Spring 2019

Newsworthy Migrants: Sentiment and Text Analysis of Dutch Newspapers

Nicholas J. Schmitz

SIT

UNIVERSITY OF NOTRE DAME

NEWSWORTHY MIGRANTS

Sentiment & Text Analysis of Dutch Newspapers

Nicholas J Schmitz

Published: 2019-04-30

Table of Contents

Introduction:

Human migration is the movement by people from one place to another with the intentions of settling, permanently or temporarily in a new location. For the entire existence of the human race, peoples have moved across the world in search of new opportunities, escaping persecution and many other reasons. These movements of people have often had significant economic, social and political impacts in both the place of departure and destination. Over time, as political borders became more strictly defined and governments started to keep track of the people residing in the country, the desire to be able to estimate migratory populations developed. Currently, in the 21st century, where the speed of human mobility is unprecedented and citizenship/resident rights come with many economic, social and political liberties and benefits, the ability to track migrant peoples has become an increasingly important task. However, given the numbers and fluid nature of migrants, estimating this population, as well as understanding how they are viewed in society, has proved to be a difficult project.

Background:

In order to do any quantitative analysis in the migration context, it is important to understand the types of data that are available to be used in research. Looking at the three different types of data, Statistical, Administrative and Innovative, we can see each of the advantages, deficiency's, and type of information that is able to be gather from each source.

Statistical Data:

Statistical data is comprised of national population and housing surveys. Data from these sources are cleaned, imputed, and aggregated to produce official figures that are then published to the public. Three of the most common types of sources for the statistical data are Censuses, Household Surveys and Labor Force Surveys.

Censuses are large data collection procedures that can provide information about migrant stock, migration flows, socio-economic characteristics of migrant populations and emigration figures. Migrants stock is defined by the UN as an estimate of the *"total number of*

international migrants present in a given country at a particular point in time". Migrant flows, as defined by the UN, is the "*number of migrants entering and leaving (inflow and outflow) a country over the course of a specific period, such as one year*" (UN Stats 2010). Because of the universal nature of a census, they often provide detailed information on groups in a country and show trends which can be easily compared across countries. The universality of a census also comes with a large financial and time cost which results in them being done infrequently. This limits researchers to use this data for real time migratory trends and does not provide any information as to why or how migrants have come into a country. As a result, census data is very rarely used in migration research and is reserved for large time frame analysis.

Surveys, both Household and Labor, are another data collection process for obtain statistical data. The information that can be gained from surveys include migrant stocks, socio-economic characteristics, drivers and impact of migration and internal migration. One of the strengths of surveys is the ability to ask specific questions to gain insight on why and how a migrant has chosen to immigrate. As surveys are smaller scale, a researcher can more precisely target a specific group and in addition, the data can be collected in a much timelier fashion than a census. However, ensuring a proper, representative sample can prove to be a large issue in surveys. In sensitive issues such as migration, answers may not be entirely representative of the actual state and in practice, most surveys ask questions about migration in an indirect way, which also does not provide clear information. Furthermore, surveys are also very costly to conduct which limits the amount of participants and making sample selection difficult, which can affect results.

Administrative Data:

Administrative procedures to register foreigners, visa applicants and students are the main sources for Administrative data. The data is collected by governments for organizational purposes and is not published or reported like the statistical data and many countries do not recognize this type of data as a legitimate source of statistics. However, there is still value in this information, especially in identifying migrant stocks and flows in real time.

One of the sources of administrative data is the Visa, residence and work permit applications. This data is collected on a regular basis as it is used to track individuals resulting in this data being timely and inexpensive to obtain. There is coverage of a broad swath of issues and groups and is a comprehensive source of information on an individual basis. However, there are many issues in using administrative data, one of which, is the inconsistency between countries and even internal administrations. As this data is not regulated by UN or other another organization, there is not a clear, universal approach as to how visas are handled across countries. Another large flaw with this data is that it completely ignores undocumented migrants, which make up a large part of many migrant populations, as well as ignoring naturalized citizens. In addition, while more timely than statistical data, it is still offset from the real migration event. The information gained from this type of data is very similar to that of statistical data such as migrant stocks, flows, drivers and socio-economic characteristics.

Another way administrative data is collected is by border data collection systems. Border data collection systems are at official border crossing points, often called customs, and track people as they enter and leave a country. As this data is collected on a continuous basis, it is very timely as well as cheaper to obtain again compared to statistical data. It also provides detailed information about those who cross using these official border posts which can be used to estimate migration flows in different areas and reasons for crossing. While this data can be valuable for analyzing migratory patterns, it only covers those who cross at an official post and does not represent any undocumented migration. It can also be difficult to distinguish migrants from tourists and data quality can be affected by citizenship type.

Innovative Data:

According to the International Organization of Migration, there are significant gaps in the data collected related to migration. Currently, the focus of the data collection is related to students, migrant stocks, remittances and human trafficking. Topics such as irregular migration, migration flows, smuggling, impact of policies as well as factors such as migrant health and integration are considered to be the major data gaps in analyzing migration. These gaps mean that data on these topics are either incomplete, inaccessible or not disaggregated. This can

result from an inability to collect data on these topics to privacy concerns. These gaps in migration data have made using a quantitative approach to analyzing migration very difficult as the quality of the statistical and administrative data is often not granular enough to create valuable analyses. However, because of the increasing desire to be able to understand more about migration from governments, researchers and other organizations, a new source of data has arisen.

Innovative Data as it is referred to by the IOM, is a catch all term refers to any type of big data available such as social media records, online banking transactions, mobile phones and GIS satellite data. This data is called big, not only because of its size, but the velocity that it can be collected at and complexity of information that is available. Big data does not refer to a random sample of individuals but to the entire population using these platforms. This increased complexity and granularity comes at a prices as more advanced and technical processes are need to obtain information from this data (Global Migration Group). The previously mentioned sources of big data have been helping to fill the data gaps and provide real time information about migration. In addition to assisting in migration stock and flow estimates, innovative data can provide insight as to populations of people that are likely to migrant, exact migration routes, drivers and impact of migration, public perception, and internal movements and displacement. Some of the most notable recent developments in this domain have been in Mobile Phone Call Detail Records, Geo-located social media activity, Social media marketing data and online search data.

Mobile Phone Call Detail Records (CDR) are records containing information such as approximate location of the calling and receiving end, time and duration of the call, as well as the calling and receiving number, which can be used as an anonymous identifier for the caller and receiver. A lot of the uses for CDR has been in analyzing internal migration/movements of people. A study done funded by the Swedish National Board of Health and Welfare used position data of subscriber identity module (SIM) cards from the largest mobile phone company in Haiti (Digicel) to estimate the magnitude and trends of population movements following the Haiti 2010 earthquake and cholera outbreak (Bengtsson 2010). The goal of the study was to demonstrate the feasibility of rapid estimates of movements of people in emergency situations.

They were able to that around 630,000 people left the capital Port-au-Prince within 19 days of the disaster which corresponded to around 20% of the population. Another group from the SoBigData Consortium has also been using similar CDR data, paired with geo-referenced Twitter data and official labor force statistics to model integration levels of Syrian Refugees (Connor 2017). The objective of this type of research is to use CDR data to help understand transnationality among migrants by creating something like an integration index to better understand populations who have and have not successfully integrated into their respective host countries.

Geo-located social media activity from platforms such as Twitter, Instagram and LinkedIn have become popular sources of Innovative data as they provide disaggregated information about people including age, sex and skill levels which is user provided. According to Statista, in January of 2018, there were 3.2 billion active social media users of which just over 2 billion were Facebook users. API data has been used by many people such as Stefano Iacus, to show the movements of tourists in cities like Milan and Rome as well as showing the travels of Italian twitter accounts across the world. (Iacus 2017). However, with growing concern of both privacy and ethical issues, companies like Instagram and Twitter have severely restricted access to the API data, making it increasingly difficult for researchers to access this information.

Another source of big data used to help under migration has been social media marketing data. For targeted advertising purposes, companies like Facebook have created individual profiles for users that can include age, sex, home country, country of residence, education level, personal interests and occupation. This can be viewed as real-time digital census available cross countries and significantly helps to fill the data gaps mentioned before. (Zagheni 2017). In 2018, the European Commission Joint Research Centre estimated the number of ‘expatriates’ residing in 17 different European countries and were able to show an increase in Venezuelan immigrants in Spain which was then subsequently conformed by the Spanish National Statistic Office later that year. One of the concerns noted however, was that social media data may over/under represent certain segments of the population creating a bias as younger people are more likely to have an active social media account (Spyratos 2018).

Profile data could also be used to show public sentiments about migrants and as well as the sentiments of the migrants themselves.

Finally, one of the last sources of big data that has been used in a migration context is online search results. The Pew Research Center published a study that used Google Trends data which tracks the number of queries related to a certain topic, to analyze whether online search records could be a useful tool to predict migratory flows. This data can also be manipulated to show migration intentions as well as expected destinations (Bohme 2018). Because of the large scope of the Google Trends data, over 1 billion users worldwide, it is thought to be highly representative of the population and, combined with other sources of data, could be a highly powerful tool to estimate future migratory flows.

With all the benefits of using big data, it may seem that this approach offers a near perfect solution to the current problem with collecting data on migration. However, there are large concerns within the community about issues such as privacy. Much of the data collected on the users is without the explicit consent of the user and it can provide detailed information about very personal topics. Many people feel uncomfortable with the fact that their social media interactions can be used in a surveillance context by government or other bodies. Because of this, many nations have addressed this concern and have created committees to draft legislation for appropriate and ethical use of this data for the future. The EU Agency for Fundamental Rights has a project called "Artificial Intelligence, Big Data and Fundamental Rights" with the goal of assessing the advantages and disadvantages in terms of fundamental human rights of using artificial intelligence, machine learning and big data for public policy and business purposes. It hopes to be able to make policy recommendation and guidelines for people using these sources and technologies in the future. Another concern with big data is that it may be inherently biased as it does not represent people from different socio-economic, rural/urban and age groups (EU 2019). Currently, much work is being done to address these concerns as to improve analysis and predicative model performances. Lastly, there are some technical and legal challenges to using big data. Access to data can be difficult as many of these sources are private companies who have no obligation to share this data and in fact have been under recent pressure by both governments and users to tighten the distribution of this

information. Additionally, there are significant technological and methodological barriers in using this information as it requires a knowledge of advanced statistical and analytical methods in order to convert the 'noisy' data into information. Many of the researchers in this field have backgrounds in Statistics, Data Science and Computer Modeling which limits the number of people who can take advantage of big data.

Research Questions:

How are migrants displayed in Dutch newspapers and how does sentiment change over time?

Abstract:

Understanding how a migrant population is viewed and displayed by the host country has been a struggle for a long period of time for anyone studying migration. Traditional methods of collecting information were tedious, time intensive and expensive. However, Big data has been providing unique solutions to gaps in information in many different fields across the world. Media plays an important part in developing and assessing the public opinion on a topic. With the availability of a large number of online articles from historical time periods, it is possible to use quantitative analysis, such as text analytics, to can see how the 'migrant' is presented in local Dutch newspapers using online records of articles. Average sentiment of the analyzed articles were in general negative but also varied over time, with negative sentiment peaking in 2015 which is correlated with high unemployment in the country. Words most commonly associated with articles with keyword 'migrant' included 'work', 'policy' and 'jobs' suggesting that migrants are often portrayed in an economic or political context.

Methodology:

Data Collection:

Given the linguistic limitations of the researcher, only news sources available and originally written in English were selected. Data was collected from one news source based in the Netherlands, the nltimes.com. It was selected as the articles were in English and the website had a convenient search feature that allowed for articles to be queried using keywords. The keyword used to find the articles was 'migrant' and all articles from the source were used in this analysis. In total 65 articles were collected from 2014-2019. Articles were then copy and pasted, along with the date, article title and URL into an Excel spreadsheet to create a manufactured data source.

Data Pre-processing:

After the data was collected and assembled into an Excel spreadsheet, the data was then cleaned to ensure ease of analysis. All special characters such as Ö, é, á and ø were removed as they would be unreadable by the packages. All dates were reformatted to be identical. Since the dataset was built, density was ensured and no missing values or incomplete features existed.

Data Analysis:

Sentiment:

Conducting sentiment analysis was done using a package called 'SentimentR', which runs through a block of text, such as an article and assigns a 'positivity value' to each word. This is done by importing the r package, lexicon, (a collection of lexical hash tables), dictionaries, and word lists that is currently maintained by Tyler Rinker at the University of Notre Dame. Not only does this package provide a simple and effective way to conduct sentiment analysis with fewer lines of code, it also corrects for inversions. If a normal package was given the sentence, "I am not good", it would judge this text token as positive because of the use of the word 'good'. However, SentimentR is able to recognize that the preceding not in the token flips the meaning

of ‘good’ and therefore would assign a negative value to the token. Using the sentiment_by command, one is able to collect aggregated sentiment of the given block of text entered. The result for each article was an *average sentiment score*, *standard deviation of average* and *word count*. The average sentiment score expresses the valence and the polarity of the sentiment. This value was used for displaying the sentiment of each article.

Text:

Conducting text analysis was done with the assistance of packages tm, SnowballC, textmineR. The first step in doing basic text analytics on these articles was to turn the data into a Document Term Matrix (DTM). A DTM is a mathematical matrix that describes the frequency of terms that occur in a collection of documents. An example of can be seen in the following picture with two short documents, D1 and D2:

- D1 = "I like databases"
- D2 = "I hate databases",

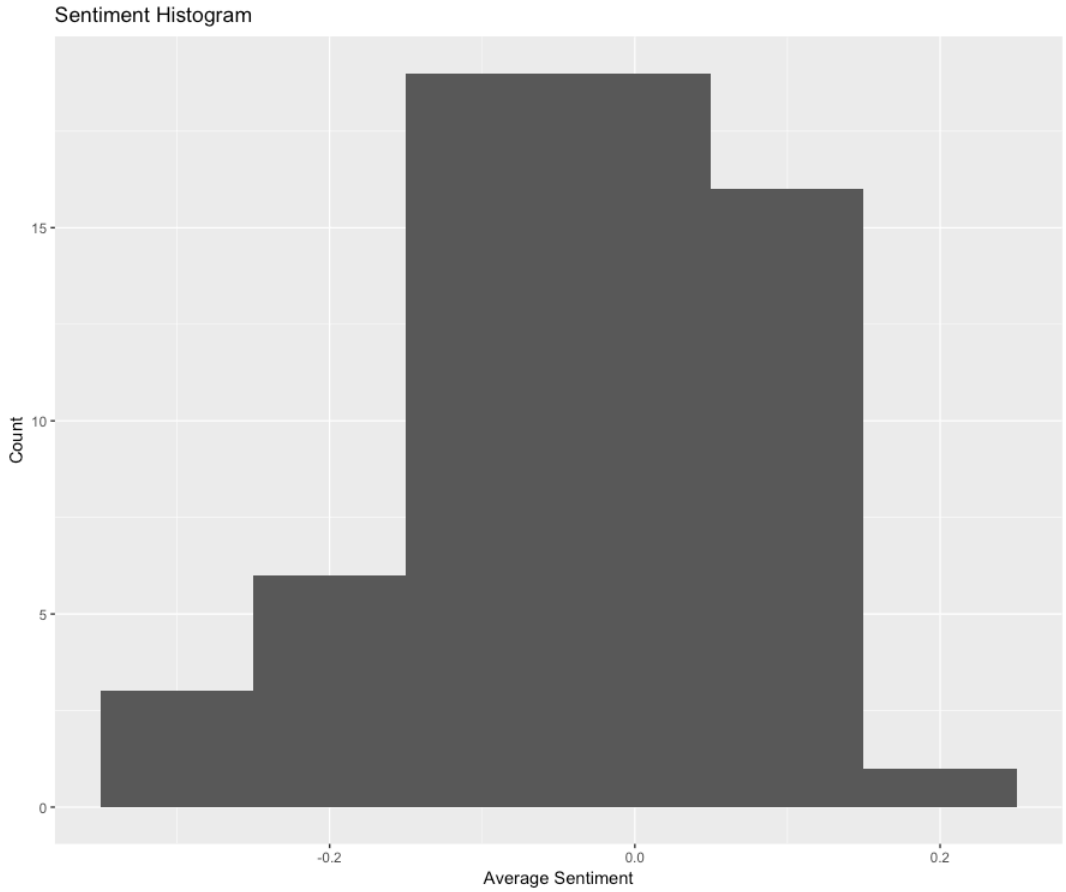
then the document-term matrix would be:

	I	like	hate	databases
D1	1	1	0	1
D2	1	0	1	1

The creation of the DTM was done using the package SnowballC and the corpus function. Then using the tm package, the DTM was cleaned to adjust all words to all lower case to ensure the ‘Migrant’ and ‘migrant’ were treated as the same word. Punctuation, numbers and English stop words (Ex. and, the, so, was etc.). The stop word dictionary is based MIT’s appendix for Natural Language Processing. Words were then stemmed to further consolidate the matrix. Stemming is the process in which the root word is identified and then the rest of the word is dropped. An example would that or ‘workers’ in which the stemming function would identify the root word ‘work’ and transform the ‘workers’ words into ‘work’. This further helps with text analysis as it helps to identify the topics and subjects of the text being analyzed rather than just word choice. From the DTM many different metrics and analysis can come to fruition as will be explained in the next section.

Findings & Evaluations:

Sentiment Distribution:



One of the first steps taken in analyzing average sentiment of the articles was to see how the distribution of sentiment occurred. This histogram above shows the number of articles in each given range of sentiment with bin sizes of 0.1. The histogram shows that the data is slightly left skewed, meaning that there is more data points or articles on the left side of 0 axis of average sentiment. Since the average sentiment ranges goes from negative to positive on the X-axis, this means that there are more articles that have a negative average sentiment than positive. However, in looking at the histogram, we can see that most articles are right around neutral to moderately negative sentiment implying that a majority of the articles analyzed display migrants negatively. The ends of the histogram are also interesting in that there is only one article with greater than 0.15 average sentiment yet there are 9 articles with less than -0.15 sentiment and 3 articles that have less than -0.25 showing that moderate to highly

negative sentiment articles are more common and extreme than those articles who display migrants in a positive sense. In summary, this histogram shows that articles containing the keyword 'migrant' are more likely to be of negative sentiment and that when articles are negative, they are more extremely negative than their positive counterparts.

In order to get a better idea of the types of articles that were most positive and negative, a list was created to show the titles of the most extreme articles. The titles of the five most negative articles were as follows with the most negative article being listed first.

1. HUMAN TRAFFICKING ARREST IN AMSTERDAM FREES EGYPTIAN MAN
2. REPORT: POLISH MIGRANT WORKERS STILL BEING EXPLOITED
3. UNDOCUMENTED MIGRANTS OFTEN HELD IN SOLITARY CONFINEMENT
4. DISTURBED MAN ARRESTED FOR ROTTERDAM AIRPORT TERRORISM THREAT
5. SOME 60 SYRIAN CHILD BRIDES, AS YOUNG AS 14, ENTERED NETHERLANDS

One observation of these negative articles is the fact that most of these articles are not explicitly inflammatory regarding migrants, but actually discussing gruesome and horrendous things that are happening to the migrant population. The sentiment is very negative because these topics are heavy and use words that carry a negative connotation. This was an interesting observation as I had originally thought more of these titles would be like number 4, displaying the 'migrant' as more explicitly negative. To highlight the differences between the positive and negative articles, the five most positive article titles are listed below with the most positive being listed first.

1. HIGHLY-EDUCATED YOUNG MIGRANTS UPBEAT ON NETHERLANDS JOB PROSPECTS
2. DUTCH POPULATION TO JUMP TO 18 MILLION BY 2060
3. MIGRATION PLAYED SMALL PART IN DUTCH POPULATION GROWTH
4. OVER A THOUSAND TEACHERS NEEDED FOR MIGRANT CHILDREN
5. NETHERLANDS EXPECTS 70,000 ASYLUM CASES IN 2016

These positive articles seem to highlight some of the benefits of migrants in Dutch society. Two of the articles discuss the population growth of the Netherlands which apparently is displayed

with positive sentiment. The most positive article highlights the success that migrants are finding within the country and how they are adapting to life in the Netherlands.

In summary, these findings in conjuncture with the histogram show that migrants are more frequently displayed negatively in the news and the most extreme negatives tend to deal with exploitation done to the migrant population or terror events in which migrants were the offenders. Less frequently and less extremely are migrants portrayed positively in the news and when they are, it focuses more on population growth of the Netherlands.

Word Frequency:

The next step in the analysis of the articles was to determine the most frequent words within the articles using the DTM. The following words appeared more than 50 times total within all of the articles:

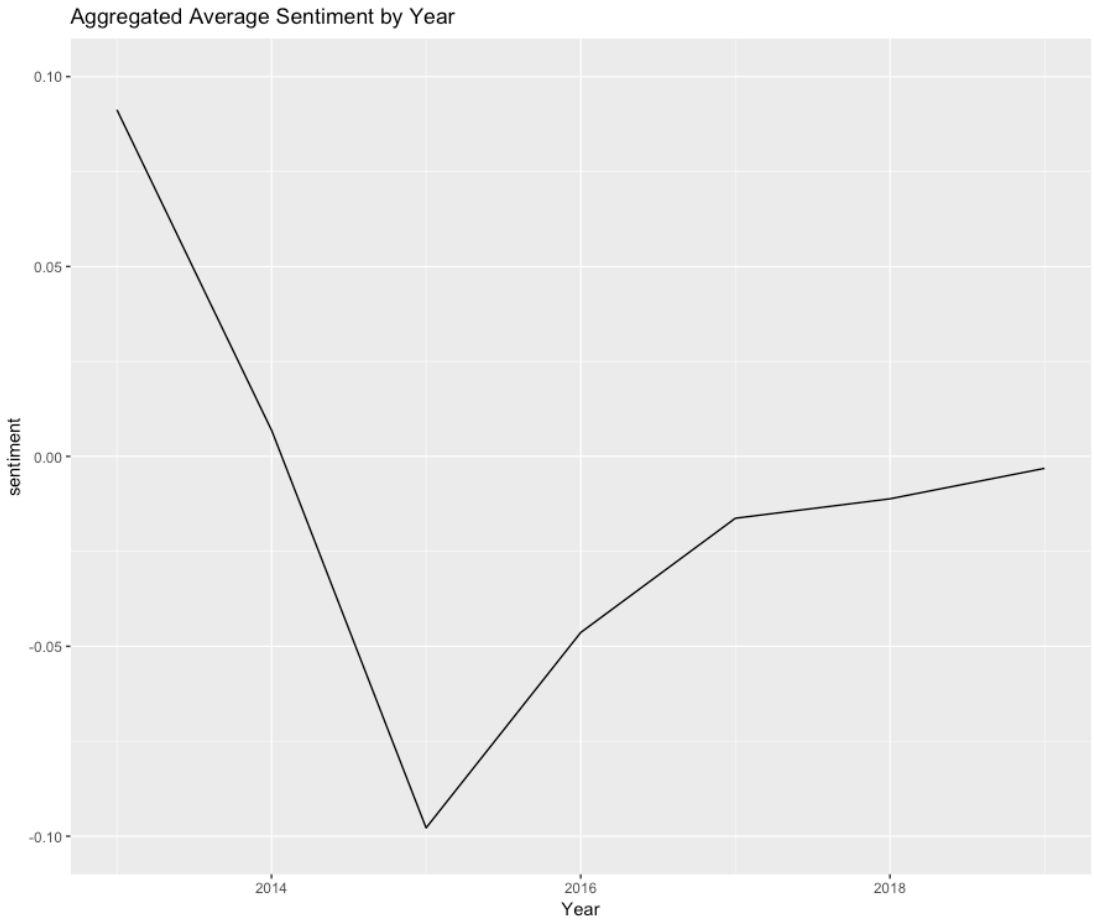
"asylum" "dutch" "migrant" "netherland" "number" "peopl" "polic" "report" "seeker" "thousand" "year" "work"

Looking at these words, one can see that many of the articles analyzed were related to employment reports concerning migrants. As I built the dataset myself and saw all the article titles as well, I can confirm this previous statement. One of the interesting implications of this is that in the public sphere, migrants are most likely to be discussed in a work setting with articles either talking about how many migrants are working in the country in a year, or voicing concern of whether or not migrants took jobs away from local works. There is greater focus around the employment of migrants versus other markers of integration such as education, language skills, or housing. Another interesting observation is that the words ‘asylum’ and ‘seeker’ appear very frequently which indicates that within the news, migrants and asylum seekers are either mentioned together very often or those two terms are used interchangeably to refer to the same population of people. This help to show the narrative that many of these articles paint of migrants as ‘asylum seekers’ rather than just foreign nationals residing within the country. The

term migrant is being used in the context of economic or political migrants rather than just people who have moved into the country.

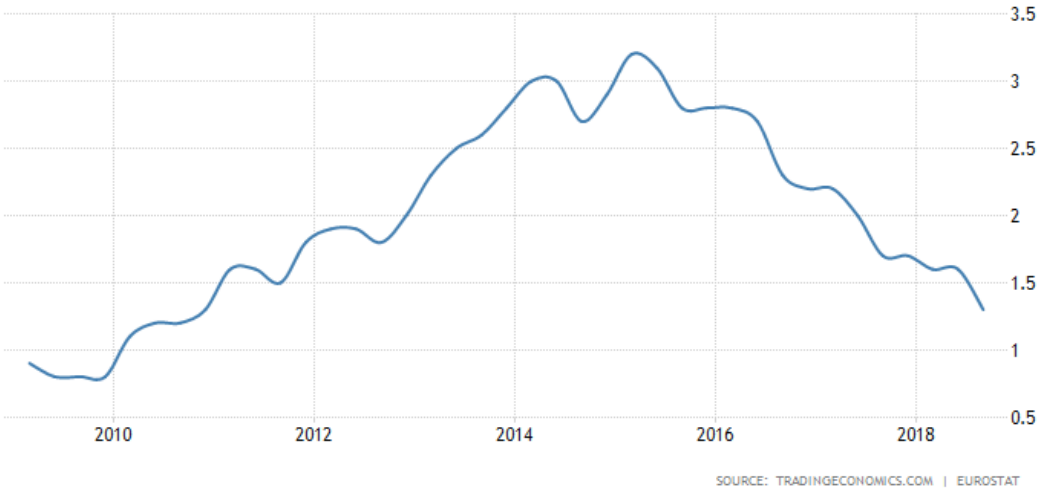
Sentiment Over Time:

The next step in the analysis was to look at average sentiment of articles across time. The following charts display mean average sentiment aggregated by both year and month.



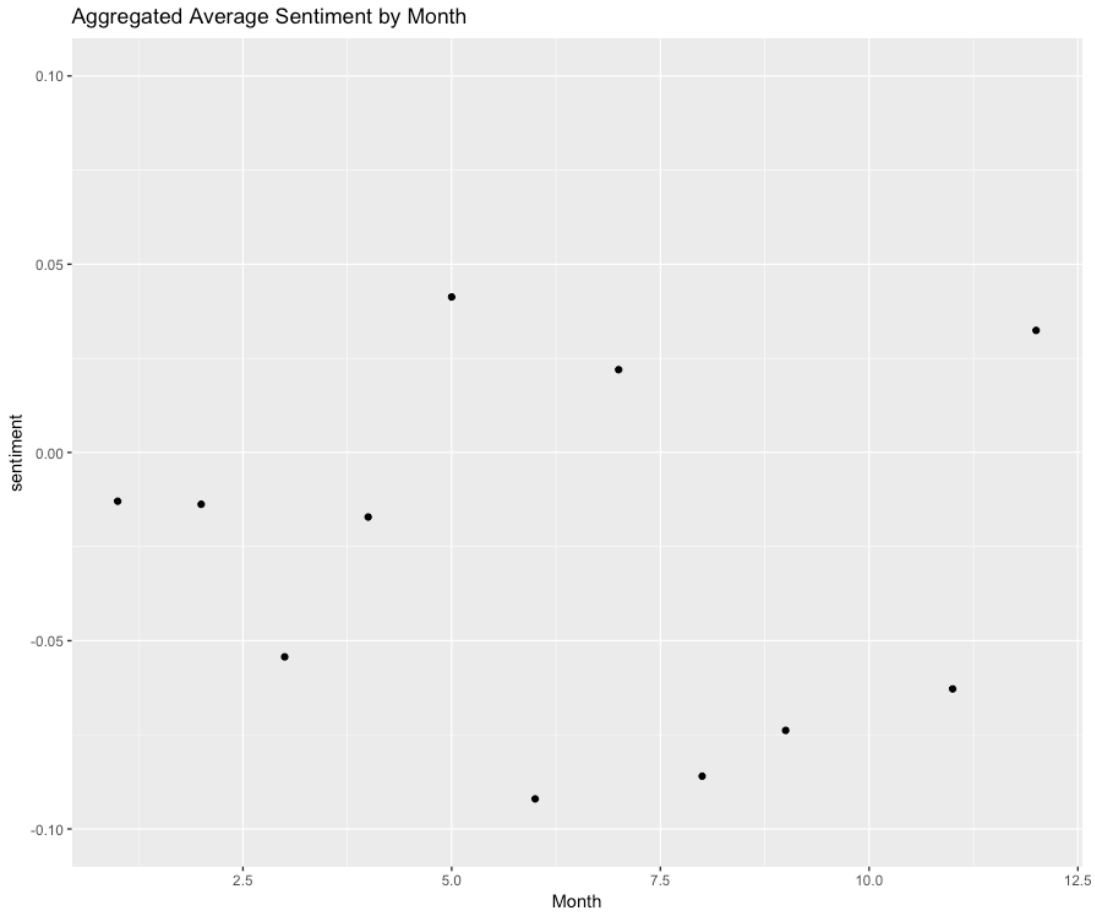
The chart showing yearly sentiment very clearly displays a differing average sentiment over time. In 2013, sentiment was high but started to fall in 2014 and hit a low in 2015. After 2015, sentiment once again rose but still remained negative. The 2013 year was an outlier as there were only two articles from that particular year, however, the count for the rest of the years was even and above 10, which means that we can make reasonable assumptions that this is representative of the actual attitude of the people. While there are plenty of reasons that

could be responsible for increased negative sentiment on articles about migrants, domestic unemployment could help to show why sentiment was falling since most of the articles about migrants were concerning work. The chart below shows the unemployment rate over the past 10 years in the Netherlands.



As one can see, unemployment hit its peak in 2015, the same year that sentiment was at its most negative level. While there is not enough proof to prove causation, it is an interesting correlation that could be further investigated in later research to see if rising unemployment rates cause the public to more negatively view migrant populations as they may view them as taking the jobs away from the native population.

Aggregated average sentiment was also done by month to see if any trends existed throughout a calendar year. I had thought that sentiment might be more extreme in a particular season where more temporary workers were working in the country. The following chart shows this average sentiment by month.



This chart while interesting, does not give any actual information as there is no clear trend. Sentiment seems to vary significantly without any pattern. Because of this, there is no evidence to support any claims that the time of the year would affect how migrants are portrayed within the newspapers.

Limitations:

Data:

In doing data analysis, unsurprisingly, the most important part and the real value comes from the data itself. I believe that in this analysis, one of the largest limitations and shortcomings of this research is within the data. Web scraping is a time consuming, tedious and difficult task to do on its own. Given the sources of data I was searching for; this process was made even more difficult as I ran into many barriers in trying to collect this data.

Initially I had wanted to use an API to collect Twitter/Facebook data but quickly realized that would not be feasible as I did not have access to the required computing power, bandwidth and internet speeds in Morocco to be able to leverage that much data in only a 4-week period. I then decided to move to newspaper articles as they have traditionally been responsible for dissemination of information to people and can be reasonably assumed to be representative of the general attitude about a given topic. However, this assumption is another limitation of the research as it may not show the whole attitude of the country or people as not everyone reads every source nor does everyone agree with what is written.

In finding newspaper articles, the first and largest issue that stood in my way was the language barrier. Europe is a very diverse continent with many cultures and peoples that speak many different languages. Being an American student who is only fluent in English, I could only reasonably conduct analysis on articles that were in English. While I thought about translating articles into English, I ran into logistical issues in that it would be extremely inefficient to have a human translate them, and using an online translator, while good for extracting rough meaning, would not be sufficient as specific word choice was important for detailed analysis. I was then reduced to sources of articles that had been originally published in English. This made finding sources very difficult as many of the reputable and most read publishers could not be used.

Another issue that I ran into was that many online publishers have restricted websites that do not allow a non-subscribing guest to access more than 3 articles per week. This proved to be a huge restriction as many of the original sources I planned to use had this restriction in place and made collecting historical articles on migrants futile and pointless as I would not be

able to gather enough articles to adequately represent them in my analysis. Furthermore, many sources rather than having a search function to look at past articles published, had incorporated a custom Google search function on their site that while queried articles with the given search, resulted in articles from other news sources coming up as well. It was almost impossible for me to see the source until I had entered the article and also expanded the article content beyond what I wanted such as including articles on migration in the US, Asia and Australia. These barriers further reduced the sites that I could use for data collection.

In the end, I ended up with only one source that ended up meeting these requirements, NLtimes.com. This website is an English newspaper written in the Netherlands that writes and reports only about events happening in the Netherlands. It had a convenient search function that allowed me to type in me keyword 'migrant' and queried all of the articles ever published on that site with that keyword and in total I was able to get 65 articles. One the advantages of using Innovative Data is the scale of information that can be used and while I was able to gather more data than using traditional qualitative methods (interviewing authors), I was disappointed on how few articles I was able to use. While this one site was able to provide me with technically enough articles, all analysis is extremely biased as all abstractions are based off just one of many newspapers within the country. While I was not able to explicitly find any political affiliation on the newspapers website, there is a chance that implicit political ideals of the authors could have affected way they write articles.

Analysis:

One of the limitations of the analysis done in this report is the simplicity of the used NLP techniques. Sentiment and DTM's are valuable and important things to use in quantitative analysis of text, but they do not give the whole picture. Articles that are written with 'positive' words may be conveying a different message to the reader given the whole context of the article. More advanced NLP techniques could have been employed to gather a more holistic understand of what the articles were really trying to convey such as preceding word selection of the keyword, lexical and distributional semantics. However, given my limited experience in

text analytics and the time constraint, many of these things were not practical to do in the context of this project.

Ethics:

As the data collected was done using quantitative methods instead of qualitative techniques, the ethical concerns are not quite as strong. However, with that being said, there is a concern regarding the collection of data. There was no explicit consent from the nltimes.nl to use the newspaper articles and even though articles were publicly available on the internet and were free for anyone to access, it is unclear if the source would consent to having their articles analyzed in this fashion. In previous experiences doing text mining, some websites will not let a user access a certain amount of records without explicit consent. Most of the reasons they do this is to protect proprietary information and I do not believe that using the articles from nltimes.nl falls under that scope.

Acknowledgements:

I would first like to thank my Professor Tibari and my advisor Chegraoui for their guidance throughout the semester. Secondly, I would like to thank Badrdine Boulaid for all his assistance in my everyday life in Morocco that made conducting this research possible. From helping me get a sim card to taking me to the doctor and sitting in the waiting with me for what felt like hours, he has made my Moroccan experience much better. Special thanks goes to Dr. Fred Nwanganga at the University of Notre Dame for his technical advice in this project and helping me over the past couple of years to become the data analyst I am today. Finally, I would like to thank my friends I have met on the abroad program, especially Zander, Will, Jack, Miranda and Liz for making the adjustment to a new lifestyle so much easier. Their friendship is something that made every day brighter and without them I do not think my experience would have been nearly as positive.

Citations:

Ardittis, S., & Laczko, F. (2017). Improving data on migration – towards the 2030 Agenda for Sustainable Development and the Global Compact on Migration. *MIGRATION POLICY PRACTICE,VII*.

Bengtsson, L., Lu, X., Thorson, A., Garfield, R., & Schreeb, J. V. (2011). Improved Response to Disasters and Outbreaks by Tracking Population Movements with Mobile Phone Network Data: A Post-Earthquake Geospatial Study in Haiti. *PLoS Medicine,8*(8). doi: 10.1371/journal.pmed.1001083

Bohme, M., Gröger, A., & Stöhr, T. (2018). Searching for a Better Life: Predicting International Migration with Online Search Keywords. *Organisation for Economic Co-operation and Development*.

Connor, P. (2017). Can Google Trends Forecast International Migration Flows? Perhaps, but Under Certain Conditions. *Pew Research Center*. Retrieved March 9, 2019.

EU. (2019, February 12). Artificial Intelligence, Big Data and Fundamental Rights. Retrieved from https://fra.europa.eu/en/project/2018/artificial-intelligence-big-data-and-fundamental-rights

Global Migration Group. (n.d.). Handbook for Improving the Production and Use of Migration Data for Development. *Global Knowledge Partnership on Migration and Development*. Retrieved from http://www.un.org/en/development/desa/population/migration/events/coordination/15/documents/Final Handbook 30.06.16_AS4.pdf

Iacus, S. M. (n.d.). Big Data and Migration: Limits and Perspectives. Retrieved from https://bluehub.jrc.ec.europa.eu/bigdata4migration/uploads/attachments/cjb4tjlhq006bcigf2b24zr97-s-iacus.pdf.

Silm, S., & Tiru, M. (2017). Tracking Trans–nationalism with Mobile Telephone Data. Retrieved March 9, 2019, from https://inimareng.ee/en/open-to-the-world/tracking-trans-nationalism-with-mobile-telephone-data/

Spyratos, S. (2018). Migration data using social media. *Joint Research Centre (EU Commision)*.

Statista. (n.d.). Facebook users worldwide 2018. Retrieved from https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/

UN Stats. (2010). Principles and Recommendations for Population and Housing Censuses. *United Nations Statistics Division*.

Zagheni, E., Garimella, V. R., Weber, I., & State, B. (2014). Inferring international and internal migration patterns from Twitter data. *Proceedings of the 23rd International Conference on World Wide Web - WWW 14 Companion*. doi:10.1145/2567948.2576930

Zagheni, E., Weber, I., & Gummadi, K. (2017). Leveraging Facebooks Advertising Platform to Monitor Stocks of Migrants. *Population and Development Review,43*(4), 721-734. doi:10.1111/padr.12102