

level-3

June 28, 2024

1 Task 1

Restaurant Reviews 1. Analyze the text reviews to identify the most common positive and negative keywords.

```
[2]: %pip install nltk
```

```
Defaulting to user installation because normal site-packages is not
writeableNote: you may need to restart the kernel to use updated packages.
```

```
Collecting nltk
```

```
  Downloading nltk-3.8.1-py3-none-any.whl.metadata (2.8 kB)
```

```
Requirement already satisfied: click in c:\users\nagap\appdata\local\packages\py
thonsoftwarefoundation.python.3.12_qbz5n2kfra8p0\localcache\local-
packages\python312\site-packages (from nltk) (8.1.7)
```

```
Requirement already satisfied: joblib in c:\users\nagap\appdata\local\packages\p
ythonsoftwarefoundation.python.3.12_qbz5n2kfra8p0\localcache\local-
packages\python312\site-packages (from nltk) (1.3.2)
```

```
Collecting regex>=2021.8.3 (from nltk)
```

```
  Downloading regex-2024.5.15-cp312-cp312-win_amd64.whl.metadata (41 kB)
```

```
----- 0.0/42.0 kB ? eta -:--:--
```

```
----- 41.0/42.0 kB 960.0 kB/s eta 0:00:01
```

```
----- 42.0/42.0 kB 675.2 kB/s eta 0:00:00
```

```
Requirement already satisfied: tqdm in c:\users\nagap\appdata\local\packages\pyt
honsoftwarefoundation.python.3.12_qbz5n2kfra8p0\localcache\local-
packages\python312\site-packages (from nltk) (4.66.4)
```

```
Requirement already satisfied: colorama in c:\users\nagap\appdata\local\packages
\pythonsoftwarefoundation.python.3.12_qbz5n2kfra8p0\localcache\local-
packages\python312\site-packages (from click->nltk) (0.4.6)
```

```
Downloading nltk-3.8.1-py3-none-any.whl (1.5 MB)
```

```
----- 0.0/1.5 MB ? eta -:--:--
```

```
----- 0.3/1.5 MB 6.5 MB/s eta 0:00:01
```

```
----- 1.0/1.5 MB 10.1 MB/s eta 0:00:01
```

```
----- 1.5/1.5 MB 10.6 MB/s eta 0:00:00
```

```
Downloading regex-2024.5.15-cp312-cp312-win_amd64.whl (268 kB)
```

```
----- 0.0/268.5 kB ? eta -:--:--
```

```
----- 268.5/268.5 kB 17.2 MB/s eta 0:00:00
```

```
Installing collected packages: regex, nltk
```

Successfully installed nltk-3.8.1 regex-2024.5.15

[notice] A new release of pip is available: 24.0 -> 24.1.1

[notice] To update, run: C:\Users\nagap\AppData\Local\Microsoft\WindowsApps\PythonSoftwareFoundation.Python.3.12_qbz5n2kfra8p0\python.exe -m pip install --upgrade pip

```
[12]: # import libraries
import pandas as pd

import nltk

from nltk.sentiment.vader import SentimentIntensityAnalyzer

from nltk.corpus import stopwords

from nltk.tokenize import word_tokenize

from nltk.stem import WordNetLemmatizer

# download nltk corpus (first time only)
import nltk

nltk.download('all')
```

```
[nltk_data] Downloading collection 'all'
[nltk_data] |
[nltk_data] | Downloading package abc to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\abc.zip.
[nltk_data] | Downloading package alpino to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\alpino.zip.
[nltk_data] | Downloading package averaged_perceptron_tagger to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping taggers\averaged_perceptron_tagger.zip.
[nltk_data] | Downloading package averaged_perceptron_tagger_ru to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping
[nltk_data] | taggers\averaged_perceptron_tagger_ru.zip.
[nltk_data] | Downloading package basque_grammars to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping grammars\basque_grammars.zip.
[nltk_data] | Downloading package bcp47 to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Downloading package biocreative_ppi to
```

```

[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\biocreative_ppi.zip.
[nltk_data] | Downloading package bllip_wsj_no_aux to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping models\bllip_wsj_no_aux.zip.
[nltk_data] | Downloading package book_grammars to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping grammars\book_grammars.zip.
[nltk_data] | Downloading package brown to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\brown.zip.
[nltk_data] | Downloading package brown_tei to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\brown_tei.zip.
[nltk_data] | Downloading package cess_cat to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\cess_cat.zip.
[nltk_data] | Downloading package cess_esp to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\cess_esp.zip.
[nltk_data] | Downloading package chat80 to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\chat80.zip.
[nltk_data] | Downloading package city_database to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\city_database.zip.
[nltk_data] | Downloading package cmudict to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\cmudict.zip.
[nltk_data] | Downloading package comparative_sentences to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\comparative_sentences.zip.
[nltk_data] | Downloading package comtrans to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Downloading package conll2000 to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\conll2000.zip.
[nltk_data] | Downloading package conll2002 to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\conll2002.zip.
[nltk_data] | Downloading package conll2007 to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Downloading package crubadan to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\crubadan.zip.
[nltk_data] | Downloading package dependency_treebank to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\dependency_treebank.zip.

```

```

[nltk_data] | Downloading package dolch to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\dolch.zip.
[nltk_data] | Downloading package europarl_raw to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\europarl_raw.zip.
[nltk_data] | Downloading package extended_omw to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Downloading package floresta to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\floresta.zip.
[nltk_data] | Downloading package framenet_v15 to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\framenet_v15.zip.
[nltk_data] | Downloading package framenet_v17 to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\framenet_v17.zip.
[nltk_data] | Downloading package gazetteers to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\gazetteers.zip.
[nltk_data] | Downloading package genesis to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\genesis.zip.
[nltk_data] | Downloading package gutenber to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\gutenberg.zip.
[nltk_data] | Downloading package ieer to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\ieer.zip.
[nltk_data] | Downloading package inaugural to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\inaugural.zip.
[nltk_data] | Downloading package indian to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\indian.zip.
[nltk_data] | Downloading package jeita to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Downloading package kimmo to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\kimmo.zip.
[nltk_data] | Downloading package knbc to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Downloading package large_grammars to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping grammars\large_grammars.zip.
[nltk_data] | Downloading package lin_thesaurus to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\lin_thesaurus.zip.

```

```

[nltk_data] | Downloading package mac_morpho to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\mac_morpho.zip.
[nltk_data] | Downloading package machado to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Downloading package masc_tagged to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Downloading package maxent_ne_chunker to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping chunkers\maxent_ne_chunker.zip.
[nltk_data] | Downloading package maxent_treebank_pos_tagger to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping taggers\maxent_treebank_pos_tagger.zip.
[nltk_data] | Downloading package mooses_sample to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping models\mooses_sample.zip.
[nltk_data] | Downloading package movie_reviews to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\movie_reviews.zip.
[nltk_data] | Downloading package mte_teip5 to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\mte_teip5.zip.
[nltk_data] | Downloading package mwa_ppdb to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping misc\mwa_ppdb.zip.
[nltk_data] | Downloading package names to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\names.zip.
[nltk_data] | Downloading package nombank.1.0 to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Downloading package nonbreaking_prefixes to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\nonbreaking_prefixes.zip.
[nltk_data] | Downloading package nps_chat to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\nps_chat.zip.
[nltk_data] | Downloading package omw to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Downloading package omw-1.4 to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Downloading package opinion_lexicon to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\opinion_lexicon.zip.
[nltk_data] | Downloading package panlex_swadesh to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Downloading package paradigms to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\paradigms.zip.

```

```

[nltk_data] | Downloading package pe08 to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\pe08.zip.
[nltk_data] | Downloading package perluniprops to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping misc\perluniprops.zip.
[nltk_data] | Downloading package pil to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\pil.zip.
[nltk_data] | Downloading package pl196x to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\pl196x.zip.
[nltk_data] | Downloading package porter_test to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping stemmers\porter_test.zip.
[nltk_data] | Downloading package ppattach to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\ppattach.zip.
[nltk_data] | Downloading package problem_reports to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\problem_reports.zip.
[nltk_data] | Downloading package product_reviews_1 to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\product_reviews_1.zip.
[nltk_data] | Downloading package product_reviews_2 to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\product_reviews_2.zip.
[nltk_data] | Downloading package propbank to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Downloading package pros_cons to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\pros_cons.zip.
[nltk_data] | Downloading package ptb to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\ptb.zip.
[nltk_data] | Downloading package punkt to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping tokenizers\punkt.zip.
[nltk_data] | Downloading package qc to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\qc.zip.
[nltk_data] | Downloading package reuters to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Downloading package rslp to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping stemmers\rslp.zip.
[nltk_data] | Downloading package rte to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...

```

```

[nltk_data] | Unzipping corpora\rte.zip.
[nltk_data] | Downloading package sample_grammars to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping grammars\sample_grammars.zip.
[nltk_data] | Downloading package semcor to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Downloading package senseval to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\senseval.zip.
[nltk_data] | Downloading package sentence_polarity to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\sentence_polarity.zip.
[nltk_data] | Downloading package sentiwordnet to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\sentiwordnet.zip.
[nltk_data] | Downloading package shakespeare to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\shakespeare.zip.
[nltk_data] | Downloading package sinica_treebank to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\sinica_treebank.zip.
[nltk_data] | Downloading package smultron to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\smultron.zip.
[nltk_data] | Downloading package snowball_data to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Downloading package spanish_grammars to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping grammars\spanish_grammars.zip.
[nltk_data] | Downloading package state_union to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\state_union.zip.
[nltk_data] | Downloading package stopwords to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\stopwords.zip.
[nltk_data] | Downloading package subjectivity to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\subjectivity.zip.
[nltk_data] | Downloading package swadesh to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\swadesh.zip.
[nltk_data] | Downloading package switchboard to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\switchboard.zip.
[nltk_data] | Downloading package tagsets to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping help\tagsets.zip.
[nltk_data] | Downloading package timit to

```

```

[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\timit.zip.
[nltk_data] | Downloading package toolbox to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\toolbox.zip.
[nltk_data] | Downloading package treebank to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\treebank.zip.
[nltk_data] | Downloading package twitter_samples to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\twitter_samples.zip.
[nltk_data] | Downloading package udhr to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\udhr.zip.
[nltk_data] | Downloading package udhr2 to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\udhr2.zip.
[nltk_data] | Downloading package unicode_samples to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\unicode_samples.zip.
[nltk_data] | Downloading package universal_tagset to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping taggers\universal_tagset.zip.
[nltk_data] | Downloading package universal_treebanks_v20 to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Downloading package vader_lexicon to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Downloading package verbnet to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\verbnet.zip.
[nltk_data] | Downloading package verbnet3 to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\verbnet3.zip.
[nltk_data] | Downloading package webtext to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\webtext.zip.
[nltk_data] | Downloading package wmt15_eval to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping models\wmt15_eval.zip.
[nltk_data] | Downloading package word2vec_sample to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping models\word2vec_sample.zip.
[nltk_data] | Downloading package wordnet to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Downloading package wordnet2021 to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Downloading package wordnet2022 to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...

```



```

[nltk_data] | Unzipping corpora\wordnet2022.zip.
[nltk_data] | Downloading package wordnet31 to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Downloading package wordnet_ic to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\wordnet_ic.zip.
[nltk_data] | Downloading package words to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\words.zip.
[nltk_data] | Downloading package ycoe to
[nltk_data] | C:\Users\nagap\AppData\Roaming\nltk_data...
[nltk_data] | Unzipping corpora\ycoe.zip.
[nltk_data] |
[nltk_data] Done downloading collection all

```

[12]: True

```
[10]: df1 = pd.read_csv('Dataset .csv')
```

```
[13]: df1.columns
```

```

[13]: Index(['Restaurant ID', 'Restaurant Name', 'Country Code', 'City', 'Address',
            'Locality', 'Locality Verbose', 'Longitude', 'Latitude', 'Cuisines',
            'Average Cost for two', 'Currency', 'Has Table booking',
            'Has Online delivery', 'Is delivering now', 'Switch to order menu',
            'Price range', 'Aggregate rating', 'Rating color', 'Rating text',
            'Votes'],
            dtype='object')

```

```

[15]: # create preprocess_text function
def preprocess_text(text):

    # Tokenize the text

    tokens = word_tokenize(text.lower())

    # Remove stop words

    filtered_tokens = [token for token in tokens if token not in stopwords.
↪words('english')]

    # Lemmatize the tokens

    lemmatizer = WordNetLemmatizer()

    lemmatized_tokens = [lemmatizer.lemmatize(token) for token in ↪
↪filtered_tokens]

```

```

# Join the tokens back into a string

processed_text = ' '.join(lemmatized_tokens)

return processed_text

# apply the function df

df1['Rating text'] = df1['Rating text'].apply(preprocess_text)
df1

```

```

[15]:
Restaurant ID      Restaurant Name  Country Code      City \
0          6317637      Le Petit Souffle      162      Makati City
1          6304287      Izakaya Kikufuji      162      Makati City
2          6300002      Heat - Edsa Shangri-La      162      Mandaluyong City
3          6318506                        Ooma      162      Mandaluyong City
4          6314302      Sambo Kojin      162      Mandaluyong City
...          ...          ...          ...          ...
9546        5915730      Naml Gurme      208      stanbul
9547        5908749      Ceviz A ac      208      stanbul
9548        5915807                        Huqqa      208      stanbul
9549        5916112      A k Kahve      208      stanbul
9550        5927402      Walter's Coffee Roastery      208      stanbul

```

```

Address \
0      Third Floor, Century City Mall, Kalayaan Avenu...
1      Little Tokyo, 2277 Chino Roces Avenue, Legaspi...
2      Edsa Shangri-La, 1 Garden Way, Ortigas, Mandal...
3      Third Floor, Mega Fashion Hall, SM Megamall, O...
4      Third Floor, Mega Atrium, SM Megamall, Ortigas...
...          ...
9546      Kemanke Karamustafa Pa a Mahallesi, Rht m ...
9547      Ko uyolu Mahallesi, Muhittin st_nda Cadd...
9548      Kuru_e me Mahallesi, Muallim Naci Caddesi, N...
9549      Kuru_e me Mahallesi, Muallim Naci Caddesi, N...
9550      Cafea a Mahallesi, Bademalt Sokak, No 21/B, ...

```

```

Locality \
0      Century City Mall, Poblacion, Makati City
1      Little Tokyo, Legaspi Village, Makati City
2      Edsa Shangri-La, Ortigas, Mandaluyong City
3      SM Megamall, Ortigas, Mandaluyong City
4      SM Megamall, Ortigas, Mandaluyong City
...          ...
9546      Karak_y
9547      Ko uyolu

```

9548	Kuru_e me
9549	Kuru_e me
9550	Moda

	Locality Verbose	Longitude \
0	Century City Mall, Poblacion, Makati City, Mak...	121.027535
1	Little Tokyo, Legaspi Village, Makati City, Ma...	121.014101
2	Edsa Shangri-La, Ortigas, Mandaluyong City, Ma...	121.056831
3	SM Megamall, Ortigas, Mandaluyong City, Mandal...	121.056475
4	SM Megamall, Ortigas, Mandaluyong City, Mandal...	121.057508
...
9546	Karak_y, stanbul	28.977392
9547	Ko uyolu, stanbul	29.041297
9548	Kuru_e me, stanbul	29.034640
9549	Kuru_e me, stanbul	29.036019
9550	Moda, stanbul	29.026016

	Latitude	Cuisines ...	Currency \
0	14.565443	French, Japanese, Desserts ...	Botswana Pula(P)
1	14.553708	Japanese ...	Botswana Pula(P)
2	14.581404	Seafood, Asian, Filipino, Indian ...	Botswana Pula(P)
3	14.585318	Japanese, Sushi ...	Botswana Pula(P)
4	14.584450	Japanese, Korean ...	Botswana Pula(P)
...
9546	41.022793	Turkish ...	Turkish Lira(TL)
9547	41.009847	World Cuisine, Patisserie, Cafe ...	Turkish Lira(TL)
9548	41.055817	Italian, World Cuisine ...	Turkish Lira(TL)
9549	41.057979	Restaurant Cafe ...	Turkish Lira(TL)
9550	40.984776	Cafe ...	Turkish Lira(TL)

	Has Table booking	Has Online delivery	Is delivering now \
0	Yes	No	No
1	Yes	No	No
2	Yes	No	No
3	No	No	No
4	Yes	No	No
...
9546	No	No	No
9547	No	No	No
9548	No	No	No
9549	No	No	No
9550	No	No	No

	Switch to order menu	Price range	Aggregate rating	Rating color \
0	No	3	4.8	Dark Green
1	No	3	4.5	Dark Green
2	No	4	4.4	Green

3	No	4	4.9	Dark Green
4	No	4	4.8	Dark Green
...
9546	No	3	4.1	Green
9547	No	3	4.2	Green
9548	No	4	3.7	Yellow
9549	No	4	4.0	Green
9550	No	2	4.0	Green

	Rating	text	Votes
0	excellent		314
1	excellent		591
2	good		270
3	excellent		365
4	excellent		229
...	
9546	good		788
9547	good		1034
9548	good		661
9549	good		901
9550	good		591

[9551 rows x 21 columns]

```
[30]: # initialize NLTK sentiment analyzer

analyzer = SentimentIntensityAnalyzer()

# Create empty lists to store positive and negative words
positive_words = []
negative_words = []

# Analyze each review and extract positive and negative keywords
for review in df1['Rating text']:
    tokens = word_tokenize(review)
    sentiment_scores = analyzer.polarity_scores(review)
    if sentiment_scores['compound'] >= 0.2:
        positive_words.extend(tokens)
    elif sentiment_scores['compound'] <= -0.2:
        negative_words.extend(tokens)

# Calculate the most common positive and negative keywords
common_positive_keywords = Counter(positive_words).most_common(10)
common_negative_keywords = Counter(negative_words).most_common(10)

# Print the results
```

```
print("Most Common Positive Keywords:")
print(common_positive_keywords)

print("\nMost Common Negative Keywords:")
print(common_negative_keywords)
```

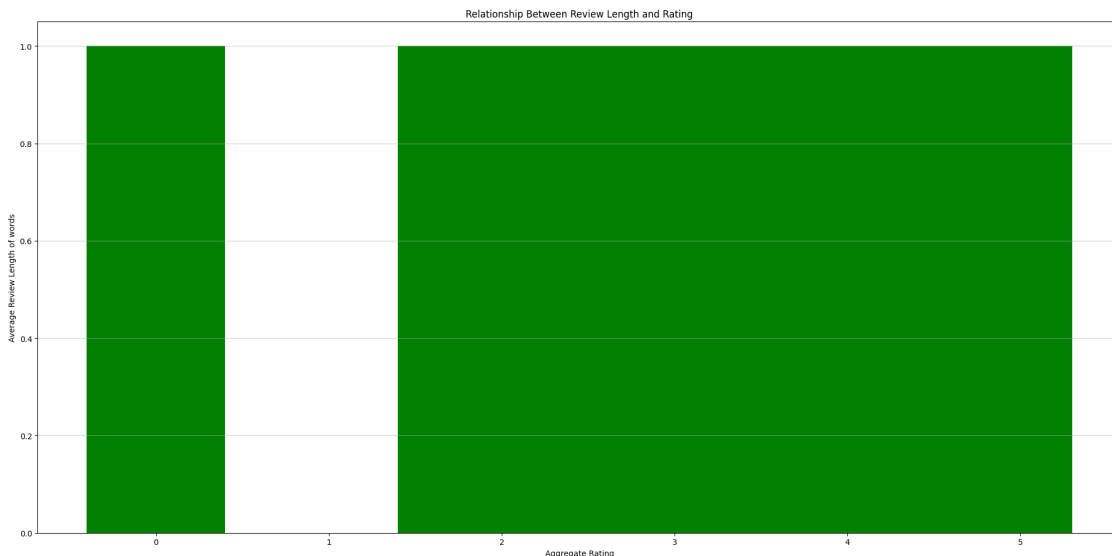
Most Common Positive Keywords:
[('good', 3179), ('excellent', 301)]

Most Common Negative Keywords:
[('poor', 186)]

2. Calculate the average length of reviews and explore if there is a relationship between review length and rating.

```
[20]: df1['Review Length'] = df1['Rating text'].apply(lambda x: len(x.split()))
Rating_by_Avg_len = df1.groupby('Aggregate rating')['Review Length'].mean()
```

```
[27]: import matplotlib.pyplot as plt
plt.figure(figsize=(20,10))
plt.bar(Rating_by_Avg_len.index, Rating_by_Avg_len.values, color='green')
plt.xlabel('Aggregate Rating')
plt.ylabel('Average Review Length of words')
plt.title('Relationship Between Review Length and Rating')
plt.grid(axis='y', linestyle='--', alpha=0.6)
plt.tight_layout()
plt.show()
```



2 Task 2

Votes Analysis 1. Identify the restaurants with the highest and lowest number of votes.

```
[31]: ch_votes = df1.groupby("Restaurant Name")["Votes"].sum().reset_index(name="vote_
      ↪count")
```

```
[32]: sort_votes = ch_votes.sort_values(by="vote count", ascending=False)
      Top_15 = sort_votes.head(15)
      Top_15
```

```
[32]:
```

	Restaurant Name	vote count
663	Barbeque Nation	28142
101	AB's - Absolute Barbecues	13400
6943	Toit	10934
785	Big Chill	10853
2297	Farzi Cafe	10098
6988	Truffles	9682
1510	Chili's	8156
2879	Hauz Khas Social	7931
3261	Joey's Pizza	7807
4902	Peter Cat	7574
796	Big Yellow Door	7511
5571	Saravana Bhavan	7238
6080	Starbucks	7139
4941	Pirates of Grill	7091
3405	Karim's	6878

```
[34]: Least_15 = sort_votes.tail(15)
      Least_15
```

```
[34]:
```

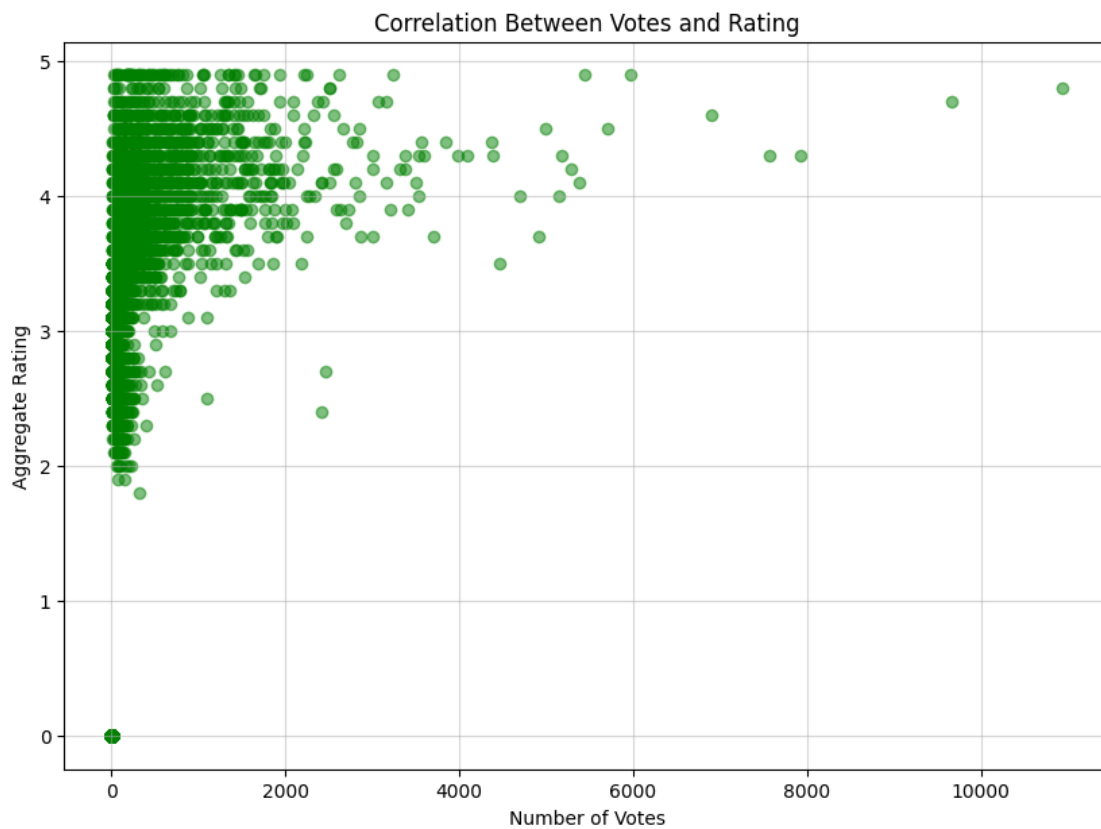
	Restaurant Name	vote count
4353	Mughlai Darbar Muradabadi	0
2094	Dolphin - The Food Cafe	0
4360	Mukesh Bhojnalaya	0
349	Angaar	0
743	Bhai Ji Dhaba	0
3008	Hotel Green View Palace	0
6210	Sweetcake.in	0
3006	Hotel Delhi 43	0
6224	TAG	0
6220	Szoun Chinese Food	0
4375	Muncheezz	0
6217	Swiss Gourmessa	0
2110	Dosa and Pizza Corner	0
6215	Sweets n Treats	0
2433	Food Station	0

2. Analyze if there is a correlation between the number of votes and the rating of a restaurant.

```
[41]: # Create a scatter plot to analyze the correlation
plt.figure(figsize=(10, 7))
plt.scatter(df1['Votes'], df1['Aggregate rating'], alpha=0.5, color='green')
plt.xlabel('Number of Votes')
plt.ylabel('Aggregate Rating')
plt.title('Correlation Between Votes and Rating')
plt.grid(axis='both', linestyle='-', alpha=0.5)
plt.show()

# Calculate the correlation between number of votes and rating
correlation = df1[['Votes', 'Aggregate rating']].corr()

print("Correlation Between Number of Votes and Rating:", correlation)
```



Correlation Between Number of Votes and Rating:

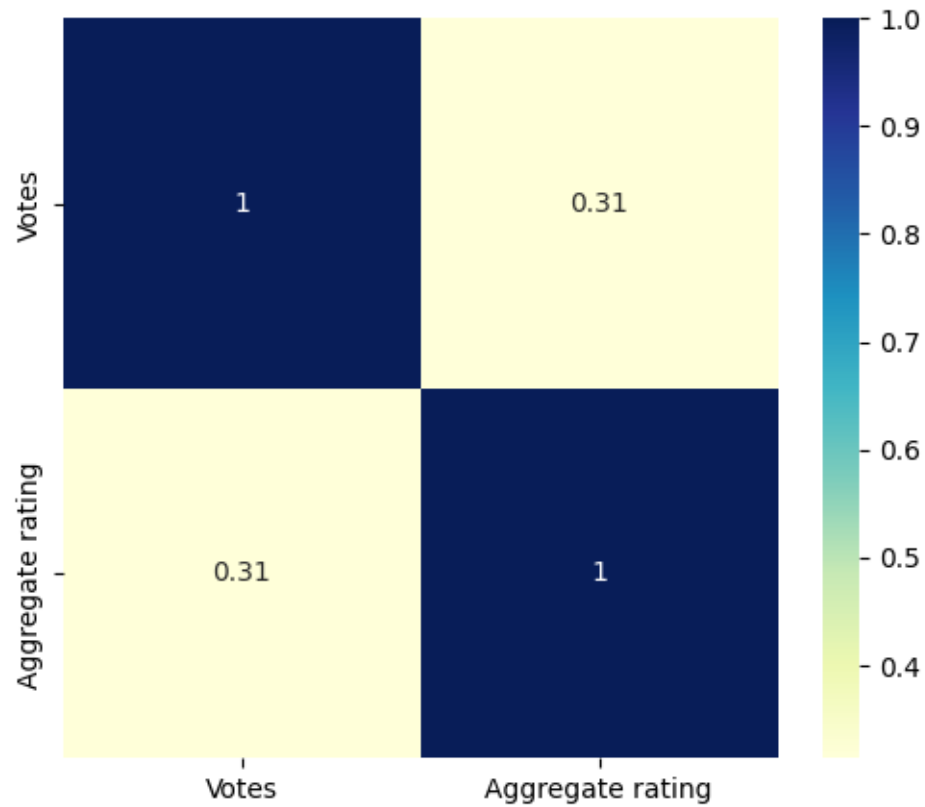
Aggregate rating

Votes 1.000000 0.313691

Aggregate rating 0.313691 1.000000

Votes

```
[42]: import seaborn as sns
plt.figure(figsize=(6, 5))
sns.heatmap(correlation, cmap="YlGnBu", annot=True)
plt.show()
```



```
[ ]:
```