Statistical Analysis on Dream Valence Using Sentiment Analysis Package VADER

Matti Key
University of California, San Diego
Department of Mathematics

Abstract

Dreams are essential to our overall health and well-being. In particular, negative dreams experienced during the REM stage of sleep can have serious mental and physical repercussions. Here, we present a unique approach to analyzing data of dreams through textual data and analysis. The first portion of our analysis focused on exploring the data and applying VADER to our data set. Utilizing the Autoregressive (AR), Moving Averages (MA), and the combined ARMA models, we attempt, with mixed results, to predict the subsequent compound score of randomly selected individuals.

1 Introduction

While the importance of sleep for overall health has been well-established, the significance of dreams is often overlooked. Research has shown that dreams foster creativity and problem-solving, process emotions to reduce anxiety, and solidify long-term memories **[4]**. Dreaming occurs during the rapid eye movement (REM) stage of sleep, which is when the body repairs physical health by increasing blood flow to the brain, regulating our growth hormones, and nourishing metabolic and immune functions **[5]**. Notably, the occurrence of bad dreams, specifically recurring nightmares, is linked to heightened anxiety and increased avoidance of REM sleep in the following days **[6]**. This severe REM sleep deprivation is heavily correlated with mental disorder development **[4]**. Therefore, it is crucial to study both positive and negative dream experiences to promote better health and well-being.

In this paper, we analyze the dream diaries of 94 individuals across the 1940s to 2000s. Specifically, we identified and investigated the valence scores of dream entries using the VADER library, tested the relationship between the sentiment scores and the individual dreamers, and utilized a statistical analysis model to predict future dream sentiment. On the granular level, we scrutinized certain dreamers to identify recurring themes. This paper aims to identify patterns and predict future valence scores of dreams to uncover underlying causes and hopefully mitigate the occurrence of negative dreams.

2 Background/related works

Most research that has been done on dreams involves using brain scan technology on dreamers to analyze psychological patterns and activity in the brain. Still, none involve analyzing textual patterns of the dream records themselves (check sources again).

3 Data Source and Intake

The source of data is the DreamBank website created by Adam Schneider and G. William Domhoff from the Psychology Department at UC Santa Cruz **[1]**. It is a collection of over 20,000 dream reports from a wide range of unique individuals. The diverse demographic of each dreamer provided a fantastic sample to explore dreaming across a broad spectrum of people. Furthermore, the expressive nature of natural language as a medium for evaluating a person's emotional state was critical for this project, and DreamBank delivered on this.

To extract the records contained in the website, we utilized the Selenium package for Python to write a script that would go through the website's pages, pulling the information and converting it into comma-separated files (CSV) and pipe-separated files (PSV) due to the use of commas within many of the records. Below is a sample of the CSVs created from the data within DreamBank.

id	diary_ref	person	description	sex	sex_code	year	entry_count
29	diary29.csv	german dreams (m)	These dream reports -- in German -- were colle...	male	0	1990s	140
26	diary26.csv	emma: 48 years of dreams	Emma is an elderly woman who wrote down her dr...	female	1	1949-1997	1221
55	diary55.csv	madeline 2: college dorms	Madeline is a young woman in her 20s who gave ...	female	1	2000-2001	186
7	diary07.csv	bay area girls: grades 4-6	The 388 dreams in this set were collected in N...	female	1	1996-1997	234
5	diary05.csv	barb sanders #2	1138 more dreams from Barb Sanders, written do...	female	1	1997-2001	1138

Figure 1

raw_number	content	negative	neutral	positive	compound
#009 (07/04/2007)	My dream last night was that it was spring gal...	0.048	0.899	0.052	0.3071
#013 (07/25/2007)	I had the weirdest dream last night! I was jus...	0.134	0.747	0.119	0.2967
#015 (07/31/2007)	Last night I had the worst dream that my mothe...	0.192	0.681	0.127	-0.9329
#024 (12/24/2007)	I had a dream last night that there was a seco...	0.050	0.910	0.040	-0.3594
#052 (02/09/2010)	I had such a vivid dream about graduation in M...	0.052	0.844	0.104	0.9682

Figure 2

Figure 1 shows the details associated with each person. Figure 2 is an example of a single person's dreams. To get the negative, neutral, and positive valence proportions for each dream, the text of each diary entry was fed through the VADER package, which is a sentiment analysis tool that provides accurate and credible scoring of a text's valence **[2]**. The compound score (valence) was critical for many of the analyses to follow and thus needed to be robust. While the VADER package is often utilized for applications within the industry and academia, its reliability in creating a bedrock for our analysis was a question that was important to answer during a portion of our exploratory analysis.

4 Methods

4.1 Sentiment analysis

VADER (Valence Aware Dictionary and Sentiment Reasoner) is a lexicon sentiment analysis tool specifically designed for analyzing text. It is best known for analyzing social media because of its ability to handle slang, emoticons, and other informal language. Once given a text to input, the tool gives a numerical value from the range -1 to +1, where any number towards -1 has more of a negative connotation, +1 has more of a positive connotation, and 0 has a neutral connotation. Vader then gives a negative, neutral, positive, and compound score, which is all of these scores added together but still based in this range. The table with Vader is shown below (Figure 3). A density plot distribution of Vader's dream analysis can also be found below (Figure 4).

	Unnamed: 0	dreams	date	tokenized	stop_words	stemmed	cleaned_sentence
0	0	the one at the meads's house\twhere it's bigge...	1957	[the, one, at, the, meads, s, house, where, it...	[one, meads, house, bigger, inside, european, ...	[one, mead, hous, bigger, insid, european, vil...	one meads house bigger inside european village...
1	1	i'm at a family reunion in a large fine house ...	8/11/67	[i, m, at, a, family, reunion, in, a, large, f...	[family, reunion, large, fine, house, grounds,...	[famili, reunion, larg, fine, hous, ground, ma...	family reunion large fine house grounds maybe ...
2	2	i watch a plane fly past and shortly realize i...	8/1/85	[i, watch, a, plane, fly, past, and, shortly, ...	[watch, plane, fly, past, shortly, realize, lo...	[watch, plane, fli, past, shortli, realiz, low...	watch plane fly past shortly realize low crash...
3	3	me pulling the green leaves and berries off so...	1985?	[me, pulling, the, green, leaves, and, berries...	[pulling, green, leaves, berries, branches, bu...	[pull, green, leav, berri, branch, bush, live,...	pulling green leaves berries branches bush liv...
4	4	i'm in a room that reminds me of (but definite...	1985?	[i, m, in, a, room, that, reminds, me, of, but...	[room, reminds, definitely, living, room, hous...	[room, remind, definit, live, room, hous, stre...	room reminds definitely living room house stre...

Figure 3

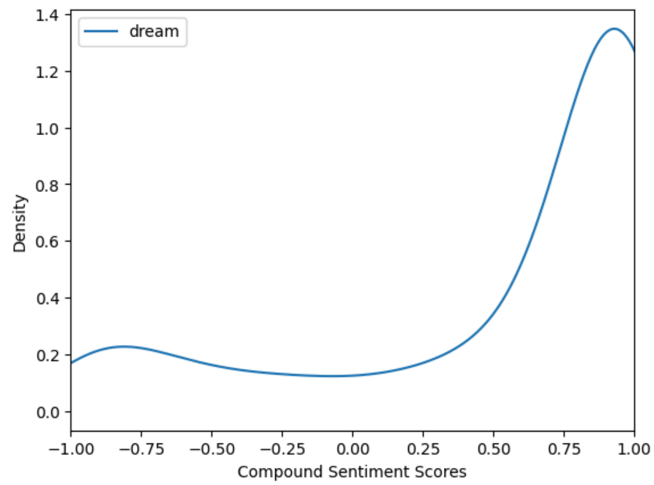

Figure 4

TextBlob is a Python Library that provides a simple API for common NLP tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, and translation. Once given an input, Text Blob will produce two numerical factors: the polarity and subjectivity scores. The polarity score ranges from -1 to +1, with 0 indicating a neutral sentiment, similar to Vader, while the subjectivity score ranges from 0 to +1, with 0 being very objective (factorial) and +1 being subjective (opinionated). Below is the density plot of Text Blob's polarity score (Figure 5) and the table with all sentiment analysis scores (Figure 6).

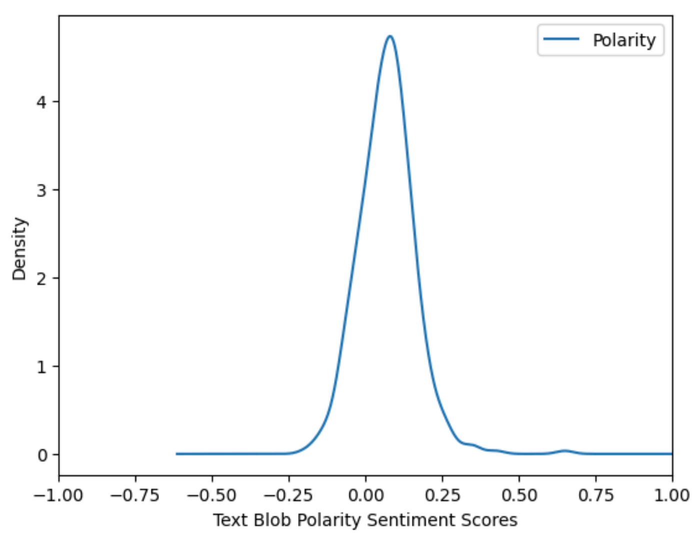

Figure 5

	vader	text_blob_pol	text_blob_sub
count	422.000000	422.000000	422.000000
mean	0.554464	0.070272	0.441186
std	0.629383	0.091867	0.097045
min	-0.985200	-0.192424	0.100000
25%	0.406800	0.013914	0.386881
50%	0.889750	0.072234	0.437046
75%	0.976675	0.123440	0.493844
max	0.999000	0.650000	0.783333

Figure 6

4.1.1 Sentiment Analysis Results

Since Text Blob's polarity score is the most related to Vader's compound score, I will compare them for further analysis. During a correlation analysis, the two sentiment analyses had a correlation coefficient of 0.476. I also checked the percentage of texts where the methods agreed or disagreed by sentiment in sentiment agreement calculations, which is 82.23%. During an error analysis, where I found the mean squared error between the two methods, there is about a 0.583 difference, considering the range of the two sentiment scores is large. Due to these facts and the fact that Vader just gives more information about the dream contents we are trying to analyze, we will continue using Vader sentiment analysis for further modeling.

4.2 Broad Exploratory Analysis

To preface our analysis, we wished to take a look at some general trends in our data. We first took a look at some of the trends of our entire dataset and then analyzed a single individual. For both analyses in this section, we focused on the compound score, as we felt it provided the best assessment of overall sentiment.

We first took a mean of compound scores for the entire dataset, yielding a score of -0.012; on average, our dreams were neutral in sentiment.

The distribution of compound scores for the entire dataset follows as below:

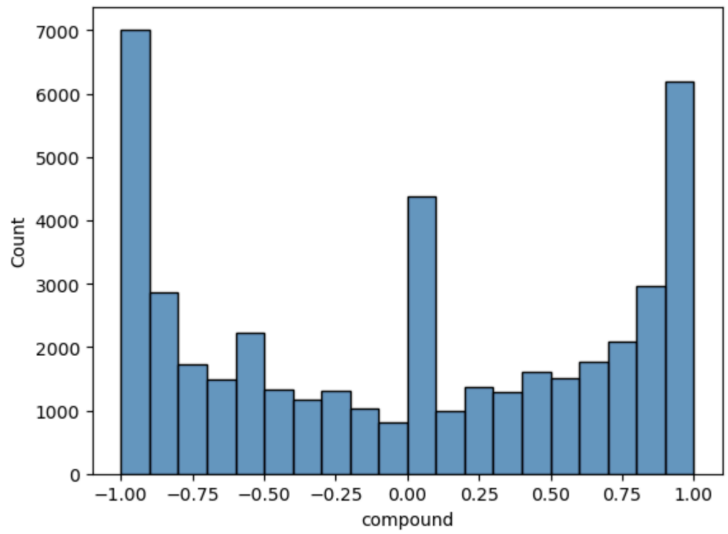

Figure 7

Our data is distributed trimodality, with peaks at -1, 0, and 1. This may be because our dream sentiments are distributed in this manner, or because VADER tends to categorize dreams as very positive, negative, and neutral.

Our data also contains individuals who recorded their dreams in languages besides English. Our sentiment analysis tended to score these dreams as negative. For instance, on average, a German individual's dreams were given a compound score of -0.7886. This may also explain why our data has such an apparent peak at -1.

For this analysis, we examined Barb Sanders, a middle-aged woman who composed 4452 dream entries between her two diaries (4 and 5).

	Unnamed: 0	raw_number	content	negative	neutral	positive	compound
0	0	#0000 (1960-05-03)	I had the neatest dream about Blake, me, Reta ...	0.000	0.680	0.320	0.5719
1	1	#0001 (1960-05-04)	I had another neat dream about Blake.	0.000	0.500	0.500	0.6124
2	2	#0002 (1960-07-16)	I had a dream that Nate came back and I felt j...	0.000	0.875	0.125	0.3164
3	3	#0003 (1960-08-04)	For the second night in a row, I dreamed of Jo...	0.044	0.853	0.104	0.8513
4	4	#0004 (1960-12-02)	I Didn't dream as last night before I woke up!...	0.079	0.730	0.191	0.5349

Figure 8

Sanders had a mean compound score of 0.252, making her dreams more positive than the average individuals in our dataset. Her compound score distribution is as follows:

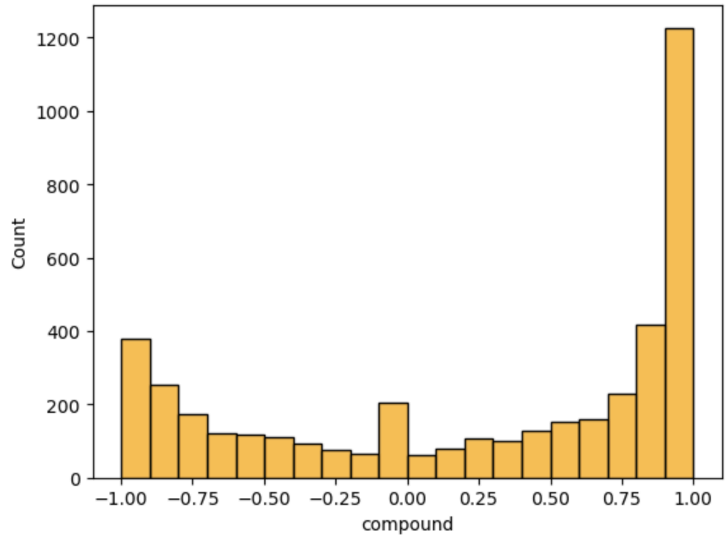

Figure 9

Sanders tended to have fewer neutral and negative dreams than our overall dataset and had many more very positive dreams than our overall dataset. Sanders' distribution is different from our overall distribution. Running a Kolmogorov-Smirnov test allows us to quantify this difference. We decided upon a significance level of .01.

H_0: Barb Sanders' distribution is the same as the overall distribution
H_A: Barb Sanders' distribution differs from the overall distribution
P value: $8.8 * 10^{-130}$

Our P value of $8.8 * 10^{-130}$ is much less than our significance level of .01. We reject our null hypothesis, Sanders' distribution seems to be different from the overall distribution of compound scores.

Sanders was selected for analysis partially due to how well her dream's dates were recorded. We were able to convert her raw_number variable into datetime objects with regex, and plot heatmaps for each month, as well as each day:

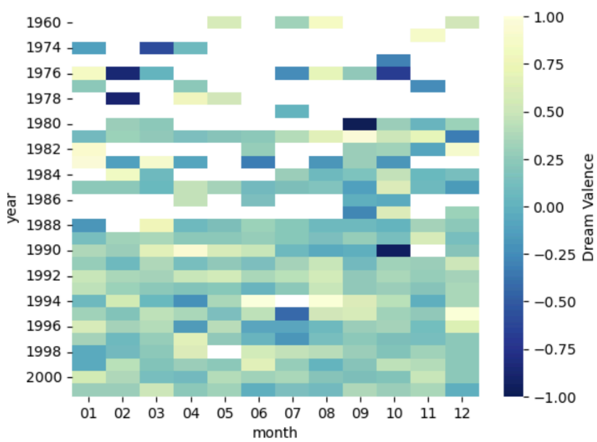

Figure 10

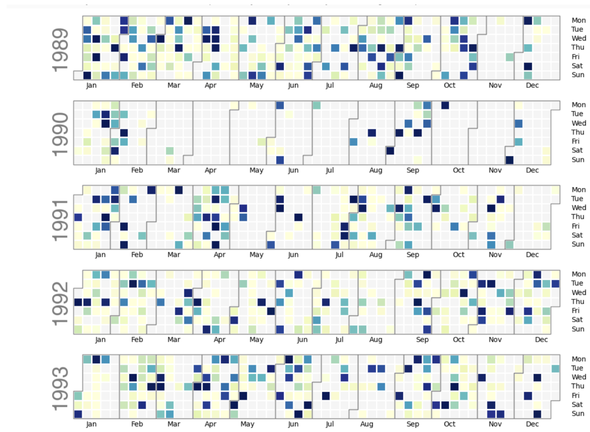

Figure 11

Just from a glance, there do not appear to be any visible trends in our heatmaps, especially for our daily heatmap. This is in trend for what we know about dreams; they can be erratic, and especially hard to predict for very fine-grained data.

4.2.1 Broad Exploratory Analysis Results

The data set ranges from -1 to +1 where there are trimodal peaks at the -1, 0, and +1 marks, with a majority of the distribution being on the negative side of the graph. This could be due to most dreams being described as extreme rather than a more neutral sentiment, or this could be because Vader tends to calculate compound scores as extreme. The mean of the compound scores of the dreams at -0.012. Looking more closely at Barb Sanders, her dreams tend to be more positive with a 0.252 compound score, and her dream collection dates were scattered from 1960-1987, then become more consistent in collecting from 1988-2001.

4.3 Narrow Exploratory Analysis

For this piece of exploratory analysis, we were particularly interested in the positive and negative aspects of the dreams the participants reported. More specifically, we were interested in the positive and negative scores that VADER yielded in Barb Sander's diaries. As the broad analysis pointed out, VADER tended to either group dreams as really positive, negative, or neutral. I had access to the sample means of each of the 94 participant's diary entries for the positive and negative scores produced by VADER. My objective was to analyze a middle-aged woman, Barb Sanders, and analyze whether or not her dreams were more positive or negative compared to the rest of the 94 participants. After constructing the necessary data frame to perform the desired analysis, we were particularly interested in the sample means of Barb Sander's positive and negative valence compared to each participant. Then, we explored the normality of these positive and negative sample means among each of our 94 participants. These positive and negative sample means across the 94 samples in our dataset turned out to be normally distributed, as is shown below in Figure 13.

The magnitude of Barb Sander's diary entries was fairly large with 4452 entries across diaries four and five. These diary entries were recorded from the years 1960 to the end of 2001. With such a large sample size of dreams in her diary, Barb Sanders was a more than suitable candidate for analysis. I began by examining the positive mean scores for Diary Four and Diary Five. Assuming that the data was normally distributed, I went ahead and performed a t-test, where my null hypothesis was the positive sample mean of Barb Sander's fourth diary. I performed a two-sided and left-sided t-test, as seen below, to test my null hypothesis. I did the same process for the fifth diary entry.

In this case, it turns out that the true population mean of our sample means for the positive scores from our model VADER are less than the sample mean for the fourth and fifth diary entries. This means that we can reject the null hypothesis, and it suggests that Barb Sander's dreams were likely more positive than the rest of the 93 other participants.

To run these one-sample t-tests for both the positive and negative samples, it is imperative to check that the sample means of each of our 94 participants are normally distributed. To check this, I used the seaborn and SciPy packages in Python to plot a scatter plot, Q-Q plot, and histogram overlaid with a normal distribution.

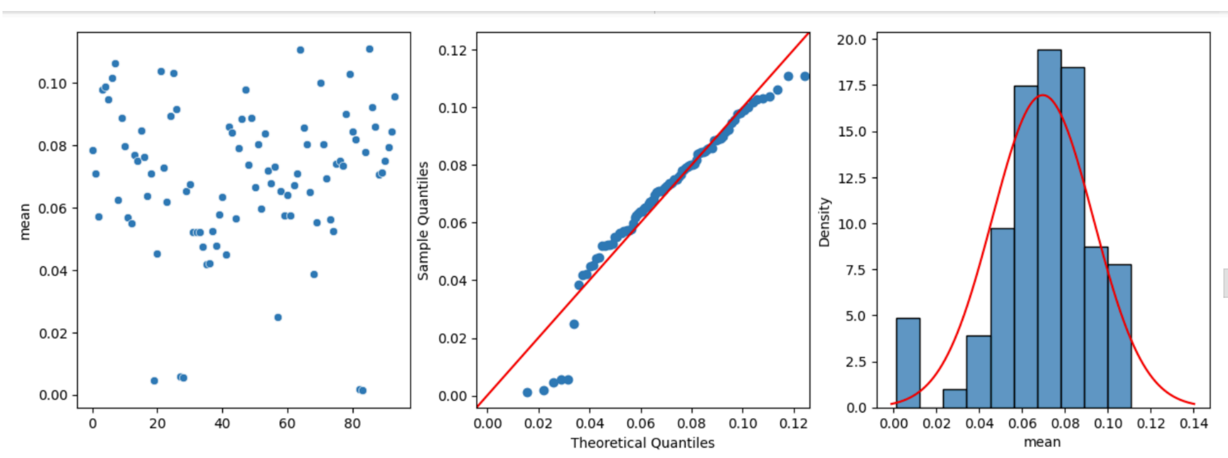

Figure 12

4.3.1 Narrow Exploratory Analysis Results

By the normality of the data and the extraordinarily low p-value associated with the one-sample t-tests, we can conclude that Barb Sander's dreams were generally more positive than the other 93 participants in our data with confidence level $100(1-\alpha) = 95\%$

4.4 Data Visualizations Time Plots

On the individual level, identifying patterns in valence scores was also a goal. There were a handful of interesting dreamers to analyze: a vivid and detailed dreamer, a Vietnam war veteran, a wife and her husband, and a college girl through high school and college. Visualizing the positive and negative valence scores through time offered insights into potential patterns created by individual characteristics.

The initial data cleaning involved transforming every diary's "raw_number" column, which was a column that contained the hand-typed, and usually inconsistent, dates for each diary entry recorded. They were mainly cleaned using regex and then converted to legible, usable dates. With these cleaned dates, time plots at both the individual and group levels for average sentiment scores were created.

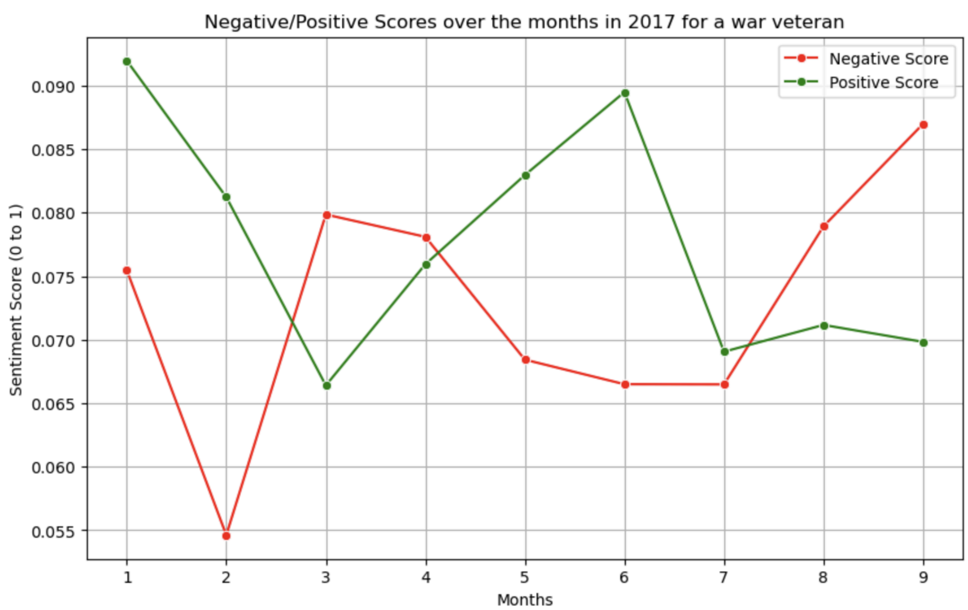

Figure 13

The figure above shows the average positive and negative sentiment scores by month during 2017 for a Vietnam war veteran. From the plot, we can see that early in the year, his dreams are more positive. However, from March to April, his dreams became more negative until he had another spike in positivity around June. Unfortunately though, after July, his dreams became increasingly negative until September, when his last diaries were recorded.

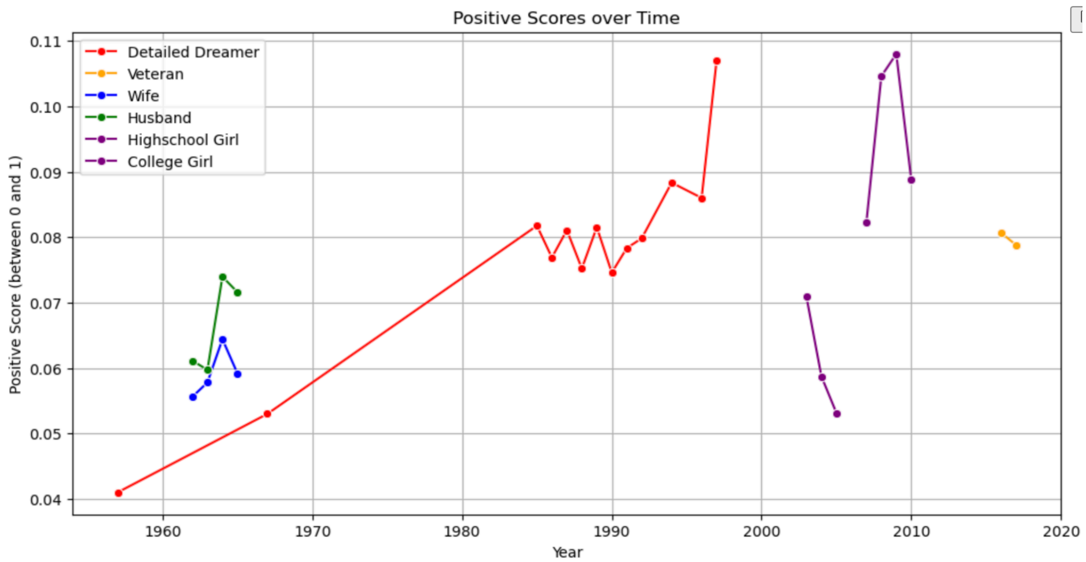

Figure 14

The relative positive sentiment scores for the select individuals can be seen in the figure above. Starting with Atla, the detailed dreamer, who had the most entries across the years. Her positive sentiment scores increased steadily until they fluctuated around 1985 and spiked up in positivity around 1995. The Vietnam War veteran, unfortunately, did not get as many diary entries recorded but has had a decline in positive sentiment scores from 2016 to 2017. For the wife and husband, they got data collected at the same time, but it seems like the husband has higher positive scores on average compared to the wife. The high school and college girl timeline follow the same girl, Bea. For Bea, she had a considerable decrease in positive scores throughout high school, however starting college, she started with a higher positive score, spiked in the middle, and then spiked back down at the end.

4.4.1 Data Visualizations Time Plots

Through looking into individual people's dreams, we can see that outside influences actively correlate to dreaming sentiment. This is clearly shown in the college girls' dream content being consistently negative, then when summer/college begins, their dreams increase exponentially. This further proves why studying dreams and their sentiment is important to everyone's livelihood.

5 Modeling and Prediction

Attempting to predict future states of dream valence is a lofty goal. But creating a model that is generalizable to all people is even more ambitious **[3]**. The reality of such a complex task became quickly apparent during the exploratory phases of this project. Ultimately, the scope was narrowed and focus was given to gaining some insight into valence patterns at the individual level. Furthermore, due to the many difficulties encountered with erroneous dates and sparse timelines among many dream sequences, some simplifications and assumptions were made. In particular, the date association was dropped and each row was assumed to be a daily entry in the right order. The justification for this decision is based on the observations of a few people with consistent daily recordings. With these records, the overall erratic pattern of valence scores was similar with, or without date associations.

Given the decisions above, the models investigated were the Autoregressive (AR), Moving Averages (MA), and the combined ARMA models. These are time series models that seek to predict future sequences of values based on past sequences of values. Of the three models, the AR model performed the best. It was able to capture general shifts in weekly valence averages among several individuals. For example, below is a 10-week prediction for Alta, a detailed dreamer with over 420 emotionally rich dreams.

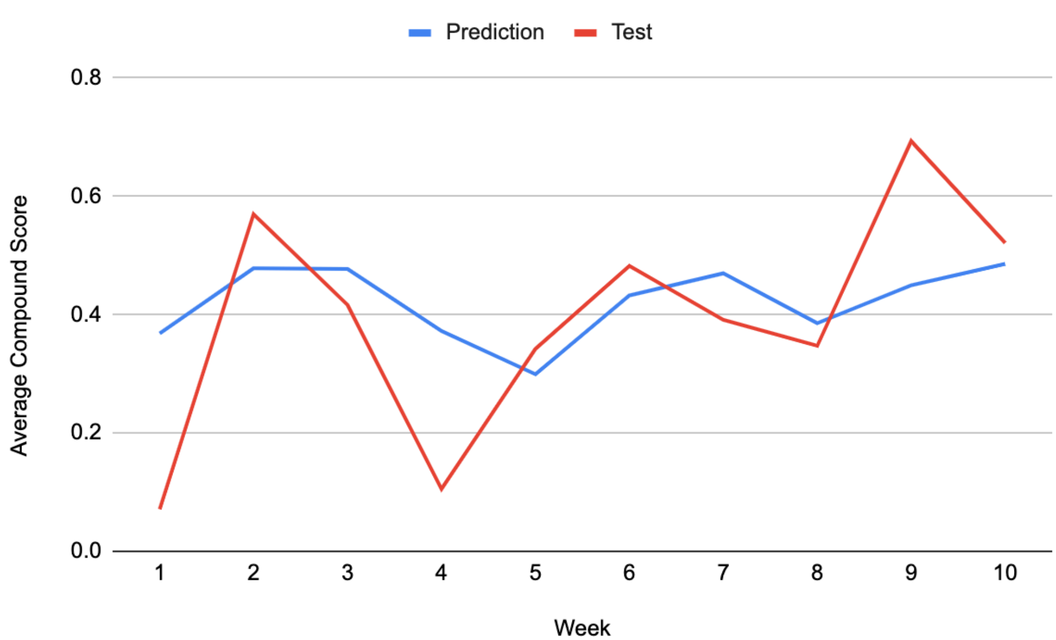

Figure 15

However, while the AR model performed the best receiving a Root Mean Square Error (RMSE) of 0.1563, it comes with the aforementioned assumptions, and arguably, a questionable preprocessing step of a weekly valence average—the compound score was a bimodal distribution. This muddies the interpretation and inference possible about the original data. Furthermore, none of the coefficient p-values during the AR fitting process were statistically significant ($p < 0.05$) based on the Ljung-Box test for autocorrelation. On top of this, the Akaike Information Criterion (AIC) resulted in 18.525 and the Bayesian Information Criterion (BIC) resulted in 33.185. This means that the model had a tough time fitting the data. Figure 16 illustrates this fact with the wide confidence intervals indicating a lack of precision within the fit.

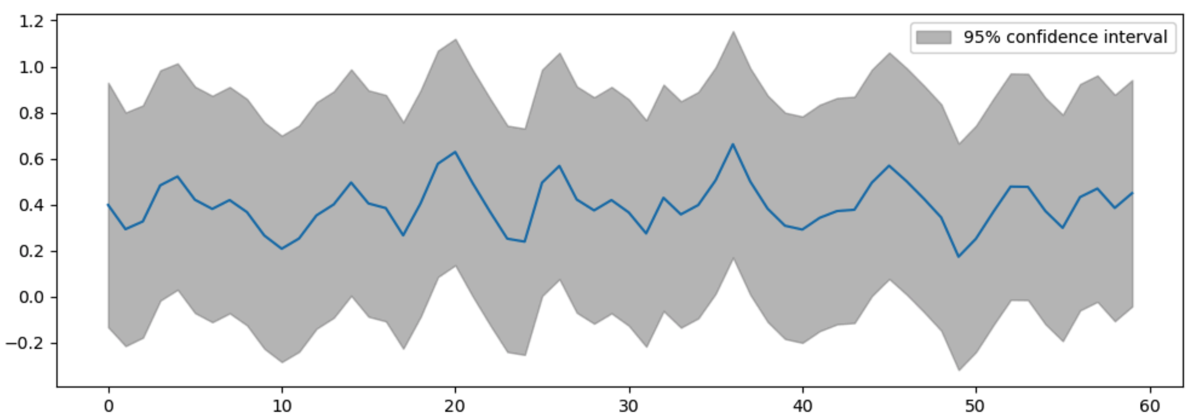

Figure 16

5.1 Modeling and Prediction Results

Ultimately, there was no real success in the prediction attempts. A wide selection of model parameters were used, but due to the close resemblance to white noise, it appears unlikely that dream valence could be predicted with either the AR, MA, or combined ARMA models. The erratic nature of dream valence and the lack of correlation over time, regardless of lag, made the task difficult for these models.

6 Conclusion and Future Work

In summary, we analyzed for this paper how a human's dream works in our brain and saw the emotional difference. We used 20,000 CSV files that contain individual diaries of dreams to do a sentimental analysis, analyzing them to have compound scores and future predictions. The libraries we used are Textblob and Vader. The libraries are NLTK modules for classifying positive, negative, and neutral words. We compared both libraries and selected Vader because the sentiment scores were easier to use for the analysis and more accurate than Textblob.

Regarding extracting and cleaning data, we had difficulties with consistency because there were format differences in their timelines. Avoiding the date columns decreased the compound scores' accuracy. Using spreadsheet applications or regular expressions might be better for modifying and correcting the format. Still, having complete reliability from all of the data is insecure because of the number of diaries and the fact that the exact date is not specified for many of them.

When it comes to data analysis, we have done the t-test from individual diary sample data to have an average dream workflow to compare with a particular person and implemented a goodness-fit test to see the compound values and a Shapiro-Wilks test to see whether the data is normally distributed. Using QQ-plot is for data visualization. Time-plot is for looking at different people's sentiment scores over time. We observed standout results in terms of how overall compound scores were distributed. However, it turns out that even though Vader is powerful in evaluating the overall sentiment, there are some limitations to some domains' analysis of the data. It is often used for social media's sentimental analysis, and it might not be the best library to analyze our diary data and see dreams' compound scores accurately.

About machine learning, we used the ARMA model to predict the future dream workflow based on the intensity of compound scores that show positive and negative feelings from individual dreams. Unfortunately, figuring out the specific pattern to predict the future dream workflow was unsuccessful. That is not only because of the lack of consistent data sources but also because the dream will be affected by many other complex factors, such as psychological conditions. To get rid of the ambiguity in defining the compound scores, having the domain knowledge, in this case, knowledge for sentiment analysis, will be helpful to understand the concept and rules behind the data and enrich the performance of the analysis results by capturing the small nuances instead of relying on only the machine learning model's preciseness.

Our concerns throughout this paper were cleaning the dataset, figuring out the best library for the dream's sentimental analysis, and having more domain knowledge. By Improving what we learned, we would be able to have more accurate results and use the skills for future analysis.

References

[1] DreamBank by Adam Schneider and G. William Domhoff at the Psychology Department, UC Santa Cruz

[2] VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text by C.J. Hutto and Eric Gilbert at the Georgia Institute of Technology

[3] Experimental Research on Dreaming: State of the Art and Neuropsychoanalytic Perspectives by Perrine M. Ruby at the Neuroscience Research Center, Brain Dynamics and Cognition Team, Lyon, France

[4] The Science Behind Dreaming by Sander Van Der Linden at Scientific American

[5] Why Your Brain Needs to Dream by Matthew Walker at Berkeley

[6] Nightmares by Eric Suni and Dr. Abhinav Singh at Sleep Foundation