Sentiment Analysis: A Comparison of the Instagram Accounts of PLNU and Other Universities

Alexa Do

Point Loma Nazarene University

Honors Research Project

April 2, 2022

Committee: Michael Leih, Ph.D. (Mentor); Jamie McIlwaine, Ph.D.; Michael Wiese, Ph.D.

Table of Contents

Abstract:

Sentiment analysis is a machine learning, Natural Language Processing technique that associates text with their given emotion. Marketers often use sentiment analysis to gain insights into their customers, especially in discovering areas that they can improve on or should continue doing. Sentiment analysis is often performed on social media platforms so that marketers can modify their content based on the positive sentiments found. Through a collaboration with PLNU's Marketing Department, sentiment analysis was conducted on the Instagram accounts of PLNU and their competitors. Using the VADER Python library, sentiment scores were created, ranging in numerical values from -1 to +1. These values were then translated as positive, negative, or neutral. By putting the Instagram Posts into categories, the colleges were ranked based on positive sentiment. Each category was then analyzed as to the reasons why a particular college received the top rank. Then sentiment scores were conducted for the universities collectively, demonstrating that Pepperdine received the highest sentiment overall.

Introduction:

In today's digital world, we use electronic devices to perform numerous tasks. Whether it be as our form of entertainment, transportation, or profession, the Internet is part of our everyday life. An October 2021 report by DataReportal revealed that there are 4.88 billion individuals that use the Internet across the globe (Kemp, 2021). Of these individuals, around 4.55 billion people use social media. Social media has managed to penetrate all industries, particularly politics and the way people receive health information. Because of this, businesses – especially marketing firms – have taken an interest in the data that can be attained from these platforms. Now, social media platforms offer analytics for marketers to analyze. But the problem with social media analytics is its subjectivity, as expressed in reviews, blogs, and comments. What if there was a way to analyze this data objectively? Wouldn't it be interesting for a company to find out if their product or service was positively received by their followers? Another problem with the data related to social media analytics is that the amount of it is constantly increasing, due to its presence in our lives.

One solution to these problems is sentiment analysis, a machine learning technique that detects the emotions behind a particular text on a large scale. This analytical technique involves the use of Big Data, which means that programmers can enter thousands of unstructured text and receive the results in an easy-to-analyze, organized format. This technique of ascertaining public opinion can provide useful insights into the perceptions consumers have about a brand, and help firms make decisions on altering their image. It is also beneficial for companies to utilize because social media users discuss various subjects, making it difficult to categorize them. This way, companies can categorize the topics social media users discuss based on their associated emotions. With social media significantly affecting consumer behavior, it would be beneficial for companies to invest in a sentiment analysis tool to learn more about their customers.

Background:

What is Sentiment Analysis?

Sentiment analysis falls under Natural Language Processing (NLP), a computer science field that combines artificial intelligence and linguistics to allow computers to understand the spoken and written word the way humans can (IBM Cloud Education, 2020). Examples of NLP that you commonly encounter are digital assistants, such as Apple's Siri; customer service chatbots; and speech-to-text dictation software. NLP also falls under the category of "big data," meaning that it can summarize large volumes of text, via many different sources. This means that NLP can offer real-time data to companies, thereby streamlining day-to-day operations and increasing employee productivity. The data retrieved from social media platforms is an example of "large volumes of text," where the data lacks structure.

Information such as the number of people that viewed a Post, the times that they viewed it, and how they interacted with the Post (ex. Likes and Comments) is indistinguishable, thus programmers need to organize the information into fixed categories in some way. One problem that arises from this is that social media users tend to showcase their different perspectives on a matter, making it hard for programmers to precisely categorize them (Sant, 2021). A solution to this dilemma is sentiment analysis, a tool developed to extract value from this unstructured text (*What is Text Mining? A Beginner's Guide,* n.d.). Businesses in a variety of industries will find it valuable to invest in sentiment analysis, due to its ability to detect patterns, such as common emotions, from the unstructured data (Joshi, 2018).

How does NLP allow sentiment analysis to take place? Known as *word embedding*, words are represented as vectors, where their value is constructed of real-valued numbers (Genc, 2019). These "real-valued numbers" are formed from inspecting large amounts of data. Each number that makes up the word vector is associated with the English dictionary's definition of the word, along with the value of semantically similar words. For example, “car” and “auto” would have close vectors. Word embedding is useful to sentiment analysis because its numerical values mean that mathematical operations can be used to decipher their relative meaning to one another.

Sentiment analysis determines if a particular piece of content is negative, positive, or neutral. It is especially useful in the field of social media because it helps firms gauge public opinion on a specific topic. For example, marketers can understand customer attitudes and preferences if they can see what people think about their products or their competitors’ products.

Different Sentiment Analysis Techniques

The first sentiment analysis technique is known as lexicon, or rules-based sentiment, where the rules used to evaluate the sentiment of a particular entity (nouns) are based on its association with known positive and negative words (adjectives and adverbs) (*Sentiment Analysis Explained,* n.d.). In other words, nouns are considered entities while adjectives and adverbs are considered emotionally-charged terms. Because most rules-based sentiment models form a relationship between a noun and an adjective (or adverb), the score may be artificially low if the adjective or adverb has a high-intensity level. For instance, the phrase *“I don’t have any problem with it, but I was treated rudely”* may have a low score, because the word “*bad*” has a negative connotation. A programmer may dictate that the sentiment model should overlook contractions such as “*wasn’t*,” which means that this phrase will receive a low score, even though “*wasn’t*” negates the negative association with “*bad*.”

A notable disadvantage of rules-based sentiment systems is that they are inherently naïve, as they can lead you to false conclusions about the data, such as in the case above (Lee, 2021). Thus, a drawback of rules-based sentiment is that the guidelines are written at the programmer’s discretion, so if the programmer does not have complete knowledge of the English language, they may not include syntax conditions. Therefore, a disadvantage of rules-based sentiment is the nature of the rules themselves – the maintenance of them may become very tedious, especially as the English language adds acronyms, double-meanings, and abbreviations into its vocabulary (Lee, 2021). Similarly, a limitation to the rules-based system is that it only evaluates the words themselves, not the context they are written in. Thus, it may not pick up sarcastic or ironic remarks, leading to an inaccurate score.

The other classification of sentiment analysis is a multi-layered model, where individual entities are given their own sentiment score (*Sentiment Analysis: A Definitive Guide*, n.d.). In the above example, the "*bad*" and "*wasn't*" could cancel each other out, potentially offering a more neutral score. Thus, the multi-layered approach offers deeper insight into the writer's emotions, as it can be used on documents from a particular person or event. The multi-layered model is often viewed in machine learning algorithms. Many machine learning algorithms involve training data. Training data is the data used to teach a machine learning model to predict the outcomes the programmer designs, by having the algorithm associate input texts to corresponding tags. This is similar to training for a chess game – you must consider all the various moves your opponent may make before the actual match. In supervised machine learning, data must be labeled, which means that any subjectivity and bias in the data is reflected in the system. In comparing the rules-based model and machine learning, the rules-based offers more control over the output and is most appropriate for specific applications (Lee, 2021).

Choosing a Sentiment Analysis Tool

The first glimpse into the realm of sentiment analysis was through Brand24, a media monitoring platform that offers insights into social media trends, such as sentiment analysis

(Tom, 2018). A unique ability Brand24 had was that it offered insights into *all* forms of media, including Twitter, Instagram, Facebook, News, Podcasts, Blogs, and the Web in general. By inputting a couple of keywords associated with your company, you would be able to view the general sentiment for content regarding the words in question. But one thing it lacked was its inability to allow users to *choose* the content – the only thing that can be controlled are the keywords.

An evaluation of the Application Program Interfaces (APIs) was completed to determine the flexibility desired. Upon looking up "sentiment analysis APIs," one can see many technology firms with their own text processing techniques. Text processing encompasses many techniques found in Natural Language Processing (NLP), such as detecting the language a piece of content was written in, extracting keywords, and sentiment analysis. A few of these APIs include Amazon Comprehend, Google Cloud Natural Language, and Microsoft Azure (ActiveWizards Group LLC, n.d.).

Amazon Comprehend was first used, but the functionalities did not meet the required needs for the project. The "required needs" were to not manually enter each comment one at a time, which Amazon Comprehend would require.

The Google Cloud NLP API offered various text processing methods, such as language detection, categorizing the text, analyzing the syntax, and analyzing the "entities" (i.e. declaring if the text was about a figure, landmark, etc.) (*Cloud Natural Language Documentation,* n.d.). Upon working with Google Cloud NLP, it took time to realize that its functionalities did not meet the required needs for the project. Similar to the reasons why Amazon Comprehend was not used, Google Cloud NLP offered techniques that were not necessary to this project. For example, the "categorizing the text" feature of Google Cloud NLP was irrelevant, as categories were made by the researcher. For instance, if the content was about the majors that PLNU offered, Google Cloud NLP would categorize it as "education." Furthermore, Google Cloud NLP would be beneficial if you were analyzing an Instagram user's entire profile, rather than individual Posts, due to this categorization feature the API offers.

Then, Valence Aware Dictionary and sEntiment Reasoner (VADER) was found. VADER is an NLP algorithm, available in the NLTK Python package. The Natural Language Processing Toolkit (NLTK) consists of a collection of NLP algorithms. VADER solely focuses on <u>general</u> sentiment analysis, in comparison to Google NLP. It utilizes grammatical and syntactical rules to enact its sentiment criteria. The dictionary consists of words, phrases (such as "the bomb"), emoticons (such as "😊"), and abbreviations (such as "LOL"). It then uses a scale, ranging from -1 for extremely negative to +1 for extremely positive, offering users the polarity and intensity of the emotion evoked. Then it takes the average score as the indicator of the sentiment. In other words, it reveals how much it associates a text with "positive," "negative," and "neutral," and its value. It also offers a collective – or compound - score of the entire text, thereby modifying the polarity of the score.

An advantage VADER had over the other sentiment analysis techniques is that it does not require any training data. Because VADER does not require any training data, it decreases the preparation time needed to set up the algorithm. VADER is also able to handle real-time streaming data, which may be useful for social media data, which is constantly changing. The latter benefit may be handy for businesses, that must make decisions in real-time.

Limitations of Sentiment Analysis

A significant limitation of sentiment analysis is that it does not understand the context of a situation, particularly in the way that humans can (*Sentiment analysis: A definitive guide,* n.d.).

For example, any sentiment analysis API or library cannot detect sarcasm. While this was not necessarily a problem for this dataset, the ability to pick up nuances of human language why affects the sentiment scores of other texts. Similarly, another limitation is that an individual may be using an Emoji ironically, which sentiment analysis – as a machine learning technique – would not be able to determine. Another limitation of sentiment analysis regards the use of Emojis. Emojis may be used in a context that a computer cannot pick up. For example, the snake Emoji may be viewed by a computer as simply a reptile, but a snake may also represent death or an evil being.

Why Python?

As forementioned, sentiment analysis can be conducted via a web application or an API. APIs, such as VADER, can be applied through programming languages. Similar to the spoken word, there are numerous programming languages to choose from. Programmers often decide which one they should use based on their needs. For example, Python is often used for data analysis because of its built-in data structures. This makes it extremely useful for each stage of data analysis (data mining, data processing and modeling, and data visualization) (Ezra, 2021). Data analysis is the process of gathering raw data and converting it into information that can be used for decision-making. Using libraries such as Scrapy and BeautifulSoup offers users the ability to collect data. While Scrapy allows programmers to collect already-structured data from the Internet or APIs, BeautifulSoup cannot collect data from APIs, instead of being able to scrape data and then structure it. Python aids in the data processing and modeling stage by arranging data sets. For instance, the NumPy library is great for performing mathematical operations on big data sets. The Pandas library converts data to the preferred data frame or series. The last stage of data analytics that Python assists with is data visualization, where Matplotlib and Seaborn are often used, due to their easy-to-understand graphics. Thus, an advantage Python offers over other programming languages is that it has a multitude of free libraries that offer unique solutions. Python's simple syntax also makes it ideal for application development, creating software, and scriptwriting. Furthermore, Python can streamline the processing of massive, complex data sets.

A Closer Look at Social Media

<u>The Customer</u>

This project is done with the Marketing Department at Point Loma Nazarene University (PLNU). Nate Hamill, the Senior Content Strategy Manager, was the advisor throughout this project. The Marketing Department's goal is to communicate and visually show the look, life, and culture of PLNU to an outside audience. Their secondary goal was to recruit prospective students. These two goals are demonstrated via five main platforms (LinkedIn, TikTok, Instagram, Twitter, and Facebook). Nate stated that his interest in the project largely depended on the audiences (in other words, the target market) that each social media platform held. LinkedIn mainly consisted of alumni, current students, and the Loma community; TikTok of undergrads; Instagram of prospective and current students; Twitter of current and alumni students; and Facebook of the family of current students. Then, each platform was eliminated as we thought of the presence that PLNU had on each platform. TikTok generally consisted of undergraduate students, which meant that it did not target prospective students, and we wanted to understand the opinions of *all* of PLNU's audience, not just a niche portion of it. Similarly, Facebook did not target prospective students either. Twitter was the least performing of the aforementioned sites. We were left with LinkedIn and Instagram, as both platforms were used by the target market, but ultimately Nate wanted to go with Instagram.

An advantage of using Instagram Posts for sentiment analysis is it is written by many users, thereby eliminating the likelihood of the experiment suffering from small sample bias.

A Closer Look at Instagram

Instagram has four main ways users can engage with their content – IGTV, Reels, Stories, and Posts (Figure 1). IGTV allows users to upload sixty-minute-long videos and is similar to YouTube, in the sense that it generally offers users an avenue to explain a concept in finer detail (Blunt, 2021)). In comparison to the other three types of content, IGTV is harder to discover, as the easiest way to find IGTV content is to go directly to someone's profile. Reels are similar to TikTok, in that they are short videos that are limited to 15 seconds in length. They are more discoverable than IGTV because there is a Reels tab on the navigation portion of the app. On the other hand, Stories are more popular than IGTV and Reels due to the level of interaction users can engage with (Moeller, 2020). They are commonly found on the top of the screen when you open Instagram, so they are hard to miss. Unless a user "pins" the Story, the content will be removed from Instagram after twenty-four hours. In comparison, Posts remain on a user's account forever, until they manually delete them. Users can include a maximum of ten pictures or videos on a single Post, offering them numerous methods to express themselves.

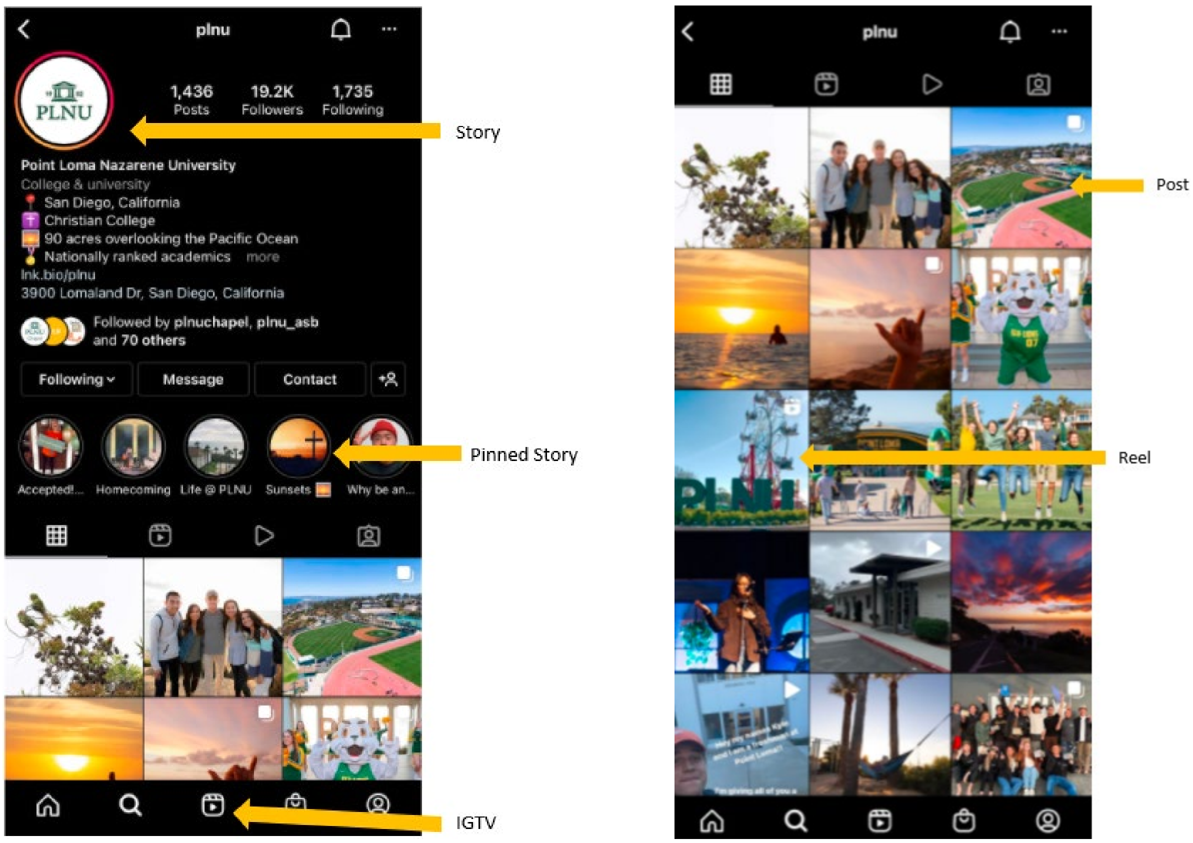


Figure 1. Ways to Post on Instagram.

Web Scraping

To quickly grab Instagram data – specifically the usernames of the commentors, the comments themselves, and any other additional information relating to the Post the comments were from – the concept of web scraping is used. Here, users can gain access to structured web data through an automated system. Also known as data scraping, this automated process means that programmers do not need to waste valuable time manually looking for leads and collecting data. Instead, data scraping collects the data you are looking for and structures it in an easy-to-understand format.

Towards the beginning of the project, sentiment analysis was first implemented on LinkedIn, a networking social media platform that connects job seekers and professionals.

LinkedIn declared that it is completely legal to scrape public data from its platform. In 2019, LinkedIn lost a Supreme Court case against San Francisco company hiQ Labs, a firm that hopes to improve how the human resources department in companies make decisions regarding their employees via data science (Woollacott, 2019). The Court ruled that data scraping was not violating any privacy rights because “people who make data publicly available on a social site don’t have a reasonable expectation of privacy.” The *hiQ Labs v LinkedIn* case set a precedent that scraping publicly available user data would not result in you getting in trouble with the platform.

Data from LinkedIn was scraped using numerous libraries. Libraries are similar to the extension package for a board game – Python provides the foundation, but you can add on more packages if desired. Adding on more libraries also increases the capabilities programmers have.

Once it was decided that this project would focus on Instagram instead of LinkedIn, the next step was to determine how to scrape data on this platform. Instagram has strict guidelines regarding web scraping – as of 2013, Instagram declared that individuals cannot “crawl, scrape, or otherwise cache any content from Instagram” (*Terms of Use • Instagram*, n.d.).

Because of this, it was vital to find a web scraping platform that scraped data from Instagram legally. Apify (a membership-web scraping tool) is a platform that does this. Depending on the “Actor” that you choose, one can discover a platform’s users, commentors, and comments. The Apify Actor “Instagram Comment Scraper,” by the developer Zuzka Pelechová, was used because it offered the all the necessary details required– an ID that corresponds to each comment, the comment’s username, the actual comment, a timestamp for the Post, and more (Figure 2). One benefit of Apify is that it presents the data in a variety of file formats (such as CSV, JSON, and HTML), allowing users many ways to download and analyze the data.

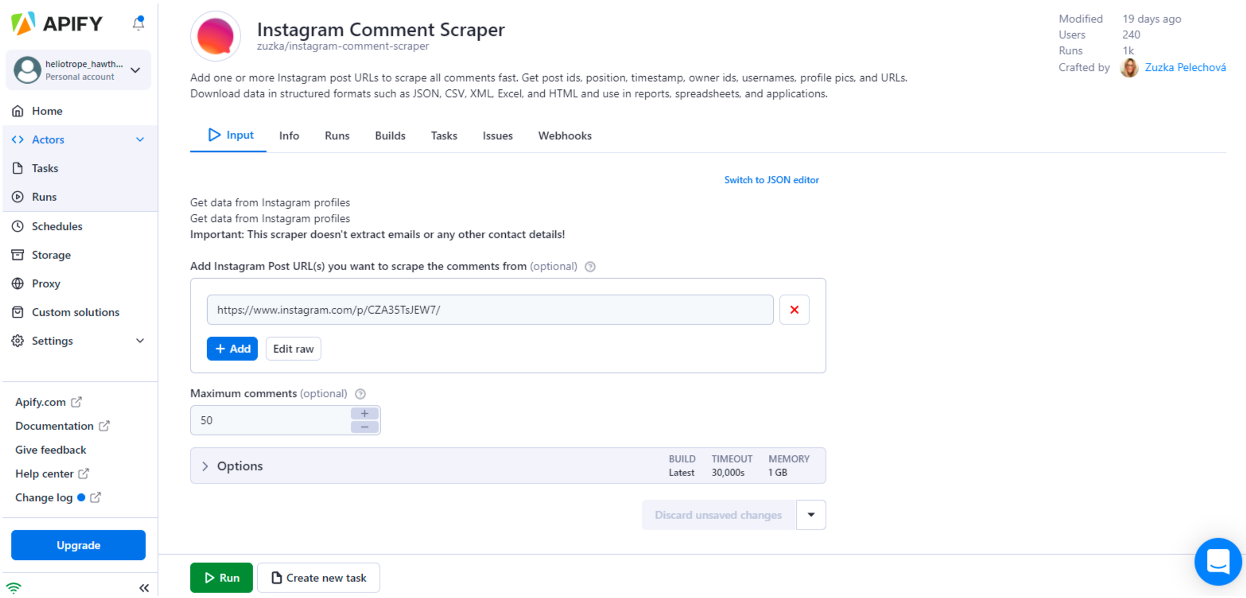


Figure 2. Apify Platform, using the One-Month Free Trial

Ethics of Social Media Research

Research on social media websites has exponentially increased due to the rise in its uses, and its ability to gain a glimpse into people's lives. This raised the question of whether or not this project needed to be approved by the Institutional Research Boards (IRBs), who review research methods to ensure that they are ethically conducted. There are three main research approaches - observational, interactive, and survey/interview (Moreno, 2013).

Observational research is exempt from IRB review if there is public access to the social media website, the information can be attributed to a particular person, and the information gathering does not require interacting with the person who posted it (Moreno, 2013). In this case, "public access" is defined by the decision participants make in deciding if their profile should be public (allowing any user access to the profile) or private (the owner approves the individuals who can access their profile). When social media websites were initially used for research, some

IRBs decided that if a username and password were required to access the website, the site was considered private, therefore consent was needed to view the content. But newer platforms, such as YouTube, only require usernames and passwords to allow those over 18-years-old to post and view videos of adult content. In contrast, Facebook only requires usernames and passwords for marketing purposes and to ensure that the profile owner solely posts content to their page. Still, other websites, such as Instagram, allow the profile owners themselves to make the information available to the public, so the IRB guidelines may need to be reconsidered. Observational research can also be exempt from IRB review if the study observes public behavior, except when information obtained is recorded in a manner where subjects are indirectly or directly identified, any disclosure of subjects' responses outside of research purposes could reasonably place them at risk of criminal or civil liability, or be harmful to the subjects' finances, employment, or reputation (Moreno, 2013).

The research conducted in this project falls under observational, due to the nature of the Instagram - while you do need a username and password to view content, the profile owner must grant you access to view their content, either by being a public or private account. The comments used are identifiable, but not private, as one does not need to follow all the profiles of commenters to view their interaction with a Post. The information-gathering process also required no direct interactions with the subjects involved.

Preprocessing Data

Instagram users have taken to using Hashtags and Emojis in both their comments and captions. Hashtags are often used to increase traffic to a particular comment or caption, or to express the opinions or feelings a user has about a particular matter, such as "#hatersgonnahate" (Warren, 2021). They can also be used to track related posts about a specific event or person, such as “#comedy” and “#fashion.” Similarly, Emojis are used to describe a user’s emotional state when it is too difficult to express in words. A single Emoji character can even enhance the expressivity of a message, especially when supplemented by additional words (Alrumaih, 2020). In Shiha’s exploration of the role Emojis have in sentiment analysis, they found that they improve the overall sentiment scores of positive opinions, in comparison to negative opinions. Furthermore, Alrumaih points out the importance of analyzing Emojis in conjunction with their respective text, because they work together in offering the intended meaning (2020). In other words, Emojis need to be included in the analysis of the text because it increases the accuracy of sentiment analysis.

Liu describes how they cleaned up text messages and status updates from social media platforms Twitter and Facebook (2011). Liu’s team replaced “non-standard” tokens with their respective English word. Liu’s research discovered that one approach classified the non-standard tokens (abbreviation, prefix-clipping, etc.). Liu’s approach did not consider acronyms (ex. “TYSM” for “thank you so much”), because they assumed that each non-standard token is only dependent on one English word. For this reason, the acronyms and text abbreviations were left as is.

Instagram comments also tend to have multiword Hashtags. Removing the spaces between the various words included in a particular Hashtag increases the accuracy of NLP applications (Koehn, 2003). Known as “compound splitting,” this technique breaks up foreign words, translating them into a format that machine learning algorithms can understand. Koehn describes how to properly compound split, you “consider all splits into known words” and use filler words. These filler words are similar to adding spaces in between each word in a Hashtag.

Koehn portrays a limitation of compound splitting – the part-of-speech, because you want to split a compound into nouns, adverbs, adjectives, and verbs, rather than prepositions or determiners.

Methodology:

1. The first thing was to determine which Instagram accounts to analyze. Because the goal was to discover something that the marketing team at a different university was doing that PLNU's was not, it was imperative to find universities with a similar demographic – private, Christian, and in Southern California. Collectively, the following colleges shared the same demographic, similar undergraduate majors, and similar campus structures. The colleges picked were:

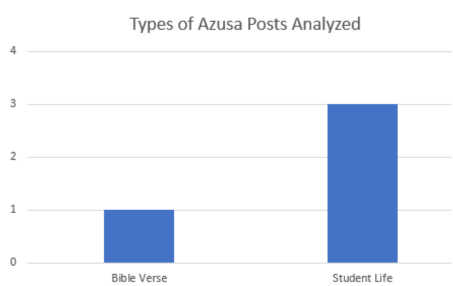
College	Public or Private	Religious Affiliation	Student Size	Location	Instagram Handle
Azusa Pacific University (Azusa)	Private	Evangelical Christian	Around 4,500	Azusa, California	azusapacific
Biola University (Biola)	Private	Nondenominational Christian	Around 3,700	La Mirada, California	biolauniversity
Cal Baptist University (CBU)	Private	Church of Christ	Around 8,100	Riverside, California	calbaptist
Pepperdine University (Pepperdine)	Private	Christian	Around 7,600	Malibu, California	pepperdine
Point Loma Nazarene University (PLNU)	Private	Church of the Nazarene	Around 4,600	San Diego, California	plnu
University of San Diego (USD)	Private	Roman Catholic	Around 8,300	San Diego, California	uofsandiego

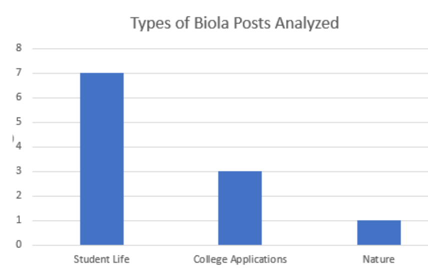
2. The next task was to determine the medium that should be analyzed – Posts, compared to Reels, Stories, or IGTV. Because Stories only last 24 hours on an account before

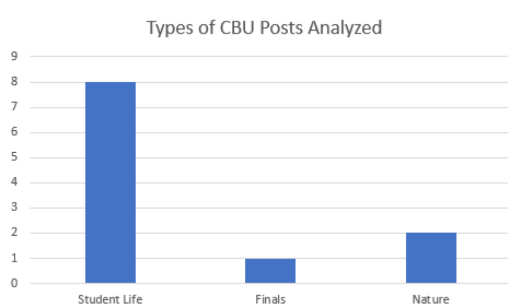
disappearing, and often do not receive comments, they were not analyzed. IGTV was also not analyzed since they may not receive the same level of interactions as Posts and Reels. While it would be interesting to analyze the sentiments of each medium category, it would also be difficult to determine how many people interact with Posts, compared to Reels, Stories, or IGTV.

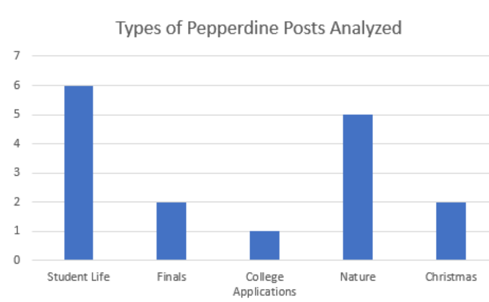
3. The next step was to find which Instagram posts should be analyzed. After performing some preliminary testing, the criteria were settled – a Post needed to have more than 10 comments and a particular post was uploaded between November 1, 2021 and February 12, 2022. A Post having at least 10 Comments was vital because Posts with less than this offered a somewhat low sentiment score, that was not necessarily consistent, therefore having more than 10 comments increased the accuracy of the score.

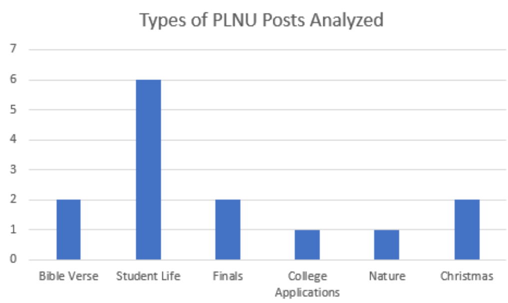
The Posts were put into six different categories: 1) Bible verses, 2) Student Life, 3) Finals, 4) College Applications, 5) Nature, and 6) Christmas. The second category, Student Life, includes student life, student athletics, and programs offered on campus (ex. Study Abroad). The types of Posts analyzed are:

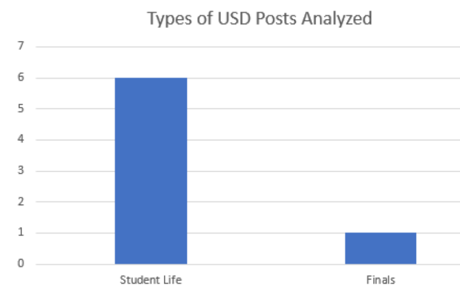












4. The Instagram Posts were analyzed using the Apify Actor Instagram Comment Scraper. This Actor allows you to scrape data from multiple Posts at the same time, easing the process of manually entering the Post URLs. This Actor also presents you with the post ID, the commentor's ID, whether or not the owner is Verified, the URL of the commentor's profile picture, the commentor's username, the "position" on the Post, a "shortCode" (another way to identify a particular comment), the comment itself, and the timestamp of when the comment was posted (Figure 3). Make sure to download the scraped data as a CSV file, because Microsoft Excel files do not work with this particular code (because the CSV module is being used).

	A	B	C	D	E	F	G	H	I	J
1	id	ownerId	ownerIsVerified	ownerProfilePicUrl	ownerUsername	position	postId	shortCode	text	timestamp
2	1.79E+16	25891295	FALSE	https://scontent-mi	[redacted]	1	CYsMO	CYsMO9vPl	@jack__sample where I want you to go ðŸ¤—	2022-01-15T19:
3	1.83E+16	2.23E+08	FALSE	https://scontent-mi	[redacted]	2	CYsMO	CYsMO9vPl	Incredible!!	2022-01-15T00:
4	1.79E+16	1.73E+09	FALSE	https://scontent-mi	[redacted]	3	CYsMO	CYsMO9vPl	Lightning Bolt on the Greek!!! All spectacular ! âš¡ï¸ðŸŒ©	2022-01-14T17:
5	1.80E+16	8.46E+09	FALSE	https://scontent-mi	[redacted]	4	CYsMO	CYsMO9vPl	I like the ðŸ¦-	2022-01-14T16:
6	1.83E+16	1.39E+09	FALSE	https://scontent-mi	[redacted]	5	CYsMO	CYsMO9vPl	Spectacular captures!	2022-01-14T15:
7	1.79E+16	43435550	FALSE	https://scontent-mi	[redacted]	6	CYsMO	CYsMO9vPl	âš¡ï¸	2022-01-14T07:
8	1.80E+16	3.64E+08	FALSE	https://scontent-mi	[redacted]	7	CYsMO	CYsMO9vPl	They are all amazing	2022-01-14T06:
9	1.82E+16	3.28E+08	FALSE	https://scontent-mi	[redacted]	8	CYsMO	CYsMO9vPl	ðŸ˜	2022-01-14T03:
10	1.79E+16	4.81E+09	FALSE	https://scontent-mi	[redacted]	9	CYsMO	CYsMO9vPl	@therna.raw	2022-01-14T03:
11	1.79E+16	1.44E+08	FALSE	https://scontent-mi	[redacted]	10	CYsMO	CYsMO9vPl	The flying turtle!!	2022-01-14T03:
12	1.79E+16	8.09E+09	FALSE	https://scontent-mi	[redacted]	11	CYsMO	CYsMO9vPl	3ðŸ’›	2022-01-14T02:
13	1.80E+16	54289442	FALSE	https://scontent-mi	[redacted]	12	CYsMO	CYsMO9vPl	#3	2022-01-14T01:
14	1.82E+16	3.23E+09	FALSE	https://scontent-mi	[redacted]	13	CYsMO	CYsMO9vPl	The storm	2022-01-14T00:
15	1.80E+16	3.01E+08	FALSE	https://scontent-mi	[redacted]	14	CYsMO	CYsMO9vPl	â¤ï¸	2022-01-14T00:

Figure 3. Example Data from Apify Actor. For anonymity, the owner's Username is hidden.

A notable restriction of the Apify Actor used was that it did not scrape any "hidden" comments (Figure 4). Because of this, the "hidden" comments were manually added. Another criticism of the Apify Actor is that there is a limitation to the number of comments it can scrape. For example, if one specifies that you want 50 comments to be scraped, it may scrape 10 or so. To circumvent this problem, any comments that were not added by the Apify Actor were also manually added.

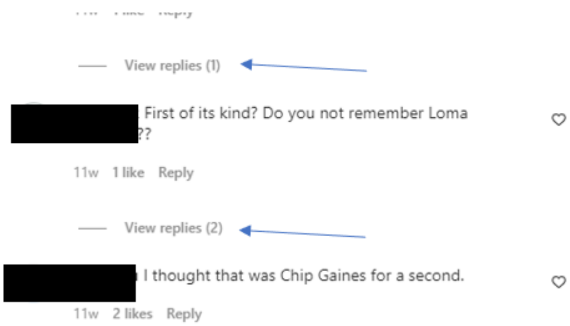


Figure 4. Hidden Comments on an Instagram Post. For anonymity, the owner's Username is hidden.

5. After Apify extracted the necessary data from the Posts, the comments need to be cleaned up before the sentiment score can be determined. In line 13, change the CSV file to the file that you would like to obtain the sentiment scores. Whenever a comment tagged another user via an "@" symbol, the "a" is replaced with the phrase "mentioning" (Figure 5). This was also done to increase the accuracy of the sentiment score. Because Instagram users typically reply with short comments, it was imperative to analyze all the characters, rather than remove any of them. Using the *emoji* library, the Emojis were replaced by their description. Likewise, the Hashtag ("#") symbol was removed and replaced by the term "Hashtag." This was also done to remove a symbol that did not have any actual meaning to VADER. Because Hashtags can appear in multiple words, such as "pointlomanazareneuniversity," spaces were manually created between the words that

created a Hashtag. In the above example, "pointlomanazareneuniversity" would turn into "point loma nazarene university."

emojiCleanup.py > ...

```python
#program cleans up the postComments, replacing the @ mentions (also known as tags) and # hashtags.

import pandas as pd
import emoji
import re
import csv

#create lists that you will append to. comments contains the original text, and demojizedComments contains the preprocessed text
comments = []
demojizedComments = []

#create file that you want to clean data from. Put the 'text' column into the pandas dataframe
df = pd.read_csv('PLNU New Semester.csv')
comments = df['text'].tolist()

#preprocess the comments before analysis
for c in comments:
    t = emoji.demojize(c, delimiters=(" ", " ")) #convert an emoji into a text description of what the emoji is
    newString = t.replace("@", "mentioning ").replace("#", "hashtag ") #replace @ with 'mentioning ' and # with 'hashtag '
    demojizedComments.append(newString)

#add the demojizedComments into the csv file
df['demojizedText'] = demojizedComments
df.to_csv('PLNU Finals.csv', index=False)
```

Figure 5. Python program that cleans up the data

	A	B	C	D	E	F	G	H	I	J	K
1	id	ownerId	ownerIsVerified	ownerProfilePicUrl	ownerUsername	position	postId	shortCode	text	timestamp	demojizedText
2	1.79E+16	25891295	FALSE	https://scontent-mi	[redacted]	1	CYsMO	CYsMO9vPI	@jack__sample where I want you to go ðŸ¤—	2022-01-15T19:	mentioning jack__sample where I want you to go smiling_face_with_open_hands
3	1.83E+16	2.23E+08	FALSE	https://scontent-mi	[redacted]	2	CYsMO	CYsMO9vPI	Incredible!!	2022-01-15T00:	Incredible!!
4	1.79E+16	1.73E+09	FALSE	https://scontent-mi	[redacted]	3	CYsMO	CYsMO9vPI	Lightning Bolt on the Greek!!! All spectacular ! âš¡ï¸ðŸŒ©	2022-01-14T17:	Lightning Bolt on the Greek!!! All spectacular ! high_voltage cloud_with_lightning
5	1.80E+16	8.46E+09	FALSE	https://scontent-mi	[redacted]	4	CYsMO	CYsMO9vPI	I like the ðŸ¦-	2022-01-14T16:	I like the seal
6	1.83E+16	1.39E+09	FALSE	https://scontent-mi	[redacted]	5	CYsMO	CYsMO9vPI	Spectacular captures!	2022-01-14T15:	Spectacular captures!
7	1.79E+16	43435550	FALSE	https://scontent-mi	[redacted]	6	CYsMO	CYsMO9vPI	âš¡ï¸	2022-01-14T07:	high_voltage
8	1.80E+16	3.64E+08	FALSE	https://scontent-mi	[redacted]	7	CYsMO	CYsMO9vPI	They are all amazing	2022-01-14T06:	They are all amazing
9	1.82E+16	3.28E+08	FALSE	https://scontent-mi	[redacted]	8	CYsMO	CYsMO9vPI	ðŸ˜	2022-01-14T03:	smiling_face_with_heart-eyes
10	1.79E+16	4.81E+09	FALSE	https://scontent-mi	[redacted]	9	CYsMO	CYsMO9vPI	@therna.raw	2022-01-14T03:	mentioning therna.raw
11	1.79E+16	1.44E+08	FALSE	https://scontent-mi	[redacted]	10	CYsMO	CYsMO9vPI	The flying turtle!!	2022-01-14T03:	The flying turtle!!
12	1.79E+16	8.09E+09	FALSE	https://scontent-mi	[redacted]	11	CYsMO	CYsMO9vPI	3ðŸ’›	2022-01-14T02:	3 yellow_heart
13	1.80E+16	54289442	FALSE	https://scontent-mi	[redacted]	12	CYsMO	CYsMO9vPI	#3	2022-01-14T01:	hashtag 3
14	1.82E+16	3.23E+09	FALSE	https://scontent-mi	[redacted]	13	CYsMO	CYsMO9vPI	The storm	2022-01-14T00:	The storm
15	1.80E+16	3.01E+08	FALSE	https://scontent-mi	[redacted]	14	CYsMO	CYsMO9vPI	â¤ï¸	2022-01-14T00:	red_heart

Figure 6. Example Output of Figure 5

6. If a Post had any multiword Hashtags, they were manually spaced out. To do so, copy all the contents of Column K into Column L. Then change the Column Name to "hashtagText." Next, manually space the words that make up a specific Hashtag.
7. To conduct sentiment analysis on a particular Post, line 11 (Figure 7) was changed to the appropriate file. If Step 6 was followed, use line 14, commenting line 13 out, so it is not used. If it is was not, use line 13, commenting line 14 out.

sentimentVader.py > ...

```python
1   #program assigns each preprocessedComment with a 'negative, neutral, positive, compound' score
2
3   from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
4   import pandas as pd
5   import csv
6   import os
7   import argparse
8
9   #create file that you want to clean data from. Put the 'text' column into the pandas dataframe
10  file = 'PLNU New Semester.csv'
11  df = pd.read_csv(file)
12  demojiComments = df['demojizedText'].tolist() #use when there are no hashtags
13  #demojiComments = df['hashtagText'].tolist() #use when there are hashtags. I manually broke up the multiword hashtags
14
15  scores = []
16
17  #find the sentiment scores for each post. It will output a 'negative, neutral, positive, compound' score
18  sid = SentimentIntensityAnalyzer()
19  for c in demojiComments:
20      sen = sid.polarity_scores(c)
21      scores.append(sen)
22
23  #add the score into the csv file
24  df['scores'] = scores
25  df.to_csv(file, index=False)
```

Figure 7. Python Program that Finds Sentiment Scores of a particular CSV file

	A	B	C	D	E	F	G	H	I	J	K	L
1	id	ownerId	ownerIsVerified	ownerProfilePicUrl	ownerUsername	position	postId	shortCode	text	timestamp	demojizedText	scores
2	1.79E+16	25891295	FALSE	https://scontent-mi	[redacted]	1	CYsMO	CYsMO9vPl	@jack__sample where	2022-01-15T19:	mentioning jack__sample wher	{'neg': 0.0, 'neu': 0.86, 'pos': 0.14, 'compound': 0.0772}
3	1.83E+16	2.23E+08	FALSE	https://scontent-mi	[redacted]	2	CYsMO	CYsMO9vPl	Incredible!!	2022-01-15T00:	Incredible!!	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
4	1.79E+16	1.73E+09	FALSE	https://scontent-mi	[redacted]	3	CYsMO	CYsMO9vPl	Lightning Bolt on the (	2022-01-14T17:	Lightning Bolt on the Greek!!! A	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
5	1.80E+16	8.46E+09	FALSE	https://scontent-mi	[redacted]	4	CYsMO	CYsMO9vPl	I like the ðŸ¦-	2022-01-14T16:	I like the seal	{'neg': 0.0, 'neu': 0.545, 'pos': 0.455, 'compound': 0.3612}
6	1.83E+16	1.39E+09	FALSE	https://scontent-mi	[redacted]	5	CYsMO	CYsMO9vPl	Spectacular captures!	2022-01-14T15:	Spectacular captures!	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
7	1.79E+16	43435550	FALSE	https://scontent-mi	[redacted]	6	CYsMO	CYsMO9vPl	âš¡ï¸	2022-01-14T07:	high_voltage	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
8	1.80E+16	3.64E+08	FALSE	https://scontent-mi	[redacted]	7	CYsMO	CYsMO9vPl	They are all amazing	2022-01-14T06:	They are all amazing	{'neg': 0.0, 'neu': 0.441, 'pos': 0.559, 'compound': 0.5859}
9	1.82E+16	3.28E+08	FALSE	https://scontent-mi	[redacted]	8	CYsMO	CYsMO9vPl	ðŸ˜	2022-01-14T03:	smiling_face_with_heart-eyes	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
10	1.79E+16	4.81E+09	FALSE	https://scontent-mi	[redacted]	9	CYsMO	CYsMO9vPl	@therna.raw	2022-01-14T03:	mentioning therna.raw	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
11	1.79E+16	1.44E+08	FALSE	https://scontent-mi	[redacted]	10	CYsMO	CYsMO9vPl	The flying turtle!!	2022-01-14T03:	The flying turtle!!	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
12	1.79E+16	8.09E+09	FALSE	https://scontent-mi	[redacted]	11	CYsMO	CYsMO9vPl	3ðŸ’›	2022-01-14T02:	3 yellow_heart	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
13	1.80E+16	54289442	FALSE	https://scontent-mi	[redacted]	12	CYsMO	CYsMO9vPl	#3	2022-01-14T01:	hashtag 3	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
14	1.82E+16	3.23E+09	FALSE	https://scontent-mi	[redacted]	13	CYsMO	CYsMO9vPl	The storm	2022-01-14T00:	The storm	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
15	1.80E+16	3.01E+08	FALSE	https://scontent-mi	[redacted]	14	CYsMO	CYsMO9vPl	â¤ï¸	2022-01-14T00:	red_heart	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}

Figure 8. Example Output of Figure 7

After retrieving the sentiment scores, clean up the score itself so you can separate the compound score. This way, you can analyze the compound score itself. To do this, change line 18 to the appropriate file (Figure 9). This program will also determine if a particular post is positive ("pos"), neutral ("neu"), or negative ("neg").

splitText.py > ...

```python
#program separates the compound score from the rest of the score. The compound score will be used to create pie charts

from decimal import InvalidContext
from operator import index
import pandas as pd
import csv
import os
import argparse

#create lists that you will append to
newScores = [] #contains the scores that no longer have the single quotation marks and brackets
textScores = [] #contains everything to the right of the compound score (the phrase "compound" and its actual score)
compoundScores = [] #removes the word "compound" from the textScores array
compS = [] #convert the compoundScore into a float
compScores = [] #state that the compoundScore is positive, negative or neutral

#create file that you want to clean data from. Put the 'scores' column into the pandas dataframe
file = 'PLNU New Semester.csv'
df = pd.read_csv(file)
scores = df['scores'].tolist()

#remove the single quotation marks and brackets
for s in scores:
    newString = s.replace("'", "").replace('{',"").replace('}', "")
    newScores.append(newString)

#delete the neg, neu, pos scores and only keep the compound score (everything to the right of compound:)
for s in newScores:
    index = s.find('compound: ')
    if index !=-1:
        textScores.append(s[index:])

#remove the word "compound"
for c in textScores:
    newC = c.replace("compound: ", "")
    compoundScores.append(newC)

#change the format of the compound score to act as floats
for c in compoundScores:
    newNum = float(c)
    compS.append(newNum)

#add the score into the csv file
df['compound scores'] = compS
df.to_csv(file, index=False)

#find the compound score (positive/negative) of the comment
for cs in compS:
    if cs >= 0.05 :
        compScores.append('pos')
    elif cs > -0.05 and cs < 0.05 :
        compScores.append('neu')
    elif cs <= -0.05 :
        compScores.append('neg')

#add the compScore into the csv file
df['comp scores'] = compScores
df.to_csv(file, index=False)
```

Figure 9. Python Program that cleans up the sentiment scores of a particular CSV file

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	id	ownerId	ownerIsVerified	ownerProfilePicUrl	ownerUsername	position	postId	shortCode	text	timestamp	demojizedText	scores	compound scores	comp scores
2	1.79E+16	25891295	FALSE	https://scontent-mi	[redacted]	1	CYsMO	CYsMO9vPI	@jack__sample where	2022-01-15T19:	mentioning jack__sample wher	{'neg': 0.0, 'neu': 0.86, 'pos': 0.14, 'compound': 0.0772}	0.0772	pos
3	1.83E+16	2.23E+08	FALSE	https://scontent-mi		2	CYsMO	CYsMO9vPI	Incredible!!	2022-01-15T00:	Incredible!!	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}	0	neu
4	1.79E+16	1.73E+09	FALSE	https://scontent-mi		3	CYsMO	CYsMO9vPI	Lightning Bolt on the (	2022-01-14T17:	Lightning Bolt on the Greek!!! A	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}	0	neu
5	1.80E+16	8.46E+09	FALSE	https://scontent-mi		4	CYsMO	CYsMO9vPI	I like the ðŸ¦-	2022-01-14T16:	I like the seal	{'neg': 0.0, 'neu': 0.545, 'pos': 0.455, 'compound': 0.3612}	0.3612	pos
6	1.83E+16	1.39E+09	FALSE	https://scontent-mi		5	CYsMO	CYsMO9vPI	Spectacular captures!	2022-01-14T15:	Spectacular captures!	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}	0	neu
7	1.79E+16	43435550	FALSE	https://scontent-mi		6	CYsMO	CYsMO9vPI	âš¡ï¸	2022-01-14T07:	high_voltage	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}	0	neu
8	1.80E+16	3.64E+08	FALSE	https://scontent-mi		7	CYsMO	CYsMO9vPI	They are all amazing	2022-01-14T06:	They are all amazing	{'neg': 0.0, 'neu': 0.441, 'pos': 0.559, 'compound': 0.5859}	0.5859	pos
9	1.82E+16	3.28E+08	FALSE	https://scontent-mi		8	CYsMO	CYsMO9vPI	ðŸ˜	2022-01-14T03:	smiling_face_with_heart-eyes	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}	0	neu
10	1.79E+16	4.81E+09	FALSE	https://scontent-mi		9	CYsMO	CYsMO9vPI	@therna.raw	2022-01-14T03:	mentioning therna.raw	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}	0	neu
11	1.79E+16	1.44E+08	FALSE	https://scontent-mi		10	CYsMO	CYsMO9vPI	The flying turtle!!	2022-01-14T03:	The flying turtle!!	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}	0	neu
12	1.79E+16	8.09E+09	FALSE	https://scontent-mi		11	CYsMO	CYsMO9vPI	3ðŸ’›	2022-01-14T02:	3 yellow_heart	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}	0	neu
13	1.80E+16	54289442	FALSE	https://scontent-mi		12	CYsMO	CYsMO9vPI	#3	2022-01-14T01:	hashtag 3	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}	0	neu
14	1.82E+16	3.23E+09	FALSE	https://scontent-mi		13	CYsMO	CYsMO9vPI	The storm	2022-01-14T00:	The storm	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}	0	neu
15	1.80E+16	3.01E+08	FALSE	https://scontent-mi		14	CYsMO	CYsMO9vPI	â¤ï¸	2022-01-14T00:	red_heart	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}	0	neu

Figure 10. Example Output of Figure 9

<u>Limitations of the Methodology</u>

One limitation was that the Posts were chosen based on criteria, meaning that the composite sentiment score may not encompass all of the opinions of Instagram users that interact with a particular university's post. Another limitation is the criteria itself – numerous posts were not analyzed only because they did not have enough comments. This means that these commentors will not contribute to the overall sentiment score. Because the requirements were built on how many comments there were, an issue with the data is the accuracy of it, since some posts generated more comments than others.

A critique of the Posts used is that commentors were often negatively triggered by specific post captions and responded accordingly. For example, a university post illustrates their decision to have the first few weeks of school be completely online. This led to numerous, relatively furious, people stating their dislike of this act, often using extremely negative words to voice their opinions, in particular, concerning COVID-19. Because VADER is not able to understand the context of the situation, these comments may offer an inaccurate depiction of the sentiments felt by all people at a particular university. This also means that the people that, for example, may have appreciated this decision to delay the return to school may not have an accurate score either.

Furthermore, an issue with the Apify Actor used was that the comments used may have been out of order. Although this did not affect the overall sentiment score, this may affect the results if someone was to perform the sentiment score on each comment individually. Another issue with Apify is that it did not scrape all the comments corresponding to a particular Post. My solution was to manually add the comments that were not scraped, including "hidden" comments. The problem with manually adding comments is that I may have missed some, which could contribute to an inaccurate score.

Results and Analysis:

General Thoughts

One thing that frequently came up was that asking more open-ended questions in an Instagram post caption, or a pinned comment, can increase engagement. This is something that PLNU often didn't do, so they could work on this. Another interesting fact is that the posts from the schools I analyzed often received more neutral scores than positive or negative. This is most likely because commenters would tag their friends and say something like "I see you" and little else. It was also interesting to see that a single word such as "very" and "no" could alter the sentiment score significantly. In general, the time period used (November to February) was where PLNU didn't post anything controversial, while the other schools did. A controversial issue that received a lot of negative scores had to do with mask-wearing and a school's COVID-19 policies.

Analysis of Universities' Bible Verses Posts

When it came to which college had the highest positive sentiment, Azusa received this honor. Azusa ranked higher than PLNU, by about 24%, indicating that their Posts generate more Comments than PLNU's do (Figures 11 and 12). This may have to do with the content of the Posts themselves (rather than the Verse chosen) – Azusa's had to do with the campus itself, while PLNU's depicted students. It could also be due to the relationship that the Post had with the Bible quotation, and if they correlated or not. For example, Azusa's Post depicted the mountains close to their campus and the Bible Verse described the interactions the author of the Psalms had with a mountain. The results may indicate that Bible Verses are still good ways for PLNU's Marketing Department to interact with their followers, but is not the best way to generate positive remarks. It may also suggest that PLNU's Marketing Department could increase engagement by generating interactive content around the Bible Verses themselves, rather than simply portraying students. A way this can be done is by posing questions like "Which Bible Verse do you live by?"

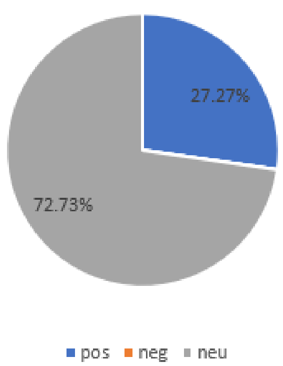


Figure 11. Sentiment Scores for Azusa's Bible Verses Posts. Analyzes 11 Comments.

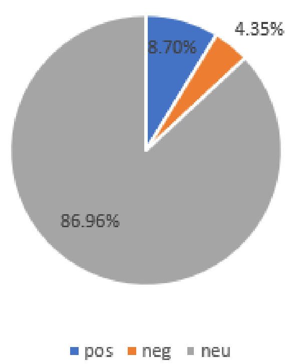


Figure 12. Sentiment Scores for PLNU's Bible Verses Posts. Analyzes 23 Comments.

Analysis of Universities' Student Life Posts

Biola had the highest positive sentiment in this category, with 51%, PLNU coming in second at 48% (Figures 14 and 17). Despite the category being so broad, Biola's Posts showing study abroad and campus tours generated a lot of engagement in their audience. In comparison, PLNU depicted campus events (such as Homecoming) and athletics. This suggests that student takeovers, where the Marketing Department reveals what students are doing, and campus tours could increase the number of Comments the university can receive. Through these takeovers and tours, the target audience could learn more about what students do on campus. Biola's Posts were more interactive, using a range of Reels and Posts to depict the interactions students had with faculty. This variety in posting styles may also interest the target audience more, so the Marketing Department could benefit by doing this.

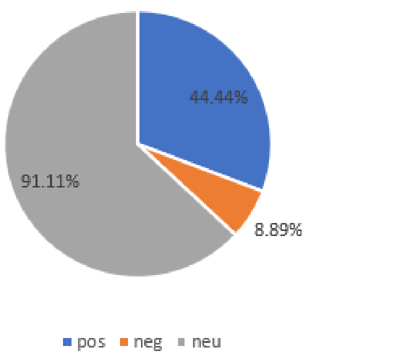


Figure 13. Sentiment Scores for Azusa's Student Life Posts. Analyzes 45 Comments.

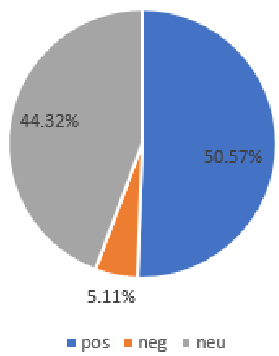


Figure 14. Sentiment Scores for Biola's Student Life Posts. Analyzes 176 Comments.

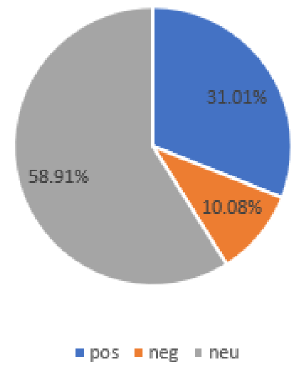


Figure 15. Sentiment Scores for CBU's Student Life Posts. Analyzes 129 Comments.

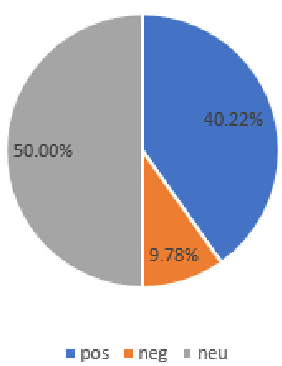


Figure 16. Sentiment Scores for Pepperdine's Student Life Posts. Analyzes 92 Comments.

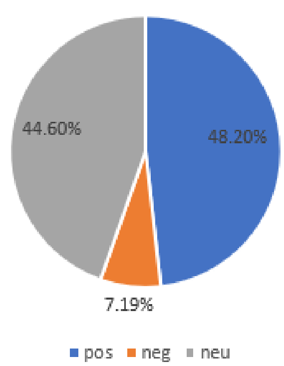


Figure 17. Sentiment Scores for PLNU's Student Life Posts. Analyzes 139 Comments.

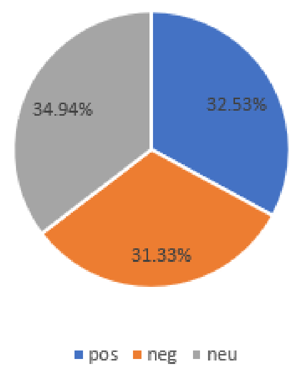


Figure 18. Sentiment Scores for USD's Student Life Posts. Analyzes 105 Comments.

Analysis of Universities' College Applications Posts

PLNU ranked first in the College Applications category, implying that our current posts depicting students committing to our school are working (Figure 21). Biola's ranking (as third) is a really good example of how many people would tag their friends but not do much else (Figure

19). While Biola did have about 18% of positive sentiment, they also had 75% of neutral sentiment.

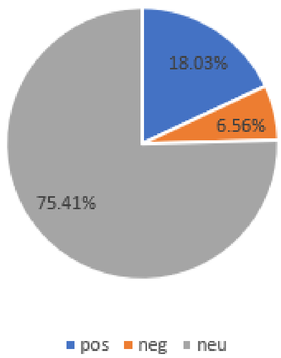


Figure 19. Sentiment Scores for Biola's College Applications Posts. Analyzes 122 Comments.

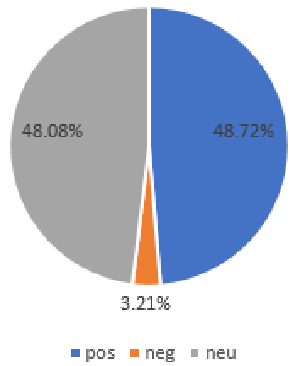


Figure 20. Sentiment Scores for Pepperdine's College Applications Posts. Analyzes 156 Comments.

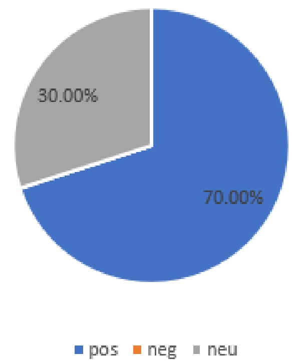


Figure 21. Sentiment Scores for PLNU's College Applications Posts. Analyzes 20 Comments.

Analysis of Universities' Nature Posts

In the Nature category, Biola ranked first, with PLNU coming in fourth (Figures 22 and 25). Biola's Post was a Reel of the scenic spots around their campus, while PLNU's was a series of shots, ranging from athletics to pictures of the wildlife around campus. This may suggest that people respond better to Reels, and therefore have positive sentiment, or may not like nature pictures where people are featured.

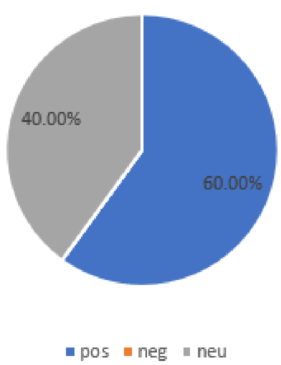


Figure 22. Sentiment Scores for Biola's Nature Posts. Analyzes 15 Comments.

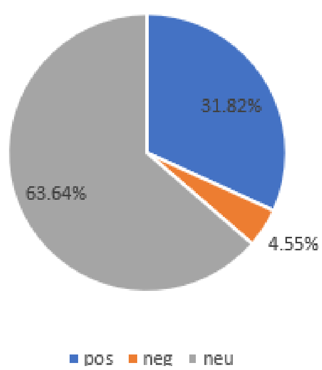


Figure 23. Sentiment Scores for CBU's Nature Posts. Analyzes 44 Comments.

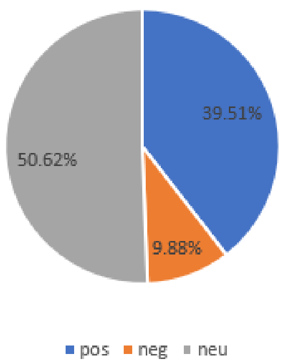


Figure 24. Sentiment Scores for Pepperdine's Nature Posts. Analyzes 81 Comments.

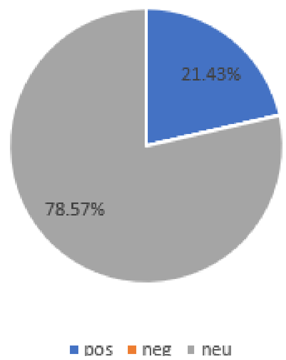


Figure 25. Sentiment Scores for PLNU's Nature Posts. Analyzes 14 Comments.

Analysis of Universities' Finals Posts

Pepperdine ranked first in the Finals category, with PLNU coming in second (Figures 27 and 28). In their posts, Pepperdine asked their alumni for tips on study habits and self-care, and also featured their university chaplain to pray over the students. But in PLNU's, the Marketing Department featured a prayer for students to read, in addition to saying goodbye to their students. This suggests that people respond more to questions that they can answer and interact with. An idea that could increase engagement is to ask for study tips. The level of engagement could also be increased if PLNU faculty were involved, similar to what Pepperdine did. It is also significant to note that CBU received the lowest sentiment score in this category – at 13% (Figure 26). This is another example of how a post can receive a high neutral sentiment score when individuals mention their friends and little else. This implies that PLNU should not create posts like CBU's.

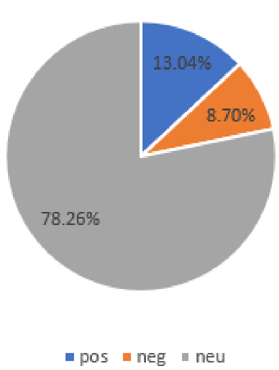


Figure 26. Sentiment Scores for CBU's Finals Posts. Analyzes 46 Comments.

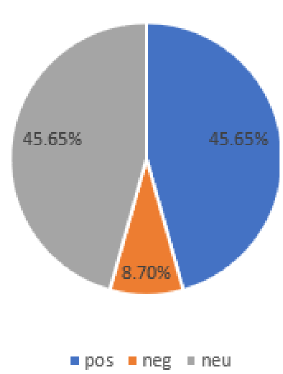


Figure 27. Sentiment Scores for Pepperdine's Finals Posts. Analyzes 46 Comments.

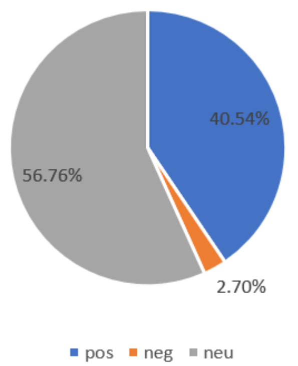


Figure 28. Sentiment Scores for PLNU Finals Posts. Analyzes 37 Comments.

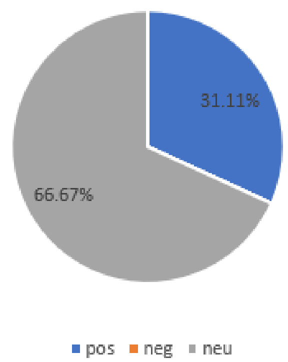


Figure 29. Sentiment Scores for USD's Finals Posts. Analyzes 21 Comments.

Analysis of Universities' Christmas Posts

It was extremely interesting to look at this category, especially since Pepperdine's positive sentiment score ranked significantly higher than PLNU's – 64% compared to 18% (Figures 30 and 31). Pepperdine's post encompassed their Christmas tree, while Point Loma's had more to do with campus events surrounding Christmas. This implies that people respond more when they feel like they are part of the community, as they see the Christmas tree being built, rather than something that they can expect - the Marketing Department depicting another campus event. This suggests that Instagram users appreciated the construction of an enormous Christmas tree more than events specific to a university.

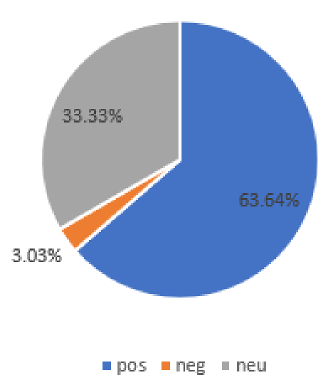


Figure 30. Sentiment Scores for Pepperdine's Christmas Posts. Analyzes 33 Comments.

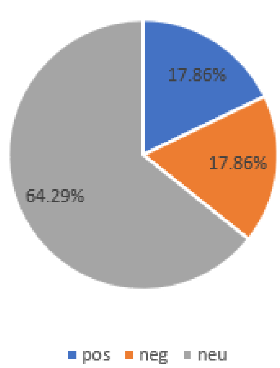


Figure 31. Sentiment Scores for PLNU's Christmas Posts. Analyzes 28 Comments.

Analysis of Overall Sentiment Scores

When taking a look at the sentiment scores for each university, rather than by category, Pepperdine had the highest positive sentiment – at 46% (Figure 35). When analyzing the overall sentiment scores for the colleges, Pepperdine was ranked first, thereby implying that the target audience gravitated towards the categories of Christmas and Finals, in comparison to PLNU. This suggests that Pepperdine's posts where students can interact with their faculty and events unique to the college can increase the number of positive comments created. But Biola ranked first in two categories as well (Student Life and Nature), suggesting that the PLNU Marketing Department could increase engagement by separating campus events from nature shots. For instance, Posts in the Student Life category should primarily be about campus life, while Nature Posts only feature Mother Nature and wildlife.

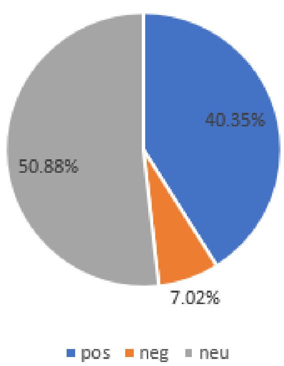


Figure 32. Overall Sentiment Scores for Azusa. Analyzes 239 Comments.

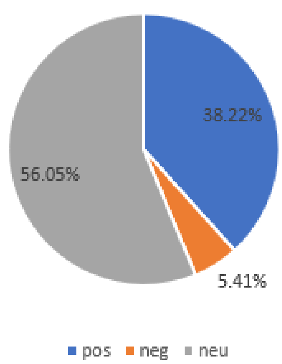


Figure 33. Overall Sentiment Scores for Biola. Analyzes 314 Comments.

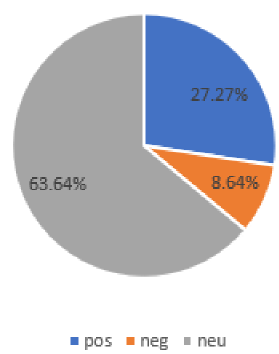


Figure 34. Overall Sentiment Scores for CBU. Analyzes 220 Comments.

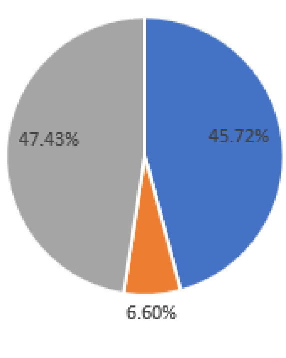


Figure 35. Overall Sentiment Scores for Pepperdine. Analyzes 409 Comments.

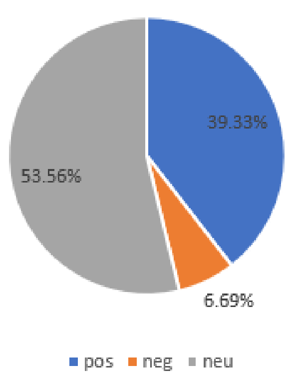


Figure 36. Overall Sentiment Scores for PLNU. Analyzes 239 Comments.

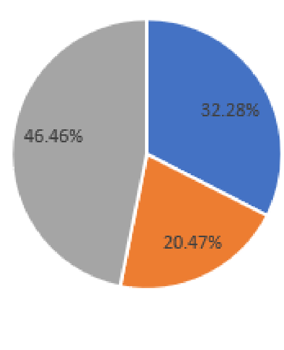


Figure 37. Overall Sentiment Scores for USD. Analyzes 127 Comments.

Conclusion:

The primary research objective was to find a sentiment analysis tool that was effective. Furthermore, the marketing objectives were to determine areas that PLNU's Marketing Department that could increase engagement, based on their sentiment scores. VADER is an extremely effective tool that resulted in consistent sentiment scores. Here, "consistent" is defined as being fair or accurate. But "consistent" scores can also apply to the initial hypotheses created upon first viewing the Comments. For example, many sentiment scores appeared to be compatible with the Comment itself. The Marketing Department also gained numerous insights into the Instagram Posts that gain their users' attention, and content that did not. Through this research, the Marketing Department was introduced to a different way of creating their content – through categories. The six categories created in this research, and the recommendations provided in each, could introduce new levels of engagement offered to consumers.

Next Steps:

In the future, it would be interesting to analyze the username tagged in a post. Does including the username increase the accuracy of the sentiment score? Excluding? For instance, @johnSmith135 created a post, stating "@plnu I love this!!!," if "plnu" was removed, could removing the "plnu" portion of the content alter the sentiment score? It would also be interesting to see if a post's sentiment score is increased when "expanding" text abbreviations. It would also be interesting to detect sentiment analysis scores based on a particular Hashtag, or posts that received many likes.

References:

ActiveWizards Group LLC. (n.d.). *Comparison of the Most Useful Text Processing APIs*. ActiveWizards. Retrieved March 22, 2022, from https://activewizards.com/blog/comparison-of-the-most-useful-text-processing-apis/.

Alrumaih, A., Al-Sabbagh, A., Alsabah, R., Kharrufa, H., & Baldwin, J. (2020). Sentiment Analysis of Comments in Social Media. *International Journal of Electrical and Computer Engineering (IJECE)*, *10*(6), 5917–5922. https://doi.org/10.11591/ijece.v10i6.pp5917-5922.

Blunt, W. (2021, July 27). *IGTV vs. Reels vs. Stories: How to Leverage Each Instagram Video Type*. ShortStack. Retrieved March 22, 2022, from https://www.shortstack.com/blog/igtv-vs-reels-vs-stories/.

cjhutto. (2021, March 15). *VADER-Sentiment Analysis.* GitHub. Retrieved March 17, 2022, from https://github.com/cjhutto/vaderSentiment.

Cloud Natural Language Documentation. (n.d.). Google. Retrieved March 22, 2022, from https://cloud.google.com/natural-language/docs#docs

Ezra, M. (2021, May 4). *Why Python is Essential for Data Analysis?* Section. Retrieved March 22, 2022, from https://www.section.io/engineering-education/why-python-is-essential-for-data-analysis/#:~:text=Python's%20built%2Din%20analytics%20tools,critical%20matrices%20in%20evaluating%20performance.

Genc, O. (2019, April 21). *The Basics of NLP and Real Time Sentiment Analysis with Open Source Tools*. Medium. Retrieved February 16, 2022, from https://towardsdatascience.com/real-time-sentiment-analysis-on-social-media-with-open-source-tools-f864ca239afe.

IBM Cloud Education. (2020, July 2). *What is Natural Language Processing?* IBM. Retrieved March 17, 2022, from https://www.ibm.com/cloud/learn/natural-language-processing.

Instagram. (n.d.). *Terms of Use*. Terms of Use • Instagram. Retrieved March 3, 2022, from https://www.instagram.com/about/legal/terms/before-january-19-2013/.

Joshi, Naveen. (2018, March 12). *5 Reasons Why You Must Care About Customer Sentiment Analysis.* Allerin Tech PVT LTD. Retrieved March 22, 2022, from https://www.allerin.com/blog/5-reasons-why-you-must-care-about-customer-sentiment-analysis.

Kemp, S. (2021, December 1). *Digital 2021 October Global Statshot Report*. DataReportal. Retrieved March 17, 2022, from https://datareportal.com/reports/digital-2021-october-global-statshot.

Koehn, P., & Knight, K. (2003). Empirical Methods for Compound Splitting, *10th Conference of the European Chapter of the Association for Computational Linguistics*. Retrieved from https://aclanthology.org/E03-1076.pdf.

Lee, K. C. (2021, May 29). *Sentiment Analysis - Comparing 3 Common Approaches: Naive Bayes, LSTM, and VADER*. Medium. Retrieved February 1, 2022, from https://towardsdatascience.com/sentiment-analysis-comparing-3-common-approaches-naive-bayes-lstm-and-vader-ab561f834f89#:~:text=More%20than%20ever%2C%20organizations%20are,Trained%20Rule%2DBased%20VADER%20Models.

Liu, F., Weng, F., Wang, B., & Liu, Y. (2011). Insertion, Deletion, or Substitution? Normalizing Text Messages Without Pre-categorization nor Supervision. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, 71–76.

Moeller, S. (2020, October 7). *The Ultimate Guide to Instagram features: Stories, Feed, IGTV, Reels*. Content Marketing Institute. Retrieved March 22, 2022, from https://contentmarketinginstitute.com/2020/10/instagram-stories-feed-igtv-reels/.

Moreno, M. A., Goniu, N., Moreno, P. S., & Diekema, D. (2013, September). *Ethics of Social Media Research: Common Concerns and Practical Considerations*. NCBI. Retrieved February 3, 2022, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3942703/.

Sant, V. (2021, January 31). *The Big-Data Boom Explained in More Straightforward Terms*. Medium. Retrieved March 21, 2022, from https://medium.com/swlh/the-big-data-boom-explained-in-more-straightforward-terms-954f4eae3585.

Sentiment Analysis: A Definitive Guide. MonkeyLearn. (n.d.). Retrieved March 22, 2022, from https://monkeylearn.com/sentiment-analysis/.

Sentiment Analysis Explained. Lexalytics. (n.d.). Retrieved March 3, 2022, from https://www.lexalytics.com/technology/sentiment-analysis.

Terms of Use • Instagram. Instagram. (n.d.). Retrieved March 23, 2022, from https://www.instagram.com/about/legal/terms/before-january-19-2013/#:~:text=You%20must%20not%20crawl%2C%20scrape,written%20consent%20from%20Instagram%2C%20inc.

Tom. (2018, May 22). *How to Use Sentiment Analysis in Brand24 for Brand Building*. Brand24 Blog. Retrieved March 22, 2022, from https://brand24.com/blog/how-to-use-sentiment-analysis-in-brand24-for-brand-building/. *What is text mining? A beginner's guide.*

MonkeyLearn. (n.d.). Retrieved March 22, 2022, from https://monkeylearn.com/text-mining/.

Warren, J. (2021, November 18). *Instagram Hashtags: Everything You Need to Know in 2022*. Later. Retrieved March 23, 2022, from https://later.com/blog/ultimate-guide-to-using-instagram-hashtags/#:~:text=Later's%20data%20report%20shows%20that,your%20reach%20and%20engagement%20rates.

Woollacott, E. (2019, September 10). *LinkedIn Data Scraping Ruled Legal*. Forbes. Retrieved March 3, 2022, from https://www.forbes.com/sites/emmawoollacott/2019/09/10/linkedin-data-scraping-ruled-legal/?sh=6c02979e1b54.