

# Restauarnt\_Reviews

December 11, 2022

```
[ ]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

```
[ ]: df = pd.read_csv('https://raw.githubusercontent.com/subhashdixit/NLP/main/
↳Restaurant_Reviews/Restaurant_Reviews.tsv',delimiter = "\t")
```

```
[ ]: df
```

```
[ ]:
```

	Review	Liked
0	Wow... Loved this place.	1
1	Crust is not good.	0
2	Not tasty and the texture was just nasty.	0
3	Stopped by during the late May bank holiday of...	1
4	The selection on the menu was great and so wer...	1
..	...	...
995	I think food should have flavor and texture an...	0
996	Appetite instantly gone.	0
997	Overall I was not impressed and would not go b...	0
998	The whole experience was underwhelming, and I ...	0
999	Then, as if I hadn't wasted enough of my life ...	0

[1000 rows x 2 columns]

## 0.1 Text Cleaning

- Remove Punctuations, Numbers: Punctuations, Numbers don't help much in processing the given text, if included, they will just increase the size of a bag of words that we will create as the last step and decrease the efficiency of an algorithm
- Stemming: Take roots of the word
- Convert each word into its lower case: For example, it is useless to have some words in different cases (eg 'good' and 'GOOD')

```
[ ]: df.isnull().sum()
```

```
[ ]: Review    0
     Liked     0
     dtype: int64
```

**Stopwords** \* Stopwords are the words in any language which does not add much meaning to a sentence \* They can safely be ignored without sacrificing the meaning of the sentence. For some search engines, these are some of the most common, short function words, such as the, is, at, which, and on \* stop words can cause problems when searching for phrases that include them, particularly in names such as “The Who” or “Take That”

```
[ ]: """library to clean data"""
import re
"""Natural Language Tool Kit"""
import nltk
nltk.download('stopwords')
"""to remove stopwords"""
from nltk.corpus import stopwords
"""for Stemming propose"""
from nltk.stem.porter import PorterStemmer
```

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
```

```
[nltk_data] Package stopwords is already up-to-date!
```

```
[ ]: # Initialize empty array
     # to append clean text
     corpus = []

     for i in range(0, 1000):
         # column : "Review", row ith
         review = re.sub('[^a-zA-Z]', ' ', df['Review'][i])

         ## convert all cases to lower cases
         review = review.lower()

         ## split to array(default delimiter is " ")
         review = review.split()

         """creating PorterStemmer object to take main stem of each word"""
         ps = PorterStemmer()

         """loop for stemming each word in string array at ith row"""
         review = [ps.stem(word) for word in review
                    if not word in set(stopwords.words('english'))]

         """rejoin all string array elements to create back into a string"""
         review = ' '.join(review)

         """append each string to create array of clean text"""
```

```
corpus.append(review)
print(corpus)
```

['wow love place', 'crust good', 'tasti textur nasti', 'stop late may bank holiday rick steve recommend love', 'select menu great price', 'get angri want damn pho', 'honeslti tast fresh', 'potato like rubber could tell made ahead time kept warmer', 'fri great', 'great touch', 'servic prompt', 'would go back', 'cashier care ever say still end wayyy overpr', 'tri cape cod ravoli chicken cranberri mmmm', 'disgust pretti sure human hair', 'shock sign indic cash', 'highli recommend', 'waitress littl slow servic', 'place worth time let alon vega', 'like', 'burritto blah', 'food amaz', 'servic also cute', 'could care less interior beauti', 'perform', 'right red velvet cake ohhh stuff good', 'never brought salad ask', 'hole wall great mexican street taco friendli staff', 'took hour get food tabl restaur food luke warm sever run around like total overwhelm', 'worst salmon sashimi', 'also combo like burger fri beer decent deal', 'like final blow', 'found place accid could happier', 'seem like good quick place grab bite familiar pub food favor look elsewhere', 'overall like place lot', 'redeem qualiti restaur inexpens', 'ampl portion good price', 'poor servic waiter made feel like stupid everi time came tabl', 'first visit hiro delight', 'servic suck', 'shrimp tender moist', 'deal good enough would drag establish', 'hard judg whether side good gross melt styrofoam want eat fear get sick', 'posit note server attent provid great servic', 'frozen puck disgust worst peopl behind regist', 'thing like prime rib dessert section', 'bad food damn gener', 'burger good beef cook right', 'want sandwich go firehous', 'side greek salad greek dress tasti pita hummu refresh', 'order duck rare pink tender insid nice char outsid', 'came run us realiz husband left sunglass tabl', 'chow mein good', 'horribl attitud toward custom talk one custom enjoy food', 'portion huge', 'love friendli server great food wonder imagin menu', 'heart attack grill downtown vega absolut flat line excus restaur', 'much seafood like string pasta bottom', 'salad right amount sauc power scallop perfectli cook', 'rip banana rip petrifi tasteless', 'least think refil water struggl wave minut', 'place receiv star appet', 'cocktail handmad delici', 'definit go back', 'glad found place', 'great food servic huge portion give militari discount', 'always great time do gringo', 'updat went back second time still amaz', 'got food appar never heard salt batter fish chewi', 'great way finish great', 'deal includ tast drink jeff went beyond expect', 'realli realli good rice time', 'servic meh', 'took min get milkshak noth chocol milk', 'guess known place would suck insid excalibur use common sens', 'scallop dish quit appal valu well', 'time bad custom servic', 'sweet potato fri good season well', 'today second time lunch buffet pretti good', 'much good food vega feel cheat wast eat opportun go rice compani', 'come like experienc underwhelm relationship parti wait person ask break', 'walk place smell like old greas trap other eat', 'turkey roast beef bland', 'place', 'pan cake everyon rave tast like sugari disast tailor palat six year old', 'love pho spring roll oh yummi tri', 'poor batter meat ratio made chicken tender unsatisfi', 'say food amaz', 'omelet die', 'everyth fresh delici', 'summari larg disappoint dine experi', 'like realli sexi parti mouth outrag flirt hottest person parti', 'never hard rock casino never ever step forward', 'best breakfast

buffet', 'say bye bye tip ladi', 'never go', 'back', 'food arriv quickli',  
 'good', 'side cafe serv realli good food', 'server fantast found wife love roast  
 garlic bone marrow ad extra meal anoth marrow go', 'good thing waiter help kept  
 bloddi mari come', 'best buffet town price cannot beat', 'love mussel cook wine  
 reduct duck tender potato dish delici', 'one better buffet', 'went tigerlilli  
 fantast afternoon', 'food delici bartend attent person got great deal', 'ambien  
 wonder music play', 'go back next trip', 'sooooo good', 'real sushi lover let  
 honest yama good', 'least min pass us order food arriv busi', 'realli fantast  
 thai restaur definit worth visit', 'nice spici tender', 'good price', 'check',  
 'pretti gross', 'better atmospher', 'kind hard mess steak', 'although much like  
 look sound place actual experi bit disappoint', 'know place manag serv blandest  
 food ever eaten prepar indian cuisin', 'worst servic boot least worri', 'servic  
 fine waitress friendli', 'guy steak steak love son steak best worst place said  
 best steak ever eaten', 'thought ventur away get good sushi place realli hit  
 spot night', 'host staff lack better word bitch', 'bland like place number  
 reason want wast time bad review leav', 'phenomen food servic ambianc',  
 'return', 'definit worth ventur strip pork belli return next time vega', 'place  
 way overpr mediocr food', 'penn vodka excel', 'good select food includ massiv  
 meatloaf sandwich crispy chicken wrap delish tuna melt tasti burger', 'manag  
 rude', 'delici nyc bagel good select cream chees real lox caper even', 'great  
 subway fact good come everi subway meet expect', 'serious solid breakfast', 'one  
 best bar food vega', 'extrem rude realli mani restaur would love dine weekend  
 vega', 'drink never empti made realli great menu suggest', '', 'waiter help  
 friendli rare check us', 'husband ate lunch disappoint food servic', 'red curri  
 much bamboo shoot tasti', 'nice blanket moz top feel like done cover subpar  
 food', 'bathroom clean place well decor', 'menu alway chang food qualiti go  
 servic extrem slow', 'servic littl slow consid serv peopl server food come slow  
 pace', 'give thumb', 'watch waiter pay lot attent tabl ignor us', 'fianc came  
 middl day greet seat right away', 'great restaur mandalay bay', 'wait forti five  
 minut vain', 'crostini came salad stale', 'highlight great qualiti nigiri',  
 'staff friendli joint alway clean', 'differ cut piec day still wonder tender  
 well well flavor', 'order voodoo pasta first time realli excel pasta sinc go  
 gluten free sever year ago', 'place good', 'unfortun must hit bakeri leftov day  
 everyth order stale', 'came back today sinc reloc still impress', 'seat immedi',  
 'menu divers reason price', 'avoid cost', 'restaur alway full never wait',  
 'delici', 'place hand one best place eat phoenix metro area', 'go look good  
 food', 'never treat bad', 'bacon hella salti', 'also order spinach avocado salad  
 ingredi sad dress liter zero tast', 'realli vega fine dine use right menu hand  
 ladi price list', 'waitress friendli', 'lordi khao soi dish miss curri lover',  
 'everyth menu terrif also thrill made amaz accomod vegetarian daughter',  
 'perhap caught night judg review inspir go back', 'servic leav lot desir',  
 'atmospher modern hip maintain touch cozi', 'weekli haunt definit place come  
 back everi', 'liter sat minut one ask take order', 'burger absolut flavor meat  
 total bland burger overcook charcoal flavor', 'also decid send back waitress  
 look like verg heart attack', 'dress treat rude', 'probabl dirt', 'love place  
 hit spot want someth healthi lack quantiti flavor', 'order lemon raspberri ice  
 cocktail also incred', 'food suck expect suck could imagin', 'interest decor',  
 'realli like crepe station', 'also serv hot bread butter home made potato chip

bacon bit top origin good', 'watch prepar delici food', 'egg roll fantast',  
 'order arriv one gyro miss', 'salad wing ice cream dessert left feel quit  
 satisfi', 'realli sure joey vote best hot dog valley reader phoenix magazin',  
 'best place go tasti bowl pho', 'live music friday total blow', 'never insult  
 felt disrespect', 'friendli staff', 'worth drive', 'heard good thing place  
 exceed everi hope could dream', 'food great servic', 'warm beer help', 'great  
 brunch spot', 'servic friendli invit', 'good lunch spot', 'live sinc first last  
 time step foot place', 'worst experi ever', 'must night place', 'side delish mix  
 mushroom yukon gold pure white corn beateou', 'bug never show would given sure  
 side wall bug climb kitchen', 'minut wait salad realiz come time soon', 'friend  
 love salmon tartar', 'go back', 'extrem tasti', 'waitress good though', 'soggi  
 good', 'jamaican mojito delici', 'small worth price', 'food rich order  
 accordingli', 'shower area outsid rins take full shower unless mind nude everyon  
 see', 'servic bit lack', 'lobster bisqu bussel sprout risotto filet need salt  
 pepper cours none tabl', 'hope bode go busi someon cook come', 'either cold  
 enough flavor bad', 'love bacon wrap date', 'unbeliev bargain', 'folk otto alway  
 make us feel welcom special', 'main also uninspir', 'place first pho amaz',  
 'wonder experi made place must stop whenev town', 'food bad enough enjoy deal  
 world worst annoy drunk peopl', 'fun chef', 'order doubl cheeseburg got singl  
 patti fall apart pictur upload yeah still suck', 'great place coupl drink watch  
 sport event wall cover tv', 'possibl give zero star', 'descript said yum yum  
 sauc anoth said eel sauc yet anoth said spici mayo well none roll sauc', 'say  
 would hardest decis honestli dish tast suppos tast amaz', 'roll eye may stay  
 sure go back tri', 'everyon attent provid excel custom servic', 'horribl wast  
 time money', 'dish quit flavour', 'time side restaur almost empti excus', 'busi  
 either also build freez cold', 'like review said pay eat place', 'drink took  
 close minut come one point', 'serious flavor delight folk', 'much better ayc  
 sushi place went vega', 'light dark enough set mood', 'base sub par servic  
 receiv effort show gratitud busi go back', 'owner realli great peopl', 'noth  
 privileg work eat', 'greek dress creami flavor', 'overall think would take parent  
 place made similar complaint silent felt', 'pizza good peanut sauc tasti', 'tabl  
 servic pretti fast', 'fantast servic', 'well would given godfath zero star  
 possibl', 'know make', 'tough short flavor', 'hope place stick around', 'bar  
 vega ever recal charg tap water', 'restaur atmospher exquisit', 'good servic  
 clean inexpens boot', 'seafood fresh gener portion', 'plu buck', 'servic par  
 either', 'thu far visit twice food absolut delici time', 'good year ago', 'self  
 proclaim coffe cafe wildli disappoint', 'veggitarian platter world', 'cant go  
 wrong food', 'beat', 'stop place madison ironman friendli kind staff', 'chef  
 friendli good job', 'better dedic boba tea spot even jenni pho', 'like patio  
 servic outstand', 'goat taco skimp meat wow flavor', 'think', 'mac salad pretti  
 bland get', 'went bachi burger friend recommend disappoint', 'servic stink',  
 'wait wait', 'place qualiti sushi qualiti restaur', 'would definit recommend  
 wing well pizza', 'great pizza salad', 'thing went wrong burn saganaki', 'wait  
 hour breakfast could done time better home', 'place amaz', 'hate disagree fellow  
 yelper husband disappoint place', 'wait hour never got either pizza mani around  
 us came later', 'know slow', 'staff great food delish incred beer select', 'live  
 neighborhood disappoint back conveni locat', 'know pull pork could soooo  
 delici', 'get incred fresh fish prepar care', 'go gave star rate pleas know

third time eat bachi burger write review', 'love fact everyth menu worth',  
 'never dine place', 'food excel servic good', 'good beer drink select good food  
 select', 'pleas stay away shrimp stir fri noodl', 'potato chip order sad could  
 probabl count mani chip box probabl around', 'food realli bore', 'good servic  
 check', 'greedi corpor never see anoth dime', 'never ever go back', 'much like  
 go back get pass atroci servic never return', 'summer dine charm outdoor patio  
 delight', 'expect good', 'fantast food', 'order toast english muffin came  
 untoast', 'food good', 'never go back', 'great food price high qualiti hous  
 made', 'bu boy hand rude', 'point friend basic figur place joke mind make  
 publicli loudli known', 'back good bbq lighter fare reason price tell public  
 back old way', 'consid two us left full happi go wrong', 'bread made hous',  
 'downsid servic', 'also fri without doubt worst fri ever', 'servic except food  
 good review', 'coupl month later return amaz meal', 'favorit place town  
 shawarrrrrrma', 'black eye pea sweet potato unreal', 'disappoint', 'could serv  
 vinaigrett may make better overal dish still good', 'go far mani place never  
 seen restaur serv egg breakfast espec', 'mom got home immedi got sick bite  
 salad', 'server pleasant deal alway honor pizza hut coupon', 'truli unbeliev  
 good glad went back', 'fantast servic pleas atmospher', 'everyth gross', 'love  
 place', 'great servic food', 'first bathroom locat dirti seat cover replenish  
 plain yucki', 'burger got gold standard burger kind disappoint', 'omg food  
 delicioso', 'noth authent place', 'spaghetti noth special whatsoev', 'dish  
 salmon best great', 'veget fresh sauc feel like authent thai', 'worth drive  
 tucson', 'select probabl worst seen vega none', 'pretti good beer select',  
 'place like chipotl better', 'classi warm atmospher fun fresh appet succul steak  
 basebal steak', 'star brick oven bread app', 'eaten multipl time time food  
 delici', 'sat anoth ten minut final gave left', 'terribl', 'everyon treat equal  
 special', 'take min pancak egg', 'delici', 'good side staff genuin pleasant  
 enthusiast real treat', 'sadli gordon ramsey steak place shall sharpli avoid  
 next trip vega', 'alway even wonder food delici', 'best fish ever life',  
 'bathroom next door nice', 'buffet small food offer bland', 'outstand littl  
 restaur best food ever tast', 'pretti cool would say', 'definit turn doubt back  
 unless someon els buy', 'server great job handl larg rowdi tabl', 'find wast  
 food despic food', 'wife lobster bisqu soup lukewarm', 'would come back sushi  
 crave vega', 'staff great ambianc great', 'deserv star', 'left stomach ach felt  
 sick rest day', 'drop ball', 'dine space tini elegantli decor comfort', 'custom  
 order way like usual eggplant green bean stir fri love', 'bean rice mediocr  
 best', 'best taco town far', 'took back money got outta', 'interest part town  
 place amaz', 'rude inconsider manag', 'staff friendli wait time serv horribl one  
 even say hi first minut', 'back', 'great dinner', 'servic outshin definit  
 recommend halibut', 'food terribl', 'never ever go back told mani peopl happen',  
 'recommend unless car break front starv', 'come back everi time vega', 'place  
 deserv one star food', 'disgrac', 'def come back bowl next time', 'want healthi  
 authent ethic food tri place', 'continu come ladi night andddd date night highli  
 recommend place anyon area', 'sever time past experi alway great', 'walk away  
 stuf happi first vega buffet experi', 'servic excel price pretti reason consid  
 vega locat insid crystal shop mall aria', 'summar food incred nay transcend noth  
 bring joy quit like memori pneumat condiment dispens', 'probabl one peopl ever  
 go ian like', 'kid pizza alway hit lot great side dish option kiddo', 'servic

perfect famili atmospher nice see', 'cook perfect servic impecc', 'one simpli  
 disappoint', 'overall disappoint qualiti food bouchon', 'account know get screw',  
 'great place eat remind littl mom pop shop san francisco bay area', 'today first  
 tast bulldogi gourmet hot dog tell ever thought possibl', 'left frustrat',  
 'definit soon', 'food realli good got full petti fast', 'servic fantast', 'total  
 wast time', 'know kind best ice tea', 'come hungri leav happi stuf', 'servic  
 give star', 'assur disappoint', 'take littl bad servic food suck', 'gave tri eat  
 crust teeth still sore', 'complet gross', 'realli enjoy eat', 'first time go  
 think quickli becom regular', 'server nice even though look littl overwhelm need  
 stay profession friendli end', 'dinner companion told everyth fresh nice textur  
 tast', 'ground right next tabl larg smear step track everywher pile green bird  
 poop', 'furthermor even find hour oper websit', 'tri like place time think  
 done', 'mistak', 'complaint', 'serious good pizza expert connisseur topic',  
 'waiter jerk', 'strike want rush', 'nicest restaur owner ever come across',  
 'never come', 'love biscuit', 'servic quick friendli', 'order appet took minut  
 pizza anoth minut', 'absolutley fantast', 'huge awkward lb piec cow th gristl  
 fat', 'definit come back', 'like steiner dark feel like bar', 'wow spici  
 delici', 'familiar check', 'take busi dinner dollar elsewher', 'love go back',  
 'anyway fs restaur wonder breakfast lunch', 'noth special', 'day week differ  
 deal delici', 'mention combin pear almond bacon big winner', 'back', 'sauc  
 tasteless', 'food delici spici enough sure ask spicier prefer way', 'ribey steak  
 cook perfectli great mesquit flavor', 'think go back anytim soon', 'food  
 gooodd', 'far sushi connoisseur definit tell differ good food bad food certainli  
 bad food', 'insult', 'last time lunch bad', 'chicken wing contain driest chicken  
 meat ever eaten', 'food good enjoy everi mouth enjoy relax venu coupl small  
 famili group etc', 'nargil think great', 'best tater tot southwest', 'love  
 place', 'definit worth paid', 'vanilla ice cream creami smooth profiterol choux  
 pastri fresh enough', 'im az time new spot', 'manag worst', 'insid realli quit  
 nice clean', 'food outstand price reason', 'think run back carli anytim soon  
 food', 'due fact took minut acknowledg anoth minut get food kept forget thing',  
 'love margarita', 'first vega buffet disappoint', 'good though', 'one note  
 ventil could use upgrad', 'great pork sandwich', 'wast time', 'total letdown  
 would much rather go camelback flower shop cartel coffe', 'third chees friend  
 burger cold', 'enjoy pizza brunch', 'steak well trim also perfectli cook',  
 'group claim would handl us beauti', 'love', 'ask bill leav without eat bring  
 either', 'place jewel la vega exactli hope find nearli ten year live', 'seafood  
 limit boil shrimp crab leg crab leg definit tast fresh', 'select food best',  
 'delici absolut back', 'small famili restaur fine dine establish', 'toro tartar  
 cavier extraordinari like thinli slice wagyu white truffl', 'dont think back  
 long time', 'attach ga station rare good sign', 'awesom', 'back mani time soon',  
 'menu much good stuff could decid', 'wors humili worker right front bunch  
 horribl name call', 'conclus fill meal', 'daili special alway hit group',  
 'tragedi struck', 'pancak also realli good pretti larg', 'first crawfish experi  
 delici', 'monster chicken fri steak egg time favorit', 'waitress sweet funni',  
 'also tast mom multi grain pumpkin pancak pecan butter amaz fluffi delici',  
 'rather eat airlin food serious', 'cant say enough good thing place', 'ambianc  
 incred', 'waitress manag friendli', 'would recommend place', 'overall impress  
 noca', 'gyro basic lettuc', 'terribl servic', 'thoroughli disappoint', 'much

pasta love homemad hand made pasta thin pizza', 'give tri happi', 'far best  
 cheesecurd ever', 'reason price also', 'everyth perfect night', 'food good typic  
 bar food', 'drive get', 'first glanc love bakeri cafe nice ambianc clean  
 friendli staff', 'anyway think go back', 'point finger item menu order  
 disappoint', 'oh thing beauti restaur', 'gone go', 'greasi unhealthi meal',  
 'first time might last', 'burger amaz', 'similarli deliveri man say word apolog  
 food minut late', 'way expens', 'sure order dessert even need pack go tiramisu  
 cannoli die', 'first time wait next', 'bartend also nice', 'everyth good tasti',  
 'place two thumb way', 'best place vega breakfast check sat sun', 'love authent  
 mexican food want whole bunch interest yet delici meat choos need tri place',  
 'terribl manag', 'excel new restaur experienc frenchman', 'zero star would give  
 zero star', 'great steak great side great wine amaz dessert', 'worst martini  
 ever', 'steak shrimp opinion best entre gc', 'opportun today sampl amaz pizza',  
 'wait thirti minut seat although vacant tabl folk wait', 'yellowtail carpaccio  
 melt mouth fresh', 'tri go back even empti', 'go eat potato found stranger  
 hair', 'spici enough perfect actual', 'last night second time dine happi decid  
 go back', 'even hello right', 'dessert bit strang', 'boyfriend came first time  
 recent trip vega could pleas qualiti food servic', 'realli recommend place go  
 wrong donut place', 'nice ambianc', 'would recommend save room', 'guess mayb  
 went night disgrac', 'howev recent experi particular locat good', 'know like  
 restaur someth', 'avoid establish', 'think restaur suffer tri hard enough',  
 'tapa dish delici', 'heart place', 'salad bland vinegrett babi green heart  
 palm', 'two felt disgust', 'good time', 'believ place great stop huge belli  
 hanker sushi', 'gener portion great tast', 'never go back place never ever  
 recommend place anyon', 'server went back forth sever time even much help',  
 'food delici', 'hour serious', 'consid theft', 'eew locat need complet  
 overhaul', 'recent wit poor qualiti manag toward guest well', 'wait wait wait',  
 'also came back check us regularli excel servic', 'server super nice check us  
 mani time', 'pizza tast old super chewi good way', 'swung give tri deepli  
 disappoint', 'servic good compani better', 'staff also friendli effici', 'servic  
 fan quick serv nice folk', 'boy sucker dri', 'rate', 'look authent thai food go  
 els', 'steak recommend', 'pull car wait anoth minut acknowledg', 'great food  
 great servic clean friendli set', 'assur back', 'hate thing much cheap qualiti  
 black oliv', 'breakfast perpar great beauti present giant slice toast lightli  
 dust powder sugar', 'kid play area nasti', 'great place fo take eat', 'waitress  
 friendli happi accomod vegan veggi option', 'omg felt like never eaten thai food  
 dish', 'extrem crumbi pretti tasteless', 'pale color instead nice char flavor',  
 'crouton also tast homemad extra plu', 'got home see driest damn wing ever',  
 'regular stop trip phoenix', 'realli enjoy crema caf expand even told friend  
 best breakfast', 'good money', 'miss wish one philadelphia', 'got sit fairli  
 fast end wait minut place order anoth minut food arriv', 'also best chees crisp  
 town', 'good valu great food great servic', 'ask satisfi meal', 'food good',  
 'awesom', 'want leav', 'made drive way north scottsdal one bit disappoint',  
 'eat', 'owner realli realli need quit soooooo cheap let wrap freak sandwich two  
 paper one', 'check place coupl year ago impress', 'chicken got definit reheat ok  
 wedg cold soggi', 'sorri get food anytim soon', 'absolut must visit', 'cow tongu  
 cheek taco amaz', 'friend like bloodi mari', 'despit hard rate busi actual rare  
 give star', 'realli want make experi good one', 'return', 'chicken pho tast



bland', 'disappoint', 'grill chicken tender yellow saffron season', 'drive thru mean want wait around half hour food somehow end go make us wait wait', 'pretti awesom place', 'ambien perfect', 'best luck rude non custom servic focus new manag', 'grandmoth make roast chicken better one', 'ask multipl time wine list time ignor went hostess got one', 'staff alway super friendli help especi cool bring two small boy babi', 'four star food guy blue shirt great vibe still let us eat', 'roast beef sandwich tast realli good', 'even drastic sick', 'high qualiti chicken chicken caesar salad', 'order burger rare came done', 'promptli greet seat', 'tri go lunch madhous', 'proven dead wrong sushi bar qualiti great servic fast food impecc', 'wait hour seat greatest mood', 'good joint', 'macaron insan good', 'eat', 'waiter attent friendli inform', 'mayb cold would somewhat edibl', 'place lot promis fail deliv', 'bad experi', 'mistak', 'food averag best', 'great food', 'go back anytim soon', 'disappoint order big bay plater', 'great place relax awesom burger beer', 'perfect sit famili meal get togeth friend', 'much flavor poorli construct', 'patio seat comfort', 'fri rice dri well', 'hand favorit italian restaur', 'scream legit book somethat also pretti rare vega', 'fun experi', 'atmospher great love duo violinist play song request', 'person love hummu pita baklava falafel baba ganoush amaz eggplant', 'conveni sinc stay mgm', 'owner super friendli staff courteou', 'great', 'eclect select', 'sweet potato tot good onion ring perfect close', 'staff attent', 'chef gener time even came around twice take pictur', 'owner use work nobu place realli similar half price', 'googl mediocr imagin smashburg pop', 'dont go', 'promis disappoint', 'sushi lover avoid place mean', 'great doubl cheeseburg', 'awesom servic food', 'fantast neighborhood gem', 'wait go back', 'plantain worst ever tast', 'great place highli recommend', 'servic slow attent', 'gave star give star', 'staff spend time talk', 'dessert panna cotta amaz', 'good food great atmospher', 'damn good steak', 'total brunch fail', 'price reason flavor spot sauc home made slaw drench mayo', 'decor nice piano music soundtrack pleasant', 'steak amaz rge fillet relleno best seafood plate ever', 'good food good servic', 'absolut amaz', 'probabl back honest', 'definit back', 'sergeant pepper beef sandwich auju sauc excel sandwich well', 'hawaiian breez mango magic pineappl delight smoothi tri far good', 'went lunch servic slow', 'much say place walk expect amaz quickli disappoint', 'mortifi', 'needless say never back', 'anyway food definit fill price pay expect', 'chip came drip greas mostli edibl', 'realli impress strip steak', 'go sinc everi meal awesom', 'server nice attent serv staff', 'cashier friendli even brought food', 'work hospit industri paradis valley refrain recommend cibo longer', 'atmospher fun', 'would recommend other', 'servic quick even go order like like', 'mean realli get famou fish chip terribl', 'said mouth belli still quit pleas', 'thing', 'thumb', 'read pleas go', 'love grill pizza remind legit italian pizza', 'pro larg seat area nice bar area great simpl drink menu best brick oven pizza homemad dough', 'realli nice atmospher', 'tonight elk filet special suck', 'one bite hook', 'order old classic new dish go time sore disappoint everyth', 'cute quaint simpl honest', 'chicken delici season perfect fri outsid moist chicken insid', 'food great alway compliment chef', 'special thank dylan recommend order yummi tummi', 'awesom select beer', 'great food awesom servic', 'one nice thing ad gratuiti bill sinc parti larger expect tip', 'fli appl juic fli', 'han nan chicken also tasti', 'servic thought good', 'food bare lukewarm must sit wait server bring

us', 'ryan bar definit one edinburgh establish revisit', 'nicest chines restaur', 'overall like food servic', 'also serv indian naan bread hummu spici pine nut sauc world', 'probabl never come back recommend', 'friend pasta also bad bare touch', 'tri airport experi tasti food speedi friendli servic', 'love decor chines calligraphi wall paper', 'never anyth complain', 'restaur clean famili restaur feel', 'way fri', 'sure long stood long enough begin feel awkwardli place', 'open sandwich impress good way', 'back', 'warm feel servic felt like guest special treat', 'extens menu provid lot option breakfast', 'alway order vegetarian menu dinner wide array option choos', 'watch price inflat portion get smaller manag attitud grow rapidli', 'wonder lil tapa ambienc made feel warm fuzzzi insid', 'got enjoy seafood salad fabul vinegrett', 'wonton thin thick chewi almost melt mouth', 'level spici perfect spice whelm soup', 'sat right time server get go fantast', 'main thing enjoy crowd older crowd around mid', 'side town definit spot hit', 'wait minut get drink longer get arepa', 'great place eat', 'jalapeno bacon soooo good', 'servic poor that nice', 'food good servic good price good', 'place clean food oh stale', 'chicken dish ok beef like shoe leather', 'servic beyond bad', 'happi', 'tast like dirt', 'one place phoenix would defin go back', 'block amaz', 'close hous low key non fanci afford price good food', 'hot sour egg flower soup absolut star', 'sashimi poor qualiti soggi tasteless', 'great time famili dinner sunday night', 'food tasti say real tradit hunan style', 'bother slow servic', 'flair bartend absolut amaz', 'frozen margarita way sugari tast', 'good order twice', 'nutshel restaraunt smell like combin dirti fish market sewer', 'girlfriend veal bad', 'unfortun good', 'pretti satifi experi', 'join club get awesom offer via email', 'perfect someon like beer ice cold case even colder', 'bland flavorless good way describ bare tepid meat', 'chain fan beat place easili', 'nacho must', 'come back', 'mani word say place everyth pretti well', 'staff super nice quick even crazi crowd downtown juri lawyer court staff', 'great atmospher friendli fast servic', 'receiv pita huge lot meat thumb', 'food arriv meh', 'pay hot dog fri look like came kid meal wienerschnitzel idea good meal', 'classic main lobster roll fantast', 'brother law work mall ate day guess sick night', 'good go review place twice herea tribut place tribut event held last night', 'chip salsa realli good salsa fresh', 'place great', 'mediocr food', 'get insid impress place', 'super pissd', 'servic super friendli', 'sad littl veget overcook', 'place nice surpris', 'golden crispi delici', 'high hope place sinc burger cook charcoal grill unfortun tast fell flat way flat', 'could eat bruschetta day devin', 'singl employe came see ok even need water refil final serv us food', 'lastli mozzarella stick best thing order', 'first time ever came amaz experi still tell peopl awesom duck', 'server neglig need made us feel unwelcom would suggest place', 'servic terribl though', 'place overpr consist boba realli overpr', 'pack', 'love place', 'say dessert yummi', 'food terribl', 'season fruit fresh white peach pure', 'kept get wors wors offici done', 'place honestli blown', 'definit would eat', 'wast money', 'love put food nice plastic contain oppos cram littl paper takeout box', 'cr pe delic thin moist', 'aw servic', 'ever go', 'food qualiti horribl', 'price think place would much rather gone', 'servic fair best', 'love sushi found kabuki price hip servic', 'favor stay away dish', 'poor servic', 'one tabl thought food averag worth wait', 'best servic food ever maria server good friendli made day', 'excel', 'paid bill tip felt server terribl

job', 'lunch great experi', 'never bland food surpris consid articl read focus much spice flavor', 'food way overpr portion fuck small', 'recent tri caballero back everi week sinc', 'buck head realli expect better food', 'food came good pace', 'ate twice last visit especi enjoy salmon salad', 'back', 'could believ dirti oyster', 'place deserv star', 'would recommend place', 'fact go round star awesom', 'disbelief dish qualifi worst version food ever tast', 'bad day low toler rude custom servic peopl job nice polit wash dish otherwis', 'potato great biscuit', 'probabl would go', 'flavor perfect amount heat', 'price reason servic great', 'wife hate meal coconut shrimp friend realli enjoy meal either', 'fella got huevo ranchero look appeal', 'went happi hour great list wine', 'may say buffet pricey think get pay place get quit lot', 'probabl come back', 'worst food servic', 'place pretti good nice littl vibe restaur', 'talk great custom servic cours back', 'hot dish hot cold dish close room temp watch staff prepar food bare hand glove everyth deep fri oil', 'love fri bean', 'alway pleasur deal', 'plethora salad sandwich everyth tri get seal approv', 'place awesom want someth light healthi summer', 'sushi strip place go', 'servic great even manag came help tabl', 'feel dine room colleg cook cours high class dine servic slow best', 'start review two star edit give one', 'worst sushi ever eat besid costco', 'excel restaur highlight great servic uniqu menu beauti set', 'boyfriend sat bar complet delight experi', 'weird vibe owner', 'hardli meat', 'better bagel groceri store', 'go place gyro', 'love owner chef one authent japanes cool dude', 'burger good pizza use amaz doughi flavorless', 'found six inch long piec wire salsa', 'servic terribl food mediocr', 'defin enjoy', 'order albondiga soup warm tast like tomato soup frozen meatbal', 'three differ occas ask well done medium well three time got bloodiest piec meat plate', 'two bite refus eat anymor', 'servic extrem slow', 'minut wait got tabl', 'serious killer hot chai latt', 'allergi warn menu waitress absolut clue meal contain peanut', 'boyfriend tri mediterranean chicken salad fell love', 'rotat beer tap also highlight place', 'price bit concern mellow mushroom', 'worst thai ever', 'stay vega must get breakfast least', 'want first say server great perfect servic', 'pizza select good', 'strawberri tea good', 'highli unprofession rude loyal patron', 'overall great experi', 'spend money elsewher', 'regular toast bread equal satisfi occasion pat butter mmmmm', 'buffet bellagio far anticip', 'drink weak peopl', 'order correct', 'also feel like chip bought made hous', 'disappoint dinner went elsewher dessert', 'chip sal amaz', 'return', 'new fav vega buffet spot', 'serious cannot believ owner mani unexperienc employe run around like chicken head cut', 'sad', 'felt insult disrespect could talk judg anoth human like', 'call steakhous properli cook steak understand', 'impress concept food', 'thing crazi guacamol like pur ed', 'realli noth postino hope experi better', 'got food poison buffet', 'brought fresh batch fri think yay someth warm', 'hilari yummi christma eve dinner rememb biggest fail entir trip us', 'needless say go back anytim soon', 'place disgust', 'everi time eat see care teamwork profession degre', 'ri style calamari joke', 'howev much garlic fondu bare edibl', 'could bare stomach meal complain busi lunch', 'bad lost heart finish', 'also took forev bring us check ask', 'one make scene restaur get definit lost love one', 'disappoint experi', 'food par denni say good', 'want wait mediocr food downright terribl servic place', 'waaaaaayyyyyyyyyy rate say', 'go back', 'place fairli clean food simpli worth', 'place lack style', 'sangria

half glass wine full ridicul', 'bother come', 'meat pretti dri slice brisket pull pork', 'build seem pretti neat bathroom pretti trippi eat', 'equal aw', 'probabl hurri go back', 'slow seat even reserv', 'good stretch imagin', 'cashew cream sauc bland veget undercook', 'chipolt ranch dip saus tasteless seem thin water heat', 'bit sweet realli spici enough lack flavor', 'disappoint', 'place horribl way overpr', 'mayb vegetarian fare twice thought averag best', 'busi know', 'tabl outsid also dirti lot time worker alway friendli help menu', 'ambianc feel like buffet set douchey indoor garden tea biscuit', 'con spotti servic', 'fri hot neither burger', 'came back cold', 'food came disappoint ensu', 'real disappoint waiter', 'husband said rude even apolog bad food anyth', 'reason eat would fill night bing drink get carb stomach', 'insult profound deuchebaggeri go outsid smoke break serv solidifi', 'someon order two taco think may part custom servic ask combo ala cart', 'quit disappoint although blame need place door', 'rave review wait eat disappoint', 'del taco pretti nasti avoid possibl', 'hard make decent hamburg', 'like', 'hell go back', 'gotten much better servic pizza place next door servic receiv restaur', 'know big deal place back ya', 'immedi said want talk manag want talk guy shot firebal behind bar', 'ambianc much better', 'unfortun set us disappoint entre', 'food good', 'server suck wait correct server heimer suck', 'happen next pretti put', 'bad caus know famili own realli want like place', 'overpr get', 'vomit bathroom mid lunch', 'kept look time soon becom minut yet still food', 'place eat circumst would ever return top list', 'start tuna sashimi brownish color obvious fresh', 'food averag', 'sure beat nacho movi would expect littl bit come restaur', 'ha long bay bit flop', 'problem charg sandwich bigger subway sub offer better amount veget', 'shrimp unwrap live mile brushfir liter ice cold', 'lack flavor seem undercook dri', 'realli impress place close', 'would avoid place stay mirag', 'refri bean came meal dri crusti food bland', 'spend money time place els', 'ladi tabl next us found live green caterpillar salad', 'present food aw', 'tell disappoint', 'think food flavor textur lack', 'appetit instantli gone', 'overall impress would go back', 'whole experi underwhelm think go ninja sushi next time', 'wast enough life pour salt wound draw time took bring check']

```
[ ]: """Old and New Text"""
pd.concat([df,pd.DataFrame(corpus, columns=['New'])], axis = 1)
```

```
[ ]:
Review Liked \
0          Wow... Loved this place.          1
1          Crust is not good.                0
2          Not tasty and the texture was just nasty.  0
3  Stopped by during the late May bank holiday of...  1
4  The selection on the menu was great and so wer...  1
..          ...                               ...
995 I think food should have flavor and texture an...  0
996          Appetite instantly gone.          0
997 Overall I was not impressed and would not go b...  0
998 The whole experience was underwhelming, and I ...  0
999 Then, as if I hadn't wasted enough of my life ...  0
```

```

                                New
0                               wow love place
1                               crust good
2                               tasti textur nasti
3    stop late may bank holiday rick steve recommen...
4                               select menu great price
..                               ...
995                             think food flavor textur lack
996                             appetit instantli gone
997                             overal impress would go back
998 whole experi underwhelm think go ninja sushi n...
999 wast enough life pour salt wound draw time too...

```

[1000 rows x 3 columns]

## 0.2 Tokenization

- Involves splitting sentences and words from the body of the text
- Corpus – Body of text, singular. Corpora is the plural of this.
- Lexicon – Words and their meanings.
- Token – Each “entity” that is a part of whatever was split up based on rules. \* For examples, each word is a token when a sentence is “tokenized” into words. Each sentence can also be a token, if you tokenized the sentences out of a paragraph. So basically tokenizing involves splitting sentences and words from the body of the text.

```
[ ]: # libraries
from nltk.tokenize import sent_tokenize, word_tokenize
nltk.download("all")
```

```

[nltk_data] Downloading collection 'all'
[nltk_data] |
[nltk_data] | Downloading package abc to /root/nltk_data...
[nltk_data] | Package abc is already up-to-date!
[nltk_data] | Downloading package alpino to /root/nltk_data...
[nltk_data] | Package alpino is already up-to-date!
[nltk_data] | Downloading package averaged_perceptron_tagger to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package averaged_perceptron_tagger is already up-
[nltk_data] | to-date!
[nltk_data] | Downloading package averaged_perceptron_tagger_ru to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package averaged_perceptron_tagger_ru is already
[nltk_data] | up-to-date!
[nltk_data] | Downloading package basque_grammars to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package basque_grammars is already up-to-date!

```

```

[nltk_data] | Downloading package bcp47 to /root/nltk_data...
[nltk_data] | Package bcp47 is already up-to-date!
[nltk_data] | Downloading package biocreative_ppi to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package biocreative_ppi is already up-to-date!
[nltk_data] | Downloading package bllip_wsj_no_aux to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package bllip_wsj_no_aux is already up-to-date!
[nltk_data] | Downloading package book_grammars to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package book_grammars is already up-to-date!
[nltk_data] | Downloading package brown to /root/nltk_data...
[nltk_data] | Package brown is already up-to-date!
[nltk_data] | Downloading package brown_tei to /root/nltk_data...
[nltk_data] | Package brown_tei is already up-to-date!
[nltk_data] | Downloading package cess_cat to /root/nltk_data...
[nltk_data] | Package cess_cat is already up-to-date!
[nltk_data] | Downloading package cess_esp to /root/nltk_data...
[nltk_data] | Package cess_esp is already up-to-date!
[nltk_data] | Downloading package chat80 to /root/nltk_data...
[nltk_data] | Package chat80 is already up-to-date!
[nltk_data] | Downloading package city_database to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package city_database is already up-to-date!
[nltk_data] | Downloading package cmudict to /root/nltk_data...
[nltk_data] | Package cmudict is already up-to-date!
[nltk_data] | Downloading package comparative_sentences to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package comparative_sentences is already up-to-
[nltk_data] | date!
[nltk_data] | Downloading package comtrans to /root/nltk_data...
[nltk_data] | Package comtrans is already up-to-date!
[nltk_data] | Downloading package conll2000 to /root/nltk_data...
[nltk_data] | Package conll2000 is already up-to-date!
[nltk_data] | Downloading package conll2002 to /root/nltk_data...
[nltk_data] | Package conll2002 is already up-to-date!
[nltk_data] | Downloading package conll2007 to /root/nltk_data...
[nltk_data] | Package conll2007 is already up-to-date!
[nltk_data] | Downloading package crubadan to /root/nltk_data...
[nltk_data] | Package crubadan is already up-to-date!
[nltk_data] | Downloading package dependency_treebank to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package dependency_treebank is already up-to-date!
[nltk_data] | Downloading package dolch to /root/nltk_data...
[nltk_data] | Package dolch is already up-to-date!
[nltk_data] | Downloading package europarl_raw to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package europarl_raw is already up-to-date!

```

```

[nltk_data] | Downloading package extended_omw to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package extended_omw is already up-to-date!
[nltk_data] | Downloading package floresta to /root/nltk_data...
[nltk_data] | Package floresta is already up-to-date!
[nltk_data] | Downloading package framenet_v15 to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package framenet_v15 is already up-to-date!
[nltk_data] | Downloading package framenet_v17 to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package framenet_v17 is already up-to-date!
[nltk_data] | Downloading package gazetteers to /root/nltk_data...
[nltk_data] | Package gazetteers is already up-to-date!
[nltk_data] | Downloading package genesis to /root/nltk_data...
[nltk_data] | Package genesis is already up-to-date!
[nltk_data] | Downloading package gutenber to /root/nltk_data...
[nltk_data] | Package gutenber is already up-to-date!
[nltk_data] | Downloading package ieer to /root/nltk_data...
[nltk_data] | Package ieer is already up-to-date!
[nltk_data] | Downloading package inaugural to /root/nltk_data...
[nltk_data] | Package inaugural is already up-to-date!
[nltk_data] | Downloading package indian to /root/nltk_data...
[nltk_data] | Package indian is already up-to-date!
[nltk_data] | Downloading package jeita to /root/nltk_data...
[nltk_data] | Package jeita is already up-to-date!
[nltk_data] | Downloading package kimmo to /root/nltk_data...
[nltk_data] | Package kimmo is already up-to-date!
[nltk_data] | Downloading package knbc to /root/nltk_data...
[nltk_data] | Package knbc is already up-to-date!
[nltk_data] | Downloading package large_grammars to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package large_grammars is already up-to-date!
[nltk_data] | Downloading package lin_thesaurus to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package lin_thesaurus is already up-to-date!
[nltk_data] | Downloading package mac_morpho to /root/nltk_data...
[nltk_data] | Package mac_morpho is already up-to-date!
[nltk_data] | Downloading package machado to /root/nltk_data...
[nltk_data] | Package machado is already up-to-date!
[nltk_data] | Downloading package masc_tagged to /root/nltk_data...
[nltk_data] | Package masc_tagged is already up-to-date!
[nltk_data] | Downloading package maxent_ne_chunker to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package maxent_ne_chunker is already up-to-date!
[nltk_data] | Downloading package maxent_treebank_pos_tagger to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package maxent_treebank_pos_tagger is already up-
[nltk_data] | to-date!

```

```

[nltk_data] | Downloading package moses_sample to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package moses_sample is already up-to-date!
[nltk_data] | Downloading package movie_reviews to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package movie_reviews is already up-to-date!
[nltk_data] | Downloading package mte_teip5 to /root/nltk_data...
[nltk_data] | Package mte_teip5 is already up-to-date!
[nltk_data] | Downloading package mwa_ppdb to /root/nltk_data...
[nltk_data] | Package mwa_ppdb is already up-to-date!
[nltk_data] | Downloading package names to /root/nltk_data...
[nltk_data] | Package names is already up-to-date!
[nltk_data] | Downloading package nombank.1.0 to /root/nltk_data...
[nltk_data] | Package nombank.1.0 is already up-to-date!
[nltk_data] | Downloading package nonbreaking_prefixes to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package nonbreaking_prefixes is already up-to-date!
[nltk_data] | Downloading package nps_chat to /root/nltk_data...
[nltk_data] | Package nps_chat is already up-to-date!
[nltk_data] | Downloading package omw to /root/nltk_data...
[nltk_data] | Package omw is already up-to-date!
[nltk_data] | Downloading package omw-1.4 to /root/nltk_data...
[nltk_data] | Package omw-1.4 is already up-to-date!
[nltk_data] | Downloading package opinion_lexicon to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package opinion_lexicon is already up-to-date!
[nltk_data] | Downloading package panlex_swadesh to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package panlex_swadesh is already up-to-date!
[nltk_data] | Downloading package paradigms to /root/nltk_data...
[nltk_data] | Package paradigms is already up-to-date!
[nltk_data] | Downloading package pe08 to /root/nltk_data...
[nltk_data] | Package pe08 is already up-to-date!
[nltk_data] | Downloading package perluniprops to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package perluniprops is already up-to-date!
[nltk_data] | Downloading package pil to /root/nltk_data...
[nltk_data] | Package pil is already up-to-date!
[nltk_data] | Downloading package pl196x to /root/nltk_data...
[nltk_data] | Package pl196x is already up-to-date!
[nltk_data] | Downloading package porter_test to /root/nltk_data...
[nltk_data] | Package porter_test is already up-to-date!
[nltk_data] | Downloading package ppattach to /root/nltk_data...
[nltk_data] | Package ppattach is already up-to-date!
[nltk_data] | Downloading package problem_reports to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package problem_reports is already up-to-date!
[nltk_data] | Downloading package product_reviews_1 to

```



```

[nltk_data] | /root/nltk_data...
[nltk_data] | Package product_reviews_1 is already up-to-date!
[nltk_data] | Downloading package product_reviews_2 to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package product_reviews_2 is already up-to-date!
[nltk_data] | Downloading package propbank to /root/nltk_data...
[nltk_data] | Package propbank is already up-to-date!
[nltk_data] | Downloading package pros_cons to /root/nltk_data...
[nltk_data] | Package pros_cons is already up-to-date!
[nltk_data] | Downloading package ptb to /root/nltk_data...
[nltk_data] | Package ptb is already up-to-date!
[nltk_data] | Downloading package punkt to /root/nltk_data...
[nltk_data] | Package punkt is already up-to-date!
[nltk_data] | Downloading package qc to /root/nltk_data...
[nltk_data] | Package qc is already up-to-date!
[nltk_data] | Downloading package reuters to /root/nltk_data...
[nltk_data] | Package reuters is already up-to-date!
[nltk_data] | Downloading package rslp to /root/nltk_data...
[nltk_data] | Package rslp is already up-to-date!
[nltk_data] | Downloading package rte to /root/nltk_data...
[nltk_data] | Package rte is already up-to-date!
[nltk_data] | Downloading package sample_grammars to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package sample_grammars is already up-to-date!
[nltk_data] | Downloading package semcor to /root/nltk_data...
[nltk_data] | Package semcor is already up-to-date!
[nltk_data] | Downloading package senseval to /root/nltk_data...
[nltk_data] | Package senseval is already up-to-date!
[nltk_data] | Downloading package sentence_polarity to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package sentence_polarity is already up-to-date!
[nltk_data] | Downloading package sentiwordnet to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package sentiwordnet is already up-to-date!
[nltk_data] | Downloading package shakespeare to /root/nltk_data...
[nltk_data] | Package shakespeare is already up-to-date!
[nltk_data] | Downloading package sinica_treebank to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package sinica_treebank is already up-to-date!
[nltk_data] | Downloading package smultron to /root/nltk_data...
[nltk_data] | Package smultron is already up-to-date!
[nltk_data] | Downloading package snowball_data to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package snowball_data is already up-to-date!
[nltk_data] | Downloading package spanish_grammars to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package spanish_grammars is already up-to-date!
[nltk_data] | Downloading package state_union to /root/nltk_data...

```

```

[nltk_data] | Package state_union is already up-to-date!
[nltk_data] | Downloading package stopwords to /root/nltk_data...
[nltk_data] | Package stopwords is already up-to-date!
[nltk_data] | Downloading package subjectivity to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package subjectivity is already up-to-date!
[nltk_data] | Downloading package swadesh to /root/nltk_data...
[nltk_data] | Package swadesh is already up-to-date!
[nltk_data] | Downloading package switchboard to /root/nltk_data...
[nltk_data] | Package switchboard is already up-to-date!
[nltk_data] | Downloading package tagsets to /root/nltk_data...
[nltk_data] | Package tagsets is already up-to-date!
[nltk_data] | Downloading package timit to /root/nltk_data...
[nltk_data] | Package timit is already up-to-date!
[nltk_data] | Downloading package toolbox to /root/nltk_data...
[nltk_data] | Package toolbox is already up-to-date!
[nltk_data] | Downloading package treebank to /root/nltk_data...
[nltk_data] | Package treebank is already up-to-date!
[nltk_data] | Downloading package twitter_samples to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package twitter_samples is already up-to-date!
[nltk_data] | Downloading package udhr to /root/nltk_data...
[nltk_data] | Package udhr is already up-to-date!
[nltk_data] | Downloading package udhr2 to /root/nltk_data...
[nltk_data] | Package udhr2 is already up-to-date!
[nltk_data] | Downloading package unicode_samples to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package unicode_samples is already up-to-date!
[nltk_data] | Downloading package universal_tagset to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package universal_tagset is already up-to-date!
[nltk_data] | Downloading package universal_treebanks_v20 to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package universal_treebanks_v20 is already up-to-
[nltk_data] | date!
[nltk_data] | Downloading package vader_lexicon to
[nltk_data] | /root/nltk_data...
[nltk_data] | Package vader_lexicon is already up-to-date!
[nltk_data] | Downloading package verbnet to /root/nltk_data...
[nltk_data] | Package verbnet is already up-to-date!
[nltk_data] | Downloading package verbnet3 to /root/nltk_data...
[nltk_data] | Package verbnet3 is already up-to-date!
[nltk_data] | Downloading package webtext to /root/nltk_data...
[nltk_data] | Package webtext is already up-to-date!
[nltk_data] | Downloading package wmt15_eval to /root/nltk_data...
[nltk_data] | Package wmt15_eval is already up-to-date!
[nltk_data] | Downloading package word2vec_sample to
[nltk_data] | /root/nltk_data...

```

```

[nltk_data] | Package word2vec_sample is already up-to-date!
[nltk_data] | Downloading package wordnet to /root/nltk_data...
[nltk_data] | Package wordnet is already up-to-date!
[nltk_data] | Downloading package wordnet2021 to /root/nltk_data...
[nltk_data] | Package wordnet2021 is already up-to-date!
[nltk_data] | Downloading package wordnet31 to /root/nltk_data...
[nltk_data] | Package wordnet31 is already up-to-date!
[nltk_data] | Downloading package wordnet_ic to /root/nltk_data...
[nltk_data] | Package wordnet_ic is already up-to-date!
[nltk_data] | Downloading package words to /root/nltk_data...
[nltk_data] | Package words is already up-to-date!
[nltk_data] | Downloading package ycoe to /root/nltk_data...
[nltk_data] | Package ycoe is already up-to-date!
[nltk_data] |
[nltk_data] Done downloading collection all

```

```
[ ]: True
```

```

[ ]: # print(sent_tokenize(corpus))
# print(word_tokenize(corpus))
# word_tokenize(corpus[0])
# for i in corpus:
#     print(i)
# word_tokenize(str(corpus))
# X = df.iloc[:, :1]
# y = df.iloc[:, -1:]

```

### 0.3 Bag of words

- Bag of Words model is used to preprocess the text by converting it into a bag of words, which keeps a count of the total occurrences of most frequently used words

```

[ ]: # Creating the Bag of Words model
from sklearn.feature_extraction.text import CountVectorizer

# To extract max 1500 feature.
# "max_features" is attribute to
# experiment with to get better results
cv = CountVectorizer(max_features = 5)

# X contains corpus (dependent variable)
X = cv.fit_transform(corpus).toarray()

# y contains answers if review
# is positive or negative
y = df.iloc[:, 1].values

```

## 0.4 Train Test Split

```
[ ]: from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.25,
↪random_state = 1)
```

## 0.5 Random Forest

```
[ ]: from sklearn.ensemble import RandomForestClassifier
model = RandomForestClassifier(n_estimators = 501, criterion = 'entropy')
model.fit(X_train, y_train)
```

```
[ ]: RandomForestClassifier(criterion='entropy', n_estimators=501)
```

```
[ ]: y_pred = model.predict(X_test)
```

## 0.6 Accuracy

- TRUE POSITIVE : measures the proportion of actual positives that are correctly identified.
- TRUE NEGATIVE : measures the proportion of actual positives that are not correctly identified.
- FALSE POSITIVE : measures the proportion of actual negatives that are correctly identified.
- FALSE NEGATIVE : measures the proportion of actual negatives that are not correctly identified.
- Note: True or False refers to the assigned classification being Correct or Incorrect, while Positive or Negative refers to assignment to the Positive or the Negative Category

```
[ ]: from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test, y_pred)
cm
```

```
[ ]: array([[119, 15],
          [ 69, 47]])
```

```
[ ]: from sklearn.metrics import classification_report
```

```
[ ]: print(classification_report(y_test, y_pred))
```

	precision	recall	f1-score	support
0	0.63	0.89	0.74	134
1	0.76	0.41	0.53	116
accuracy			0.66	250
macro avg	0.70	0.65	0.63	250

weighted avg          0.69          0.66          0.64          250

## 0.7 Sample Prediction

```
[ ]: def Predict_Sentiment(sample_review):  
    word = []  
    sample_review = re.sub('[^a-zA-Z]', ' ', sample_review)  
    sample_review = sample_review.lower()  
    sample_review = sample_review.split()  
    """creating PorterStemmer object to take main stem of each word"""  
    ps = PorterStemmer()  
    """loop for stemming each word in string array at ith row"""  
    sample_review = [ps.stem(word) for word in sample_review if not word in   
↪set(stopwords.words('english'))]  
    """rejoin all string array elements to create back into a string"""  
    sample_review = ' '.join(sample_review)  
    """append each string to create array of clean text"""  
    word.append(sample_review)  
    temp = cv.transform(word).toarray()  
    return model.predict(temp)[0]  
  
[ ]: if (Predict_Sentiment("Food was very good an dit was pretty as well")):  
    print("Good Review")  
else:  
    print("Bad Review")
```

Good Review

## 0.8 Save Model

```
[ ]: import pickle  
pickle.dump(model, open("model_random.pkl", "wb"))
```

# 1 THE END