

Data preparation

Import and review data:

```
In [ ]: !pip install pandas
import pandas as pd

Requirement already satisfied: pandas in c:\users\andrea\anaconda3\lib\site-packages (2.0.3)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\andrea\anaconda3\lib\site-packages (from pandas) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\users\andrea\anaconda3\lib\site-packages (from pandas) (2023.3.post1)
Requirement already satisfied: tzdata>=2022.1 in c:\users\andrea\anaconda3\lib\site-packages (from pandas) (2023.3)
Requirement already satisfied: numpy>=1.21.0 in c:\users\andrea\anaconda3\lib\site-packages (from pandas) (1.24.3)
Requirement already satisfied: six>=1.5 in c:\users\andrea\anaconda3\lib\site-packages (from python-dateutil>=2.8.2->pandas) (1.16.0)

In [ ]: #import csv file (from Local machine)

data = pd.read_csv("C:/Users/Andrea/Documents/CIND 820/WELFake_Dataset.csv")

In [ ]: #check first 5 rows

pd.set_option('max_colwidth', None)
data.head()
```

Out[]:	Unnamed: 0	title	text	label
0	0	LAW ENFORCEMENT ON HIGH ALERT Following Threats Against Cops And Whites On 9-11By #BlackLivesMatter And #FYF911 Terrorists [VIDEO]	<p>No comment is expected from Barack Obama Members of the #FYF911 or #FukYoFlag and #BlackLivesMatter movements called for the lynching and hanging of white people and cops. They encouraged others on a radio show Tuesday night to turn the tide and kill white people and cops to send a message about the killing of black people in America. One of the F***YoFlag organizers is called Sunshine. She has a radio blog show hosted from Texas called, Sunshine s F***ing Opinion Radio Show. A snapshot of her #FYF911 @LOLatWhiteFear Twitter page at 9:53 p.m. shows that she was urging supporters to Call now!! #fyf911 tonight we continue to dismantle the illusion of white Below is a SNAPSHOT Twitter Radio Call Invite #FYF911The radio show aired at 10:00 p.m. eastern standard time. During the show, callers clearly call for lynching and killing of white people. A 2:39 minute clip from the radio show can be heard here. It was provided to Breitbart Texas by someone who would like to be referred to as Hannibal. He has already received death threats as a result of interrupting #FYF911 conference calls. An unidentified black man said when those mother f**kers are by themselves, that s when when we should start f***ing them up. Like they do us, when a bunch of them ni**ers takin one of us out, that s how we should roll up. He said, Cause we already roll up in gangs anyway. There should be six or seven black mother f**kers, see that white person, and then lynch their ass. Let s turn the tables. They conspired that if cops started losing people, then there will be a state of emergency. He speculated that one of two things would happen, a big-ass [R s?????] war, or ni**ers, they are going to start backin up. We are already getting killed out here so what the f**k we got to lose? Sunshine could be heard saying, Yep, that s true. That s so f***ing true. He said, We need to turn the tables on them. Our kids are getting shot out here. Somebody needs to become a sacrifice on their side. He said, Everybody ain t down for that s***, or whatever, but like I say, everybody has a different position of war. He continued, Because they don t give a f**k anyway. He said again, We might as well utilized them for that s***t and turn the tables on these n**ers. He said, that way we can start lookin like we ain t havin that many casualties, and there can be more causalities on their side instead of ours. They are out their killing black people, black lives don t matter, that s what those mother f**kers so we got to make it matter to them. Find a mother f**ker that is alone. Snap his ass, and then f***in hang him from a damn tree. Take a picture of it and then send it to the mother f**kers. We just need one example, and then people will start watchin . This will turn the tables on s***, he said. He said this will start a trickle-down effect. He said that when one white person is hung and then they are just flat-hanging, that will start the trickle-down effect. He continued, Black people are good at starting trends. He said that was how to get the upper-hand. Another black man spoke up saying they needed to kill cops that are killing us. The first black male said, That will be the best method right there. Breitbart Texas previously reported how Sunshine was upset when racist white people infiltrated and disrupted one of her conference calls. She subsequently released the phone number of one of the infiltrators. The veteran immediately started receiving threatening calls. One of the #F***YoFlag movement supporters allegedly told a veteran who infiltrated their publicly posted conference call, We are going to rape and gut your pregnant wife, and your f***ing piece of sh*t unborn creature will be hung from a tree. Breitbart Texas previously encountered Sunshine at a Sandra Bland protest at the Waller County Jail in Texas, where she said all white people should be killed. She told journalists and photographers, You see this nappy-ass hair on my head? That means I am one of those more militant Negroes. She said she was at the protest because these redneck mother-f**kers murdered Sandra Bland because she had nappy hair like me. #FYF911 black radicals say they will be holding the imperial powers that are actually responsible for the terrorist attacks on September 11th accountable on that day, as reported by Breitbart Texas. There are several websites and Twitter handles for the movement. Palmetto Star describes himself as one of the head organizers. He said in a YouTube video that supporters will be burning their symbols of the illusion of their superiority, their false white supremacy, like the American flag, the British flag, police uniforms, and Ku Klux Klan hoods. Sierra McGrone or Nocturnus Libertus posted, you too can help a young Afrikan clean their a** with the rag of oppression. She posted two photos, one that appears to be herself, and a photo of a black man, wiping their naked butts with the American flag. For entire story: Breitbart News</p>	1
1	1	NaN	Did they post their votes for Hillary already?	1
2	2	UNBELIEVABLE! OBAMA'S ATTORNEY GENERAL SAYS MOST CHARLOTTE RIOTERS WERE "PEACEFUL" PROTESTERS...In Her Home State Of North Carolina [VIDEO]	Now, most of the demonstrators gathered last night were exercising their constitutional and protected right to peaceful protest in order to raise issues and create change. Loretta Lynch aka Eric Holder in a skirt	1
3	3	Bobby Jindal, raised Hindu, uses story of Christian conversion to woo evangelicals for potential 2016 bid	A dozen politically active pastors came here for a private dinner Friday night to hear a conversion story unique in the context of presidential politics: how Louisiana Gov. Bobby Jindal traveled from Hinduism to Protestant Christianity and, ultimately, became what he calls an "evangelical Catholic." Over two hours, Jindal, 42, recalled talking with a girl in high school who wanted to "save my soul," reading the Bible in a closet so his parents would not see him and feeling a stir while watching a movie during his senior year that depicted Jesus on the cross. "I was struck, and struck hard," Jindal told the pastors. "This was the Son of God, and He had died for our sins." Jindal's session with the Christian clergy, who lead congregations in the early presidential battleground states of Iowa and South Carolina, was part of a behind-the-scenes effort by the Louisiana governor to find a political base that could help propel him into the top tier of Republican candidates seeking to run for the White House in 2016. Known in GOP circles mostly for his mastery of policy issues such as health care, Jindal, a Rhodes Scholar and graduate of the Ivy League's Brown University, does not have an obvious pool of activist supporters to help drive excitement outside his home state. So he is harnessing his religious experience in a way that has begun to appeal to parts of the GOP's influential core of religious conservatives, many of whom have yet to find a favorite among the Republicans eyeing the presidential race. Other potential 2016 GOP candidates are wooing the evangelical base, including Sens. Rand Paul (Ky.) and Ted Cruz (Tex.) and Indiana Gov. Mike Pence. But over the weekend in Lynchburg — a mecca of sorts for evangelicals as the home of Liberty University, founded in the 1970s by the Rev. Jerry Falwell — Jindal appeared to make progress. In addition to his dinner with the pastors, he delivered a well-received "call to action" address to 40,000 Christian conservatives gathered for Liberty's commencement ceremony, talking again about his faith while assailing what he said was	0

Unnamed: 0	title	text label
		<p>President Obama's record of attacking religious liberty.\n\nThe pastors who came to meet Jindal said his intimate descriptions of his experiences stood out.\n\n"He has the convictions, and he has what it takes to communicate them," said Brad Sherman of Solid Rock Christian Church in Coralville, Iowa. Sherman helped former Arkansas governor Mike Huckabee in his winning 2008 campaign for delegates in Iowa.\n\nAnother Huckabee admirer, the Rev. C. Mitchell Brooks of Second Baptist Church in Belton, S.C., said Jindal's commitment to Christian values and his compelling story put him "on a par" with Huckabee, who was a Baptist preacher before entering politics.\n\nThe visiting pastors flew to Lynchburg over the weekend at the invitation of the American Renewal Project, a well-funded nonprofit group that encourages evangelical Christians to engage in the civic arena with voter guides, get-out-the-vote drives and programs to train pastors in grass-roots activism. The group's founder, David Lane, has built a pastor network in politically important states such as Iowa, Missouri, Ohio and South Carolina and has led trips to Israel with Paul and others seeking to make inroads with evangelical activists.\n\nThe group that Lane invited to Lynchburg included Donald Wildmon, a retired minister and founder of the American Family Association, a prominent evangelical activist group that has influence through its network of more than 140 Christian radio stations.\n\nMost of the pastors that Lane's organization brought to Lynchburg had not met Jindal. But they said he captured their interest recently when he stepped forward to defend Phil Robertson, patriarch of the "Duck Dynasty" television-show family, amid a controversy over disparaging remarks he made about gays in an interview with GQ magazine.\n\nThroughout his Lynchburg visit, Jindal presented himself as a willing culture warrior.\n\nDuring his commencement address Saturday, he took up the cause of twin brothers whose HGTV reality series about renovating and reselling houses, "Flip It Forward," was canceled last week after a Web site revealed that they had protested against same-sex marriage at the 2012 Democratic National Convention in Charlotte.\n\nThe siblings, Jason and David Benham, both Liberty graduates, attended the graduation and a private lunch with Jindal, who called the action against them "another demonstration of intolerance from the entertainment industry."\n\n"If these guys had protested at the Republican Party convention, instead of canceling their show, HGTV would probably have given them a raise," Jindal said as the Liberty crowd applauded.\n\nHe cited the Hobby Lobby craft store chain, which faced a legal challenge after refusing to provide employees with insurance coverage for contraceptives as required under the Affordable Care Act. Members of the family that owns Hobby Lobby, who have become heroes to many religious conservatives, have said that they are morally opposed to the use of certain types of birth control and that they considered the requirement a violation of their First Amendment right to religious freedom.\n\nThe family was "committed to honor the Lord by being generous employers, paying well above minimum wage and increasing salaries four years in a row even in the midst of the enduring recession," Jindal told the Liberty graduates. "None of this matters to the Obama administration."\n\nBut for the pastors who came to see Jindal in action, the governor's own story was the highlight of the weekend. And in many ways, he was unlike any other aspiring president these activists had met.\n\nPiyush Jindal was born in 1971, four months after his parents arrived in Baton Rouge, La., from their native India. He changed his name to Bobby as a young boy, adopting the name of a character on a favorite television show, "The Brady Bunch."\n\nHis decision to become a Christian, he told the pastors, did not come in one moment of lightning epiphany. Instead, he said, it happened in phases, growing from small seeds planted over time.\n\nJindal recalled that his closest friend from grade school gave him a Bible with his name emblazoned in gold on the cover as a Christmas present. It struck him initially as an unimpressive gift, Jindal told the pastors.\n\n"Who in the world would spend good money for a Bible when everyone knows you can get one free in any hotel?" he recalled thinking at the time. "And the gold lettering meant I couldn't give it away or return it."\n\nHis religious education reached a higher plane during his junior year in high school, he told his dinner audience. He wanted to ask a pretty girl on a date during a hallway conversation, and she started talking about her faith in God and her opposition to abortion. The girl invited him to visit her church.\n\nJindal said he was skeptical and set out to "investigate all these fanciful claims" made by the girl and other friends. He started reading the Bible in his closet at home. "I was unsure how my parents would react," he said.\n\nAfter the stirring moment when he saw Christ depicted on the cross during the religious movie, the Bible and his very existence suddenly seemed clearer to him, Jindal told the pastors.\n\nJindal did not dwell on his subsequent conversion to Catholicism just a few years later in college, where he said he immersed himself in the traditions of the church.\n\nHe touched on it briefly during the commencement address, noting in passing that "I am best described as an evangelical Catholic." Mostly, he sought to showcase the ways in which he shares values with other Christian conservatives.\n\n"I read the words of Jesus Christ, and I realized that they were true," Jindal told the graduates Saturday, offering a less detailed accounting of his conversion than he had done the night before with the pastors. "I used to think that I had found God, but I believe it is more accurate to say that He found me."</p>
4	4 SATAN 2: Russia unveils an image of its terrifying new 'SUPERNUKE' – Western world takes notice	<p>The RS-28 Sarmat missile, dubbed Satan 2, will replace the SS-18 Flies at 4.3 miles (7km) per sec and with a range of 6,213 miles (10,000km) The weapons are perceived as part of an increasingly aggressive Russia It could deliver a warhead of 40 megatons – 2,000 times as powerful as the atom bombs dropped on Hiroshima and Nagasaki in 1945 By LIBBY PLUMMER and GARETH DAVIE S Russia has unveiled chilling pictures of its largest ever nuclear missile, capable of destroying an area the size of France. The RS-28 Sarmat missile, dubbed Satan 2 by Nato, has a top speed of 4.3 miles (7km) per second and has been designed to outfox anti-missile shield systems. The new Sarmat missile could deliver warheads of 40 megatons – 2,000 times as powerful as the atom bombs dropped on Hiroshima and Nagasaki in 1945. Scroll down for video Russian President Vladimir Putin is reportedly planning to replace the country's older SS-18 Satan weapons with the new missiles amid a string of recent disagreements with the West. The Kremlin has stepped up the rhetoric against the West and carried a series of manoeuvres that has infuriated politicians in the US and UK. The pictures were revealed online by chief designers from the Makeyev Rocket Design Bureau. A message posted alongside the picture said: 'In accordance with the Decree of the Russian Government 'On the State Defense Order for 2010 and the planning period 2012-2013', the Makeyev Rocket Design Bureau was instructed to start design and development work on the Sarmat.' The RS-28 Sarmat missile is said to contain 16 nuclear warheads and is capable of destroying an area the size of France or Texas, according to Russian news network Zvezda, which is owned by Russia's ministry of defence. The weapon is also able to evade radar. It is expected to have a range of 6,213 miles (10,000 km), which would allow Moscow to attack London and FOR ENTIRE ARTICLE CLICK LINK</p>

In []: #check class counts

```
data.label.value_counts()
```

```
Out[ ]: label
1    37106
0    35028
Name: count, dtype: int64

In [ ]: #check for null values

data.isnull().sum()

Out[ ]: Unnamed: 0      0
title       558
text        39
label        0
dtype: int64

In [ ]: #drop null values

data = data.dropna()

In [ ]: #check for duplicate titles

data['title'].duplicated().sum()

Out[ ]: 9229

In [ ]: #check for duplicate text

data['text'].duplicated().sum()

Out[ ]: 9337

In [ ]: #remove duplicates, leaving the first instance

data = data.drop_duplicates(subset=['title'], keep='first')
data = data.drop_duplicates(subset=['text'], keep='first')

In [ ]: #new data counts after dropping null values and duplicates

data.label.value_counts()

Out[ ]: label
0    34237
1    27156
Name: count, dtype: int64

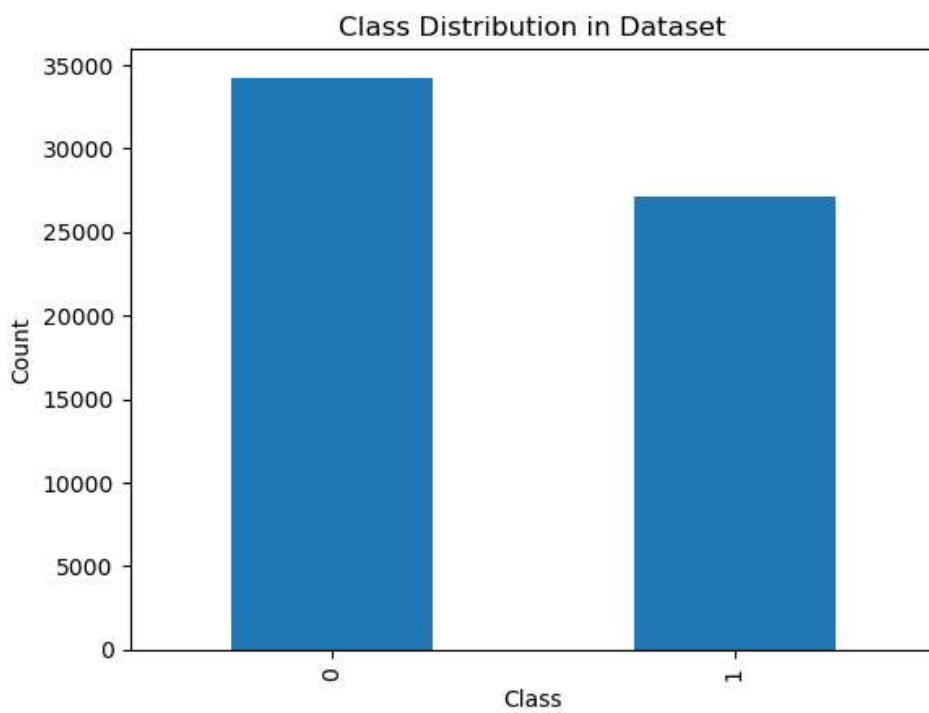
In [ ]: #plot class counts

!pip install matplotlib

import matplotlib.pyplot as plot

data.label.value_counts().plot.bar()
plot.title("Class Distribution in Dataset")
plot.ylabel("Count")
plot.xlabel("Class")

Requirement already satisfied: matplotlib in c:\users\andrea\anaconda3\lib\site-packages (3.7.2)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\andrea\anaconda3\lib\site-packages (from matplotlib) (1.0.5)
Requirement already satisfied: cycler>=0.10 in c:\users\andrea\anaconda3\lib\site-packages (from matplotlib) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\andrea\anaconda3\lib\site-packages (from matplotlib) (4.25.0)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\andrea\anaconda3\lib\site-packages (from matplotlib) (1.4.4)
Requirement already satisfied: numpy>=1.20 in c:\users\andrea\anaconda3\lib\site-packages (from matplotlib) (1.24.3)
Requirement already satisfied: packaging>=20.0 in c:\users\andrea\anaconda3\lib\site-packages (from matplotlib) (23.1)
Requirement already satisfied: pillow>=6.2.0 in c:\users\andrea\anaconda3\lib\site-packages (from matplotlib) (9.4.0)
Requirement already satisfied: pyparsing<3.1,>=2.3.1 in c:\users\andrea\anaconda3\lib\site-packages (from matplotlib) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\andrea\anaconda3\lib\site-packages (from matplotlib) (2.8.2)
Requirement already satisfied: six>=1.5 in c:\users\andrea\anaconda3\lib\site-packages (from python-dateutil>=2.7->matplotlib) (1.16.0)
Text(0.5, 0, 'Class')
```



Clean the data:

```
In [ ]: !pip install nltk
import nltk

Requirement already satisfied: nltk in c:\users\andrea\anaconda3\lib\site-packages (3.8.1)
Requirement already satisfied: click in c:\users\andrea\anaconda3\lib\site-packages (from nltk) (8.0.4)
Requirement already satisfied: joblib in c:\users\andrea\anaconda3\lib\site-packages (from nltk) (1.2.0)
Requirement already satisfied: regex>=2021.8.3 in c:\users\andrea\anaconda3\lib\site-packages (from nltk) (2022.7.9)
Requirement already satisfied: tqdm in c:\users\andrea\anaconda3\lib\site-packages (from nltk) (4.65.0)
Requirement already satisfied: colorama in c:\users\andrea\anaconda3\lib\site-packages (from click->nltk) (0.4.6)

In [ ]: #make new copy of data frame to clean
cleandata = data.copy()

In [ ]: #change all text to Lowercase
cleandata['title'] = cleandata['title'].str.lower()
cleandata['text'] = cleandata['text'].str.lower()

In [ ]: #remove urls
cleandata['title'] = cleandata['title'].str.replace(r'https?:\/\/\S+', '', regex=True)
cleandata['text'] = cleandata['text'].str.replace(r'https?:\/\/\S+', '', regex=True)

In [ ]: #install package to detect languages
!pip install langdetect

from langdetect import detect, DetectorFactory

Requirement already satisfied: langdetect in c:\users\andrea\anaconda3\lib\site-packages (1.0.9)
Requirement already satisfied: six in c:\users\andrea\anaconda3\lib\site-packages (from langdetect) (1.16.0)

In [ ]: #create function to add a column to the dataframe indicating the language of the text and title

def detect_lang(x):
    DetectorFactory.seed = 42
    try:
        lang = detect(x)
    except:
        lang = 'other'
    return lang

In [ ]: #add columns using function
cleandata['text_lang'] = cleandata['text'].apply(detect_lang)
```

```
In [ ]: cleandata['title_lang'] = cleandata['title'].apply(detect_lang)

In [ ]: #check how many rows are not in english

sum(cleandata['text_lang'] != 'en')

Out[ ]: 560

In [ ]: sum(cleandata['title_lang'] != 'en')

Out[ ]: 2724

In [ ]: #drop any rows with non-English text

cleandata.drop(cleandata[(cleandata['text_lang'] != 'en') | (cleandata['title_lang'] != 'en')].index, inplace=True)

In [ ]: #check results

sum(cleandata['text_lang'] != 'en')

Out[ ]: 0

In [ ]: sum(cleandata['text_lang'] != 'en')

Out[ ]: 0

In [ ]: #install package to expand contractions

!pip install contractions

import contractions

Requirement already satisfied: contractions in c:\users\andrea\anaconda3\lib\site-packages (0.1.73)
Requirement already satisfied: textsearch>=0.0.21 in c:\users\andrea\anaconda3\lib\site-packages (from contractions) (0.0.2)
Requirement already satisfied: anyascii in c:\users\andrea\anaconda3\lib\site-packages (from textsearch>=0.0.21->contractions) (0.3.2)
Requirement already satisfied: pyahocorasick in c:\users\andrea\anaconda3\lib\site-packages (from textsearch>=0.0.21->contractions) (2.0.0)

In [ ]: #expand contractions

cleandata['title'] = cleandata['title'].apply(lambda x: [contractions.fix(word) for word in str(x).split()])
cleandata['text'] = cleandata['text'].apply(lambda x: [contractions.fix(word) for word in str(x).split()])

In [ ]: #remove punctuation and numbers, except for hyphens in compound words

cleandata['title'] = cleandata['title'].astype(str)
cleandata['text'] = cleandata['text'].astype(str)

cleandata['title'] = cleandata['title'].str.replace('[^A-Za-z](?![A-Za-z]-([A-Za-z]))', ' ', regex=True)
cleandata['text'] = cleandata['text'].str.replace('[^A-Za-z](?![A-Za-z]-([A-Za-z]))', ' ', regex=True)

In [ ]: #check the results of cleaning

cleandata.head()
```

0	0	<p>no comment is expected from barack obama members of the fyf or fukyoflag and blacklivesmatter movements called for the lynching and hanging of white people and cops they encouraged others on a radio show tuesday night to turn the tide and kill white people and cops to send a message about the killing of black people in america one of the f yoflag organizers is called sunshine she has a radio blog show hosted from texas called sunshine s fucking opinion radio show a snapshot of her fyf lolatwhitefear twitter page at pm shows that she was urging supporters to call now fyf tonight we continue to dismantle the illusion of white below is a snapshot twitter radio call invite fyf the radio show aired at pm eastern standard time during the show callers clearly call for lynching and killing of white people a minute clip from the radio show can be heard here it was provided to breitbart texas by someone who would like to be referred to as hannibal he has already received death threats as a result of interrupting fyf conference calls an unidentified black man said when those mother f kers are by themselves that s when when we should start f ing them up like they do us when a bunch of them niggers taking one of us out that s how we should roll up he said because we already roll up in gangs anyway there should be six or seven black mother f ckers see that white person and then lynch their ass let s turn the tables they conspired that if cops started losing people then there will be a state of emergency he speculated that one of two things would happen a big-ass r s war or niggers they are going to start backin up we are already getting killed out here so what the f k we got to lose sunshine could be heard saying yep that s true that s so f king true he said we need to turn the tables on them our kids are getting shot out here somebody needs to become a sacrifice on their side he said everybody ain t down for that s t or whatever but like i say everybody has a different position of war he continued because they don t give a f k anyway he said again we might as well utilized them for that s t and turn the tables on these niggers he said that way we can start lookin like we ain t havin that many casualties and there can be more causalities on their side instead of ours they are out their killing black people black lives don t matter that s what those mother f kers so we got to make it matter to them find a mother f ker that is alone snap his ass and then f in hang him from a damn tree take a picture of it and then send it to the mother f kers we just need one example and then people will start watchin this will turn the tables on s t he said he said this will start a trickle-down effect he said that when one white person is hung and then they are just flat-hanging that will start the trickle-down effect he continued black people are good at starting trends he said that was how to get the upper-hand another black man spoke up saying they needed to kill cops that are killing us the first black male said that will be the best method right there breitbart texas previously reported how sunshine was upset when racist white people infiltrated and disrupted one of her conference calls she subsequently released the phone number of one of the infiltrators the veteran immediately started receiving threatening calls one of the f yoflag movement supporters allegedly told a veteran who infiltrated their publicly posted conference call we are going to rape and gut your pregnant wife and your f ing piece of sh t unborn creature will be hung from a tree breitbart texas previously encountered sunshine at a sandra bland protest at the waller county jail in texas where she said all white people should be killed she told journalists and photographers you see this nappy-ass hair on my head that means i am one of those more militant negroes she said she was at the protest because these redneck mother-f kers murdered sandra bland because she had nappy hair like me fyf black radicals say they will be holding the imperial powers that are actually responsible for the terrorist attacks on september th accountable on that day as reported by breitbart texas there are several websites and twitter handles for the movement palmetto star describes himself as one of the head organizers he said in a youtube video that supporters will be burning their symbols of the illusion of their superiority their false white supremacy like the american flag the british flag police uniforms and ku klux klan hoods sierra mcgrone or nocturnus libertus posted you too can help a young afrikan clean their a with the rag of oppression she posted two photos one that appears to be herself and a photo of a black man wiping their naked butts with the american flag for entire story breitbart news</p>	1	en	en
2	2	<p>unbelievable obama s attorney general says most charlotte rioters were peaceful protesters in her home state of north carolina video</p> <p>now most of the demonstrators gathered last night were exercising their constitutional and protected right to peaceful protest in order to raise issues and create change loretta lynch aka eric holder in a skirt</p>	1	en	en
3	3	<p>bobby jindal raised hindu uses story of christian conversion to woo evangelicals for potential bid</p> <p>a dozen politically active pastors came here for a private dinner friday night to hear a conversion story unique in the context of presidential politics how louisiana gov bobby jindal traveled from hinduism to protestant christianity and ultimately became what he calls an evangelical catholic over two hours jindal recalled talking with a girl in high school who wanted to save my soul reading the bible in a closet so his parents would not see him and feeling a stir while watching a movie during his senior year that depicted jesus on the cross i was struck and struck hard jindal told the pastors this was the son of god and he had died for our sins jindal s session with the christian clergy who lead congregations in the early presidential battleground states of iowa and south carolina was part of a behind-the-scenes effort by the louisiana governor to find a political base that could help propel him into the top tier of republican candidates seeking to run for the white house in known in gop circles mostly for his mastery of policy issues such as health care jindal a Rhodes scholar and graduate of the ivy league s brown university does not have an obvious pool of activist</p>	0	en	en

Unnamed: 0	title	text	label	text_lang	title_lang
		<p>supporters to help drive excitement outside his home state so he is harnessing his religious experience in a way that has begun to appeal to parts of the gop s influential core of religious conservatives many of whom have yet to find a favorite among the republicans eyeing the presidential race other potential gop candidates are wooing the evangelical base including sens rand paul ky and ted cruz tex and indiana gov mike pence but over the weekend in lynchburg a mecca of sorts for evangelicals as the home of liberty university founded in the s by the rev jerry falwell jindal appeared to make progress in addition to his dinner with the pastors he delivered a well-received call to action address to christian conservatives gathered for liberty s commencement ceremony talking again about his faith while assailing what he said was president obama s record of attacking religious liberty the pastors who came to meet jindal said his intimate descriptions of his experiences stood out he has the convictions and he has what it takes to communicate them said brad sherman of solid rock christian church in coralville iowa sherman helped former arkansas governor mike huckabee in his winning campaign for delegates in iowa another huckabee admirer the rev c mitchell brooks of second baptist church in belton s c said jindal s commitment to christian values and his compelling story put him on a par with huckabee who was a baptist preacher before entering politics the visiting pastors flew to lynchburg over the weekend at the invitation of the american renewal project a well-funded nonprofit group that encourages evangelical christians to engage in the civic arena with voter guides get-out-the-vote drives and programs to train pastors in grass-roots activism the group s founder david lane has built a pastor network in politically important states such as iowa missouri ohio and south carolina and has led trips to israel with paul and others seeking to make inroads with evangelical activists the group that lane invited to lynchburg included donald wild xadmon a retired minister and founder of the american family association a prominent evangelical activist group that has influence through its network of more than christian radio stations most of the pastors that lane s organization brought to lynchburg had not met jindal but they said he captured their interest recently when he stepped forward to defend phil robertson patriarch of the duck dynasty television-show family amid a controversy over disparaging remarks he made about gays in an interview with gq magazine throughout his lynchburg visit jindal presented himself as a willing culture warrior during his commencement address saturday he took up the because of twin brothers whose hgvt reality series about renovating and reselling houses flip it forward was canceled last week after a web site revealed that they had protested against same-sex marriage at the democratic national convention in charlotte the siblings jason and david benham both liberty graduates attended the graduation and a private lunch with jindal who called the action against them another demonstration of intolerance from the entertainment industry if these guys had protested at the republican party convention instead of canceling their show hgvt would probably have given them a raise jindal said as the liberty crowd applauded he cited the hobby lobby craft store chain which faced a legal challenge after refusing to provide employees with insurance coverage for contraceptives as required under the affordable care act members of the family that owns hobby lobby who have become heroes to many religious conservatives have said that they are morally opposed to the use of certain types of birth control and that they considered the requirement a violation of their first amendment right to religious freedom the family was committed to honor the lord by being generous employers paying well above minimum wage and increasing salaries four years in a row even in the midst of the enduring recession jindal told the liberty graduates none of this matters to the obama administration but for the pastors who came to see jindal in action the governor s own story was the highlight of the weekend and in many ways he was unlike any other aspiring president these activists had met piyush jindal was born in four months after his parents arrived in baton rouge la from their native india he changed his name to bobby as a young boy adopting the name of a character on a favorite television show the brady bunch his decision to become a christian he told the pastors did not come in one moment of lightning epiphany instead he said it happened in phases growing from small seeds planted over time jindal recalled that his closest friend from grade school gave him a bible with his name emblazoned in gold on the cover as a christmas present it struck him initially as an unimpressive gift jindal told the pastors who in the world would spend good money for a bible when everyone knows you can get one free in any hotel he recalled thinking at the time and the gold lettering meant i could not give it away or return it his religious education reached a higher plane during his junior year in high school he told his dinner audience he wanted to ask a pretty girl on a date during a hallway conversation and she started talking about her faith in god and her opposition to abortion the girl invited him to visit her church jindal said he was skeptical and set out to investigate all these fanciful claims made by the girl and other friends he started reading the bible in his closet at home i was unsure how my parents would react he said after the stirring moment when he saw christ depicted on the cross during the religious movie the bible and his very existence suddenly seemed clearer to him jindal told the pastors jindal did not dwell on his subsequent conversion to catholicism just a few years later in college where he said he immersed himself in the traditions of the church he touched on it briefly during the commencement address noting in passing that i am best described as an evangelical catholic mostly he sought to showcase the ways in which he shares values with other christian conservatives i read the words of jesus christ and i realized that they were true jindal told the graduates saturday offering a less detailed accounting of his conversion than he had done the night before with the pastors i used to think that i had found god but i believe it is more accurate to say that he found me</p>	1	en	en
4	satan russia unveils an image of its terrifying new supernuke	the rs sarmat missile dubbed satan will replace the ss flies at miles km per sec and with a range of miles km the weapons are perceived as part of an increasingly aggressive russia it could deliver a warhead of megatons times as powerful as the atom bombs dropped on hiroshima and nagasaki in by libby plummer and gareth davie s russia has unveiled chilling	1	en	en

Unnamed: 0	title	text	label	text_lang	title_lang
	western world takes notice	pictures of its largest ever nuclear missile capable of destroying an area the size of france the rs sarmat missile dubbed satan by nato has a top speed of miles km per second and has been designed to outfox anti-missile shield systems the new sarmat missile could deliver warheads of megatons times as powerful as the atom bombs dropped on hiroshima and nagasaki in scroll down for video russian president vladimir putin is reportedly planning to replace the country s older ss satan weapons with the new missiles amid a string of recent disagreements with the west the kremlin has stepped up the rhetoric against the west and carried a series of manoeuvres that has infuriated politicians in the us and uk the pictures were revealed online by chief designers from the makeyev rocket design bureau a message posted alongside the picture said in accordance with the decree of the russian government on the state defense order for and the planning period the makeyev rocket design bureau was instructed to start design and development work on the sarmat the rs sarmat missile is said to contain nuclear warheads and is capable of destroying an area the size of france or texas according to russian news network zvezda which is owned by russia s ministry of defence the weapon is also able to evade radar it is expected to have a range of miles km which would allow moscow to attack london and for entire article click link			
5	about time christian group sues amazon and splc for designation as hate group	<p>all we can say on this one is it s about time someone sued the southern poverty law center on tuesday d james kennedy ministries djkm filed a lawsuit against the southern poverty law center splc the charity navigation organization guidestar and amazon for defamation religious discrimination and trafficking in falsehood the splc listed djkm as a hate group while guidestar also categorized it in those terms and amazon kept the ministry off of its charity donation program amazon smile we embarked today on a journey to right a terrible wrong dr frank wright president and ceo at djkm said in a statement tuesday those who knowingly label christian ministries as hate groups solely for subscribing to the historic christian faith are either woefully uninformed or willfully deceitful in the case of the southern poverty law center our lawsuit alleges the latter the splc has labeled djkm an anti-lgbt hate group for its opposition to same-sex marriage and transgenderism these false and illegal characterizations have a chilling effect on the free exercise of religion and on religious free speech for all people of faith wright declared after having given the splc an opportunity to retract we have undertaken this legal action seeking a trial by a jury of our peers to preserve our own rights under the law and to defend the religious free speech rights of all americans the djkm president concluded the lawsuit laid out charges against the splc guidestar and amazon read more pjm</p>	1	en	en

```
In [ ]: #tokenize the text (separate sentences into individual words)
```

```
from nltk.tokenize import word_tokenize
```

```
In [ ]: cleandata['text'] = cleandata['text'].apply(word_tokenize)
cleandata['title'] = cleandata['title'].apply(word_tokenize)
```

```
In [ ]: #remove stop words
```

```
from nltk.corpus import stopwords

stopwords = stopwords.words('english')

cleandata['title'] = cleandata['title'].apply(lambda words: [word for word in words if word not in stopwords])
cleandata['text'] = cleandata['text'].apply(lambda words: [word for word in words if word not in stopwords])
```

```
In [ ]: #review results of tokenization
```

```
pd.set_option('max_colwidth', None)

cleandata.head()
```

Out[]:	Unnamed: 0	title	text	label	text_lang	title_lang
	0	[law, enforcement, high, alert, following, threats, cops, whites, blacklivesmatter, fyf, terrorists, video]	[comment, expected, barack, obama, members, fyf, fukyoflag, blacklivesmatter, movements, called, lynching, hanging, white, people, cops, encouraged, others, radio, show, tuesday, night, turn, tide, kill, white, people, cops, send, message, killing, black, people, america, one, f, yofflag, organizers, called, sunshine, radio, blog, show, hosted, texas, called, sunshine, f, ing, opinion, radio, show, snapshot, fyf, lolatwhitefear, twitter, page, p, shows, urging, supporters, call, fyf, tonight, continue, dismantle, illusion, white, snapshot, twitter, radio, call, invite, fyf, radio, show, aired, p, eastern, standard, time, show, callers, clearly, call, lynching, killing, white, people, minute, clip, radio, show, heard, provided, breitbart, texas, someone, would, like, referred, ...]	1	en	en
2	2	[unbelievable, obama, attorney, general, says, charlotte, rioters, peaceful, protesters, home, state, north, carolina, video]	[demonstrators, gathered, last, night, exercising, constitutional, protected, right, peaceful, protest, order, raise, issues, create, change, loretta, lynch, aka, eric, holder, skirt]	1	en	en
3	3	[bobby, jindal, raised, hindu, uses, story, christian, conversion, woo, evangelicals, potential, bid]	[dozen, politically, active, pastors, came, private, dinner, friday, night, hear, conversion, story, unique, context, presidential, politics, louisiana, gov, bobby, jindal, traveled, hinduism, protestant, christianity, ultimately, became, calls, evangelical, catholic, two, hours, jindal, recalled, talking, girl, high, school, wanted, save, soul, reading, bible, closet, parents, would, see, feeling, stir, watching, movie, senior, year, depicted, jesus, cross, struck, struck, hard, jindal, told, pastors, son, god, died, sins, jindal, session, christian, clergy, lead, congregations, early, presidential, battleground, states, iowa, south, carolina, part, behind-the-scenes, effort, louisiana, governor, find, political, base, could, help, propel, top, tier, republican, candidates, seeking, run, white, house, known, gop, circles, ...]	0	en	en
4	4	[satan, russia, unvelis, image, terrifying, new, supernuke, western, world, takes, notice]	[rs, sarmat, missile, dubbed, satan, replace, ss, flies, miles, km, per, sec, range, miles, km, weapons, perceived, part, increasingly, aggressive, russia, could, deliver, warhead, megatons, times, powerful, atom, bombs, dropped, hiroshima, nagasaki, libby, plummer, gareth, davie, russia, unveiled, chilling, pictures, largest, ever, nuclear, missile, capable, destroying, area, size, france, rs, sarmat, missile, dubbed, satan, nato, top, speed, miles, km, per, second, designed, outfox, anti-missile, shield, systems, new, sarmat, missile, could, deliver, warheads, megatons, times, powerful, atom, bombs, dropped, hiroshima, nagasaki, scroll, video, russian, president, vladimir, putin, reportedly, planning, replace, country, older, ss, satan, weapons, new, missiles, amid, string, recent, disagreements, ...]	1	en	en
5	5	[time, christian, group, sues, amazon, splc, designation, hate, group]	[say, one, time, someone, sued, southern, poverty, law, center, tuesday, james, kennedy, ministries, djkm, filed, lawsuit, southern, poverty, law, center, splc, charity, navigation, organization, guidestar, amazon, defamation, religious, discrimination, trafficking, falsehood, splc, listed, djkm, hate, group, guidestar, also, categorized, terms, amazon, kept, ministry, charity, donation, program, amazon, smile, embarked, today, journey, right, terrible, wrong, dr, frank, wright, president, ceo, djkm, said, statement, tuesday, knowingly, label, christian, ministries, hate, groups, solely, subscribing, historic, christian, faith, either, woefully, uninformed, willfully, deceitful, case, southern, poverty, law, center, lawsuit, alleges, latter, splc, labeled, djkm, anti-lgbt, hate, group, opposition, same-sex, marriage, transgenderism, false, illegal, characterizations, ...]	1	en	en

Lemmatize the data:

```
In [ ]: #create copy of data frame prior to Lemmatization and prepare function
lemdatad = cleandata.copy()

from nltk.stem import WordNetLemmatizer

lemmatize = WordNetLemmatizer()

In [ ]: lemdatad['title'] = lemdatad['title'].apply(lambda lst:[lemmatize.lemmatize(word) for word in lst])

In [ ]: lemdatad['title'] = lemdatad['title'].apply(lambda lst:[lemmatize.lemmatize(word, pos='v') for word in lst])

In [ ]: lemdatad['text'] = lemdatad['text'].apply(lambda lst:[lemmatize.lemmatize(word) for word in lst])

In [ ]: lemdatad['text'] = lemdatad['text'].apply(lambda lst:[lemmatize.lemmatize(word, pos='v') for word in lst])

In [ ]: #review results of Lemmatization
lemdatad.head()
```

Out[]:	Unnamed: 0	title	text	label	text_lang	title_lang
	0	[law, enforcement, high, alert, follow, threat, cop, white, blacklivesmatter, fyf, terrorist, video]	[comment, expect, barack, obama, member, fyf, fukyoflag, blacklivesmatter, movement, call, lynch, hang, white, people, cop, encourage, others, radio, show, tuesday, night, turn, tide, kill, white, people, cop, send, message, kill, black, people, america, one, f, yoflag, organizer, call, sunshine, radio, blog, show, host, texas, call, sunshine, f, ing, opinion, radio, show, snapshot, fyf, lolatwhitefear, twitter, page, p, show, urge, supporter, call, fyf, tonight, continue, dismantle, illusion, white, snapshot, twitter, radio, call, invite, fyf, radio, show, air, p, eastern, standard, time, show, caller, clearly, call, lynch, kill, white, people, minute, clip, radio, show, hear, provide, breitbart, texas, someone, would, like, refer, ...]	1	en	en
2	2	[unbelievable, obama, attorney, general, say, charlotte, rioter, peaceful, protester, home, state, north, carolina, video]	[demonstrator, gather, last, night, exercise, constitutional, protect, right, peaceful, protest, order, raise, issue, create, change, loretta, lynch, aka, eric, holder, skirt]	1	en	en
3	3	[bobby, jindal, raise, hindu, us, story, christian, conversion, woo, evangelicals, potential, bid]	[dozen, politically, active, pastor, come, private, dinner, friday, night, hear, conversion, story, unique, context, presidential, politics, louisiana, gov, bobby, jindal, travel, hinduism, protestant, christianity, ultimately, become, call, evangelical, catholic, two, hour, jindal, recall, talk, girl, high, school, want, save, soul, read, bible, closet, parent, would, see, feel, stir, watch, movie, senior, year, depict, jesus, cross, strike, strike, hard, jindal, tell, pastor, son, god, die, sin, jindal, session, christian, clergy, lead, congregation, early, presidential, battleground, state, iowa, south, carolina, part, behind-the-scenes, effort, louisiana, governor, find, political, base, could, help, propel, top, tier, republican, candidate, seek, run, white, house, know, gop, circle, ...]	0	en	en
4	4	[satan, russia, unvelis, image, terrify, new, supernuke, western, world, take, notice]	[r, sarmat, missile, dub, satan, replace, s, fly, mile, km, per, sec, range, mile, km, weapon, perceive, part, increasingly, aggressive, russia, could, deliver, warhead, megaton, time, powerful, atom, bomb, drop, hiroshima, nagasaki, libby, plummer, gareth, davie, russia, unveil, chill, picture, largest, ever, nuclear, missile, capable, destroy, area, size, france, r, sarmat, missile, dub, satan, nato, top, speed, mile, km, per, second, design, outfox, anti-missile, shield, system, new, sarmat, missile, could, deliver, warhead, megaton, time, powerful, atom, bomb, drop, hiroshima, nagasaki, scroll, video, russian, president, vladimir, putin, reportedly, plan, replace, country, older, s, satan, weapon, new, missile, amid, string, recent, disagreement, ...]	1	en	en
5	5	[time, christian, group, sue, amazon, splc, designation, hate, group]	[say, one, time, someone, sue, southern, poverty, law, center, tuesday, jam, kennedy, ministry, djkm, file, lawsuit, southern, poverty, law, center, splc, charity, navigation, organization, guidestar, amazon, defamation, religious, discrimination, traffic, falsehood, splc, list, djkm, hate, group, guidestar, also, categorize, term, amazon, keep, ministry, charity, donation, program, amazon, smile, embark, today, journey, right, terrible, wrong, dr, frank, wright, president, ceo, djkm, say, statement, tuesday, knowingly, label, christian, ministry, hate, group, solely, subscribe, historic, christian, faith, either, woefully, uninformed, willfully, deceitful, case, southern, poverty, law, center, lawsuit, allege, latter, splc, label, djkm, anti-lgbt, hate, group, opposition, same-sex, marriage, transgenderism, false, illegal, characterization, ...]	1	en	en

In []: #remove any rows with empty lists in the text/title

```
lemmdata = lemmdata[lemmdata['title'].str.len() != 0]
lemmdata = lemmdata[lemmdata['text'].str.len() != 0]
```

In []: #check new value counts

```
lemmdata.label.value_counts()
```

Out[]: label
0 32570
1 25984
Name: count, dtype: int64

Create and store variables for later use:

In []: #create new dataframe with text column

```
textdata = lemmdata[['text', 'label']].copy()
```

In []: textdata.head()

Out[]:

0	[comment, expect, barack, obama, member, fyf, fukyoflag, blacklivesmatter, movement, call, lynch, hang, white, people, cop, encourage, others, radio, show, tuesday, night, turn, tide, kill, white, people, cop, send, message, kill, black, people, america, one, f, yoflag, organizer, call, sunshine, radio, blog, show, host, texas, call, sunshine, f, ing, opinion, radio, show, snapshot, fyf, lolatwhitefear, twitter, page, p, show, urge, supporter, call, fyf, tonight, continue, dismantle, illusion, white, snapshot, twitter, radio, call, invite, fyf, radio, show, air, p, eastern, standard, time, show, caller, clearly, call, lynch, kill, white, people, minute, clip, radio, show, hear, provide, breitbart, texas, someone, would, like, refer, ...]	1	
2	[demonstrator, gather, last, night, exercise, constitutional, protect, right, peaceful, protest, order, raise, issue, create, change, loretta, lynch, aka, eric, holder, skirt]	1	
3	[dozen, politically, active, pastor, come, private, dinner, friday, night, hear, conversion, story, unique, context, presidential, politics, louisiana, gov, bobby, jindal, travel, hinduism, protestant, christianity, ultimately, become, call, evangelical, catholic, two, hour, jindal, recall, talk, girl, high, school, want, save, soul, read, bible, closet, parent, would, see, feel, stir, watch, movie, senior, year, depict, jesus, cross, strike, strike, hard, jindal, tell, pastor, son, god, die, sin, jindal, session, christian, clergy, lead, congregation, early, presidential, battleground, state, iowa, south, carolina, part, behind-the-scenes, effort, louisiana, governor, find, political, base, could, help, propel, top, tier, republican, candidate, seek, run, white, house, know, gop, circle, ...]	0	
4	[r, sarmat, missile, dub, satan, replace, s, fly, mile, km, per, sec, range, mile, km, weapon, perceive, part, increasingly, aggressive, russia, could, deliver, warhead, megaton, time, powerful, atom, bomb, drop, hiroshima, nagasaki, libby, plummer, gareth, davie, russia, unveil, chill, picture, largest, ever, nuclear, missile, capable, destroy, area, size, france, r, sarmat, missile, dub, satan, nato, top, speed, mile, km, per, second, design, outfox, anti-missile, shield, system, new, sarmat, missile, could, deliver, warhead, megaton, time, powerful, atom, bomb, drop, hiroshima, nagasaki, scroll, video, russian, president, vladimir, putin, reportedly, plan, replace, country, older, s, satan, weapon, new, missile, amid, string, recent, disagreement, ...]	1	
5	[say, one, time, someone, sue, southern, poverty, law, center, tuesday, jam, kennedy, ministry, djkm, file, lawsuit, southern, poverty, law, center, splc, charity, navigation, organization, guidestar, amazon, defamation, religious, discrimination, traffic, falsehood, splc, list, djkm, hate, group, guidestar, also, categorize, term, amazon, keep, ministry, charity, donation, program, amazon, smile, embark, today, journey, right, terrible, wrong, dr, frank, wright, president, ceo, djkm, say, statement, tuesday, knowingly, label, christian, ministry, hate, group, solely, subscribe, historic, christian, faith, either, woefully, uninformed, willfully, deceitful, case, southern, poverty, law, center, lawsuit, allege, latter, splc, label, djkm, anti-lgbt, hate, group, opposition, same-sex, marriage, transgenderism, false, illegal, characterization, ...]	1	

In []: #create new dataframe with title column

titledata = lemmdata[['title', 'label']].copy()

In []: titledata.head()

Out[]:

		title	label
0	[law, enforcement, high, alert, follow, threat, cop, white, blacklivesmatter, fyf, terrorist, video]	1	
2	[unbelievable, obama, attorney, general, say, charlotte, rioter, peaceful, protester, home, state, north, carolina, video]	1	
3	[bobby, jindal, raise, hindu, us, story, christian, conversion, woo, evangelicals, potential, bid]	0	
4	[satan, russia, unvelis, image, terrify, new, supernuke, western, world, take, notice]	1	
5	[time, christian, group, sue, amazon, splc, designation, hate, group]	1	

In []: #store new data frames for use in next notebook

%store textdata

%store titledata

Stored 'textdata' (DataFrame)

Stored 'titledata' (DataFrame)