

Read in Data and NLTK Basics

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
plt.style.use('ggplot')
import nltk
```

```
C:\Users\nithi\anaconda3\lib\site-packages\scipy\_init_.py:146: UserWarning: A NumPy version >=1.16.5 and <1.23.0 is required for this version of SciPy (detected version 1.24.3
warnings.warn(f"A NumPy version >={np_minversion} and <{np_maxversion}")
```

```
In [2]: df=pd.read_excel('C:\\Users\\nithi\\review-details.xlsx')
```

```
In [3]: df
```

Out[3]:	report_date	online_store	upc	retailer_product_code	brand	category	sub_category	product_description	review_da
0	2019-01-02	FRESHAMAZON	8718114216478	B0142CI6FC	Dove Men+Care	Personal Care	Deos	Dove Men+Care Extra Fresh Anti-perspirant Deod...	2019-01-
1	2019-01-03	FRESHAMAZON	5000184201199	B014DFNNRY	Marmite	Foods	Savoury	Marmite Spread Yeast Extract 500g	2019-01-
2	2019-01-03	FRESHAMAZON	5000184201199	B014DFNNRY	Marmite	Foods	Savoury	Marmite Spread Yeast Extract 500g	2019-01-
3	2019-01-03	FRESHAMAZON	8712566479726	B014DFKELC	Knorr	Foods	Savoury	Knorr Beef Stock Pot 8 x 28g	2019-01-
4	2019-01-03	FRESHAMAZON	8717163536476	B014G37I7E	Cif	Homecare	HHC	Cif Citrus Bathroom Mousse 500ml	2019-01-
...
2496	2019-05-27	OCADO	8714100873885	399390011	Cornetto	Foods	Ice Cream	Cornetto Made with Soy and Gluten Free Ice Cre...	2019-04-
2497	2019-05-27	OCADO	50184439	11120011	Bovril	Foods	Savoury	Bovril Extract Beef and Yeast 250g	2019-05-
2498	2019-05-27	OCADO	8722700479451	13958011	Hellmann's	Foods	Dressings	Hellmann's Light Squeezy Mayonnaise 430ml	2019-05-
2499	2019-05-27	OCADO	8714100873885	399390011	Cornetto	Foods	Ice Cream	Cornetto Made with Soy and Gluten Free Ice Cre...	2019-05-
2500	2019-05-27	OCADO	8714100873885	399390011	Cornetto	Foods	Ice Cream	Cornetto Made with Soy and Gluten Free Ice Cre...	2019-04-

2501 rows × 32 columns

```
In [4]: df.head()
```

Out[4]:	report_date	online_store	upc	retailer_product_code	brand	category	sub_category	product_description	review_date
0	2019-01-02	FRESHAMAZON	8718114216478	B0142CI6FC	Dove Men+Care	Personal Care	Deos	Dove Men+Care Extra Fresh Anti-perspirant Deod...	2019-01-01
1	2019-01-03	FRESHAMAZON	5000184201199	B014DFNNRY	Marmite	Foods	Savoury	Marmite Spread Yeast Extract 500g	2019-01-02
2	2019-01-03	FRESHAMAZON	5000184201199	B014DFNNRY	Marmite	Foods	Savoury	Marmite Spread Yeast Extract 500g	2019-01-02
3	2019-01-03	FRESHAMAZON	8712566479726	B014DFKELC	Knorr	Foods	Savoury	Knorr Beef Stock Pot 8 x 28g	2019-01-02
4	2019-01-03	FRESHAMAZON	8717163536476	B014G37I7E	Cif	Homecare	HHC	Cif Citrus Bathroom Mousse 500ml	2019-01-02

5 rows × 32 columns

```
In [5]: df.tail()
```

```
Out[5]:
```

	report_date	online_store	upc	retailer_product_code	brand	category	sub_category	product_description	review_date
2496	2019-05-27	OCADO	8714100873885	399390011	Cornetto	Foods	Ice Cream	Cornetto Made with Soy and Gluten Free Ice Cre...	2019-04-27
2497	2019-05-27	OCADO	50184439	11120011	Bovril	Foods	Savoury	Bovril Extract Beef and Yeast 250g	2019-05-02
2498	2019-05-27	OCADO	8722700479451	13958011	Hellmann's	Foods	Dressings	Hellmann's Light Squeezy Mayonnaise 430ml	2019-05-05
2499	2019-05-27	OCADO	8714100873885	399390011	Cornetto	Foods	Ice Cream	Cornetto Made with Soy and Gluten Free Ice Cre...	2019-05-17
2500	2019-05-27	OCADO	8714100873885	399390011	Cornetto	Foods	Ice Cream	Cornetto Made with Soy and Gluten Free Ice Cre...	2019-04-27

5 rows × 32 columns

Quick EDA

```
In [6]: ax = df['review_rating'].value_counts().sort_index() \
        .plot(kind='bar',
              title='Count of Reviews by Customer Rating',
              figsize=(10, 5))
ax.set_xlabel('Review Rating')
plt.show()
```



Basic NLTK

```
In [7]: review_example = df['review_text'][50]
print(review_example)
```

Have used a lot of make-up brands and in the past bought their more expensive make-up remover products, but none have ever removed mascara, eye-shadow, foundation bases etc as gently and as well as Simple Cleansing Lotion. Its consistency is not runny, you have to squeeze the bottle appropriately onto the cotton wool, which is ideal for controlling how much is dispensed. Also great for tidying up eyeliner or mascara errors when applied sparsely on a cotton wool bud. Always keep this now in my bathroom cabinet. Why pay more?

```
In [8]: import nltk
nltk.download('all')
```

```
[nltk_data] Downloading collection 'all'
[nltk_data] |
[nltk_data] | Downloading package abc to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package abc is already up-to-date!
[nltk_data] | Downloading package alpino to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package alpino is already up-to-date!
[nltk_data] | Downloading package averaged_perceptron_tagger to
[nltk_data] | C:\Users\nithi\AppData\Roaming\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] | C:\Users\nithi\AppData\Roaming\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] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package basque_grammars is already up-to-date!
[nltk_data] | Downloading package bcp47 to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package bcp47 is already up-to-date!
[nltk_data] | Downloading package biocreative_ppi to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package biocreative_ppi is already up-to-date!
[nltk_data] | Downloading package bllip_wsj_no_aux to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package bllip_wsj_no_aux is already up-to-date!
[nltk_data] | Downloading package book_grammars to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package book_grammars is already up-to-date!
[nltk_data] | Downloading package brown to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package brown is already up-to-date!
[nltk_data] | Downloading package brown_tei to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package brown_tei is already up-to-date!
[nltk_data] | Downloading package cess_cat to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package cess_cat is already up-to-date!
[nltk_data] | Downloading package cess_esp to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package cess_esp is already up-to-date!
[nltk_data] | Downloading package chat80 to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package chat80 is already up-to-date!
[nltk_data] | Downloading package city_database to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package city_database is already up-to-date!
[nltk_data] | Downloading package cmudict to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package cmudict is already up-to-date!
[nltk_data] | Downloading package comparative_sentences to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package comparative_sentences is already up-to-
[nltk_data] | date!
[nltk_data] | Downloading package comtrans to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package comtrans is already up-to-date!
[nltk_data] | Downloading package conll2000 to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package conll2000 is already up-to-date!
[nltk_data] | Downloading package conll2002 to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package conll2002 is already up-to-date!
[nltk_data] | Downloading package conll2007 to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package conll2007 is already up-to-date!
[nltk_data] | Downloading package crubadan to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package crubadan is already up-to-date!
[nltk_data] | Downloading package dependency_treebank to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package dependency_treebank is already up-to-date!
[nltk_data] | Downloading package dolch to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package dolch is already up-to-date!
[nltk_data] | Downloading package europarl_raw to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package europarl_raw is already up-to-date!
[nltk_data] | Downloading package extended_omw to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package extended_omw is already up-to-date!
[nltk_data] | Downloading package floresta to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package floresta is already up-to-date!
[nltk_data] | Downloading package framenet_v15 to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package framenet_v15 is already up-to-date!
[nltk_data] | Downloading package framenet_v17 to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package framenet_v17 is already up-to-date!
[nltk_data] | Downloading package gazetteers to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package gazetteers is already up-to-date!
[nltk_data] | Downloading package genesis to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package genesis is already up-to-date!
[nltk_data] | Downloading package gutenber to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package gutenber is already up-to-date!
[nltk_data] | Downloading package ieer to
```

```
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package ieer is already up-to-date!
[nltk_data] | Downloading package inaugural to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package inaugural is already up-to-date!
[nltk_data] | Downloading package indian to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package indian is already up-to-date!
[nltk_data] | Downloading package jeita to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package jeita is already up-to-date!
[nltk_data] | Downloading package kimmo to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package kimmo is already up-to-date!
[nltk_data] | Downloading package knbc to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package knbc is already up-to-date!
[nltk_data] | Downloading package large_grammars to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package large_grammars is already up-to-date!
[nltk_data] | Downloading package lin_thesaurus to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package lin_thesaurus is already up-to-date!
[nltk_data] | Downloading package mac_morpho to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package mac_morpho is already up-to-date!
[nltk_data] | Downloading package machado to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package machado is already up-to-date!
[nltk_data] | Downloading package masc_tagged to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package masc_tagged is already up-to-date!
[nltk_data] | Downloading package maxent_ne_chunker to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package maxent_ne_chunker is already up-to-date!
[nltk_data] | Downloading package maxent_treebank_pos_tagger to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package maxent_treebank_pos_tagger is already up-
to-date!
[nltk_data] | Downloading package mooses_sample to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package mooses_sample is already up-to-date!
[nltk_data] | Downloading package movie_reviews to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package movie_reviews is already up-to-date!
[nltk_data] | Downloading package mte_teip5 to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package mte_teip5 is already up-to-date!
[nltk_data] | Downloading package mwa_ppdb to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package mwa_ppdb is already up-to-date!
[nltk_data] | Downloading package names to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package names is already up-to-date!
[nltk_data] | Downloading package nombank.1.0 to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package nombank.1.0 is already up-to-date!
[nltk_data] | Downloading package nonbreaking_prefixes to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package nonbreaking_prefixes is already up-to-date!
[nltk_data] | Downloading package nps_chat to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package nps_chat is already up-to-date!
[nltk_data] | Downloading package omw to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package omw is already up-to-date!
[nltk_data] | Downloading package omw-1.4 to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package omw-1.4 is already up-to-date!
[nltk_data] | Downloading package opinion_lexicon to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package opinion_lexicon is already up-to-date!
[nltk_data] | Downloading package panlex_swadesh to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package panlex_swadesh is already up-to-date!
[nltk_data] | Downloading package paradigms to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package paradigms is already up-to-date!
[nltk_data] | Downloading package pe08 to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package pe08 is already up-to-date!
[nltk_data] | Downloading package perluniprops to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package perluniprops is already up-to-date!
[nltk_data] | Downloading package pil to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package pil is already up-to-date!
[nltk_data] | Downloading package pl196x to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
```

```
[nltk_data] | Package pl196x is already up-to-date!
[nltk_data] | Downloading package porter_test to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package porter_test is already up-to-date!
[nltk_data] | Downloading package ppattach to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package ppattach is already up-to-date!
[nltk_data] | Downloading package problem_reports to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package problem_reports is already up-to-date!
[nltk_data] | Downloading package product_reviews_1 to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package product_reviews_1 is already up-to-date!
[nltk_data] | Downloading package product_reviews_2 to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package product_reviews_2 is already up-to-date!
[nltk_data] | Downloading package propbank to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package propbank is already up-to-date!
[nltk_data] | Downloading package pros_cons to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package pros_cons is already up-to-date!
[nltk_data] | Downloading package ptb to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package ptb is already up-to-date!
[nltk_data] | Downloading package punkt to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package punkt is already up-to-date!
[nltk_data] | Downloading package qc to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package qc is already up-to-date!
[nltk_data] | Downloading package reuters to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package reuters is already up-to-date!
[nltk_data] | Downloading package rslp to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package rslp is already up-to-date!
[nltk_data] | Downloading package rte to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package rte is already up-to-date!
[nltk_data] | Downloading package sample_grammars to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package sample_grammars is already up-to-date!
[nltk_data] | Downloading package semcor to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package semcor is already up-to-date!
[nltk_data] | Downloading package senseval to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package senseval is already up-to-date!
[nltk_data] | Downloading package sentence_polarity to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package sentence_polarity is already up-to-date!
[nltk_data] | Downloading package sentiwordnet to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package sentiwordnet is already up-to-date!
[nltk_data] | Downloading package shakespeare to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package shakespeare is already up-to-date!
[nltk_data] | Downloading package sinica_treebank to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package sinica_treebank is already up-to-date!
[nltk_data] | Downloading package smultron to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package smultron is already up-to-date!
[nltk_data] | Downloading package snowball_data to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package snowball_data is already up-to-date!
[nltk_data] | Downloading package spanish_grammars to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package spanish_grammars is already up-to-date!
[nltk_data] | Downloading package state_union to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package state_union is already up-to-date!
[nltk_data] | Downloading package stopwords to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package stopwords is already up-to-date!
[nltk_data] | Downloading package subjectivity to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package subjectivity is already up-to-date!
[nltk_data] | Downloading package swadesh to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package swadesh is already up-to-date!
[nltk_data] | Downloading package switchboard to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package switchboard is already up-to-date!
[nltk_data] | Downloading package tagsets to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package tagsets is already up-to-date!
[nltk_data] | Downloading package timit to
```

```

[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package timit is already up-to-date!
[nltk_data] | Downloading package toolbox to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package toolbox is already up-to-date!
[nltk_data] | Downloading package treebank to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package treebank is already up-to-date!
[nltk_data] | Downloading package twitter_samples to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package twitter_samples is already up-to-date!
[nltk_data] | Downloading package udhr to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package udhr is already up-to-date!
[nltk_data] | Downloading package udhr2 to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package udhr2 is already up-to-date!
[nltk_data] | Downloading package unicode_samples to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package unicode_samples is already up-to-date!
[nltk_data] | Downloading package universal_tagset to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package universal_tagset is already up-to-date!
[nltk_data] | Downloading package universal_treebanks_v20 to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package universal_treebanks_v20 is already up-to-
[nltk_data] | date!
[nltk_data] | Downloading package vader_lexicon to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package vader_lexicon is already up-to-date!
[nltk_data] | Downloading package verbnet to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package verbnet is already up-to-date!
[nltk_data] | Downloading package verbnet3 to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package verbnet3 is already up-to-date!
[nltk_data] | Downloading package webtext to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package webtext is already up-to-date!
[nltk_data] | Downloading package wmt15_eval to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package wmt15_eval is already up-to-date!
[nltk_data] | Downloading package word2vec_sample to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package word2vec_sample is already up-to-date!
[nltk_data] | Downloading package wordnet to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package wordnet is already up-to-date!
[nltk_data] | Downloading package wordnet2021 to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package wordnet2021 is already up-to-date!
[nltk_data] | Downloading package wordnet2022 to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package wordnet2022 is already up-to-date!
[nltk_data] | Downloading package wordnet31 to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package wordnet31 is already up-to-date!
[nltk_data] | Downloading package wordnet_ic to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package wordnet_ic is already up-to-date!
[nltk_data] | Downloading package words to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package words is already up-to-date!
[nltk_data] | Downloading package ycoe to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] | Package ycoe is already up-to-date!
[nltk_data] | Done downloading collection all

```

Out[8]: True

In [9]: `nltk.download('punkt')`

```

[nltk_data] Downloading package punkt to
[nltk_data] | C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] Package punkt is already up-to-date!

```

Out[9]: True

In [10]: `tokens = nltk.word_tokenize(review_example)`
`tokens[:10]`

Out[10]: ['Have', 'used', 'a', 'lot', 'of', 'make-up', 'brands', 'and', 'in', 'the']

In [11]: `from nltk.tokenize import word_tokenize`

In [12]: `from nltk import pos_tag`
`nltk.download('averaged_perceptron_tagger')`

```
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data] C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] Package averaged_perceptron_tagger is already up-to-
[nltk_data] date!
```

Out[12]: True

```
In [13]: tagged = nltk.pos_tag(tokens)
tagged[:10]
```

```
Out[13]: [('Have', 'VBP'),
('used', 'VBN'),
('a', 'DT'),
('lot', 'NN'),
('of', 'IN'),
('make-up', 'JJ'),
('brands', 'NNS'),
('and', 'CC'),
('in', 'IN'),
('the', 'DT')]
```

```
In [14]: nltk.download('words')
```

```
[nltk_data] Downloading package words to
[nltk_data] C:\Users\nithi\AppData\Roaming\nltk_data...
[nltk_data] Package words is already up-to-date!
```

Out[14]: True

```
In [15]: entities = nltk.chunk.ne_chunk(tagged)
entities.pprint()
```

```
(S
  Have/VBP
  used/VBN
  a/DT
  lot/NN
  of/IN
  make-up/JJ
  brands/NNS
  and/CC
  in/IN
  the/DT
  past/JJ
  bought/VBD
  their/PRP$
  more/RBR
  expensive/JJ
  make-up/NN
  remover/NN
  products/NNS
  ,/,
  but/CC
  none/NN
  have/VBP
  ever/RB
  removed/VBN
  mascara/NN
  ,/,
  eye-shadow/JJ
  ,/,
  foundation/NN
  bases/NNS
  etc/VBP
  as/IN
  gently/RB
  and/CC
  as/RB
  well/RB
  as/IN
  (PERSON Simple/NNP Cleansing/NNP Lotion/NNP)
  ./
  Its/PRP$
  consistency/NN
  is/VBZ
  not/RB
  runny/JJ
  ,/,
  you/PRP
  have/VBP
  to/TO
  squeeze/VB
  the/DT
  bottle/NN
  appropriately/RB
  onto/IN
  the/DT
  cotton/NN
  wool/NN
```



```

/,
which/WDT
is/VBZ
ideal/JJ
for/IN
controlling/VBG
how/WRB
much/JJ
is/VBZ
dispensed/VBN
./
Also/RB
great/JJ
for/IN
tidying/VBG
up/RP
eyeliner/NN
or/CC
mascara/NN
errors/NNS
when/WRB
applied/VBN
sparsely/RB
on/IN
a/DT
cotton/NN
wool/NN
bud/NN
./
Always/VBZ
keep/VB
this/DT
now/RB
in/IN
my/PRP$
bathroom/NN
cabinet/NN
./
Why/WRB
pay/NN
more/JJR
?/.)

```

VADER Seniment Scoring

```

In [16]: from nltk.sentiment import SentimentIntensityAnalyzer
from tqdm.notebook import tqdm
sia = SentimentIntensityAnalyzer()

```

```

In [17]: sia.polarity_scores('I am so happy!')

```

```

Out[17]: {'neg': 0.0, 'neu': 0.318, 'pos': 0.682, 'compound': 0.6468}

```

```

In [18]: sia.polarity_scores('This is the worst thing ever.')

```

```

Out[18]: {'neg': 0.451, 'neu': 0.549, 'pos': 0.0, 'compound': -0.6249}

```

```

In [19]: sia.polarity_scores(review_example)

```

```

Out[19]: {'neg': 0.045, 'neu': 0.792, 'pos': 0.163, 'compound': 0.9349}

```

```

In [20]: # Run the polarity score on the entire dataset
res = {}
for i, row in tqdm(df.iterrows(), total=len(df)):
    text = row['review_text']
    myid = row['review_hash_id']
    res[myid] = sia.polarity_scores(text)

```

```

0%|          | 0/2501 [00:00<?, ?it/s]

```

```

In [21]: vaders = pd.DataFrame(res).T

```

```

In [22]: vaders = vaders.reset_index().rename(columns={'index': 'review_rating'})

```

```

In [23]: vaders.head()

```



```
Out[23]:
```

	review_rating	neg	neu	pos	compound
0	3f129b02-ea76-0323-bd59-235d97a4f83f	0.097	0.703	0.200	0.5707
1	d7f3b9aa-e8b3-626d-683b-374e201c8315	0.000	0.719	0.281	0.8803
2	e58a523d-0155-a366-f107-7ac6817ac3b7	0.000	0.000	1.000	0.5719
3	aaa9bb87-4f99-bb89-65cb-3b400ebb45c0	0.000	0.459	0.541	0.7096
4	48c71b34-d7fe-5e90-51dd-239e153fb0ae	0.000	1.000	0.000	0.0000

```
In [24]: import seaborn as sns
```

```
In [25]: pip install vaderSentiment
```

Requirement already satisfied: vaderSentiment in c:\users\nithi\anaconda3\lib\site-packages (3.3.2)
Requirement already satisfied: requests in c:\users\nithi\anaconda3\lib\site-packages (from vaderSentiment) (2.27.1)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\nithi\anaconda3\lib\site-packages (from requests->vaderSentiment) (2021.10.8)
Requirement already satisfied: charset-normalizer~=2.0.0 in c:\users\nithi\anaconda3\lib\site-packages (from requests->vaderSentiment) (2.0.4)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\nithi\anaconda3\lib\site-packages (from requests->vaderSentiment) (1.26.9)
Requirement already satisfied: idna<4,>=2.5 in c:\users\nithi\anaconda3\lib\site-packages (from requests->vaderSentiment) (3.3)
Note: you may need to restart the kernel to use updated packages.

```
In [26]: # import SentimentIntensityAnalyzer class
# from vaderSentiment.vaderSentiment module.
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
```

```
In [27]: from nltk.sentiment import SentimentIntensityAnalyzer
from tqdm.notebook import tqdm

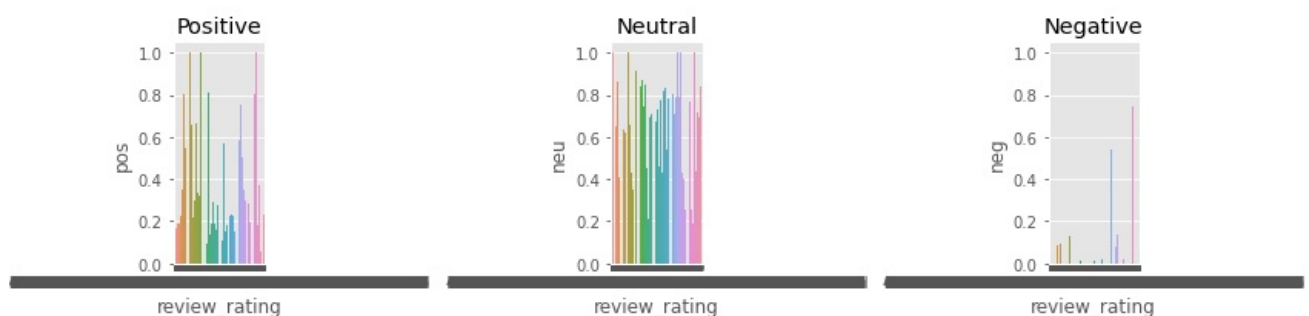
sia = SentimentIntensityAnalyzer()
```

Plot VADER results

```
In [28]: ax = sns.barplot(data=vaders, x='review_rating', y='compound')
ax.set_title('Compound Score by Amazon Customer Review')
plt.show()
```



```
In [29]: fig, axes = plt.subplots(1, 3, figsize=(12, 3))
sns.barplot(data=vaders, x='review_rating', y='pos', ax=axes[0])
sns.barplot(data=vaders, x='review_rating', y='neu', ax=axes[1])
sns.barplot(data=vaders, x='review_rating', y='neg', ax=axes[2])
axes[0].set_title('Positive')
axes[1].set_title('Neutral')
axes[2].set_title('Negative')
plt.tight_layout()
plt.show()
```



Roberta Pretrained Model

In [30]: `pip install transformers`

```
Requirement already satisfied: transformers in c:\users\nithi\anaconda3\lib\site-packages (4.30.2)
Requirement already satisfied: huggingface-hub<1.0,>=0.14.1 in c:\users\nithi\anaconda3\lib\site-packages (from transformers) (0.15.1)
Requirement already satisfied: regex!=2019.12.17 in c:\users\nithi\anaconda3\lib\site-packages (from transformers) (2022.3.15)
Requirement already satisfied: tqdm>=4.27 in c:\users\nithi\anaconda3\lib\site-packages (from transformers) (4.64.0)
Requirement already satisfied: requests in c:\users\nithi\anaconda3\lib\site-packages (from transformers) (2.27.1)
Requirement already satisfied: filelock in c:\users\nithi\anaconda3\lib\site-packages (from transformers) (3.6.0)
Requirement already satisfied: safetensors>=0.3.1 in c:\users\nithi\anaconda3\lib\site-packages (from transformers) (0.3.1)
Requirement already satisfied: numpy>=1.17 in c:\users\nithi\anaconda3\lib\site-packages (from transformers) (1.24.3)
Requirement already satisfied: tokenizers!=0.11.3,<0.14,>=0.11.1 in c:\users\nithi\anaconda3\lib\site-packages (from transformers) (0.13.3)
Requirement already satisfied: pyyaml>=5.1 in c:\users\nithi\anaconda3\lib\site-packages (from transformers) (6.0)
Requirement already satisfied: packaging>=20.0 in c:\users\nithi\anaconda3\lib\site-packages (from transformers) (21.3)
Requirement already satisfied: typing-extensions>=3.7.4.3 in c:\users\nithi\anaconda3\lib\site-packages (from huggingface-hub<1.0,>=0.14.1->transformers) (4.1.1)
Requirement already satisfied: fsspec in c:\users\nithi\anaconda3\lib\site-packages (from huggingface-hub<1.0,>=0.14.1->transformers) (2022.2.0)
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in c:\users\nithi\anaconda3\lib\site-packages (from packaging>=20.0->transformers) (3.0.4)
Requirement already satisfied: colorama in c:\users\nithi\anaconda3\lib\site-packages (from tqdm>=4.27->transformers) (0.4.4)
Requirement already satisfied: charset-normalizer~=2.0.0 in c:\users\nithi\anaconda3\lib\site-packages (from requests->transformers) (2.0.4)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\nithi\anaconda3\lib\site-packages (from requests->transformers) (2021.10.8)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\nithi\anaconda3\lib\site-packages (from requests->transformers) (1.26.9)
Requirement already satisfied: idna<4,>=2.5 in c:\users\nithi\anaconda3\lib\site-packages (from requests->transformers) (3.3)
Note: you may need to restart the kernel to use updated packages.
```

In [31]: `pip install transformers[torch]`

Requirement already satisfied: transformers[torch] in c:\users\nithi\anaconda3\lib\site-packages (4.30.2)
 Requirement already satisfied: safetensors>=0.3.1 in c:\users\nithi\anaconda3\lib\site-packages (from transformers[torch]) (0.3.1)
 Requirement already satisfied: pyyaml>=5.1 in c:\users\nithi\anaconda3\lib\site-packages (from transformers[torch]) (6.0)
 Requirement already satisfied: tokenizers!=0.11.3,<0.14,>=0.11.1 in c:\users\nithi\anaconda3\lib\site-packages (from transformers[torch]) (0.13.3)
 Requirement already satisfied: requests in c:\users\nithi\anaconda3\lib\site-packages (from transformers[torch]) (2.27.1)
 Requirement already satisfied: regex!=2019.12.17 in c:\users\nithi\anaconda3\lib\site-packages (from transformers[torch]) (2022.3.15)
 Requirement already satisfied: tqdm>=4.27 in c:\users\nithi\anaconda3\lib\site-packages (from transformers[torch]) (4.64.0)
 Requirement already satisfied: numpy>=1.17 in c:\users\nithi\anaconda3\lib\site-packages (from transformers[torch]) (1.24.3)
 Requirement already satisfied: huggingface-hub<1.0,>=0.14.1 in c:\users\nithi\anaconda3\lib\site-packages (from transformers[torch]) (0.15.1)
 Requirement already satisfied: packaging>=20.0 in c:\users\nithi\anaconda3\lib\site-packages (from transformers[torch]) (21.3)
 Requirement already satisfied: filelock in c:\users\nithi\anaconda3\lib\site-packages (from transformers[torch]) (3.6.0)
 Requirement already satisfied: accelerate>=0.20.2 in c:\users\nithi\anaconda3\lib\site-packages (from transformers[torch]) (0.20.3)
 Requirement already satisfied: torch!=1.12.0,>=1.9 in c:\users\nithi\anaconda3\lib\site-packages (from transformers[torch]) (2.0.1)
 Requirement already satisfied: psutil in c:\users\nithi\anaconda3\lib\site-packages (from accelerate>=0.20.2->transformers[torch]) (5.8.0)
 Requirement already satisfied: fsspec in c:\users\nithi\anaconda3\lib\site-packages (from huggingface-hub<1.0,>=0.14.1->transformers[torch]) (2022.2.0)
 Requirement already satisfied: typing-extensions>=3.7.4.3 in c:\users\nithi\anaconda3\lib\site-packages (from huggingface-hub<1.0,>=0.14.1->transformers[torch]) (4.1.1)
 Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in c:\users\nithi\anaconda3\lib\site-packages (from packaging>=20.0->transformers[torch]) (3.0.4)
 Requirement already satisfied: jinja2 in c:\users\nithi\anaconda3\lib\site-packages (from torch!=1.12.0,>=1.9->transformers[torch]) (2.11.3)
 Requirement already satisfied: sympy in c:\users\nithi\anaconda3\lib\site-packages (from torch!=1.12.0,>=1.9->transformers[torch]) (1.10.1)
 Requirement already satisfied: networkx in c:\users\nithi\anaconda3\lib\site-packages (from torch!=1.12.0,>=1.9->transformers[torch]) (2.7.1)
 Requirement already satisfied: colorama in c:\users\nithi\anaconda3\lib\site-packages (from tqdm>=4.27->transformers[torch]) (0.4.4)
 Requirement already satisfied: MarkupSafe>=0.23 in c:\users\nithi\anaconda3\lib\site-packages (from jinja2->torch!=1.12.0,>=1.9->transformers[torch]) (2.0.1)
 Requirement already satisfied: idna<4,>=2.5 in c:\users\nithi\anaconda3\lib\site-packages (from requests->transformers[torch]) (3.3)
 Requirement already satisfied: charset-normalizer~=2.0.0 in c:\users\nithi\anaconda3\lib\site-packages (from requests->transformers[torch]) (2.0.4)
 Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\nithi\anaconda3\lib\site-packages (from requests->transformers[torch]) (1.26.9)
 Requirement already satisfied: certifi>=2017.4.17 in c:\users\nithi\anaconda3\lib\site-packages (from requests->transformers[torch]) (2021.10.8)
 Requirement already satisfied: mpmath>=0.19 in c:\users\nithi\anaconda3\lib\site-packages (from sympy->torch!=1.12.0,>=1.9->transformers[torch]) (1.2.1)
 Note: you may need to restart the kernel to use updated packages.

In [32]: `import torch`

In [33]: `print(torch.__version__)`

2.0.1+cpu

In [34]: `from transformers import AutoTokenizer
 from transformers import AutoModelForSequenceClassification
 from scipy.special import softmax`

In [35]: `pip install emoji`

Requirement already satisfied: emoji in c:\users\nithi\anaconda3\lib\site-packages (2.6.0)
 Note: you may need to restart the kernel to use updated packages.

In [36]: `pip install emoji --upgrade`

Requirement already satisfied: emoji in c:\users\nithi\anaconda3\lib\site-packages (2.6.0)
 Note: you may need to restart the kernel to use updated packages.

In [37]: `!pip install lightgbm`

ERROR: Could not install packages due to an OSError: [WinError 5] Access is denied: 'C:\\Users\\nithi\\anaconda3\\Lib\\site-packages\\~\\mpy\\.libs\\libopenblas64_v0.3.21-gcc_10_3_0.dll'
 Consider using the '--user' option or check the permissions.

```
Requirement already satisfied: lightgbm in c:\users\nithi\anaconda3\lib\site-packages (4.0.0)
Requirement already satisfied: scipy in c:\users\nithi\anaconda3\lib\site-packages (from lightgbm) (1.7.3)
Requirement already satisfied: numpy in c:\users\nithi\anaconda3\lib\site-packages (from lightgbm) (1.24.3)
Collecting numpy
  Using cached numpy-1.22.4-cp39-cp39-win_amd64.whl (14.7 MB)
Installing collected packages: numpy
  Attempting uninstall: numpy
    Found existing installation: numpy 1.24.3
    Uninstalling numpy-1.24.3:
      Successfully uninstalled numpy-1.24.3
```

```
In [38]: import numpy as np
```

```
In [39]: pip install numpy --upgrade
```

```
Requirement already satisfied: numpy in c:\users\nithi\anaconda3\lib\site-packages (1.22.4)
Collecting numpy
  Using cached numpy-1.25.1-cp39-cp39-win_amd64.whl (15.1 MB)
Installing collected packages: numpy
  Attempting uninstall: numpy
    Found existing installation: numpy 1.22.4
    Uninstalling numpy-1.22.4:
      Successfully uninstalled numpy-1.22.4
Successfully installed numpy-1.25.1
Note: you may need to restart the kernel to use updated packages.
```

```
ERROR: pip's dependency resolver does not currently take into account all the packages that are installed. This
behaviour is the source of the following dependency conflicts.
daal4py 2021.5.0 requires daal==2021.4.0, which is not installed.
tensorflow-intel 2.13.0 requires numpy<=1.24.3,>=1.22, but you have numpy 1.25.1 which is incompatible.
scipy 1.7.3 requires numpy<1.23.0,>=1.16.5, but you have numpy 1.25.1 which is incompatible.
numba 0.55.1 requires numpy<1.22,>=1.18, but you have numpy 1.25.1 which is incompatible.
```

```
In [40]: import numpy as np
import pandas as pd
import sys
import emoji
import math
import pickle
from time import time
from nltk.corpus import stopwords
from nltk.stem import SnowballStemmer
from sklearn.feature_extraction.text import CountVectorizer, TfidfVectorizer
from sklearn.pipeline import Pipeline
from sklearn.metrics import confusion_matrix
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB, MultinomialNB
from sklearn.linear_model import SGDClassifier
from sklearn.cluster import KMeans
from sklearn import tree
from sklearn import svm
from contextlib import contextmanager
from sklearn.feature_extraction.text import TfidfVectorizer
from scipy.sparse import hstack
import time
import re
import string
from scipy.sparse import csr_matrix
from sklearn.preprocessing import MinMaxScaler
import lightgbm as lgb
from sklearn.model_selection import KFold
from sklearn.metrics import roc_auc_score
import gc
from collections import defaultdict
import os
import psutil
import os
```

```
In [41]: pip install keras
```

```
Requirement already satisfied: keras in c:\users\nithi\anaconda3\lib\site-packages (2.13.1)
Note: you may need to restart the kernel to use updated packages.
```

```
In [42]: pip install tensorflow
```

Requirement already satisfied: tensorflow in c:\users\nithi\anaconda3\lib\site-packages (2.13.0)
Requirement already satisfied: tensorflow-intel==2.13.0 in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow) (2.13.0)
Requirement already satisfied: setuptools in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (61.2.0)
Collecting numpy<=1.24.3,>=1.22
Using cached numpy-1.24.3-cp39-cp39-win-amd64.whl (14.9 MB)
Requirement already satisfied: protobuf!=4.21.0,!4.21.1,!4.21.2,!4.21.3,!4.21.4,!4.21.5,<5.0.0dev,>=3.20.3 in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (4.23.4)
Requirement already satisfied: wrapt>=1.11.0 in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (1.12.1)
Requirement already satisfied: absl-py>=1.0.0 in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (1.4.0)
Requirement already satisfied: tensorflow-io-gcs-filesystem>=0.23.1 in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (0.31.0)
Requirement already satisfied: flatbuffers>=23.1.21 in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (23.5.26)
Requirement already satisfied: grpcio<2.0,>=1.24.3 in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (1.56.0)
Requirement already satisfied: typing-extensions<4.6.0,>=3.6.6 in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (4.1.1)
Requirement already satisfied: termcolor>=1.1.0 in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (2.3.0)
Requirement already satisfied: tensorflow-estimator<2.14,>=2.13.0 in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (2.13.0)
Requirement already satisfied: libclang>=13.0.0 in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (16.0.6)
Requirement already satisfied: six>=1.12.0 in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
Requirement already satisfied: astunparse>=1.6.0 in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (1.6.3)
Requirement already satisfied: google-pasta>=0.1.1 in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (0.2.0)
Requirement already satisfied: h5py>=2.9.0 in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (3.6.0)
Requirement already satisfied: tensorboard<2.14,>=2.13 in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (2.13.0)
Requirement already satisfied: packaging in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (21.3)
Requirement already satisfied: opt-einsum>=2.3.2 in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (3.3.0)
Requirement already satisfied: keras<2.14,>=2.13.1 in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (2.13.1)
Requirement already satisfied: gast<=0.4.0,>=0.2.1 in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (0.4.0)
Requirement already satisfied: wheel<1.0,>=0.23.0 in c:\users\nithi\anaconda3\lib\site-packages (from astunparse>=1.6.0->tensorflow-intel==2.13.0->tensorflow) (0.37.1)
Requirement already satisfied: google-auth<3,>=1.6.3 in c:\users\nithi\anaconda3\lib\site-packages (from tensorboard<2.14,>=2.13->tensorflow-intel==2.13.0->tensorflow) (2.22.0)
Requirement already satisfied: tensorboard-data-server<0.8.0,>=0.7.0 in c:\users\nithi\anaconda3\lib\site-packages (from tensorboard<2.14,>=2.13->tensorflow-intel==2.13.0->tensorflow) (0.7.1)
Requirement already satisfied: google-auth-oauthlib<1.1,>=0.5 in c:\users\nithi\anaconda3\lib\site-packages (from tensorboard<2.14,>=2.13->tensorflow-intel==2.13.0->tensorflow) (1.0.0)
Requirement already satisfied: werkzeug>=1.0.1 in c:\users\nithi\anaconda3\lib\site-packages (from tensorboard<2.14,>=2.13->tensorflow-intel==2.13.0->tensorflow) (2.0.3)
Requirement already satisfied: requests<3,>=2.21.0 in c:\users\nithi\anaconda3\lib\site-packages (from tensorboard<2.14,>=2.13->tensorflow-intel==2.13.0->tensorflow) (2.27.1)
Requirement already satisfied: markdown>=2.6.8 in c:\users\nithi\anaconda3\lib\site-packages (from tensorboard<2.14,>=2.13->tensorflow-intel==2.13.0->tensorflow) (3.3.4)
Requirement already satisfied: cachetools<6.0,>=2.0.0 in c:\users\nithi\anaconda3\lib\site-packages (from google-auth<3,>=1.6.3->tensorboard<2.14,>=2.13->tensorflow-intel==2.13.0->tensorflow) (4.2.2)
Requirement already satisfied: pyasn1-modules>=0.2.1 in c:\users\nithi\anaconda3\lib\site-packages (from google-auth<3,>=1.6.3->tensorboard<2.14,>=2.13->tensorflow-intel==2.13.0->tensorflow) (0.2.8)
Requirement already satisfied: rsa<5,>=3.1.4 in c:\users\nithi\anaconda3\lib\site-packages (from google-auth<3,>=1.6.3->tensorboard<2.14,>=2.13->tensorflow-intel==2.13.0->tensorflow) (4.7.2)
Requirement already satisfied: urllib3<2.0 in c:\users\nithi\anaconda3\lib\site-packages (from google-auth<3,>=1.6.3->tensorboard<2.14,>=2.13->tensorflow-intel==2.13.0->tensorflow) (1.26.9)
Requirement already satisfied: requests-oauthlib>=0.7.0 in c:\users\nithi\anaconda3\lib\site-packages (from google-auth-oauthlib<1.1,>=0.5->tensorboard<2.14,>=2.13->tensorflow-intel==2.13.0->tensorflow) (1.3.1)
Requirement already satisfied: pyasn1<0.5.0,>=0.4.6 in c:\users\nithi\anaconda3\lib\site-packages (from pyasn1-modules>=0.2.1->google-auth<3,>=1.6.3->tensorboard<2.14,>=2.13->tensorflow-intel==2.13.0->tensorflow) (0.4.8)
Requirement already satisfied: idna<4,>=2.5 in c:\users\nithi\anaconda3\lib\site-packages (from requests<3,>=2.21.0->tensorboard<2.14,>=2.13->tensorflow-intel==2.13.0->tensorflow) (3.3)
Requirement already satisfied: charset-normalizer>=2.0.0 in c:\users\nithi\anaconda3\lib\site-packages (from requests<3,>=2.21.0->tensorboard<2.14,>=2.13->tensorflow-intel==2.13.0->tensorflow) (2.0.4)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\nithi\anaconda3\lib\site-packages (from requests<3,>=2.21.0->tensorboard<2.14,>=2.13->tensorflow-intel==2.13.0->tensorflow) (2021.10.8)
Requirement already satisfied: oauthlib>=3.0.0 in c:\users\nithi\anaconda3\lib\site-packages (from requests-oauthlib>=0.7.0->google-auth-oauthlib<1.1,>=0.5->tensorboard<2.14,>=2.13->tensorflow-intel==2.13.0->tensorflow) (3.2.2)
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in c:\users\nithi\anaconda3\lib\site-packages (from packaging->tensorflow-intel==2.13.0->tensorflow) (3.0.4)
Installing collected packages: numpy
Attempting uninstall: numpy
Found existing installation: numpy 1.25.1
Uninstalling numpy-1.25.1:
Successfully uninstalled numpy-1.25.1

ERROR: pip's dependency resolver does not currently take into account all the packages that are installed. This behaviour is the source of the following dependency conflicts.
daal4py 2021.5.0 requires daal==2021.4.0, which is not installed.
scipy 1.7.3 requires numpy<1.23.0,>=1.16.5, but you have numpy 1.24.3 which is incompatible.
numba 0.55.1 requires numpy<1.22,>=1.18, but you have numpy 1.24.3 which is incompatible.
Successfully installed numpy-1.24.3

In [43]: `pip install --upgrade tensorflow`

Requirement already satisfied: tensorflow in c:\users\nithi\anaconda3\lib\site-packages (2.13.0)

Requirement already satisfied: tensorflow-intel==2.13.0 in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow) (2.13.0)

Requirement already satisfied: astunparse>=1.6.0 in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (1.6.3)

Requirement already satisfied: h5py>=2.9.0 in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (3.6.0)

Requirement already satisfied: six>=1.12.0 in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (1.16.0)

Requirement already satisfied: numpy<=1.24.3,>=1.22 in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (1.24.3)

Requirement already satisfied: setuptools in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (61.2.0)

Requirement already satisfied: protobuf!=4.21.0,!4.21.1,!4.21.2,!4.21.3,!4.21.4,!4.21.5,<5.0.0dev,>=3.20.3 in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (4.23.4)

Requirement already satisfied: flatbuffers>=23.1.21 in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (23.5.26)

Requirement already satisfied: tensorflow-io-gcs-filesystem>=0.23.1 in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (0.31.0)

Requirement already satisfied: tensorflow-estimator<2.14,>=2.13.0 in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (2.13.0)

Requirement already satisfied: libclang>=13.0.0 in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (16.0.6)

Requirement already satisfied: keras<2.14,>=2.13.1 in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (2.13.1)

Requirement already satisfied: absl-py>=1.0.0 in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (1.4.0)

Requirement already satisfied: packaging in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (21.3)

Requirement already satisfied: grpcio<2.0,>=1.24.3 in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (1.56.0)

Requirement already satisfied: wrapt>=1.11.0 in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (1.12.1)

Requirement already satisfied: google-pasta>=0.1.1 in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (0.2.0)

Requirement already satisfied: opt-einsum>=2.3.2 in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (3.3.0)

Requirement already satisfied: termcolor>=1.1.0 in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (2.3.0)

Requirement already satisfied: typing-extensions<4.6.0,>=3.6.6 in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (4.1.1)

Requirement already satisfied: tensorboard<2.14,>=2.13 in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (2.13.0)

Requirement already satisfied: gast<=0.4.0,>=0.2.1 in c:\users\nithi\anaconda3\lib\site-packages (from tensorflow-intel==2.13.0->tensorflow) (0.4.0)

Requirement already satisfied: wheel<1.0,>=0.23.0 in c:\users\nithi\anaconda3\lib\site-packages (from astunparse>=1.6.0->tensorflow-intel==2.13.0->tensorflow) (0.37.1)

Requirement already satisfied: google-auth-oauthlib<1.1,>=0.5 in c:\users\nithi\anaconda3\lib\site-packages (from tensorboard<2.14,>=2.13->tensorflow-intel==2.13.0->tensorflow) (1.0.0)

Requirement already satisfied: werkzeug>=1.0.1 in c:\users\nithi\anaconda3\lib\site-packages (from tensorboard<2.14,>=2.13->tensorflow-intel==2.13.0->tensorflow) (2.0.3)

Requirement already satisfied: tensorboard-data-server<0.8.0,>=0.7.0 in c:\users\nithi\anaconda3\lib\site-packages (from tensorboard<2.14,>=2.13->tensorflow-intel==2.13.0->tensorflow) (0.7.1)

Requirement already satisfied: google-auth<3,>=1.6.3 in c:\users\nithi\anaconda3\lib\site-packages (from tensorboard<2.14,>=2.13->tensorflow-intel==2.13.0->tensorflow) (2.22.0)

Requirement already satisfied: markdown>=2.6.8 in c:\users\nithi\anaconda3\lib\site-packages (from tensorboard<2.14,>=2.13->tensorflow-intel==2.13.0->tensorflow) (3.3.4)

Requirement already satisfied: requests<3,>=2.21.0 in c:\users\nithi\anaconda3\lib\site-packages (from tensorboard<2.14,>=2.13->tensorflow-intel==2.13.0->tensorflow) (2.27.1)

Requirement already satisfied: rsa<5,>=3.1.4 in c:\users\nithi\anaconda3\lib\site-packages (from google-auth<3,>=1.6.3->tensorboard<2.14,>=2.13->tensorflow-intel==2.13.0->tensorflow) (4.7.2)

Requirement already satisfied: pyasn1-modules>=0.2.1 in c:\users\nithi\anaconda3\lib\site-packages (from google-auth<3,>=1.6.3->tensorboard<2.14,>=2.13->tensorflow-intel==2.13.0->tensorflow) (0.2.8)

Requirement already satisfied: urllib3<2.0 in c:\users\nithi\anaconda3\lib\site-packages (from google-auth<3,>=1.6.3->tensorboard<2.14,>=2.13->tensorflow-intel==2.13.0->tensorflow) (1.26.9)

Requirement already satisfied: cachetools<6.0,>=2.0.0 in c:\users\nithi\anaconda3\lib\site-packages (from google-auth<3,>=1.6.3->tensorboard<2.14,>=2.13->tensorflow-intel==2.13.0->tensorflow) (4.2.2)

Requirement already satisfied: requests-oauthlib>=0.7.0 in c:\users\nithi\anaconda3\lib\site-packages (from google-auth-oauthlib<1.1,>=0.5->tensorboard<2.14,>=2.13->tensorflow-intel==2.13.0->tensorflow) (1.3.1)

Requirement already satisfied: pyasn1<0.5.0,>=0.4.6 in c:\users\nithi\anaconda3\lib\site-packages (from pyasn1-modules>=0.2.1->google-auth<3,>=1.6.3->tensorboard<2.14,>=2.13->tensorflow-intel==2.13.0->tensorflow) (0.4.8)

Requirement already satisfied: certifi>=2017.4.17 in c:\users\nithi\anaconda3\lib\site-packages (from requests<3,>=2.21.0->tensorboard<2.14,>=2.13->tensorflow-intel==2.13.0->tensorflow) (2021.10.8)

Requirement already satisfied: idna<4,>=2.5 in c:\users\nithi\anaconda3\lib\site-packages (from requests<3,>=2.21.0->tensorboard<2.14,>=2.13->tensorflow-intel==2.13.0->tensorflow) (3.3)

Requirement already satisfied: charset-normalizer==2.0.0 in c:\users\nithi\anaconda3\lib\site-packages (from requests<3,>=2.21.0->tensorboard<2.14,>=2.13->tensorflow-intel==2.13.0->tensorflow) (2.0.4)

Requirement already satisfied: oauthlib>=3.0.0 in c:\users\nithi\anaconda3\lib\site-packages (from requests-oauthlib>=0.7.0->google-auth-oauthlib<1.1,>=0.5->tensorboard<2.14,>=2.13->tensorflow-intel==2.13.0->tensorflow) (3.2.2)

Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in c:\users\nithi\anaconda3\lib\site-packages (from packaging->tensorflow-intel==2.13.0->tensorflow) (3.0.4)

Note: you may need to restart the kernel to use updated packages.

In [44]: `import tensorflow as tf`

In [45]: `import tensorflow as tf
from tensorflow import keras`


```
In [46]: import tensorflow as tf
model = tf.keras.models.Sequential()
model.add(tf.keras.layers.Dense(12, input_dim=8, activation='relu'))
model.add(tf.keras.layers.Dense(8, activation='relu'))
model.add(tf.keras.layers.Dense(1, activation='sigmoid'))
```

```
In [47]: from keras.preprocessing import text as keras_text, sequence as keras_seq
from keras.callbacks import EarlyStopping, ModelCheckpoint
from sklearn.feature_extraction.text import CountVectorizer
import re, string
```

```
In [48]: from transformers import AutoTokenizer
from transformers import AutoModelForSequenceClassification
from scipy.special import softmax
```

```
In [49]: MODEL = "cardiffnlp/twitter-roberta-base-sentiment"
tokenizer = AutoTokenizer.from_pretrained(MODEL)
model = AutoModelForSequenceClassification.from_pretrained(MODEL)
```

```
In [50]: # VADER results on example
print(review_example)
sia.polarity_scores(review_example)
```

Have used a lot of make-up brands and in the past bought their more expensive make-up remover products, but none have ever removed mascara, eye-shadow, foundation bases etc as gently and as well as Simple Cleansing Lotion. Its consistency is not runny, you have to squeeze the bottle appropriately onto the cotton wool, which is ideal for controlling how much is dispensed. Also great for tidying up eyeliner or mascara errors when applied sparsely on a cotton wool bud. Always keep this now in my bathroom cabinet. Why pay more?

```
Out[50]: {'neg': 0.045, 'neu': 0.792, 'pos': 0.163, 'compound': 0.9349}
```

```
In [51]: encoded_text = tokenizer(review_example, return_tensors='pt')
output = model(**encoded_text)
scores = output[0][0].detach().numpy()
scores = softmax(scores)
scores_dict = {
    'roberta_neg' : scores[0],
    'roberta_neu' : scores[1],
    'roberta_pos' : scores[2]
}
print(scores_dict)

{'roberta_neg': 0.021018099, 'roberta_neu': 0.19694111, 'roberta_pos': 0.7820407}
```

```
In [52]: def polarity_scores_roberta(review_example):
    encoded_text = tokenizer(review_example, return_tensors='pt')
    output = model(**encoded_text)
    scores = output[0][0].detach().numpy()
    scores = softmax(scores)
    scores_dict = {
        'roberta_neg' : scores[0],
        'roberta_neu' : scores[1],
        'roberta_pos' : scores[2]
    }
    return scores_dict
```

```
In [53]: res = {}
for i, row in tqdm(df.iterrows(), total=len(df)):
    try:
        text = row['review_text']
        myid = row['review_rating']
        vader_result = sia.polarity_scores(text)
        vader_result_rename = {}
        for key, value in vader_result.items():
            vader_result_rename[f"vader_{key}"] = value
        roberta_result = polarity_scores_roberta(text)
        both = {**vader_result_rename, **roberta_result}
        res[myid] = both
    except RuntimeError:
        print(f'Broke for id {myid}')

0%|          | 0/2501 [00:00<?, ?it/s]
```

```
In [58]: results_df = pd.DataFrame(res).T
results_df = results_df.reset_index().rename(columns={'index': 'review_rating'})
```

```
In [59]: results_df = results_df.merge(df, how='left')
```

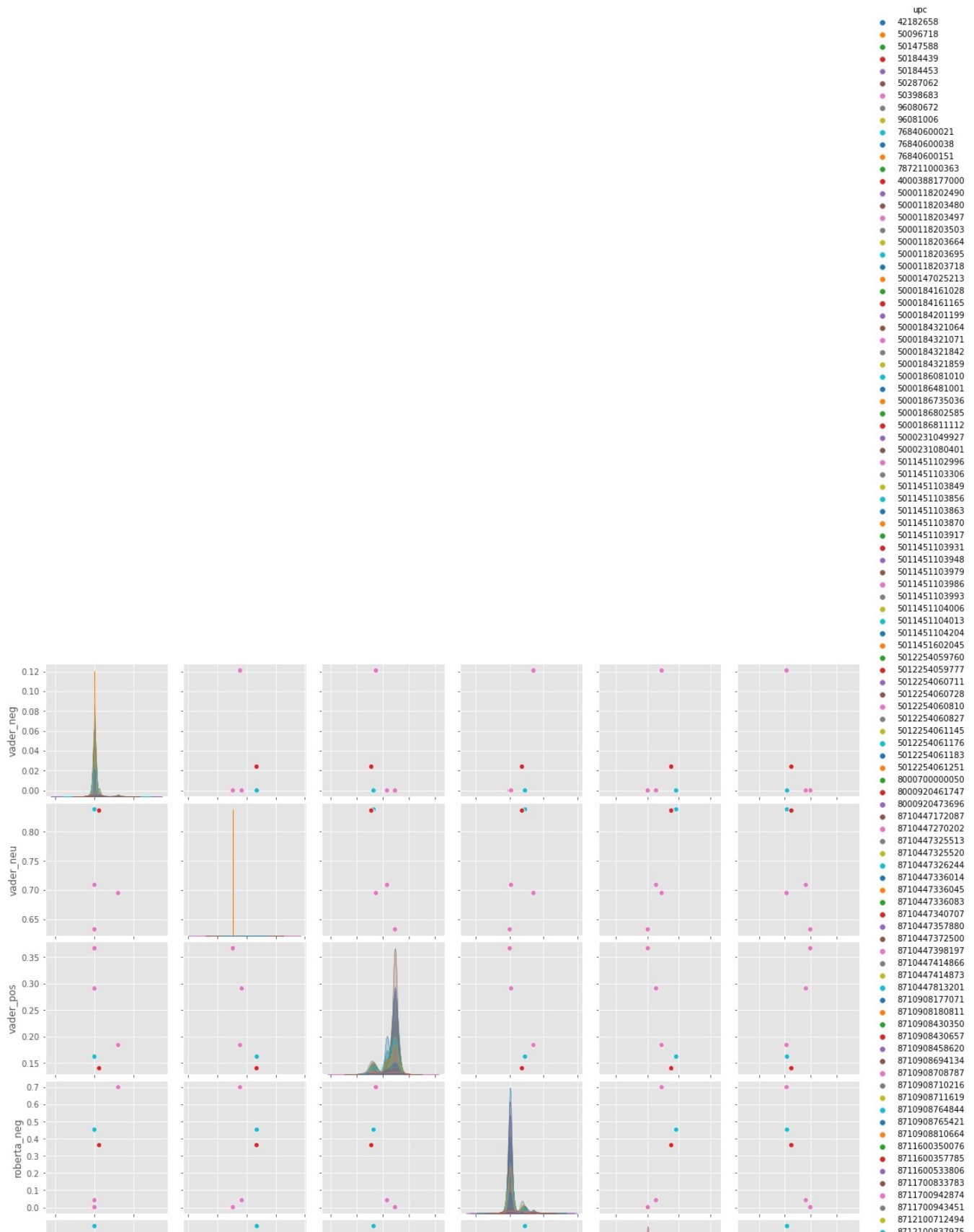
Compare Scores between models

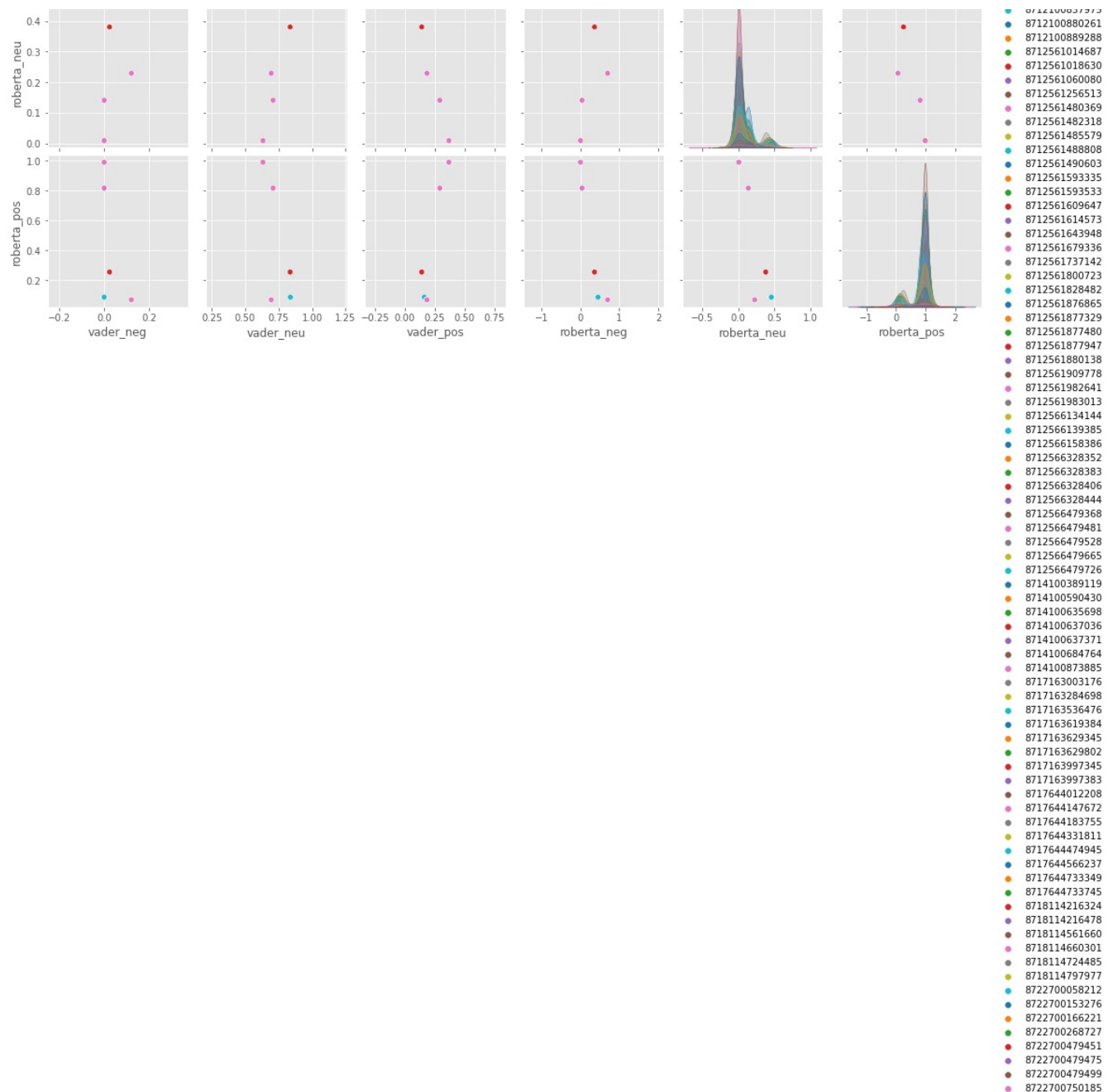
```
In [60]: results_df.columns
```

```
Out[60]: Index(['review_rating', 'vader_neg', 'vader_neu', 'vader_pos',
        'vader_compound', 'roberta_neg', 'roberta_neu', 'roberta_pos',
        'report_date', 'online_store', 'upc', 'retailer_product_code', 'brand',
        'category', 'sub_category', 'product_description', 'review_date',
        'review_title', 'review_text', 'is_competitor', 'manufacturer',
        'market', 'matched_keywords', 'time_of_publication', 'url',
        'review_type', 'parent_review', 'manufacturers_response', 'dimension1',
        'dimension2', 'dimension3', 'dimension4', 'dimension5', 'dimension6',
        'dimension7', 'dimension8', 'verified_purchase', 'helpful_review_count',
        'review_hash_id'],
        dtype='object')
```

```
In [64]: sns.pairplot(data=results_df,
        vars=['vader_neg', 'vader_neu', 'vader_pos',
        'roberta_neg', 'roberta_neu', 'roberta_pos'],
        hue='upc',
        palette='tab10')
```

```
Out[64]: <seaborn.axisgrid.PairGrid at 0x24494bdd7f0>
```





```
In [ ]: plt.show()
```

Review Examples

```
In [65]: results_df.query('upc == 42182658') \
        .sort_values('roberta_neu', ascending=False)['review_text'].values[0]
```

```
Out[65]: 'I have used Vaseline on my lips for well over 20 years and loved it. I buy a large jar of plain Vaseline and use on my lips. I loved that there is really no smell or taste to it. BUT last week, I bought a new jar, put some on, and it has a taste that I can only explain of tasting like black oil that goes into a car. It is terrible! I2019ve tried using it for a week now, and just can2019t stand it any longer. I2019m guessing there2019s been a formula or ingredient supplier change? I hope you will bring back the old version. Very disappointed!'
```

```
In [67]: results_df.query('upc ==42182658 ') \
        .sort_values('vader_pos', ascending=False)['review_text'].values[0]
```

```
Out[67]: 'I came across an article to use Vaseline on acne scars and irritated skin, so I tried it out. Never would I think to do this, I would think that this would clog my pores. However it did the opposite, I have been using this for a month now and it is life changing! My scars have visibly faded! I have gone to so many dermatologist and couldn2019t find any product that worked this well. This is a game changer my friends, please try it out!'
```

```
In [68]: results_df.query('upc== 8722700750185') \
        .sort_values('roberta_neg', ascending=False)['review_text'].values[0]
```

```
Out[68]: 'Like to make my own soups & casseroles. This product does very well for that.'
```

```
In [69]: results_df.query('upc == 8722700750185') \
        .sort_values('vader_neg', ascending=False)['review_text'].values[0]
```

```
Out[69]: 'Like to make my own soups & casseroles. This product does very well for that.'
```

The Transformers Pipeline

```
In [71]: from transformers import pipeline
```

```
sent_pipeline = pipeline("sentiment-analysis")
```

No model was supplied, defaulted to distilbert-base-uncased-finetuned-sst-2-english and revision af0f99b (<https://huggingface.co/distilbert-base-uncased-finetuned-sst-2-english>).

Using a pipeline without specifying a model name and revision in production is not recommended.

Downloading (...)lve/main/config.json: 0%| | 0.00/629 [00:00<?, ?B/s]

C:\Users\nithi\anaconda3\lib\site-packages\huggingface_hub\file_download.py:133: UserWarning: `huggingface_hub` cache-system uses symlinks by default to efficiently store duplicated files but your machine does not support them in C:\Users\nithi\.cache\huggingface\hub. Caching files will still work but in a degraded version that might require more space on your disk. This warning can be disabled by setting the `HF_HUB_DISABLE_SYMLINKS_WARNING` environment variable. For more details, see https://huggingface.co/docs/huggingface_hub/how-to-cache#limitations.

To support symlinks on Windows, you either need to activate Developer Mode or to run Python as an administrator. In order to see activate developer mode, see this article: <https://docs.microsoft.com/en-us/windows/apps/get-started/enable-your-device-for-development>

warnings.warn(message)

Downloading model.safetensors: 0%| | 0.00/268M [00:00<?, ?B/s]

Downloading (...)okenizer_config.json: 0%| | 0.00/48.0 [00:00<?, ?B/s]

Downloading (...)solve/main/vocab.txt: 0%| | 0.00/232k [00:00<?, ?B/s]

Xformers is not installed correctly. If you want to use memory_efficient_attention to accelerate training use the following command to install Xformers

pip install xformers.

```
In [76]: sent_pipeline('I love sentiment analysis!')
```

```
Out[76]: [{'label': 'POSITIVE', 'score': 0.9997853636741638}]
```

```
In [77]: sent_pipeline('I hate sentiment analysis!')
```

```
Out[77]: [{'label': 'NEGATIVE', 'score': 0.9992958307266235}]
```

```
In [78]: sent_pipeline('I may like sentiment analysis!')
```

```
Out[78]: [{'label': 'POSITIVE', 'score': 0.9128923416137695}]
```

```
In [73]: sent_pipeline('Make sure to like and subscribe!')
```

```
Out[73]: [{'label': 'POSITIVE', 'score': 0.9991742968559265}]
```

```
In [74]: sent_pipeline('booo')
```

```
Out[74]: [{'label': 'NEGATIVE', 'score': 0.9936267137527466}]
```

```
In [ ]:
```

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js