previous projects posted by teachers
normalizer=Normalizer()
normalizer.fit(X_train['teacher_number_of_previously_posted_projects'].values.reshape(1,-1))

X_train_projects_norm=normalizer.transform(X_train['teacher_number_of_previously_posted_projects'].values.reshape(1,-1))
X_cv_projects_norm=normalizer.transform(X_cv['teacher_number_of_previously_posted_projects'].values.reshape(1,-1))
X_test_projects_norm=normalizer.transform(X_test['teacher_number_of_previously_posted_projects'].values.reshape(1,-1))

print("after vectorization")
print(X_train_projects_norm.shape,y_train.shape)
print(X_cv_projects_norm.shape,y_cv.shape)
print(X_test_projects_norm.shape,y_test.shape)
```

```
after vectorization
(1, 49000) (49000,)
(1, 21000) (21000,)
(1, 30000) (30000,)
```

In [44]:

```
projects_train_norm=X_train_projects_norm.reshape(49000,1)
projects_cv_norm=X_cv_projects_norm.reshape(21000,1)
projects_test_norm=X_test_projects_norm.reshape(30000,1)
print(price_train_norm.shape)
print(price_cv_norm.shape)
print(price_test_norm.shape)
```

```
(49000, 1)
(21000, 1)
(30000, 1)
```

## 2.3 Make Data Model Ready: encoding eassay, and project\_title

In [45]:

```
# please write all the code with proper documentation, and proper titles for each subsection
# go through documentations and blogs before you start coding
# first figure out what to do, and then think about how to do.
# reading and understanding error messages will be very much helpfull in debugging your code
# make sure you featurize train and test data separatly

# when you plot any graph make sure you use
    # a. Title, that describes your plot, this will be very helpful to the reader
    # b. Legends if needed
    # c. X-axis label
    # d. Y-axis label
```

## 2.4 Dimensionality Reduction on the selected features

In [0]:

```
# please write all the code with proper documentation, and proper titles for each subsection
# go through documentations and blogs before you start coding
# first figure out what to do, and then think about how to do.
# reading and understanding error messages will be very much helpfull in debugging your code
# when you plot any graph make sure you use
    # a. Title, that describes your plot, this will be very helpful to the reader
    # b. Legends if needed
    # c. X-axis label
    # d. Y-axis label
```

In [46]:

```
essaybowvectorizer=TfidfVectorizer(min_df=10,ngram_range=(1,1))
essaybowvectorizer.fit(X_train['essay'].values)
X_train_essay=essaybowvectorizer.transform(X_train['essay'].values)
#print(X_train_essay_bow.shape)
X_cv_essay=essaybowvectorizer.transform(X_cv["essay"].values)
X_test_essay=essaybowvectorizer.transform(X_test['essay'].values)

print('AFTER VECTORIZATION')
print('='*50)
print(X_train_essay.shape, y_train.shape)
print(X_cv_essay.shape, y_cv.shape)
print(X_test_essay.shape, y_test.shape)
```

AFTER VECTORIZATION

```
=====
(49000, 12543) (49000,)
(21000, 12543) (21000,)
(30000, 12543) (30000,)
```

In [47]:

```
#encoding project title
titlevectorizer=TfidfVectorizer(min_df=10,ngram_range=(1,1))
titlevectorizer.fit(X_train['project_title'].values)
X_train_title=titlevectorizer.transform(X_train['project_title'].values)
X_cv_title=titlevectorizer.transform(X_cv['project_title'].values)
X_test_title=titlevectorizer.transform(X_test['project_title'].values)
print("after vectorization")
print(X_train_title.shape,y_train.shape)
print(X_cv_title.shape,y_cv.shape)
print(X_test_title.shape,y_test.shape)
```

```
after vectorization
(49000, 2093) (49000,)
(21000, 2093) (21000,)
(30000, 2093) (30000,)
```

In [79]:

```
from scipy.sparse import hstack
final_train=hstack((X_train_state_encoded,X_train_prefix_encoded,X_train_grade_encoded,
X_train_category_encoded,X_train_subcategories_encoded,price_train_norm,projects_train_
norm,train_quantity_norm,X_train_essay,X_train_title)).tocsr()
final_cv=hstack((X_cv_state_encoded,X_cv_prefix_encoded,X_cv_grade_encoded,X_cv_categor
y_encoded,X_cv_subcategories_encoded,price_cv_norm,projects_cv_norm,cv_quantity_norm,X_
cv_essay,X_cv_title)).tocsr()
final_test=hstack((X_test_state_encoded,X_test_prefix_encoded,X_test_grade_encoded,X_te
st_category_encoded,X_test_subcategories_encoded,price_test_norm,projects_test_norm,tes
t_quantity_norm,X_test_essay,X_test_title)).tocsr()
print(final_train.shape,y_train.shape)
print(final_cv.shape,y_cv.shape)
print(final_test.shape,y_test.shape)
```

```
(49000, 14738) (49000,)
(21000, 14738) (21000,)
(30000, 14738) (30000,)
```

## DIMENSIONALITY REDUCTION

In [80]:

```
from sklearn.feature_selection import SelectKBest, chi2
final_train_5000=SelectKBest(chi2,k=5000).fit_transform(final_train,y_train)
final_train_5000.shape
```

Out[80]:

```
(49000, 5000)
```

## 2.5 Apply Kmeans

In [0]:

```
# please write all the code with proper documentation, and proper titles for each subsection
# go through documentations and blogs before you start coding
# first figure out what to do, and then think about how to do.
# reading and understanding error messages will be very much helpfull in debugging your code
# when you plot any graph make sure you use
# a. Title, that describes your plot, this will be very helpful to the reader
# b. Legends if needed
# c. X-axis label
# d. Y-axis label
```

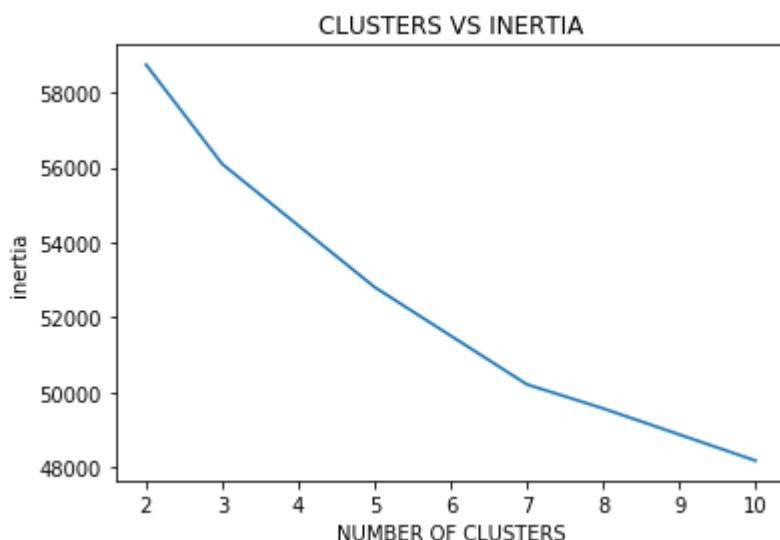
In [52]:

```
from sklearn.cluster import KMeans

clusters=[2,3,5,7,8,10]
k=[]
for i in clusters:
    kmeans=KMeans(n_clusters=i,init='k-means++', n_init=10, max_iter=100, tol=0.0001, p
recompute_distances='auto',n_jobs=-1)
    kmeans.fit(final_train_5000[0:10000],y_train[0:10000])
    k.append(kmeans.inertia_)
plt.plot(clusters,k)
plt.xlabel('NUMBER OF CLUSTERS')
plt.ylabel('inertia')
plt.title('CLUSTERS VS INERTIA')
plt.show
```

Out[52]:

```
<function matplotlib.pyplot.show(*args, **kw)>
```



OBSERVATIONS: Here for k means clustering we took tfidf vectorizer and 10k datapoints. We tried for different number of clusters and we plotted the error and we chose our number of clusters to be 5.

In [53]:

```
kmeans=KMeans(n_clusters=5,init='k-means++', n_init=10, max_iter=100, tol=0.0001, precompute_distances='auto',n_jobs=-1)
kmeans.fit(final_train_5000[0:10000],y_train[0:10000])
```

Out[53]:

```
KMeans(algorithm='auto', copy_x=True, init='k-means++', max_iter=100,
       n_clusters=5, n_init=10, n_jobs=-1, precompute_distances='auto',
       random_state=None, tol=0.0001, verbose=0)
```

In [54]:

```
kmeans.cluster_centers_
```

Out[54]:

```
array([[0.00074627, 0.02089552, 0.00522388, ..., 0.          , 0.          ,
        0.00094287],
       [0.00299529, 0.01968335, 0.01497647, ..., 0.          , 0.          ,
        0.00089277],
       [0.00406298, 0.01117318, 0.00406298, ..., 0.00090328, 0.00020248,
        0.00105266],
       [0.00389273, 0.01470588, 0.0116782 , ..., 0.00103517, 0.00068076,
        0.00021312],
       [0.00391773, 0.01469148, 0.00881489, ..., 0.0005562 , 0.          ,
        0.00068289]])
```

In [55]:

```
kmeans.labels_
```

Out[55]:

```
array([3, 4, 4, ..., 4, 1, 2])
```

In [56]:

```
X_train.columns
```

Out[56]:

```
Index(['Unnamed: 0', 'teacher_prefix', 'school_state',
       'project_submitted_datetime', 'project_title',
       'project_resource_summary',
       'teacher_number_of_previously_posted_projects', 'clean_categories',
       'clean_subcategories', 'clean_grades', 'essay', 'price', 'quantit
y'],
      dtype='object')
```

In [57]:

```
essays = X_train['essay'].values
cluster1 = []
cluster2 = []
cluster3 = []
cluster4 = []
cluster5 = []
for i in range(kmeans.labels_.shape[0]):
    if kmeans.labels_[i] == 0:
        cluster1.append(essays[i])
    elif kmeans.labels_[i] == 1:
        cluster2.append(essays[i])
    elif kmeans.labels_[i] == 2:
        cluster3.append(essays[i])
    elif kmeans.labels_[i] == 3:
        cluster4.append(essays[i])
    elif kmeans.labels_[i] == 4:
        cluster5.append(essays[i])
```

In [58]:

```
len(cluster2)
```

Out[58]:

2337

In [59]:

```
for i in range(3):  
    print((cluster1[i]))
```



My students are very hands on! They love to read, write and do math. They are very energetic. They have a lot to say and express. There is never a dull moment in my classroom!\r\n

My students are full of creativity in their own unique ways. They are very passionate about their learning and helping their peers succeed as well. Unfortunately, my students come from a low-income community which makes it hard for them to have the necessary materials they need in order to express their creativity and uniqueness. \r\nThe supplies that I am requesting for my students are meaningful. In other words, I know exactly when and how each supply will be utilized. We just need your help in getting everything into our classroom. The literacy and math supplies will be used during center time. Not only will this allow me to conduct guided reading, but it will allow the other students to engage in meaningful activities that target each of their specific needs. For example, the read along set will allow students the opportunity to listen to books on their own as they follow through and learn new vocabulary. The classroom magnetic kit set will allow students to practice spelling and writing words. All phonics activities will allow students to master the skills they need to become better readers and writers. The numbers to 100 will help students practice counting from 1-100 and better understand the concept of base ten and ones. The donations to this project will help to improve my students' understanding of both literacy and math.\r\nBy implementing centers, my students will have the opportunity to practice the skills they need in a fun and meaningful way. We will use these supplies to aide us in the process of becoming life-long readers and mathematicians. As the teacher and speaking from experience, I know the benefit and the tremendous impact centers have on students in the early grades. I truly need your support to make this a reality. Thanks!nannan

I teach at a high poverty school where several of the students receive free or reduced price lunches, and 100% receive free breakfast. \r\n\r\nEven though my second grade students have many obstacles that make learning more difficult, I always have high expectations that they will grow and learn. School is a safe and happy place for my students. Each day they come in to the classroom excited about the day and spending time exploring and learning together! For many of my students, being at school is the best part of their day.\r\n\r\nKeeping up with today's technology is an important part of education. Coding, financial literacy, along with STEAM activities are very important to me to teach my children. \r\n\r\n\r\n\r\n\r\nOur class is currently sharing ipads with Six other classes We love and adore having the iPads but we'd love even more to be able to have Chromebooks that actually stay in our room all day long. Having our own Chromebooks would give us the time and opportunity to improve in our research, typing, Google Classroom use, and many other areas of study that iPads just can not do. We would be able to incorporate technology throughout our entire day. This makes us all very excited! I would love to give my students the opportunity to learn to use technology safely and responsibly, while equipping them with skills that go well beyond the classroom. Chromebooks would help make 21st century learningChromebooks would help make 21st century learning come to life! come to life!nannan

The moment my students walk through my classroom door, they are ready to get to work. I say, "Are you ready to rock?" and they respond with, "Ready to roll!!!" I expect nothing but the best from them, hence I strive to provide the best teacher and classroom environment I can possibly give.\r\n\r\n\r\nMy school serves Latino, Polynesian, and African-American students who live in deep poverty. These are amazing students, full of life and character, who deserve nothing but the best. Most live in unfortunate situations at home, therefore the classroom must be anything but unfortunate. I strive to make sure my students have all that they need to be successful in their academics.Tip, tip, tipping those chairs! The students love to play the balancing act with those chairs. \r\n\r\n\r\n"All fours on the floor!" Is an expression I'm often saying in class to students who can't handle the use of a traditional four-legged chair. They are my students who have

a need to be constantly moving so they will play a balancing act with their four-legged chairs, tipping their chairs, rocking back and forth in order to maintain their chair as steady as possible, using only the two back legs. It truly is a frightful sight simply because it is a dangerous thing to do. Many have been the times that the child has lost his/her balancing act and have gone crashing to the floor. Yet, they will do it again and again because of the satisfaction they get from having to move their bodies ever so slightly to keep the chair steady. It truly looks like a circus act and can be as distracting as a circus-distracting to me as I have to stop what I am doing in order to remind the child to place all four legs on the floor, which then serves as a distraction to the rest of the class who stop what they are doing, and lose focus, to see what I am complaining about. \r\n\r\nGone are the days where children are to sit still because children need to move! My goal is to have all 31 of my students have the option to use a traditional chair or a wobble chair and with this project of even four more, we will have a total of twelve wobble chairs. Thanks to you, little by little, we will have a wobble chair for each child.nannan

In [60]:

```
for i in range(3):
    print((cluster2[i]))
```

I teach Math and Science to forty-four 4th grade students. Most of my students love these subjects, which makes them very easy to teach! My students come from two learning backgrounds, Montessori and Traditional learning. My students are eager to learn! They enjoy hands-on activities and like to stay engaged. Downtime can be difficult for them, as they like to stay busy. Technology, one-on-one, and small groups helps to keep my active students engaged in learning. Allowing student choice is something that I strive to accomplish each day! My students love to read and watch Science unfold, but they learn best when they get to be the scientist! My students enjoy learning through hands on Science experiments! When they are allowed to explore and investigate, they get to learn from their mistakes and tap into their higher order thinking skills to solve problems and experiment. They love "trying out" experiments they have read about. This set of Science centers will allow my students the opportunity to read about and perform numerous experiments. They will practice following directions while still maintaining the fun-factor of experimenting. When my students think about Science, I want them to think about exploring and experimenting!

The students at my school are highly motivated and eager to learn about technology and science. They are just being introduced to technology at this level. The students look forward to going to the computer lab to learn about new and excited things while having fun at the same time. With STEAM being introduced technology is playing a huge role in the classroom. These students will be the future. In the computer lab the students get reinforcement on concepts learned in the classroom. They experience what they are learning about in fun and exciting ways that they could never get in a traditional classroom. The students computers need a powerful laptop to connect to that runs a server for minecraft.edu. A laptop is best because some classes will be in different locations. The smart board will help the teachers and students understand the task and display the lesson. Technology will make anything possible and opens up learning to children that would otherwise be uninterested. With this first project I hope to set up a good foundation to build a digital learning environment that will engage students in ways never thought of before. This is the beginning of building a technology infrastructure that other more affluent schools already have in place.

My fantastic first grade students are wonderful, caring, curious students that want to have technology available in the classroom. Students at my school say this promise everyday... I promise to...make wise decisions, use kind words, show respect, take care of myself, act responsibly, notice good deeds, grow in community and solve problems. As a role model to them, I will do the same and help my students in anyway I can. Donors Choose is a way for my students to see that the world cares about them and their education! Every student deserves to feel valued and cared for. I know my school and staff does this in every way! My students need an Osmo kit to explore everything that technology can offer them! This Genius Kit and Coding Game will get students excited! First graders enjoy using technology, but still need to be able to touch and move items to truly enhance their learning. I know that my students will use these Osmo resources every day to help with reading and math! This resource is in other first grade classes and my students really want one for our classroom that they can use. New technology will keep students wanting to learn more and come to school everyday! This resource will help students prepare for the future.

In [61]:

```
for i in range(3):  
    print((cluster3[i]))
```

Every morning on the news, you hear about the budget cuts my school district is facing due to the state not passing the budget. In preparation for the upcoming school year, I will not allow the budget cuts to impact my incoming group of students. We must continue to learn and achieve greatness. Due to the budget cuts, a growing class size is a now reality and in order to prepare for them, I need your help! I teach at an awesome S.T.E.M school on the west side of Chicago. We are a Title I school in addition to having a 96.7% low income rate as well as a high mobility rate of 41.4%. Many students see my school as a safe and nurturing place due to all of the violence occurring around them daily. I want to continue to provide a engaging learning environment despite the growing class size. Today's student is more interested in video games than a book. It is extremely impossible to get a student to read, let alone comprehend a nonfiction text. Due to this, I have to become clever in my approach when it comes to teaching nonfiction text. Therefore, this year I decided that I would use Time Life for Kids. By donating Time Life for Kids to my classroom, you are providing my students with the opportunity to engage in high interest nonfiction text that is reflective of the events that occurring currently. By engaging in high interest nonfiction text, it helps improve my students reading fluency and comprehension.

Albert Einstein once said, "Creativity is intelligence having fun." This is exactly how I approach what I do every day. I use my art room as a launching pad of exploration and adventure. I want my students to continue being successful. My students don't ask for much, but they deserve everything. They are such a nice and outgoing group of kids that are eager to learn this year. We are a diverse school and city. Over 75% of our students qualify for free and reduced-price lunch. Our school has the largest ELL population in the district. They wear uniforms to school every day and go to school longer than anyone else in our district. Our students are the hardest workers in the district and hardly complain about it. A classroom rug is a great place to start our class. A rug can be warm and inviting for students to share and connect with others. By having a rug within my art room students will have a location to sit on to start class. This rug will help my students to have consistency within the art room. A rug in a classroom is a multi-functional space. It becomes a meeting area, a reading area, a work area, and a quiet area. For students, a rug within our art room will bring warmth into the room and comfort. We hope to have one soon!

I teach at a Title I urban school that has about 500 students. This year I work with 30 students that are in need of additional support in order to be able to keep up with their peers in the areas of Reading and Math. Most of my students are identified as having a Specific Learning Disability. Some of my other kiddos have Cognitive Disabilities or an Other Health Impairment. Despite, their eligibility for Special Education Services, I believe that each and every one of them have the potential to grow up and do great things. I am trying to find ways to motivate my reluctant readers and struggling students so that they will start to feel success in the academic setting. Please consider helping me financially to be able to get appropriate supplies to help motivate my students to love learning. Thanks for your support!!! There are many ways that using IPADS in my classroom will be beneficial both for my students and myself. The most important reason behind requesting IPADS is that my students enjoy working with them. I have very limited access to the IPAD cart at school but my students ALWAYS love IPAD day. At least the two groups that are able to currently use them love it. In addition to making learning fun, there is a multitude of free apps that are available to help reinforce the important skills that I am teaching such as counting money and telling time. The apps give students immediate feedback on their performance which maximizes the benefit of our short time spent together. I am able to use the data received from some of the apps to determine which students are excelling with specific skills and which students need additional small group time with me for remediation.

tion. \r\n\r\nMy students also benefit from using technology in my classroom because it helps prepare them for the real world and their future. How many times have you heard colleagues or yourself say: \"I wish they would have taught me how to use this technology in school!\"? Well, I'm trying to do just that. Please consider supporting my project. Thanks in advance for your consideration of helping me to make learning fun for my kids!nanna  
n

In [62]:

```
for i in range(3):  
    print((cluster4[i]))
```

Students at my school want to combine improving their physical fitness with the latest technology. They love to move and by using this modern technology they will strive to improve their overall fitness level. The technology that is out there today can help my students work for a personal goal that they can challenge themselves to improve on for the rest of their lives. By being able to use this technology and see how if they work hard and are consistent they will improve their overall health! By donating to The Fitbit project you are helping my students through the use of technology see how exercise can help them live better lives. Daily exercise is important for everyone. In order to teach children this they need to understand how their body functions. With the fitbit they will be able to track their steps during PE class and during a normal day. If we do not teach our children about the importance of exercise now who will teach them when they are adults? Children who exercise do better in school. Let's give them the tools so they can begin the journey of an active lifestyle!

Bringing out the best in each student is every teacher's goal, and in many instances they will do anything and everything possible to help their students reach their full potential. I affectionately call my group of quirky, innovative, rambunctious "kids" Hogan's Heroes. These fourth grade "heroes" have big hearts, big personalities, and big goals. It's my priority to help them dream big, gain confidence, and learn in the style that works best for them. A little background on my students...Our district serves a high poverty area, with more than half of our student population receiving free and reduced lunch, and many participating in our backpack food program on the weekends. Despite these economic hardships the children are eager to learn, grow, and lead. Flexible seating is a research based, best practice method to get students comfortable, learning, and involved in the classroom. Increased motivation, higher levels of engagement, and better control of excess energy are all benefits of flexible seating. Having the options of wobble stools, balance balls, or sitting wedges will help students to not only focus on their work, but will also allow them to practice problem solving, positive decision making, and accountability when choosing a spot where they can work productively and comfortably. By offering several choices involving movement (such as the wobble stools and balance balls) many of my students will be able better engage, because they will be able to focus on the concept, instead of trying to control their bodies, as they do when sitting in a traditional seat. Right now, we are exercising flexible seating in the classroom, without these wonderful options. Sitting on the floor, utilizing different size tables, and diverse work spaces has already begun to show improvement, having these other wonderful options is really going to make such a positive difference in our classroom. I am excited to see how these new options can further help my group of future entrepreneurs, coders, engineers, surgeons, teachers and presidential candidates.

Our students are very passionate about making music. Each year, I teach an average of 175 students how to play various band instruments (Flute, Oboe, Clarinet, Saxophone, Trumpet, Trombone, Baritone, Tuba, French Horn, Percussion) in the 4 elementary schools of our district. My students consistently perform competitively in All County Festivals and we are engaged in the community such as our holiday nursing home performance. Over the last several years, our demographics have shifted to a more diverse population with many immigrant families moving to the area resulting in fewer families able to spend money on music store instrument rentals as in the past. While these students are just as eager and talented to play as ever, I find myself running out of school owned instrument rentals. The majority of the Elementary school Tubas are held together with electric tape and zip ties. Some are older than I am! The only reason we still use them is we have no choice if we want Tuba players in the band. I'd love to be able to retire at least one of the Tubas that has holes and rattles when students play it. They are very expensive and I'm hoping that perhaps a benefactor



or or corporate sponsor will be looking for a big way to help out an area school with something that I honestly realize is a dream.nannan

In [63]:

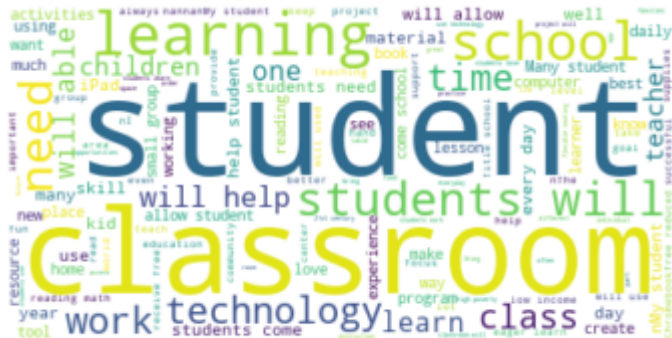
```
for i in range(3):  
    print((cluster5[i]))
```

Every student that walks through my door has their own unique story. A wide range of cultures, languages, and backgrounds are puzzled together to create a dynamic, well-rounded, and curious group of students. Though their dreams and challenges may differ, my students all share something very much in common - the desire to learn and be engaged as they prepare for their future. I hope to create a classroom environment that my students can thrive and feel inspired within. I want them to come to class ready to be shocked, amazed, intrigued, challenged, and more. Most of all, I hope my students find my classroom as a place to feel appreciated for their individuality and pushed to discover their potential. With the shift to Common Core standards, it is imperative that students justify and cite their thinking during discussion as well as in their writing. In my classroom, I use interactive notebooks with my students as a way to document information, but also process their understandings. They are an important tool for promoting literacy and writing, as well as higher order thinking. Using interactive notebooks in the classroom helps me promote a cutting edge way for students to take notes, internalize their learning, and reflect on their thinking to meet their academic needs. With this style of note-taking, students are able to use the information learned in class and use personal connections to process the information. Students have the opportunity to be creative and independent learners as they design their notebooks. In addition, the notebook is a resource students refer to all year long. At the end of the year, it is a scrapbook of the journey of their learning.

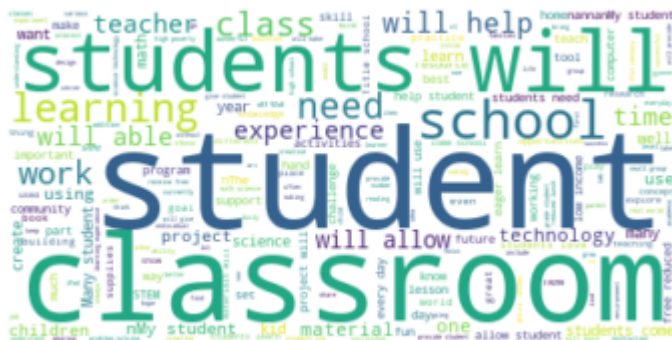
I am so impressed with our Roxana kids. Roxana is a refinery town, and a large percentage of the local jobs have recently been automated, so too many of our kids' parents have been abruptly ejected from the middle class. These are hardworking parents with respectful, intelligent kids, looking for hope and inspiration, and that's what we're striving for here in the library! Our amazing football team, in particular, needs some extra motivation to read, so that's what the One Team One Book Project aims to do! The Roxana High School football team wants to read a book that inspires them to do their best, no matter the challenges. "Imperfect: An Improbable Life" is the story of Jim Abbott, pitcher for the New York Yankees, who pitched a no-hitter in 1993 despite the fact that he only has one arm. "Imperfect" is the perfect book to inspire our football team to overcome adversity, on and off the field. This harrowing tale shows that anyone can overcome anything with enough perseverance and tenacity. The library will purchase 20 copies, but we need your help to get the 20 more copies of this book so we can read it and discuss it together.

I have 21 highly motivated first grade students. There are 10 girls and 11 boys. I was able to loop with my class from kindergarten to first grade. We have become a close family over the two years that I have been able to teach them. My students have the same interests as many six and seven year olds - playing, video games, cartoons- but also have interests that are uniquely theirs such as cars and dragons. They have a passion for reading and writing. Our school is a K4-1 early childhood center with over 600 children aged four to seven. It is located in a small town surrounded by rural communities within thirty minutes of a large city. In my classroom, we spend a good deal of time reading. I use just right books to help them. Leveled readers are a great way to help students grow as readers. My students switch books each week. These books are kept in browse bags that they take home as well as read in class. These books are "just right" books. Each student can pick out books that interest them and are on their reading level. Some of my students do not have any literature or resources to help them read in their homes. These are interesting books that the students have time to read in class and take home so they can work on reading at home with their families as well.

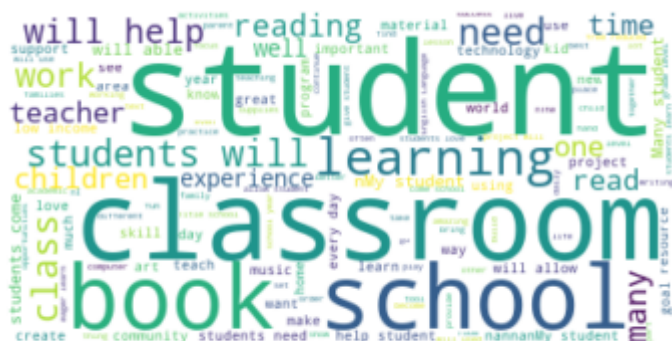
```
words=''
for i in cluster1:
    words+=str(i)
from wordcloud import WordCloud
wordcloud = WordCloud(background_color="white").generate(words)
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
```



```
words=''
for i in cluster2:
    words+=str(i)
from wordcloud import WordCloud
wordcloud = WordCloud(background_color="white").generate(words)
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
```

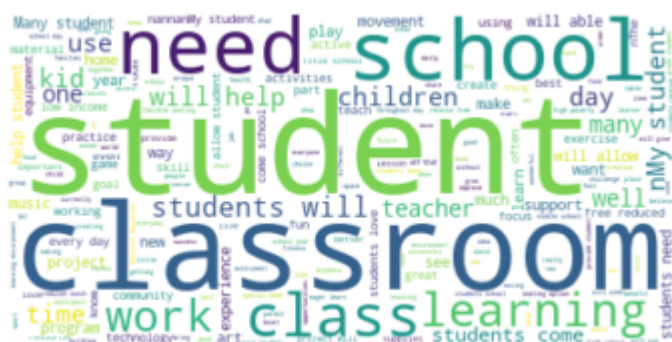


```
words=''
for i in cluster3:
    words+=str(i)
from wordcloud import WordCloud
wordcloud = WordCloud(background_color="white").generate(words)
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
```



```
words=''
for i in cluster4:
    words+=str(i)
from wordcloud import WordCloud
wordcloud = WordCloud(background_color="white").generate(words)

plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
```





In [73]:

```
agg_cl.labels_
```

Out[73]:

```
array([0, 0, 0, ..., 0, 0, 1], dtype=int64)
```

In [82]:

Out[82]:

```
(49000, 14738)
```

In [85]:

```
cluster1=[]
cluster2=[]
essays = X_train['essay'].values
for i in range(agg_cl.labels_.shape[0]):
    if agg_cl.labels_[i] == 0:
        cluster1.append(essays[i])
    elif agg_cl.labels_[i] == 1:
        cluster2.append(essays[i])
print(len(cluster1))
print(len(cluster2))
```

```
3361
```

```
1639
```

In [86]:

```
for i in range(5):  
    print((cluster1[i]))
```



Students at my school want to combine improving their physical fitness with the latest technology. They love to move and by using this modern technology they will strive to improve their overall fitness level. The technology that is out there today can help my students work for a personal goal that they can challenge themselves to improve on for the rest of their lives. By being able to use this technology and see how if they work hard and are consistent they will improve their overall health! By donating to The Fitbit project you are helping my students through the use of technology see how exercise can help them live better lives. Daily exercise is important for everyone. In order to teach children this they need to understand how their body functions. With the fitbit they will be able to track their steps during PE class and during a normal day. If we do not teach our children about the importance of exercise now who will teach them when they are adults? Children who exercise do better in school. Let's give them the tools so they can begin the journey of an active lifestyle!

annan

Every student that walks through my door has their own unique story. A wide range of cultures, languages, and backgrounds are puzzled together to create a dynamic, well-rounded, and curious group of students. Though their dreams and challenges may differ, my students all share something very much in common - the desire to learn and be engaged as they prepare for their future. I hope to create a classroom environment that my students can thrive and feel inspired within. I want them to come to class ready to be shocked, amazed, intrigued, challenged, and more. Most of all, I hope my students find my classroom as a place to feel appreciated for their individuality and pushed to discover their potential. With the shift to Common Core standards, it is imperative that students justify and cite their thinking during discussion as well as in their writing. In my classroom, I use interactive notebooks with my students as a way to document information, but also process their understandings. They are an important tool for promoting literacy and writing, as well as higher order thinking. Using interactive notebooks in the classroom helps me promote a cutting edge way for students to take notes, internalize their learning, and reflect on their thinking to meet their academic needs. With this style of note-taking, students are able to use the information learned in class and use personal connections to process the information. Students have the opportunity to be creative and independent learners as they design their notebooks. In addition, the notebook is a resource students refer to all year long. At the end of the year, it is a scrapbook of the journey of their learning.

nannan

I am so impressed with our Roxana kids. Roxana is a refinery town, and a large percentage of the local jobs have recently been automated, so too many of our kids' parents have been abruptly ejected from the middle class. These are hardworking parents with respectful, intelligent kids, looking for hope and inspiration, and that's what we're striving for here in the library! Our amazing football team, in particular, needs some extra motivation to read, so that's what the One Team One Book Project aims to do! The Roxana High School football team wants to read a book that inspires them to do their best, no matter the challenges. "Imperfect: An Improbable Life" is the story of Jim Abbott, pitcher for the New York Yankees, who pitched a no-hitter in 1993 despite the fact that he only has one arm. "Imperfect" is the perfect book to inspire our football team to overcome adversity, on and off the field. This harrowing tale shows that anyone can overcome anything with enough perseverance and tenacity. The library will purchase 20 copies, but we need your help to get the 20 more copies of this book so we can read it and discuss it together.

nannan

The students at my school are highly motivated and eager to learn about technology and science. They are just being introduced to technology at this level. The students look forward to going to the computer lab to learn about new and excited things while having fun at the same time. With STEAM being introduced technology is playing a huge role in the classroom.

These students will be the future. In the computer lab the students get reinforcement on concepts learned in the classroom. They experience what they are learning about in fun and exciting ways that they could never get in a traditional classroom. The students computers need a powerful laptop to connect to that runs a server for minecraft.edu. A laptop is best because some classes will be in different locations. The smart board will help the teachers and students understand the task and display the lesson. \r\n\r\n Technology will make anything possible and opens up learning to children that would otherwise be uninterested. With this first project I hope to set up a good foundation to build a digital learning environment that will engage students in ways never thought of before. This is the beginning of building a technology infrastructure that other more affluent schools already have in place.nannan

I have 21 highly motivated first grade students. There are 10 girls and 11 boys. I was able to loop with my class from kindergarten to first grade. We have become a close family over the two years that I have been able to teach them. \r\nMy students have the same interests as many six and seven year olds - playing, video games, cartoons- but also have interests that are uniquely theirs such as cars and dragons. They have a passion for reading and writing.\r\nOur school is a K4-1 early childhood center with over 600 children aged four to seven. It is located in a small town surrounded by rural communities within thirty minutes of a large city. In my classroom, we spend a good deal of time reading. I use just right books to help them. Leveled readers are a great way to help students grow as readers. My students switch books each week. These books are kept in browse bags that they take home as well as read in class.\r\nThese books are \"just right\" books. Each student can pick out books that interest them and are on their reading level. Some of my students do not have any literature or resources to help them read in their homes. These are interesting books that the students have time to read in class and take home so they can work on reading at home with their families as well.nannan

In [87]:

```
for i in range(5):  
    print((cluster2[i]))
```

I teach Math and Science to forty-four 4th grade students. Most of my students love these subjects, which makes them very easy to teach! My students come from two learning backgrounds, Montessori and Traditional learning. My students are eager to learn! They enjoy hands-on activities and like to stay engaged. Downtime can be difficult for them, as they like to stay busy. Technology, one-on-one, and small groups helps to keep my active students engaged in learning. Allowing student choice is something that I strive to accomplish each day! My students love to read and watch Science unfold, but they learn best when they get to be the scientist! My students enjoy learning through hands on Science experiments! When they are allowed to explore and investigate, they get to learn from their mistakes and tap into their higher order thinking skills to solve problems and experiment. They love "trying out" experiments they have read about. This set of Science centers will allow my students the opportunity to read about and perform numerous experiments. They will practice following directions while still maintaining the fun-factor of experimenting. When my students think about Science, I want them to think about exploring and experimenting!

My students are very hands on! They love to read, write and do math. They are very energetic. They have a lot to say and express. There is never a dull moment in my classroom!

My students are full of creativity in their own unique ways. They are very passionate about their learning and helping their peers succeed as well. Unfortunately, my students come from a low-income community which makes it hard for them to have the necessary materials they need in order to express their creativity and uniqueness. The supplies that I am requesting for my students are meaningful. In other words, I know exactly when and how each supply will be utilized. We just need your help in getting everything into our classroom. The literacy and math supplies will be used during center time. Not only will this allow me to conduct guided reading, but it will allow the other students to engage in meaningful activities that target each of their specific needs. For example, the read along set will allow students the opportunity to listen to books on their own as they follow through and learn new vocabulary. The classroom magnetic kit set will allow students to practice spelling and writing words. All phonics activities will allow students to master the skills they need to become better readers and writers. The numbers to 100 will help students practice counting from 1-100 and better understand the concept of base ten and ones. The donations to this project will help to improve my students' understanding of both literacy and math. By implementing centers, my students will have the opportunity to practice the skills they need in a fun and meaningful way. We will use these supplies to aide us in the process of becoming life-long readers and mathematicians. As the teacher and speaking from experience, I know the benefit and the tremendous impact centers have on students in the early grades. I truly need your support to make this a reality. Thanks!

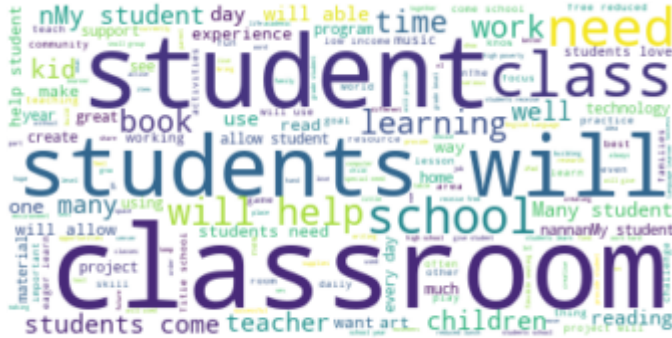
My fantastic first grade students are wonderful, caring, curious students that want to have technology available in the classroom. Students at my school say this promise everyday... I promise to...make wise decisions, use kind words, show respect, take care of myself, act responsibly, notice good deeds, grow in community and solve problems. As a role model to them, I will do the same and help my students in anyway I can. Donors Choose is a way for my students to see that the world cares about them and their education! Every student deserves to feel valued and cared for. I know my school and staff does this in every way! My students need an Osmo kit to explore everything that technology can offer them! This Genius Kit and Coding Game will get students excited! First graders enjoy using technology, but still need to be able to touch and move items to truly enhance their learning. I know that my students will use these Osmo resources every day to help with reading and math! This resource is in other first grade classes and my students really want one for our classroom that they can use. New technology will keep students wanting to learn more and

come to school everyday! This resource will help students prepare for the future.nannan

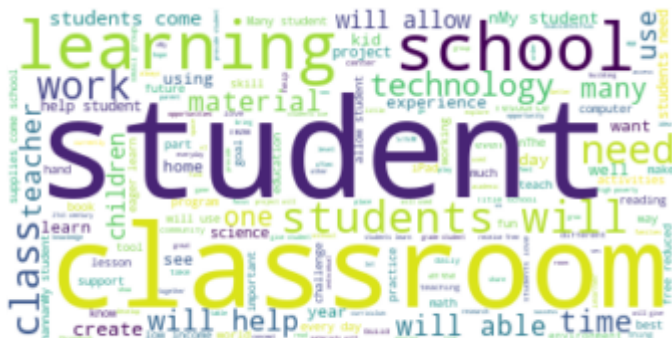
We are a curious group of fifth grade learners that do not give up and continue to search for new ways to learn and grow together. We live in suburban Long Island, New York. We make connections between subject areas and especially like to write, publish, and present our research and findings from experiments we have conducted. Next year we will be off to middle school and hope to make this last year in elementary school one to remember through hands on learning in Science!My students need materials to enable them to see, handle, measure and observe rocks, minerals and soil samples in our study of the Earth. The materials that we are requesting will provide samples, lab activities, as well as lab safety equipment for much needed hands-on science work.\r\n\r\nOur curious group of fifth grade learners will use these much needed supplies to enrich our study of our earth through grade appropriate and safe lab activities. Scales, rock and mineral samples, safety goggles and samples of soil types will enable students to experience science beyond the materials we have available through our school at this time.nannan

I teach at a high poverty school where several of the students receive free or reduced price lunches, and 100% receive free breakfast. \r\n\r\nEven though my second grade students have many obstacles that make learning more difficult, I always have high expectations that they will grow and learn. School is a safe and happy place for my students. Each day they come in to the classroom excited about the day and spending time exploring and learning together! For many of my students, being at school is the best part of their day.\r\n\r\nKeeping up with today's technology is an important part of education. Coding, financial literacy, along with STEAM activities are very important to me to teach my children. \r\n\r\n\r\n\r\nOur class is currently sharing ipads with Six other classes We love and adore having the iPads but we'd love even more to be able to have Chromebooks that actually stay in our room all day long. Having our own Chromebooks would give us the time and opportunity to improve in our research, typing, Google Classroom use, and many other areas of study that iPads just can not do. We would be able to incorporate technology throughout our entire day. This makes us all very excited! I would love to give my students the opportunity to learn to use technology safely and responsibly, while equipping them with skills that go well beyond the classroom. Chromebooks would help make 21st century learningChromebooks would help make 21st century learning come to life! come to life!nannan

```
#plotting word cloud for cluster 1
words=''
for i in cluster1:
    words+=str(i)
from wordcloud import WordCloud
wordcloud = WordCloud(background_color="white").generate(words)
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
```



```
#plotting word cloud for cluster 2
words=''
for i in cluster2:
    words+=str(i)
from wordcloud import WordCloud
wordcloud = WordCloud(background_color="white").generate(words)
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
```



## for 5 clusters

In [90]:

```
from sklearn.cluster import AgglomerativeClustering
agg_cl=AgglomerativeClustering(n_clusters=5)
agg_cl.fit(final_train_5000[0:5000].toarray(),y_train[0:5000])
```

Out[90]:

```
AgglomerativeClustering(affinity='euclidean', compute_full_tree='auto',
                        connectivity=None, linkage='ward', memory=None, n_clusters=5,
                        pooling_func=<function mean at 0x000002C4DD79D2F0>)
```

In [91]:

```
agg_cl.labels_
```

Out[91]:

```
array([4, 1, 2, ..., 2, 3, 0], dtype=int64)
```

In [92]:

```
essays = X_train['essay'].values
cluster1 = []
cluster2 = []
cluster3 = []
cluster4 = []
cluster5 = []
for i in range(agg_cl.labels_.shape[0]):
    if agg_cl.labels_[i] == 0:
        cluster1.append(essays[i])
    elif agg_cl.labels_[i] == 1:
        cluster2.append(essays[i])
    elif agg_cl.labels_[i] == 2:
        cluster3.append(essays[i])
    elif agg_cl.labels_[i] == 3:
        cluster4.append(essays[i])
    elif agg_cl.labels_[i] == 4:
        cluster5.append(essays[i])
```

In [94]:

```
print(len(cluster1))
print(len(cluster2))
print(len(cluster3))
print(len(cluster4))
print(len(cluster5))
```

1639

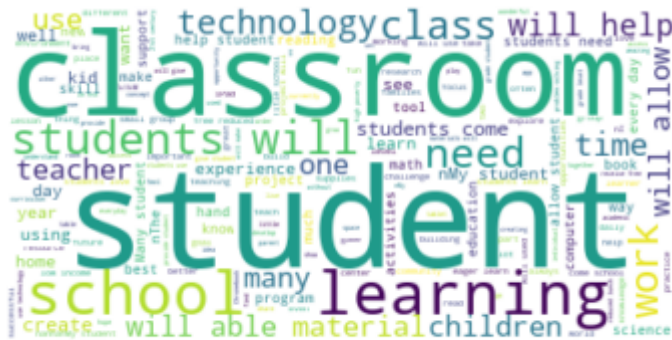
949

1196

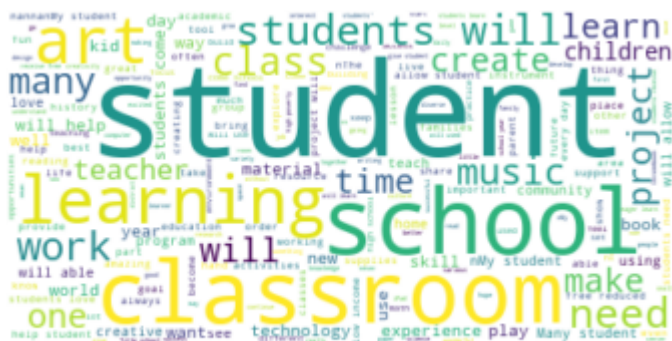
637

579

```
#plotting word cloud for cluster 1
words=''
for i in cluster1:
    words+=str(i)
from wordcloud import WordCloud
wordcloud = WordCloud(background_color="white").generate(words)
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
```

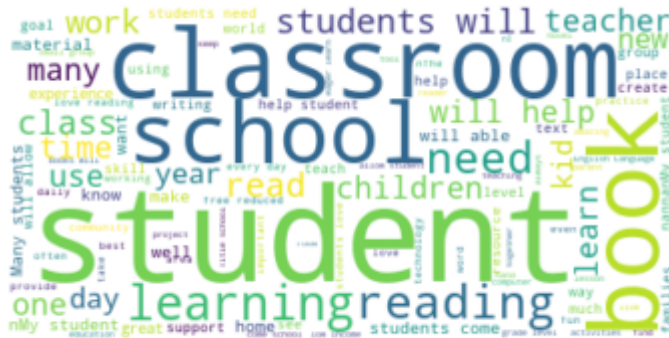


```
#plotting word cloud for cluster 2
words=''
for i in cluster2:
    words+=str(i)
from wordcloud import WordCloud
wordcloud = WordCloud(background_color="white").generate(words)
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
```

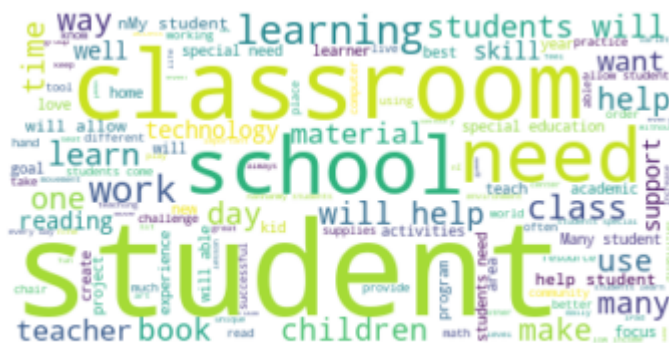




```
#plotting word cloud for cluster 3
words=''
for i in cluster3:
    words+=str(i)
from wordcloud import WordCloud
wordcloud = WordCloud(background_color="white").generate(words)
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
```



```
#plotting word cloud for cluster 4
words=''
for i in cluster4:
    words+=str(i)
from wordcloud import WordCloud
wordcloud = WordCloud(background_color="white").generate(words)
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
```



In [99]:

```
#plotting word cloud for cluster 5
words=''
for i in cluster5:
    words+=str(i)
from wordcloud import WordCloud
wordcloud = WordCloud(background_color="white").generate(words)
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
```

In [113]:

```
distance[0:5]
```

Out[113]:

```
array([[3.07096939],  
       [3.13286466],  
       [2.98826275],  
       [0.          ],  
       [2.65076333]])
```

In [112]:

```
sorted_distance=np.sort(distance)  
print(sorted_distance)  
print(len(sorted_distance))
```

```
[[3.07096939]  
 [3.13286466]  
 [2.98826275]  
 ...  
 [3.1822349 ]  
 [3.17680047]  
 [3.2526941 ]]  
5000
```



In [145]:

```
from sklearn.cluster import DBSCAN

dbscan=DBSCAN(eps=6,min_samples=50)
dbscan.fit(dbscan_5000,y_train[0:5000])
```

Out[145]:

```
DBSCAN(algorithm='auto', eps=6, leaf_size=30, metric='euclidean',
        metric_params=None, min_samples=50, n_jobs=1, p=None)
```

In [146]:

```
set(dbscan.labels_)
```

Out[146]:

```
{0}
```

In [147]:

```
cluster1=[]
for i in range(dbscan.labels_.shape[0]):
    if dbscan.labels_[i] == 0:
        cluster1.append(essays[i])
print(len(cluster1))
```

```
5000
```

In [148]:

```
for i in range(5):  
    print((cluster1[i]))
```

Students at my school want to combine improving their physical fitness with the latest technology. They love to move and by using this modern technology they will strive to improve their overall fitness level. The technology that is out there today can help my students work for a personal goal that they can challenge themselves to improve on for the rest of their lives. By being able to use this technology and see how if they work hard and are consistent they will improve their overall health! By donating to The Fitbit project you are helping my students through the use of technology see how exercise can help them live better lives. Daily exercise is important for everyone. In order to teach children this they need to understand how their body functions. With the fitbit they will be able to track their steps during PE class and during a normal day. If we do not teach our children about the importance of exercise now who will teach them when they are adults? Children who exercise do better in school. Let's give them the tools so they can begin the journey of an active lifestyle!

annan

Every student that walks through my door has their own unique story. A wide range of cultures, languages, and backgrounds are puzzled together to create a dynamic, well-rounded, and curious group of students. Though their dreams and challenges may differ, my students all share something very much in common - the desire to learn and be engaged as they prepare for their future. I hope to create a classroom environment that my students can thrive and feel inspired within. I want them to come to class ready to be shocked, amazed, intrigued, challenged, and more. Most of all, I hope my students find my classroom as a place to feel appreciated for their individuality and pushed to discover their potential. With the shift to Common Core standards, it is imperative that students justify and cite their thinking during discussion as well as in their writing. In my classroom, I use interactive notebooks with my students as a way to document information, but also process their understandings. They are an important tool for promoting literacy and writing, as well as higher order thinking. Using interactive notebooks in the classroom helps me promote a cutting edge way for students to take notes, internalize their learning, and reflect on their thinking to meet their academic needs. With this style of note-taking, students are able to use the information learned in class and use personal connections to process the information. Students have the opportunity to be creative and independent learners as they design their notebooks. In addition, the notebook is a resource students refer to all year long. At the end of the year, it is a scrapbook of the journey of their learning.

nannan

I am so impressed with our Roxana kids. Roxana is a refinery town, and a large percentage of the local jobs have recently been automated, so too many of our kids' parents have been abruptly ejected from the middle class. These are hardworking parents with respectful, intelligent kids, looking for hope and inspiration, and that's what we're striving for here in the library! Our amazing football team, in particular, needs some extra motivation to read, so that's what the One Team One Book Project aims to do! The Roxana High School football team wants to read a book that inspires them to do their best, no matter the challenges. "Imperfect: An Improbable Life" is the story of Jim Abbott, pitcher for the New York Yankees, who pitched a no-hitter in 1993 despite the fact that he only has one arm. "Imperfect" is the perfect book to inspire our football team to overcome adversity, on and off the field. This harrowing tale shows that anyone can overcome anything with enough perseverance and tenacity. The library will purchase 20 copies, but we need your help to get the 20 more copies of this book so we can read it and discuss it together.

nannan

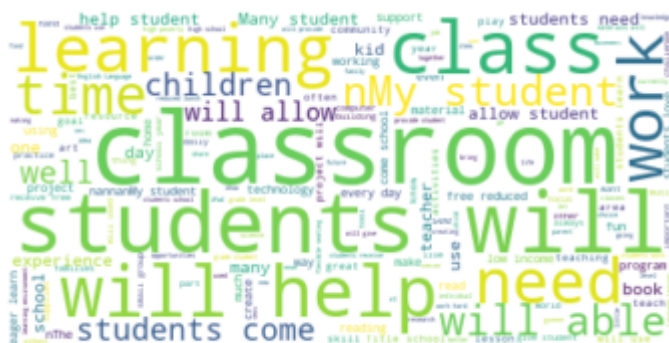
I teach Math and Science to forty-four 4th grade students. Most of my students love these subjects, which makes them very easy to teach! My students come from two learning backgrounds, Montessori and Traditional learning. My students are eager to learn! They enjoy hands-on activities and like to stay engaged. Downtime can be difficult for them, as they like to

stay busy. Technology, one-on-one, and small groups helps to keep my active students engaged in learning. Allowing student choice is something that I strive to accomplish each day! My students love to read and watch Science unfold, but they learn best when they get to be the scientist!\n\nMy students enjoy learning through hands on Science experiments! When they are allowed to explore and investigate, they get to learn from their mistakes and tap into their higher order thinking skills to solve problems and experiment. They love \"trying out\" experiments they have read about. This set of Science centers will allow my students the opportunity to read about and perform numerous experiments. They will practice following directions while still maintaining the fun-factor of experimenting. When my students think about Science, I want them to think about exploring and experimenting!\n\nannan

The students at my school are highly motivated and eager to learn about technology and science. They are just being introduced to technology at this level. The students look forward to going to the computer lab to learn about new and excited things while having fun at the same time. \n\n\n\nWith STEAM being introduced technology is playing a huge role in the classroom. These students will be the future. In the computer lab the students get reinforcement on concepts learned in the classroom. They experience what they are learning about in fun and exciting ways that they could never get in a traditional classroom. The students computers need a powerful laptop to connect to that runs a server for minecraft.edu. A laptop is best because some classes will be in different locations. The smart board will help the teachers and students understand the task and display the lesson. \n\n\n\nTechnology will make anything possible and opens up learning to children that hat would otherwise be uninterested. With this first project I hope to set up a good foundation to build a digital learning environment that will engage students in ways never thought of before. This is the beginning of building a technology infrastructure that other more affluent schools already have in place.annan

In [149]:

```
#plotting word cloud for cluster 1
words=''
for i in cluster1:
    words+=str(i)
from wordcloud import WordCloud
wordcloud = WordCloud(background_color="white").generate(words)
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
```





### 3. Cocnlusions

Please write down few lines of your observations on this assignment.

KMEANS: We used tfidf vectorizer for all the clustering techniques. We chose our number of clusters to be 5. We got clusters with different desities. AGGLOMERATIVE CLUSTERING: Here we took only 5k data points. We done clustering for number of clusters to be 2 and 5. For both clustering we got clusters with different densities. DENSITY BASED CLUSTERING: Here we took only 5k data points. We took min\_samples to be 50 and eps=60. we got only one cluster. We plotted word clouds for all the clusters.