# Yelp-Reviews-NLP-Project

Deepti Gupta, Shashank Tiwari, Bhargavi Gutta

2/25/2020

## Objective

The entrepreneurs in the restaurant business can use reviews provided by their users in order to unlock the hidden value of text to understand their customers' opinions and needs and make better, more informed, business decisions. This study is an analysis of the Yelp Reviews provided by Yelp fusion business API of Indian restaurants in New York study. This study will let us harness the power of analytics and Natural Language Processing to extract meaning from text and dive into opinions of customers and see them outside of the often-controlled environment of a survey. The analysis will include sentiment analysis using label classification and word cloud formation capturing the most important words used in the Indian food community.

## Method and Data Variables

The participants in this study are the 50 food trucks across the New York City County (Manhattan, NY) whose business data was extracted from an API request made to Yelp Fusion. The sample was retrieved by searching "Indian Food" tag which returned top 50 records per API request and is a convenience sample. The criteria specified under the request is first the business vendor should be a licensed food business, situated with in the county of New York City, i.e. Manhattan, New York, NY and business entry should exist under Yelp. The first request is using Business Search API which returned the Business Ids which was used further to request the second API - User Reviews which returned the top 3 reviews. Following are the set of columns combined from both the APIs to be used under this study. BusinessName - is the official name of the business, BusinessRating - is the overall business rating on yelp, Name - is the user name who provided the review, UserRating - is the rating provided by the respective user, and lastly and most importantly, UserReview - user review text provided by the user. Total observations in this study are 150, consisting of 50 unique businesses with 3 reviews of each of it.

```r
library(reticulate)
```

```
## Warning: package 'reticulate' was built under R version 3.6.2
```

```python
#py_install("yelp", pip = T)
#py_install("yelpapi", pip = T)

import yelpapi
import argparse

import json
import http.client
import urllib.parse
from yelp.client import Client

import spacy
nlp = spacy.load("C:\Program Files\Python38\Lib\site-packages\en_core_web_sm\en_core_web_sm-2.2.0")
```

```python
import pandas as pd
import nltk

from __future__ import print_function
import requests

import unicodedata    # Used for processing raw texts

import contractions   # Used to remove contractions
import re
from contractions import contractions_dict

from nltk.corpus import stopwords

from textblob import Word
from textblob import TextBlob
```

# Section 1. Load the Data

**Get Business IDs and user reviews for these business ids.**

```python
#
#CLIENT_KEY = "mkiXAI4ESi8xsGpizf7u0g"
#MY_API_KEY = "zUHqRv5saa5CrIhddvDahzEsETysJgFkmyESbYzopq_Zh7Ca5wvtUyYWilD91h_Ff0WkownQoP4oz4z1KwwVPfnr
#
#business_url='https://api.yelp.com/v3/businesses/search'
#
#headers = {
#        'Authorization': 'Bearer {}'.format(MY_API_KEY),
#    }
#
#def yelp_search(term): # This function launches the request for all grocery location endpoints
#    url_params = {
#        'term':term,'location':'New York City','limit':50
#     }
#
#    response = requests.get(business_url, headers=headers, params=url_params)
#    return response.json()
#
#output_json1 = yelp_search('Indian Food')
#
#df_first1 = pd.DataFrame.from_dict(output_json1['businesses'])
#
#businesses = output_json1["businesses"]
#
#cols = ['BusinessName',"BusinessRating",'UserName',"UserRating","UserReview"]
#lst = []
#
#for business in businesses:
#    id = business["id"]
#    url="https://api.yelp.com/v3/businesses/" + id + "/reviews"
#    response = requests.get(url, headers=headers)
```

```
#
#     output = json.loads(response.text)
#
#     reviews = output["reviews"]
#     for review in reviews:
#         lst.append((business["name"],business["rating"], review["user"]["name"], review["rating"], revi
#
#bus_review_sub1 = pd.DataFrame(lst,columns=cols)
#
#bus_review_sub1.to_csv("C:\\Study\\520-NaturalLanguageProcessing\\Project\\data\\business_reviews.csv"
```

## Section2. Checkpoint file

```
bus_review_sub2 = pd.read_csv("C:\\Study\\520-NaturalLanguageProcessing\\Project\\data\\business_reviews
bus_review_sub2
```

```
##                       BusinessName  ...                                         UserReview
## 0                   The MasalaWala  ...  Quite possibly the best Indian food I've had i...
## 1                   The MasalaWala  ...  A okay hipster Indian spot in the LES.  Food w...
## 2                   The MasalaWala  ...  I came to Masalawala on Christmas Eve with my ...
## 3          Bengal Tiger Indian Food  ...  What an incredible find in the heart of midtow...
## 4          Bengal Tiger Indian Food  ...  The restaurant is slightly hidden as it's on t...
## ..                             ...  ...                                                ...
## 145                       Nimbooda  ...  Solid Indian food, but nothing spectacular or ...
## 146                       Nimbooda  ...  I've ordered takeout here twice and found it t...
## 147   Clove Indian Restaurant & Bar  ...  The food is great! U can't say anything negati...
## 148   Clove Indian Restaurant & Bar  ...  I ordered the Vegetable Samosas and Tandoori C...
## 149   Clove Indian Restaurant & Bar  ...  YUM to the Clove Iced Tea! That was a spiced u...
##
## [150 rows x 5 columns]
```

## Section3. Processing raw text

### 3.1 Remove special symbols

```
#--------------------------------------------------------------------------------
# Using the library - `unicodedata` in python to remove any symbols from your text

# Will convert to lowercase after contractions since keys in contractions can have upper case letters l


def remove_accented_chars(text):
  text = unicodedata.normalize('NFKD', text).encode('ascii', 'ignore').decode('utf-8', 'ignore')
  return text

bus_review_sub2['No_Symbols'] = bus_review_sub2.UserReview.apply(lambda x: remove_accented_chars(x))
```

### 3.2 Apply Contractions

```
#Modify month keys in contraction_dict since they are throwing errors with .(fullstop)
if contractions_dict.get("jan.") != None:
```

```python
    contractions_dict['jan'] = contractions_dict.pop('jan.')

if contractions_dict.get("feb.") != None:
    contractions_dict['feb'] = contractions_dict.pop('feb.')

if contractions_dict.get("mar.") != None:
    contractions_dict['mar'] = contractions_dict.pop('mar.')

if contractions_dict.get("apr.") != None:
    contractions_dict['apr'] = contractions_dict.pop('apr.')

if contractions_dict.get("jun.") != None:
    contractions_dict['jun'] = contractions_dict.pop('jun.')

if contractions_dict.get("jul.") != None:
    contractions_dict['jul'] = contractions_dict.pop('jul.')

if contractions_dict.get("aug.") != None:
    contractions_dict['aug'] = contractions_dict.pop('aug.')

if contractions_dict.get("sep.") != None:
    contractions_dict['sep'] = contractions_dict.pop('sep.')

if contractions_dict.get("oct.") != None:
    contractions_dict['oct'] = contractions_dict.pop('oct.')

if contractions_dict.get("nov.") != None:
    contractions_dict['nov'] = contractions_dict.pop('nov.')

if contractions_dict.get("dec.") != None:
    contractions_dict['dec'] = contractions_dict.pop('dec.')

contractions_re = re.compile('|'.join(contractions_dict.keys()))

def expand_contractions(s, contractions_dict=contractions_dict):
  def replace(match):
    return contractions_dict[match.group(0)]
  return contractions_re.sub(replace, s)


bus_review_sub2['No_Contractions'] = bus_review_sub2.No_Symbols.apply(lambda x: expand_contractions(x))
```

### 3.3 Correct Misspellings

```python
Spelling_Corrected = []

for line in bus_review_sub2.No_Contractions:
    text = TextBlob(line).correct()
    Spelling_Corrected.append(str(text))

bus_review_sub2["Spelling"] = Spelling_Corrected
```

**3.4 Lemmetization**

```
library(textstem)
```

```
## Warning: package 'textstem' was built under R version 3.6.2
## Loading required package: koRpus.lang.en
## Warning: package 'koRpus.lang.en' was built under R version 3.6.2
## Loading required package: koRpus
## Warning: package 'koRpus' was built under R version 3.6.2
## Loading required package: sylly
## Warning: package 'sylly' was built under R version 3.6.2
## For information on available language packages for 'koRpus', run
##
##    available.koRpus.lang()
##
## and see ?install.koRpus.lang()
```

```
py$bus_review_sub2$Lemmatized = lemmatize_strings(py$bus_review_sub2$Spelling)
```

**3.5 Remove Stopwords**

```
stop = stopwords.words('english')
```

```
bus_review_sub2['No_Stop_Words'] = bus_review_sub2['Lemmatized'].apply(lambda x: ' '.join([word for word
stop
```

```
## ['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', "you're", "you've", "you'll", ";
```

**3.6 Convert to lowercase**

```
bus_review_sub2['Lower_Case'] = bus_review_sub2.No_Stop_Words.str.lower()
```

```
#View(py$bus_review_sub2)
head(py$bus_review_sub2)
```

```
##               BusinessName BusinessRating     UserName UserRating
## 1            The MasalaWala            4.5 Surashree K.          5
## 2            The MasalaWala            4.5         Li J.          3
## 3            The MasalaWala            4.5 Alexandra C.          3
## 4 Bengal Tiger Indian Food            4.5    Jonghan L.          5
## 5 Bengal Tiger Indian Food            4.5     Teresa D.          4
## 6 Bengal Tiger Indian Food            4.5     Heidi M.          4
##
## 1  Quite possibly the best Indian food I've had in America. We enjoyed pretty much everything we ord
## 2 A okay hipster Indian spot in the LES.  Food was perfectly fine.  Spicy but not all that flavorful
## 3   I came to Masalawala on Christmas Eve with my sister. It had come with mixed reviews from friend:
## 4    What an incredible find in the heart of midtown! Came here at 545pm on a Sunday with a party of
## 5   The restaurant is slightly hidden as it's on the second floor of what looks like an apartment wa
## 6        I wasn't completely blown away by the food here, but it was certainly very satisfactory and
##
## 1  Quite possibly the best Indian food I've had in America. We enjoyed pretty much everything we ord
```

```
## 2 A okay hipster Indian spot in the LES.  Food was perfectly fine.  Spicy but not all that flavorful
## 3   I came to Masalawala on Christmas Eve with my sister. It had come with mixed reviews from friend
## 4    What an incredible find in the heart of midtown! Came here at 545pm on a Sunday with a party of
## 5   The restaurant is slightly hidden as it's on the second floor of what looks like an apartment wal
## 6        I wasn't completely blown away by the food here, but it was certainly very satisfactory and
##
## 1 Quite possibly the best Indian food I have had in America. We enjoyed pretty much everything we ord
## 2  A okay hipster Indian spot in the LES.  Food was perfectly fine.  Spicy but not all that flavorful
## 3   I came to Masalawala on Christmas Eve with my sister. It had come with mixed reviews from friend
## 4     What an incredible find in the heart of midtown! Came here at 545pm on a Sunday with a party of
## 5   The restaurant is slightly hidden as it is on the second floor of what looks like an apartment wa
## 6        I was not completely blown away by the food here, but it was certainly very satisfactory and
##
## 1 Quite possibly the best Indian food I have had in America. He enjoyed pretty much everything we ord
## 2  A okay sister Indian spot in the LES.  Good was perfectly fine.  Pity but not all that flavorful.
## 3   I came to Masalawala on Christmas Ve with my sister. It had come with mixed reviews from friends
## 4    That an incredible find in the heart of milton! Same here at 545pm on a Sunday with a party of
## 5 The restaurant is slightly hidden as it is on the second floor of what looks like an apartment wall
## 6        I was not completely blown away by the food here, but it was certainly very satisfactory and
##
## 1 Quite possibly the good Indian food I have have in America. He enjoy pretty much everything we orde
## 2            A okay sister Indian spot in the LES. Good be perfectly fine. Pity but not all that flavo
## 3     I come to Masalawala on Christmas have with my sister. It have come with mix review from friend
## 4     That a incredible find in the heart of milton! Same here at 545pm on a Sunday with a party of
## 5        The restaurant be slightly hide as it be on the 2 floor of what look like a apartment walk up
## 6        I be not completely blow away by the food here, but it be certainly very satisfactory and ve
##
## 1 Quite possibly good Indian food I America. He enjoy pretty much everything order. I highly recommen
## 2                 A okay sister Indian spot LES. Good perfectly fine. Pity flavorful. Very clean. Me
## 3                            I come Masalawala Christmas sister. It come mix review friend fan
## 4                   That incredible find heart milton! Same 545pm Sunday party two restaurant
## 5                  The restaurant slightly hide 2 floor look like apartment walk up. Once insi
## 6                  I completely blow away food here, certainly satisfactory unfordable milton.
##
## 1 quite possibly good indian food i america. he enjoy pretty much everything order. i highly recommen
## 2                 a okay sister indian spot les. good perfectly fine. pity flavorful. very clean. me
## 3                            i come masalawala christmas sister. it come mix review friend fan
## 4                   that incredible find heart milton! same 545pm sunday party two restaurant
## 5                  the restaurant slightly hide 2 floor look like apartment walk up. once insi
## 6                  i completely blow away food here, certainly satisfactory unfordable milton.
```

6