Sentiment analysis of tweets about the crisis in Syria and Turkey

SADI Lidia, DJAODOH Ossara Elysée, DIALLO Al Hassane, AMOUWOTOR Comlan
Université Paris Descartes, 45 rue des Saints-Pères, 75006 Paris, France

Abstract:

In this report we will present two methods of sentiment analysis: TextBlob and VADER. First we will preprocess the data which are the tweets about the crisis in Syria and Turkey to facilitate the data analysis. Then we will store it in a NoSQL database and after that we will answer several questions to analyze the opinion of people towards this crisis and make the comparison between these two methods. And finally we will conclude our report with the interest of sentiment analysis and a general idea about the effect of the used methods.

1. Introduction:

Sentiment analysis is a data processing method that determines people's opinions, emotions and attitudes towards a product, brand, company or event by analyzing textual data such as comments, reviews, social media posts, news articles, etc. With the advent of Big Data technologies, sentiment analysis has become an easier, faster and more accurate task due to the ability to process large amounts of data in real time. In this project, we will use Big Data technologies to analyze people's opinions related to the crisis in Syria and Turkey.

2. State of the art: Summary of the articles read

"*Sentiment analysis using Twitter data: a comparative application of lexicon and machine-learning-based approach*" is a research article authored by Yuxing Qi and Zahratu Shabrina.
There are different approaches to sentiment analysis, namely a lexicon-based approach, machine learning-based approach, Classification models. This paper focuses on the tweets that were geotagged from the main cities of the UK during the third Covid-19 national lockdown. Major cities in the UK during the third national shutdown of Covid-19.
"*Sentiment Analysis of Twitter Data: A Survey of Techniques*" is a research article authored by Department of Computer Engg, Pune Institute of Computer Technology,Pune University of Pune (India) S.S. Sonawane"
Sentiment analysis is a widely adopted technique for analyzing and classifying people's opinions, attitudes, and emotions towards various elements.Different models and techniques, such as Naive Bayes and Support Vector Machines, Random Forest…have been used for sentiment analysis.
Pre-processing of datasets is crucial for sentiment analysis, and subtasks like subjectivity classification and sentence-level sentiment analysis are also important.
"*Twindex Fuorisalone: Social listening of Milano during Fuorisalone 2013*" is a research article authored by Marco Balduini, Emanuele Della Valle, and Daniele Dell'Aglio, Mikalai Tsytsarau, Themis Palpanas, and Cristian Confalonieri.
In the paper "Twindex Fuorisalone" the authors discuss a tool they developed to monitor social media activity related to the Milano Design Week's Fuorisalone event.The tool, which uses "RDF" as a data model and C-SPARQ for analysis, allows for real-time monitoring of social media activity, including sentiment analysis.

"*Quality of Sentiment Analysis Tools: The Reasons of Inconsistency*" is a research article authored by Wissam Mammar Kouadri, Mourad Ouziri, Salima Benbernou, Karima Echihabi, Themis Palpanas, and Iheb Ben Amor. The article addresses the challenge of inconsistency in sentiment analysis tools, where different tools can produce different sentiment labels for the same text. The authors present a study on data quality for sentiment analysis tools, highlighting the inconsistency problem and recommending different tools for different data types. They suggest CNN for short text, lexicon-based methods for explainability, and ensemble of lexicon-based methods for streaming data with long text. For social media data, they recommend character and word embedding, and for news and factual data, they suggest BERT embedding. Their insights have implications for use cases such as crowd-sourced fact checking, and they suggest research opportunities for the data management community to improve scalability properties. No previous research work has explored the problem of resolving both intra-tool and inter-tool inconsistencies.

"*Managing Diverse Sentiments at Large Scale*" is an article written by Mikalai Tsytsarau and Themis Palpanas that discusses the challenges associated with sentiment analysis and proposes a novel approach to address these challenges. The authors argue that traditional sentiment analysis approaches often overlook the diversity of opinions and perspectives that exist within a given text. To address this issue, they propose a new framework for sentiment analysis called "diversity-aware sentiment analysis" (DASA), which aims to capture and manage diverse sentiments at scale.

The article provides a detailed overview of the DASA framework, including its underlying theoretical foundations, key components, and practical applications. The authors also present the results of several case studies that demonstrate the effectiveness of the DASA approach in managing diverse sentiments in large-scale datasets. The case studies cover a range of domains, including social media, news articles, and product reviews.

The authors conclude the article by discussing the implications of the DASA approach for future research and practical applications of sentiment analysis. They argue that the DASA framework has the potential to improve the accuracy and reliability of sentiment analysis by capturing the full range of diverse sentiments and opinions that exist within a given text. The article provides a valuable contribution to the field of sentiment analysis and offers a promising direction for future research in this area.

"*The 2016 US Presidential Election on Facebook: An Exploratory Analysis of Sentiments*" is a research article published in the Proceedings of the 51st Hawaii International Conference on System Sciences in 2018 by Kandala et al. The article aims to analyze the sentiment expressed in Facebook posts during the 2016 US presidential election. The authors collected a sample of Facebook posts related to the election and used machine learning techniques to classify them into positive, negative, or neutral sentiments. The study presents the results of the sentiment analysis, including the overall sentiment patterns and variations across geographic regions, demographic groups, and election-related events. The article also discusses the implications of the findings for understanding public opinion and designing effective political strategies. The article follows a typical research article structure with sections such as introduction, literature review, data collection and preparation, sentiment analysis methodology, results and analysis, discussion, and conclusion.

« Using Social Media & Sentiment Analysis to Make Investment Decisions »

Ben Hasselgren Christos Chrysoulas, Nikolaos Pitropaki Sand William J. Buchanan

This article discusses the use of sentiment data obtained from social media (SM) to make investment decisions. The article highlights the challenge of obtaining relevant sentiment data from SM, building a system to measure the sentiment, and visualizing it to help users make investment decisions. The article also discusses the relationship between public sentiment and stock market performance and the use of sentiment analysis (SA) to identify the emotion behind text.The article presents a technical review of existing approaches for creating Sentiment Analysis models, extracting Social Media data, and evaluating results. The article proposes a novel prototype that utilizes public SM sentiment to help users make investment decisions. The prototype tracks the stock market performance of selected S&P 500 stocks, using the sentiment data obtained from SM. Additionally, the article presents a novel approach to factor SM metrics into the sentiment score effectively, which aims to measure the collective sentiment of the data accurately. The article concludes by discussing the findings and insights obtained from the prototype's evaluation.

“Beyond Sentiment Analysis: A Review of Recent Trends in Text Based Sentiment Analysis and Emotion Detection”

Lai Po Hung and Suraya Alias

This article presents a thorough overview of current developments in text-based sentiment analysis and emotion recognition. It highlights the emergence of emotion recognition as the new frontier in text classification, as emotions offer a valuable source of information in addition to author sentiments. The review examines recent works in the field, analyzing the techniques and models utilized, and explores the transition from early keyword-based approaches to more sophisticated machine learning and deep learning algorithms. The paper concludes that emotion detection has numerous potential applications in areas such as marketing, advertising, social behavioral analysis, public sentiment analysis, and human-computer interaction. It discusses the characteristics of social media text that make it different from formal text, such as non-standard word forms, fragments of sentences, and site-related markups. It also reviews recent work in sentiment analysis and emotion detection, focusing on trends in these fields over the past five years. The section on sentiment analysis notes that sentiment analysis is no longer a new task, but rather a popular tool for summarizing public sentiments, particularly in the context of big data. The section on emotion detection shows that it is gaining importance, as more recent work combines sentiment analysis and emotion detection for opinion analysis. The review concludes with a summary of the techniques used in sentiment analysis and emotion detection, and the trends in classifying techniques observed and studied. It highlights some of the challenges of sentiment analysis, such as the need to process large volumes of data in real time, and the need to detect fake reviews.

“**Improving Sentiment Analysis in social media by Handling Lengthened Words”**

ASHIMA KUKKAR 1 , RAJNI MOHANA2 , AMAN SHARMA2 , ANAND NAYYAR AND MOHD. ASIF SHAH

This research article discusses the use of sentiment analysis in social media for Emotional Recognition (ER).The authors suggest a lexicon-driven system that considers elongated words as they are, instead of omitting or normalizing them. The system uses framed lexicon rules to calculate the aggregated intensified senti-scores of lengthened words, which are then used to determine the level of sentiment of the person. The dataset utilized for evaluation consists of informal conversations between various friend groups through Facebook, Tweets, and personal chat.The proposed system outperforms traditional systems that ignore lengthened words, achieving F-measure rates of 81% to 96% for all datasets. The authors highlight the importance of social media in sharing information, ideas, and expressions through virtual networks, which can be used for citizen opinion polling, business intelligence, social contexts, and Internet of Things (IOT)-mood triggered devices.

It provides an overview of various machine learning algorithms and neural network approaches used in sentiment analysis. The article mentions that researchers have combined different methods, such as SVM and RF algorithm, to improve sentiment categorization performance. The article also mentions several common classifiers utilized in sentiment analysis, including EL, NB, ME, and SVM. The article divides neural network approaches into several types, such as ANN, CNN, RNN, GRU, LSTM, and hybrid neural network models, and summarizes recent research in each area. For example, the article mentions a study that combined ML and information retrieval approaches to improve sentiment categorization performance and another study that developed a CNN model using pre-trained word embeddings to categorize texts at the sentence level. The article concludes that while CNN was fast, new characteristics must be introduced to attain improved outcomes.

In fact, reading these papers for our work on analyzing the sentiments expressed in user comments on Twitter has allowed us to learn that Sentiment Analysis is an important field of Artificial Intelligence that uses Natural Language Processing techniques to determine the emotional tone of a text. It is used in many fields such as social media, market research, customer satisfaction surveys, etc. We also learned about the different approaches used in sentiment analysis. The most common approaches are lexicon-based analysis, rule-based analysis, and machine learning. Each of these approaches has its own advantages and limitations, and their effectiveness will depend on the nature of the text being analyzed and the objectives of the analysis.

Regarding the application areas of sentiment analysis, it can be used in many sectors such as marketing, social media, customer service, politics, market research, finance, etc. For example, in the marketing field, sentiment analysis is used to understand consumers' opinions about a company's products and services, while in politics, it is used to understand public opinion on government policies. It is also used in social networks. Sentiment analysis is particularly useful in the field of social media as it allows understanding of users' opinions, emotions, and trends on different social platforms. Companies can use sentiment analysis to monitor online conversations about their brands, products, or services and to understand people's perception of their company.

Sentiment analysis is also used to monitor social media discussions to detect emerging trends, consumer preferences, and changes in user behavior. This can help companies identify opportunities and threats, adjust their marketing strategies accordingly, and improve engagement with their target audience. Additionally, sentiment analysis is often used to measure customer satisfaction and to quickly respond to negative feedback or complaints. Companies can use this information to improve their products, services, or processes to increase customer satisfaction and strengthen their brand image. Overall, sentiment analysis is a powerful tool in the field of networks.

3. Data cleaning

Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. If data is incorrect, outcomes and algorithms are unreliable, even though they may look correct. There is no one absolute way to prescribe the exact steps in the data cleaning process because the processes will vary from dataset to dataset. But it is crucial to establish a template for your data cleaning process so you know you are doing it the right way every time[3]
The first step is to remove unwanted observations from our dataset, including duplicate or irrelevant observations, we have used the "remove_stopwords" function to remove stopwords and "clean_text" to clean our text like removing URL, separate sentences into tokens…Then fix structural errors like types.Filter unwanted outliers,handle missing data.
our data after preprocessing looks like the image below :

	date	content	hashtag	like_count	rt_count	followers_count	isverifi	languag	coordin	place	sourc
0	2023-02-13 04:07:37+00:00	desen oper busi bandogmeat dogmeatrad usa ...	['bandogmeat', 'dogmeattrade', 'USA', 'China',...	0.0	0.0	103.0	0.0	en	None		Twitter for iPhone
1	2023-02-13 04:07:29+00:00	turkeysyriaearthquak ventil , blanket , medic...	['TurkeySyriaEarthquake', '1TVideo', 'Turkey',...	0.0	1.0	6248880.0	1.0	en	None		Twitter Media Studio
2	2023-02-13 04:07:05+00:00	45 magnitud earthquak occur east central pacif...	['earthquake']	0.0	0.0	252.0	0.0	en	None		EQ FEED
3	2023-02-13 04:07:04+00:00	erdogan terrorist turkeyquak turkiy turkey	['Erdogan', 'terrorist', 'TurkeyQuake', 'Turki...	0.0	0.0	5276.0	0.0	und	None		Twitter for iPhone
4	2023-02-13 04:07:03+00:00	45 magnitud earthquak east central pacif ocean	['earthquake']	0.0	1.0	226198.0	0.0	ro	Coordinates(longitude=-110.604, latitude=16.1)		EarthquakeTrack.com
...	...	...	...	...	...	...	...	...	...	...	...
464720	2023-02-06 00:04:00+00:00	author birol bahadir busi day german turkish x...	['immigrant', 'germany', 'turkey', 'internatio...	0.0	0.0	1362.0	0.0	en	None		eClincher
464721	2023-02-06 00:03:53+00:00	earthquak terremoto m25 strike 125 km w puert...	['Earthquake', 'terremoto', 'Spain']	4.0	2.0	44544.0	0.0	es	None		emsc-csem
464722	2023-02-06 00:01:25+00:00	earthquak sismo m28 strike 72 km calama chi...	['Earthquake', 'sismo', 'Calama', 'Chile']	3.0	2.0	44544.0	0.0	en	None		emsc-csem
464723	2023-02-06 00:01:03+00:00	istanbul york discov ottoman part ottomanem...	['ottomanempire', 'ottoman', 'osman', 'osmanlıe...	0.0	0.0	66.0	0.0	en	None		Twitter for Android
464724	2023-02-06 00:00:00+00:00	the arrest case document 25 young male , 15 yo...	['Arabs', 'Kurds', 'Yazidis', 'Syria']	2.0	1.0	3891.0	0.0	en	None		Twitter Web App

464725 rows × 11 columns

Figure 1 : dataset

4. Sentiment Analysis of Customer Reviews using TextBlob and VADER methods mongodb database

4.1 Sentiment analysis based on TextBlob:

TextBlob is an NLP module on Python used for sentiment analysis. The function of TextBlob that we are interested in allows for a given text to determine the tone of the text and the sentiment of the person who wrote it.[1]It can handle tasks such as language detection, sentence and word extraction, and sentiment analysis.
TextBlob uses a rule-based classification method to perform sentiment analysis. It assigns a polarity (positive, negative, or neutral) to each sentence or word in a text based on pre-trained word dictionaries and sentence templates.

4.2 Sentiment analysis based on VADER:

VADER (Valence Aware Dictionary for Sentiment Reasoning) is a model used for textual sentiment analysis that is sensitive to both the polarity (positive/negative) and the intensity (strength) of the emotion, the latter of which can make a difference with textBlob. It is available in the NLTK package and can be applied directly to unlabeled text data. VADER sentimental analysis is based on a dictionary that associates lexical features with emotional intensities called sentiment scores. The sentiment score of a text can be obtained by summing the intensity of each word in the text.

The VADER library is used to perform a sentiment analysis on a given text.[2] it also takes into account punctuation, capitalization and emoticons to improve the accuracy of the analysis.

5. Data Management:

After analysis we decided to use a NoSQL database despite the advantages of SQL. In our case NoSQL has more advantages than SQL because :
NoSQL is designed to run efficiently on distributed systems, which makes it easy to add new servers to handle a load (Horizontal Scalability). This makes it easier to scale your NoSQL database horizontally by simply adding more nodes.
Unlike relational databases that require a fixed, predefined schema, NoSQL allows you to store unstructured or semi-structured data (Schema Flexibility). This means that you can add, modify or delete fields without having to make costly changes to your entire schema. In our case this will be very useful because our data is raw and mostly unstructured.
NoSQL databases are optimized for specific operations, such as fast insertion or reading of massive data (High Performance). They can therefore offer higher performance than relational databases for certain workloads.
In short, since we have unstructured data, which contain empty fields and especially some columns of our data set contain arrays of objects, a type of variable that supports very well the NoSQL.

```
schema({
    "date": Date,
    "content": String,
    "hashtags": Array,
    "like_count": Number,
    "rt_count": Number,
    "followers_count": Number,
    "isVerified": Boolean,
    "language": String,
    "coordinates": Array,
    "place": String,
    "source": String
})
```

Figure 2: database schema

We have tried to represent the schema of our NoSQL table. We used MongoDB in our implementation. NoSQL offers us the advantage of having fields with values of different types in the same database. This flexibility of NoSQL allows us to store all our data in the database before cleaning it. MongoDB takes care of managing primary keys for each data that will be saved. It also allows us to store array-type data in our database without the need to create additional tables or establish relationships between our different tables, which would have increased the cost of each query to the database. Therefore, in order to achieve flexibility and minimize the cost of our operations, we have chosen NoSQL.

6. Analysis of Results:

After using two Python libraries for sentiment analysis, TextBlob and VADER, it is clear that the latter is more efficient. Indeed, although TextBlob is fast in terms of execution time, it tends to produce less accurate results than those obtained with VADER. On the one hand, TextBlob is fast in terms of execution time because it uses a rule-based classification method. However, this method cannot be as accurate as the one used by VADER, which takes more time to execute but gives more accurate results by also evaluating the intensity of word and phrase polarity. Moreover, VADER takes into account elements such as punctuation, capitalization, and emoticons to improve the accuracy of the analysis, which is not always the case with TextBlob. Ultimately, the choice between the two libraries depends on the specific needs of each project. If speed is important and results do not need to be absolutely accurate, TextBlob may be a good choice. If accuracy is paramount and execution time is less important, VADER may be preferable. We conducted sentiment analysis on different subjects to evaluate people's opinions in different regions of the world. We conducted a first analysis on the crisis in Syria and Turkey, collecting tweet data on Kaggle. We used the Python libraries VADER and TextBlob to assign polarity scores to each sentence or word in the collected data. The results obtained were used to evaluate public opinion on the crisis in Syria and Turkey, in different regions of the world. In the following sections of this report, we will present the results obtained from the two methods we used, namely TextBlob and VADER. These results will be presented in the form of geographic maps to facilitate their visualization and understanding.

Firstly, Results with TextBlob:

We applied TextBlob to our dataset to perform sentiment analysis and obtained results for each tweet in terms of its polarity (positive, negative, or neutral). We then aggregated these results by region to visualize the distribution of sentiments across different parts of the world.

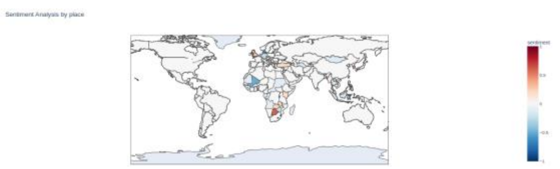

Figure 3 Sentiments analysis by place with TextBlob

The maps generated using TextBlob show the overall sentiment for each region, with areas in red indicating positive sentiment, areas in blue indicating negative sentiment, and areas in white indicating neutral sentiment. While the maps give a general sense of the sentiment in each region, they also demonstrate the limitations of TextBlob's accuracy, with some regions showing mixed sentiment or unclear results.

Secondly, Results with VADER:

We applied VADER to our dataset to perform sentiment analysis and obtained more nuanced results compared to TextBlob. VADER takes into account not only the polarity of words and phrases but also their intensity, resulting in more accurate and granular sentiment analysis.

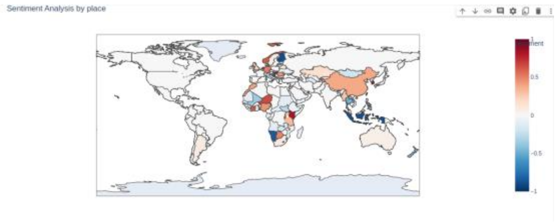

Figure 4 Sentiments analysis by place with VADER

The maps generated using VADER show a more detailed view of sentiment distribution, with a wider range of sentiment intensities represented. The areas of red, blue and white on the maps indicate positive, negative, and neutral sentiment respectively, but the shading and intensity of colors provide more nuanced information on the degree of sentiment. In comparison to the results obtained with TextBlob, the sentiment distribution shown in the VADER maps is more consistent and less mixed, indicating its superiority in terms of accuracy and precision. Overall, the results obtained with VADER demonstrate its effectiveness in providing more detailed and nuanced sentiment analysis, and highlight the need to consider it as a viable option for sentiment analysis projects.We also deduce that by changing place the color changes which means that the opinion changes too so the opinion depends on the places where people are located.

Visualizing Execution Time: TextBlob vs VADER

In addition to accuracy and precision, another important factor to consider when selecting a sentiment analysis tool is execution time. To illustrate the difference in execution time between TextBlob and VADER, we conducted a performance test using a sample dataset.To do this, we used the time module in Python to measure the amount of time it took for each analysis to run.

The results show that TextBlob is significantly faster than VADER.To visually represent the difference in execution time, we've included two images below, one for TextBlob and one for VADER. These images show the execution time for each tweet in the sample dataset, with TextBlob's image displaying a much more compressed time frame than VADER's.

```
Day
Le temps d'execution de la méthode TextBlob est:  298.10421323776245
```

Figure5

```
Le temps d'execution de la méthode VADER est:  3978.107656955719
```

Figure6

7. Difference between the two methods (Examples):

In addition to evaluating the overall performance of TextBlob and VADER, we also wanted to examine situations where the two methods produced different results. To do this, we compared the polarity scores assigned to individual phrases by each method for a set of sample data. We found that in some cases, the difference in polarity scores assigned by TextBlob and VADER was quite significant. For example, in one instance, TextBlob assigned a polarity score of 0.5 to a particular phrase, while VADER assigned a score of -0.3612. This represents a large discrepancy in the sentiment analysis of that phrase, and could have significant implications for interpreting the overall sentiment of a given text. To understand what might account for these discrepancies, we calculated the difference in polarity scores between the two methods for each phrase shown in the column "diff_polarity", and then took the average of these differences across all samples

The average difference between the polarity scores obtained using TextBlob and VADER was calculated to be 0.017. This suggests that the two methods produced somewhat similar results in terms of polarity analysis. However, it is important to note that there were some cases where the methods produced significantly different results. These cases were further examined to understand why one method worked better than the other. Overall, the difference in polarity scores between TextBlob and VADER can be attributed to the fact that they use different algorithms and lexicons for sentiment analysis. While TextBlob uses a rule-based approach, VADER employs a lexicon and rule-based approach. This could explain why VADER is better able to detect sentiment intensity and take into account nuances in language, such as sarcasm and negation. On the other hand, TextBlob may be faster in processing and may work better for simpler texts with less nuanced language.

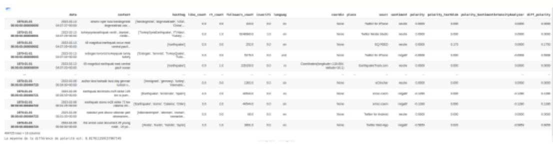

Figure 7

```
La moyenne de la différence de polarité est: 0.017811359157967145
```

Figure 8

polarity_textblob	polarity_SentimentIntensityAnalyzer	diff_polarity
0.000	0.0000	0.0000
0.000	0.0000	0.0000
0.175	0.0000	0.1750
0.000	-0.6908	0.6908
0.000	0.0000	0.0000
...	...	...
0.000	0.0000	0.0000
0.000	-0.1280	0.1280
0.000	-0.1280	0.1280
0.000	0.0000	0.0000
0.020	-0.5859	0.6059

Figure 9

We conducted sentiment analysis on a dataset of countries and identified the number of places with positive and negative sentiments using TextBlob and VADER. While both methods identified a mix of positive and negative sentiments, VADER detected more places compared to TextBlob. It is worth noting that the specific numbers may vary depending on the dataset and time period analyzed. These findings suggest the importance of choosing an appropriate sentiment analysis method depending on the specific goals and characteristics of the dataset.

These findings suggest that VADER may be more accurate in identifying countries with positive and negative sentiment, as well as detecting changes in sentiment over time. However, further analysis is needed to confirm these results.

```
451 places ont une opinion négative.
860 places ont une opinion positive.
2071 places ont une opinion neutre.
1007 places ont changé d'opinion.
394 places ont changé d'opinion de négatif à positif.
614 places ont changé d'opinion de positif à négatif.
                                                  place  sentiment
2     Place(fullName="Barberino Val d'Elsa, Toscana"...    0.00000
13    Place(fullName="Pesantren AlQur'an Nurul Falah...    0.00000
17    Place(fullName="St Patrick's Cathedral | Ardea...    0.00000
26    Place(fullName='6Th Road Station', name='6Th R...    0.00000
31    Place(fullName='Aba, Nigeria', name='Aba', typ...    0.40000
...                                                 ...        ...
3277  Place(fullName='Çarşamba, Samsun', name='Çarşa...    0.00000
3289  Place(fullName='Çelikhan, Adıyaman', name='Çel...    0.09446
3312  Place(fullName='Ören Plajı', name='Ören Plajı'...    0.00000
3321  Place(fullName='Ürgüp, Türkiye', name='Ürgüp',...    0.00000
3324  Place(fullName='İYİ Parti Genel Merkezi', name...    0.00000

[394 rows x 2 columns]
```

Figure 10 Result with TextBlob

```
783 places ont une opinion négative.
902 places ont une opinion positive.
1697 places ont une opinion neutre.
1216 places ont changé d'opinion.
585 places ont changé d'opinion de négatif à positif.
631 places ont changé d'opinion de négatif à positif.
                                                  place  sentiment
1     Place(fullName="'s-Hertogenbosch, Nederland", ...   0.244900
4     Place(fullName="Castel dell'Ovo", name="Castel...   0.000000
11    Place(fullName="L'Aquila, Abruzzo", name="L'Aq...   0.000000
13    Place(fullName="Pesantren AlQur'an Nurul Falah...   0.000000
17    Place(fullName="St Patrick's Cathedral | Ardea...   0.670500
...                                                 ...        ...
3338  Place(fullName='İskilip, Çorum', name='İskilip...   0.000000
3341  Place(fullName='İslahiye, Türkiye', name='İsla...   0.000000
3345  Place(fullName='İstanbul, Türkiye', name='İsta...   0.006917
3354  Place(fullName='Şanlıurfa Merkez, Şanlıurfa', ...   0.000000
3360  Place(fullName='Şereflikoçhisar, Türkiye', nam...   0.000000

[585 rows x 2 columns]
```

Figure 11 Result with VADER

Evolution of the average polarity of tweets for each location over time

Figure 12: Results with TextBlob on February 06 at 5am

Figure 13: Results with TextBlob on February 06 at 8am

Figure 14: Results with VADER on February 06 at 5am

Figure15: Results with VADER on February 06 at 8am

We analyzed the most common topics discussed in tweets for each place. To do this, we grouped tweets by place and extracted the top 10 most common words for each group using a function that tokenized the tweets and removed stop words. Some places had different topics that were more prevalent in their tweets.

For example, in places that experienced an earthquake, the word 'earthquake' was among the top 10 most common words as shown below.

Figure 16 : Extracted Words

Conclusion:

In conclusion, in this report we made a study of sentiment analysis which allowed us to have an idea about what people think about the crisis in Syria and Turkey using the python libraries TextBlob and VADER. Sentiment analysis is a very valuable tool that allows companies to increase their profits based on the feelings of customers. It is not 100% accurate but it gives an important percentage that helps in life to improve many things. The methods we have used have shown that sentiment analysis is based on the type of algorithm and the quality of the data used.

Références

[1]https://larevueia.fr/nlp-avec-python-analyse-de-sentiments-sur-twitter/#:~:text=TextBlob%20est%20un%20module%20NLP,la%20polarit%C3%A9%20de%20ce%20tweet.

[2]https://towardsdatascience.com/sentimental-analysis-using-vader-a3415fef7664

[3]https://www.tableau.com/learn/articles/what-is-data-cleaning#definition