25/06/2021

Analyzing Lefties Complaints on Portal das Queixas

Social Media Analytics S2
NOVA Information Management School

Ana Beatriz Cardoso	m20201286
Jordana Rodrigues	m20201289
Luciana Nassar	m20200056
Nickolas Lago	m20200097
Sofia Serzedelo	m20201232

Table of Contents

Table of Figures

1 Business Understanding

1.1 Introduction

The development of social media platforms, which can be perceived as virtual places of content creation and sharing, emulating communities and social connection between humans, created new forms for customers to express their feelings about brands, and on the brands' side, the need to efficiently find this data, extract and analyze to improve its services and measure customer's satisfaction about current and new products. The popularity of a company also passes through its ability to increase good reviews and decrease complaints; therefore, it is necessary to treat the complaints as a vital type of data.

Social media data about a company can be referred to as social network's posts about the brand, reviews left about the company on specialized websites, comments written on reports or the brand's content, and among others, what will be the focus of this paper, the complaints about the company. Since they usually come from a place of discontent with the brand, this type of data can add more insights to the company than good reviews, especially when they are direct complaints about certain services and products. The company's capability to access this feedback, respond to it, and implement its acquired knowledge to the development of new services or products means positive and satisfactory customer service.

When accessing websites that are specialized in complaints, it is possible to see many companies ignore the customer's complaints, which can lead to unsolved problems and more discontentment about a company. A brand's response rate tends to be analyzed as the company's ability to interact with customers and access its needs. The good conduct on these cases is also important since these websites count with good search engine optimization strategies that make sure, in many cases, the complaints appear on the firsts search results when looking for certain companies, hence making sure new customers can read and analyze the company's reputation and effort on treating customers complaints.

1.2 The Company: LEFTIES

Lefties is a clothing brand from the Group Inditex, one of the main distributors of fashion in the world. Lefties has 170 stores in nine countries, mainly located in Europe and the Middle East. It was initially designed to be a more accessible brand, selling last season products from the main Inditex brands and even flawed products at smaller prices. In Portugal, Lefties is present since 2006 and nowadays is present in 23 locations.

We decided to use Lefties´ complaints as the subject of our analysis based on its reputation and due to its low response and solution rates on the website we decided to use (portaldaqueixa.pt). While the solution rate, which in this case is 24,8%, corresponds to the number of complaints that were solved by the company, although doesn't allude to the satisfaction of the customer with such a solution. The response rate of 25% demonstrates not many complaints received a response from the company.

1.3 Problem Background

The development of an online store, an increasing number of complaints about the brand's quality on the products and services, and the development of the COVID-19 pandemic that enabled life online rather than offline generated more of these kinds of data about the company. And this data can be extremely important when generating new insights about business development and helping to solve issues reactively and proactively.

The analysis of the complaints data at Portal da Queixa's website has the objective of developing a more efficient manner for the company to deal with its customer relationship processes. When analyzing complaints, it is expected that the insights about the company's future ways of handling problems are based on actual feedback of the customers who were harmed by these issues.

Some of the models used in this project are meant to gather information using text mining techniques, where we try to identify the most common terms and, therefore, most common topics of a complaint. While others intend to analyze the sentiments on specific complaints that mention specific topics on them.

2 Data Understanding

2.1 Portal da Queixa

The "Portal da Queixa" is a social media for consumers to share experiences (bad or good), see complaints & recommendations of other consumers. It's more than just a place for sharing complaints but also for consumers to track the reputation of a certain brand with the Satisfaction Index. It exists to defend the consumer and as means to resolve their problems. It counts with more than 170,000 users and more than 4,000 brands.

Portal da Queixa counts with more than 13,000 complaints since the year of its creation, 2009. It's a strong reference for consumers when they are unsure to buy from a brand or not. Simple, easy to use, and free.

Analyzing now, the page of Lefties in Portal da Queixa, we can see that the Index of Satisfaction is 25.3% (as of June 24). The Index of Satisfaction is calculated based on 3 main areas: the rate of solution overall (Lefties case is 25,3%), response rate 24,4%, and the evaluation score by the consumer 3.5 out of 10.

2.2 Web Scrapping

Web Scraping is the technique to extract structured data available online, when there isn't an API, using computer programs. After the web scrapping is finalized, it's possible to analyze and act upon the data.

2.2.1 Reviews URLs and IDs Retrieve

With the web scrapping technique, we identified the URLs we wanted to scrap based on the timeframes used (January 14 of 2016 and May 5 of 2020), and then, after inserting the code, we extracted the data and stored the data on the CSV file with all the complaints.

On our dataset, we extracted the complaint ID (which represents the ID of the complaint on the Portal da Queixa), the complaint status ("Aguardar resposta", "Resolvida", "Sem Resolução"), the User name of the complaint, the complaint title, the number of views and the text of the complaint (describing the situation).

2.2.2 Reviews Dataset Creation

Using web-scraping techniques, we were able to have a dataset with 177 complaints between January 14 of 2016, and May 5 of 2020. Out of the 177 complaints that we collected, 27 are still waiting for a response, which represents 15%. Out of those 27 unresolved complaints, 20 are related to problems with the delivery process. Of the 177 complaints, 39 were resolved, which represents 22%. The remaining 111 complaints are unresolved until now, which represents 62%. And most of the unresolved complaints are related to the delivery time frame, which indicates they have a poor service of delivery and they don't follow up properly on this situations.

The complaints that have more visualizations (more than 1,000) are related to bad customer service (related to attitudes from the Lefties employees), returns outside of return dates, and the impossibility of making a return with no receipt. And the complaints with the most visualizations (total of 23 with more than 1,000 views) are unresolved.

3 Data Preparation (Data Pre-processing)

To start preparing our data, we imported the data to a pandas DataFrame df (complaints table head) with all the complaints and, performed text mining operations, to retrieve information from the complaint text.

```
df.head()
```

	complainStatus	complainUser	complainTitle	complainViews	complainText	complainDate
complainID						
59476521	Aguarda resposta	Susana	Lefties - Devolução de artigos	55	Recebi a minha encomenda no dia 19-04-2021 e i...	5 de maio 2021
58935721	Aguarda resposta	Cassia Barcelos	Lefties - Troca de peças compradas on-line	175	Boa noite, _x000D_\nVenho mostrar o meu total ...	23 de abril 2021
58754421	Aguarda resposta	Milene	Lefties - Encomenda não entregue	62	A minha encomenda que fiz no dia 19 de Março n...	20 de abril 2021
58696921	Aguarda resposta	Mariana Ferreira	Lefties - Encomenda não entregue nem resolvem ...	41	Encomenda nr 90003989775 já deveria ter sido...	19 de abril 2021
58650521	Resolvida	Olga Santos	Lefties - Encomenda não entregue	124	Bom dia. Fiz uma encomenda online na Lefties n...	17 de abril 2021

Figure 1 - Complaints table head

To extract useful information about Lefts complaints in the Portal da Queixa, a pre-processing phase comes into play with the use of unstructured textual data and transforming this data to make it clearer to be understood by applying pre-processing techniques, for example, eliminate possible ambiguity, abbreviations, slang, misspelled words and other types of issues that could change the mining of the data. With that in mind, changes may be required from time to time due to Portal da Queixa's continuous updates.

A new column "processText" was added to our data frame df during the pre-processing stage, using the function to the original data frame to be the processed output of the complainText column.

Adding a column using the function to the original data frame with the text processing

```
In [92]: df["processText"] = pd.DataFrame(data=textPreProcess(df.complainText), index=df.index)

In [93]: df
Out[93]:
```

	complainStatus	complainUser	complainTitle	complainViews	complainText	complainDate	processText
complainID							
59476521	Aguarda resposta	Susana	Lefties - Devolução de artigos	51	Recebi a minha encomenda no dia 19-04-2021 e i...	5 de maio 2021	recebi encomenda dia imediatamente dia seguir ...
58935721	Aguarda resposta	Cassia Barcelos	Lefties - Troca de peças compradas on-line	172	Boa noite, _x000D_\nVenho mostrar o meu total ...	23 de abril 2021	boa noite venho mostrar total descontentamento...
58754421	Aguarda resposta	Milene	Lefties - Encomenda não entregue	59	A minha encomenda que fiz no dia 19 de Março n...	20 de abril 2021	encomenda fiz dia março entregue devido ctt ex...
58696921	Aguarda resposta	Mariana Ferreira	Lefties - Encomenda não entregue nem resolvem ...	38	Encomenda nr 90003989775 já deveria ter sido...	19 de abril 2021	encomenda nr deveria ter sido entregue sem...
58650521	Resolvida	Olga Santos	Lefties - Encomenda não entregue	121	Bom dia. Fiz uma encomenda online na Lefties n...	17 de abril 2021	bom dia fiz encomenda online lefties dia abril...
...	...	...	...	...	...	...	...
27355319	Sem resolução	izana pontes	Lefties - Racismo na loja	900	No dia 11 de abril de 2018, fui a lefties do S...	12 de abril 2019	dia abril lefties strada odivelas parque desde...
24936119	Sem resolução	Cátia Oliveira	Lefties - Atitude de mau agrado	647	A situação foi a seguinte, recebi uma peça de ...	5 de janeiro 2019	situação seguinte recebi peça roupa tamanho er...
24875119	Sem resolução	Mafalda Pereira	Lefties - Trocas	733	Ao efetuar uma troca de uma camisa de criança ...	3 de janeiro 2019	efetuar troca camisa criança indicado valor re...
24808918	Sem resolução	Sofia Sousa	Lefties - Mau atendimento	3850	No passado dia 26 do mês corrente, pelas 19h45...	30 de dezembro 2018	passado dia mês corrente h dirigi á loja lefti...
24795518	Sem resolução	Liliana Rodrigues	Lefties - Péssimo atendimento	508	Na loja Lefties do Centro Comercial Fórum Sint...	29 de dezembro 2018	loja lefties centro comercial fórum sintra esc...

Figure 2 - Pre-processing column creation

For different steps of the data pre-processing, we have applied some python libraries to understand the Portuguese language (stop-words and name entity recognition).

3.1 Normalization

The normalization was used to help the improvement and efficiency of the text mining techniques quality, the information retrieval, as when we normalize natural language resource we can reduce randomness and get closer to standard information which the computer has to deal with.

To start the normalization, we have transformed all the words to lower case.

```
# normalize the case

textNormalized = textRaw.lower()

print(textNormalized)
```

```
boa noite, _x000d_
venho mostrar o meu total descontentamento e frustação relativamente ao serviço prestado. _x000d_
fiz uma encomenda de várias peças, uma delas um pack de 4 calças de fato de treino por 15,99€. ao chegar a encomenda deparei-me que em ve
z do pack só tinha recebido 1 calça em vez das 4. enviei e-mail para o apoio ao cliente a expor a situação, passado algumas horas respond
eram a afirmar que teriam entrado em contacto comigo várias vezes sem sucesso (o que é mentira pois confirmei o número de telefone cedido
na minha conta e não tinha nenhuma chamada não atendida), pediram para retribuir a chamada. _x000d_
contactei então por chamada telefonica, ao qual a assistente que me atendeu nunca pediu desculpa pelo sucedido mas explicou que deveria d
evolver a calça com uma etiqueta própria e que me seria enviado o pack completo. _x000d_
passado dois dias recebi o pack mas desta vez 1 das 4 calças veio com vários tamanhos abaixo do pedido. _x000d_
quero que resolvam esta situação de uma vez por todas, o meu trabalho não é compatível com o ponto picket para fazer a devolução. espero
que resolvam esta desagradável situação de uma vez por todas e concertesa não vou recomendar nem comprar mais nada a esta marca.
```

Figure 3 - Text normalization

3.2 Tokenization

We have used the library nltk.tokenize from where we have imported the function word_tokenize to apply the tokenization technique to the input text when it was divided into smaller pieces, units, or tokens. The text was transformed into separate words and became the output, a list of words also referred to as text segmentation.

```
tokenizedText = word_tokenize(textWOPontuation, language='portuguese')
print("List of words:\n",tokenizedText)
```

```
List of words:
 ['boa', 'noite', 'venho', 'mostrar', 'o', 'meu', 'total', 'descontentamento', 'e', 'frustação', 'relativamente', 'ao', 'serviço', 'prest
ado', 'fiz', 'uma', 'encomenda', 'de', 'várias', 'peças', 'uma', 'delas', 'um', 'pack', 'de', 'calças', 'de', 'fato', 'de', 'treino', 'po
r', '€', 'ao', 'chegar', 'a', 'encomenda', 'deparei-me', 'que', 'em', 'vez', 'do', 'pack', 'só', 'tinha', 'recebido', 'calça', 'em', 've
z', 'das', 'enviei', 'e-mail', 'para', 'o', 'apoio', 'ao', 'cliente', 'a', 'expor', 'a', 'situação', 'passado', 'algumas', 'horas', 'resp
onderam', 'a', 'afirmar', 'que', 'teriam', 'entrado', 'em', 'contacto', 'comigo', 'várias', 'vezes', 'sem', 'sucesso', 'o', 'que', 'é',
'mentira', 'pois', 'confirmei', 'o', 'número', 'de', 'telefone', 'cedido', 'na', 'minha', 'conta', 'e', 'não', 'tinha', 'nenhuma', 'chama
da', 'não', 'atendida', 'pediram', 'para', 'retribuir', 'a', 'chamada', 'contactei', 'então', 'por', 'chamada', 'telefonica', 'ao', 'qua
l', 'a', 'assistente', 'que', 'me', 'atendeu', 'nunca', 'pediu', 'desculpa', 'pelo', 'sucedido', 'mas', 'explicou', 'que', 'deveria', 'de
volver', 'a', 'calça', 'com', 'uma', 'etiqueta', 'própria', 'e', 'que', 'me', 'seria', 'enviado', 'o', 'pack', 'completo', 'passado', 'do
is', 'dias', 'recebi', 'o', 'pack', 'mas', 'desta', 'vez', 'das', 'calças', 'veio', 'com', 'vários', 'tamanhos', 'abaixo', 'do', 'pedid
o', 'quero', 'que', 'resolvam', 'esta', 'situação', 'de', 'uma', 'vez', 'por', 'todas', 'o', 'meu', 'trabalho', 'não', 'é', 'compatível',
'com', 'o', 'ponto', 'picket', 'para', 'fazer', 'a', 'devolução', 'espero', 'que', 'resolvam', 'esta', 'desagradável', 'situação', 'de',
'uma', 'vez', 'por', 'todas', 'e', 'concertesa', 'não', 'vou', 'recomendar', 'nem', 'comprar', 'mais', 'nada', 'a', 'esta', 'marca']
```

Figure 4 - Text tokenized

3.3 Stop Words

We have used the library nltk.corpus from where we have imported the function stopwords to apply the Stop Words technique to remove some words that do not contribute to the overall meaning of the complaints. The most common words in this case scenario in Portuguese like prepositions, articles, pronouns, etc. As examples, we have deleted stop words as "a", "e", "ao", "de", etc.

```
print(stop_words)
```

```
{'estiver', 'houver', 'houvéssemos', 'nos', 'houveriam', 'mesmo', 'você', 'às', 'tive', 'estava', 'tiveram', 'forem', 'esteve',
'houverá', 'teremos', 'houveríamos', 'houvermos', 'estivéssemos', 'que', 'te', 'serão', 'houveria', 'isso', 'hão', 'estou',
'a', 'estamos', 'eram', 'aos', 'tenho', 'havemos', 'ela', 'estejam', 'deles', 'tenhamos', 'tinha', 'os', 'estivesse', 'mais',
'um', 'fosse', 'em', 'esse', 'mas', 'vos', 'será', 'tem', 'nas', 'seriam', 'até', 'temos', 'fomos', 'tuas', 'estive', 'estivess
em', 'e', 'muito', 'pelas', 'delas', 'formos', 'tenham', 'era', 'aquele', 'com', 'houvera', 'tivermos', 'fora', 'tiver', 'tives
sem', 'houverei', 'esteja', 'já', 'estivemos', 'dele', 'serei', 'quem', 'seu', 'teus', 'por', 'seus', 'tivemos', 'de', 'este',
'hei', 'hajam', 'seremos', 'só', 'estivera', 'sem', 'ou', 'meus', 'houve', 'seja', 'aquelas', 'somos', 'nossa', 'o', 'seria',
'estiverem', 'houverem', 'fossem', 'foram', 'nem', 'tua', 'essa', 'teriam', 'as', 'tu', 'sejam', 'está', 'minha', 'for', 'tivés
semos', 'uma', 'ao', 'aquela', 'me', 'seríamos', 'há', 'tivera', 'das', 'pelo', 'sua', 'tenha', 'houvemos', 'do', 'houvesse',
'estas', 'lhe', 'éramos', 'houvessem', 'na', 'estes', 'houveremos', 'fôramos', 'eu', 'nossas', 'à', 'num', 'aquilo', 'fui', 'su
as', 'houvéramos', 'teu', 'estão', 'tinham', 'esta', 'aqueles', 'teríamos', 'fôssemos', 'pelos', 'houverão', 'isto', 'terei',
'entre', 'nossos', 'teria', 'minhas', 'tínhamos', 'elas', 'da', 'foi', 'quando', 'nosso', 'estávamos', 'tém', 'dos', 'terá', 'n
ão', 'qual', 'também', 'numa', 'sejamos', 'tiverem', 'haja', 'estavam', 'dela', 'estiveram', 'meu', 'estejamos', 'se', 'eles',
'depois', 'no', 'pela', 'tivéramos', 'esses', 'estivéramos', 'terão', 'vocês', 'ele', 'houveram', 'para', 'como', 'são', 'tives
se', 'é', 'lhes', 'nós', 'essas', 'hajamos', 'teve', 'sou', 'estivermos'}
```

Figure 5 - Portuguese stop words example

3.4 Steaming

We have used the library nltk.stem.porter from where we have imported the function PorterStemmer to apply the steaming technique and eliminate the affixes (suffixes and prefixes) from the complains texts, transforming words into a more canonical, rooted form.

As an example, we have the word “calças” in the plural, transforming the word into singular “calça”, as well as the word “noite” losing the “e” and becoming “Noit”. It will help some very similar terms, to become the same term and appearing in a higher frequency to identify the popularity and other patterns.

```
stem = PorterStemmer()
stemmedText = []
for t in cleanedText:
    stemmedWord = stem.stem(t)
    stemmedText.append(stemmedWord)
print("Stemmed text :\n",stemmedText)
```

```
Stemmed text :
 ['boa', 'noit', 'venho', 'mostrar', 'total', 'descontentamento', 'frustação', 'relativament', 'serviço', 'prestado', 'fiz', 'encomenda',
'vária', 'peça', 'pack', 'calça', 'fato', 'treino', '€', 'chegar', 'encomenda', 'deparei-m', 'vez', 'pack', 'recebido', 'calça', 'vez',
'enviei', 'e-mail', 'apoio', 'client', 'expor', 'situação', 'passado', 'alguma', 'hora', 'responderam', 'afirmar', 'entrado', 'contacto',
'comigo', 'vária', 'veze', 'sucesso', 'mentira', 'poi', 'confirmei', 'número', 'telefon', 'cedido', 'conta', 'nenhuma', 'chamada', 'atend
ida', 'pediram', 'retribuir', 'chamada', 'contactei', 'então', 'chamada', 'telefonica', 'assistent', 'atendeu', 'nunca', 'pediu', 'descul
pa', 'sucedido', 'explic', 'deveria', 'devolv', 'calça', 'etiqueta', 'própria', 'enviado', 'pack', 'completo', 'passado', 'doi', 'dia',
'recebi', 'pack', 'desta', 'vez', 'calça', 'veio', 'vário', 'tamanho', 'abaixo', 'pedido', 'quero', 'resolvam', 'situação', 'vez', 'tod
a', 'trabalho', 'compatível', 'ponto', 'picket', 'fazer', 'devolução', 'espero', 'resolvam', 'desagradável', 'situação', 'vez', 'toda',
'concertesa', 'vou', 'recomendar', 'comprar', 'nada', 'marca']
```

Figure 6 - Steamed text

3.5 Punctuation, HTML, and Unexpected Characters

In this text preprocessing phase we have decided to remove punctuation, unexpected characters, numbers, line breaks, and HTML tags.

We have used the regex library to eliminate the punctuation and unexpected characters from a string that were not removed after using the HTML parse. These were the most removed text signs (!,?,.,…,;,:) and as an example of special characters we have in this case “_x000d_”.

4 Modeling

4.1 Term Frequency Analysis and N-Grams

With the pre-processing part done, it is time for us to start analyzing our data and get insights from it. Our first analysis will tackle the question of what terms are most present on the customer complaints what are the most common bi and tri-grams.

As a first step, this will give us the hottest words and topics from the dataset. The term frequency (or uni-gram) will list the keywords that most appear in the corpus, without much context. The bi and tri-gram will then group the 2 and 3 terms that most appear together, allowing us to take extract few more insights from the text.

By looking at the term frequency analysis, or the uni-gram, we can see that the word that appears the most is “encomenda” showing, without much context, that the topic of online sales and delivery will probably be a key one for our analysis. Followed by “encomenda”, we have more generic keywords such as “dia”, “loja”, “lefties”, “cliente” and only on the 7th one, “recebi”, we can get more insights into what the complaints were.

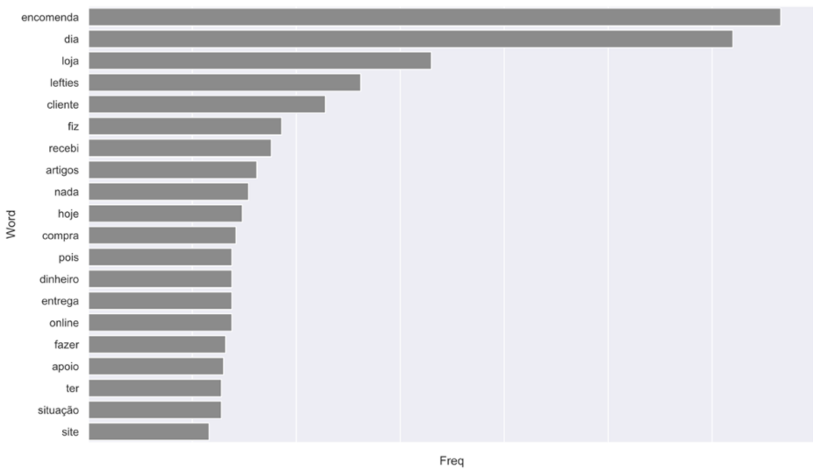

Figure 7 - Most frequent terms (Uni-Grams)

If the uni-gram could give us very short insights, its cousins bi-gram and tri-gram can definitively help us more on understanding the clients' pains. Right at first, we see that the most common bi-gram is “apoio cliente”, which shows the customer probably tries to contact Lefties before complaining on Portal das Queixa’s. Followed by it, we also see many bi-grams related to the waiting of the user for its purchase. Terms like “fiz encomenda”, “encomenda online”, “fiz compra”, “encomenda dia” show that either the waiting or the delivery is a crucial point of complaints from our clients.

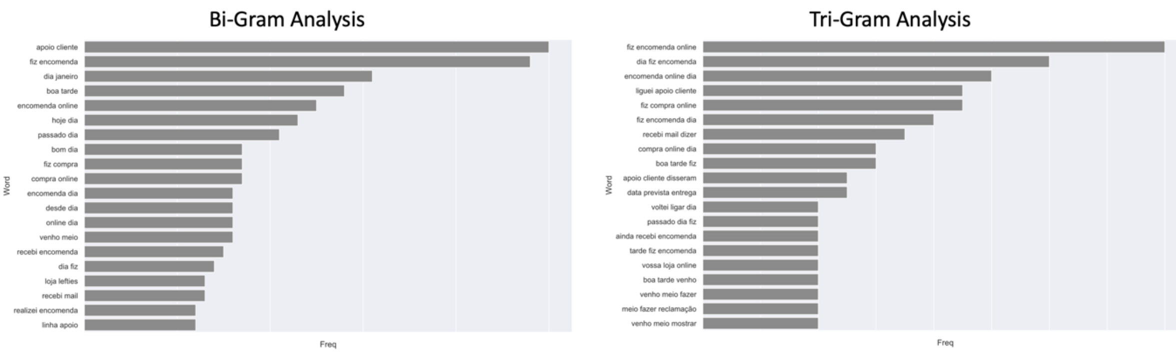

Figure 8 - Most frequent Bi-Grams and Tri-Grams

Last, when we look at the combination of three words (tri-gram) we can quite confirm the insights we saw previously. Of the eight most frequent tri-grams, six were related to online purchases and the other two were from customers who tried to contact Lefties (or got contacted, but couldn't fix its problem.

4.2 Word Cloud

The word cloud visualization shows similar data from the term frequency analysis, but visually and more easily to understand. It uses the term frequency to create the word size and plots the words in a cloud visualization (our case, a squared cloud).

At first, we can enforce the insights we had before. "Encomenda" is the word in highlight and words like "recebi", "loja", "cliente" and, "dia" also have great importance.

Figure 9 - Word cloud

4.2.1 Steamed Word Cloud

An interesting approach when creating a word cloud document is to steam the text, remembering that the pre-processing step took already the stop words away from the main corpus. That way, words that were apart before, but had the same root, can now be joined. In our case, that gave words like "compr" and "receb" to gain importance in the analysis.

Figure 10 - Steamed word cloud

4.3 Sentiment Analysis

Sentiment analysis uses natural language to identify subjective information in texts and can be very useful in analyzing the feelings of clients on messages, reviews, complaints, and in many other interactions between them and the brand. In this report, after preprocessing the texts, we broke them into sentences and used the sentiment analyzer from the package "NLTK", which provides scores of negative, neutral, positive, and the compound.

Analyzing the data can determine that some sentences are marked as completely neutral, but most of them are negative. For a better data visualization, we used the package seaboard that created the following plot, which depicts what was expected from complaints, which is that most sentences are between neutral and negative sentiments.

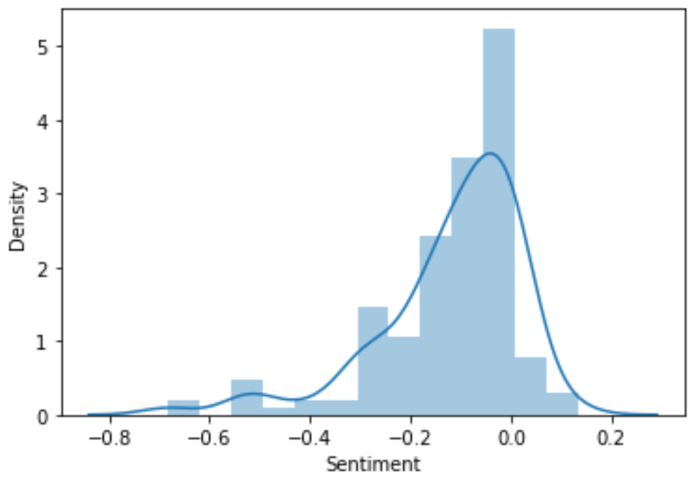

Figure 11 - Sentiment Score distribution

Using the package "JoyPy" the plot "Complains sentiment by status" shows the prevalence of negative sentiments on complaints that are solved, which can be a good strategy to maintain an effective customer relationship, solve faster complaints from users that use more negative terms on their complains.

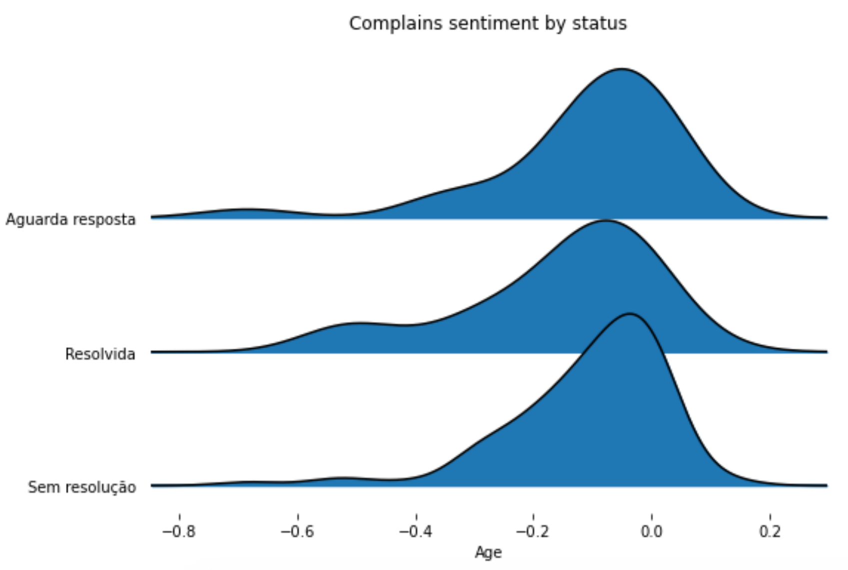

Figure 12 - Sentiment Score by Complaint Status

4.4 Named Entity Recognition

The Named Entity Recognition (NER) is useful to identify and classify entities found on the data following some categories such as people's names, organizations, locations, and others. On the code developed to analyze the complaints, we used the package "spacy" and its Portuguese version since the complaints are written in this language.

After preprocessing the complaints we were able to identify some entities on the texts, although not all of them were identified, our main goal with this model was to identify the complaints that involved two types of entities which can be more useful for the company to understand the complaints and to which type of store and location they refer to.

The last part of the code identifies the first five complaints that involve "ORG" which means an organization and it is possible to analyze that most of these refer to online shopping and delays in delivery, where the customers refer a lot to the shipping companies. Another category of entities that were identified and proved useful for the analysis was the "LOC" or location, which demonstrates again as a common subject the deliveries of products bought online.

This analysis was a demonstration of how the brand should iteratively conduct this process to keep track of the evolution on the number of complaints related to each location, which other companies are mentioned in such complaints, and understand where the customer experience is not fulfilling the expectations from the buyers.

4.5 Topic Modelling

Topic modeling is an unsupervised method of clustering the customers based on what they say in their complaints, in this case. As showed in class, we applied the LDA algorithm to cluster the customers on their reviews. First, we used the function *"compute_coherence_values(dictionary, corpus, texts, limit, start=2, step=3): "* from the notebook examples to find the optimal number of subjects from our document. This method plots the coherence score from the number of k selected for us to pick the optimal one, like what happens in K-Means.

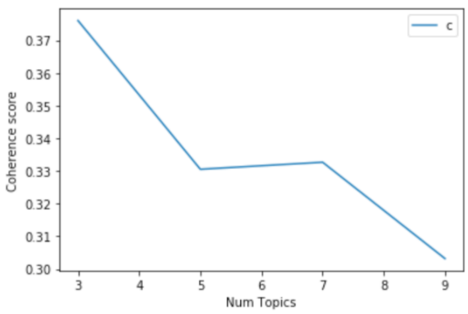

Figure 13 - Topic Modeling best k selection

We ran the LDA model to calculate its coherence with k being [3, 5, 7, 9], having the best score on k=3. With that solved, we used "genism" to plot the clusters and analyze their most relevant terms.

4.5.1 Cluster 1: Average Online Client

The first cluster kind of represents the average complaints from the analysis. Besides that, it's the only one that highlights the word "entrega" and has most of the mentions from "online" of the document, so we can assume most of the complaints are about their online purchases and deliveries.

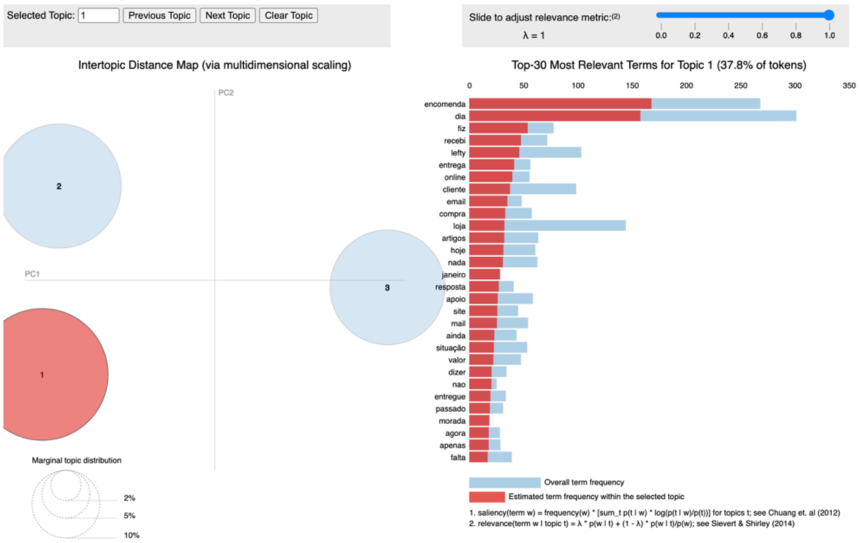

Figure 14 - Topic Modelling: Cluster 1

4.5.2 Cluster 2: Returning Items Complaints

At cluster two we can spot some interesting situations, for example, it is the cluster that holds most of the mentions of "dinheiro", "€", "devolver, or devolução". That can be cluster as clients that might need support to return unwanted items purchased. Funny that some of them mentioned store ("loja"), so it might happen that customer is having issues returning online items on stores or items bought in stores, which would need further analysis.

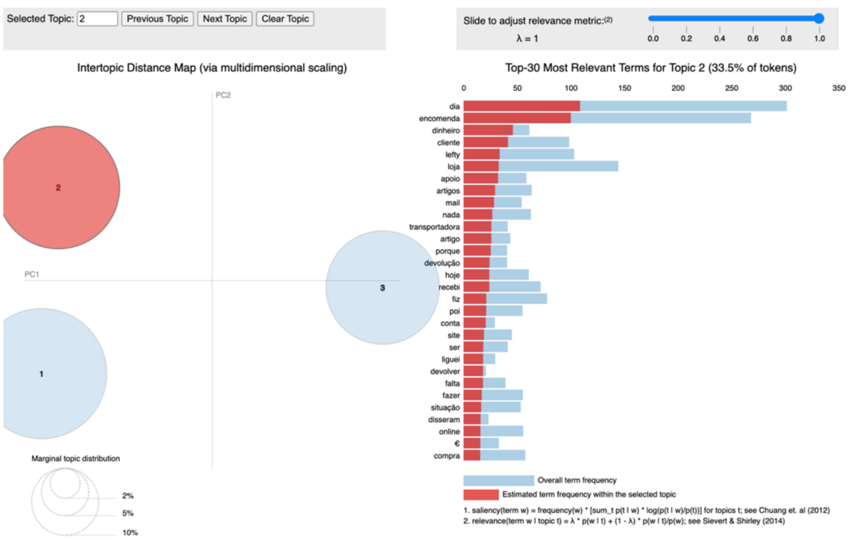

Figure 15 - Topic Modelling: Cluster 2

4.5.3 Cluster 3: Complaints from Store

By analyzing the terms from the most relevant words from the third group, it was clear to us that this group belongs to people who had contact with some Lefties store. It holds half of the store ("loja") mentions and it is the only one who highlights the employees ("funcionária") on its terms. These reviews could be used to improve the customer experience at the store and therefore for future employee training and other practices.

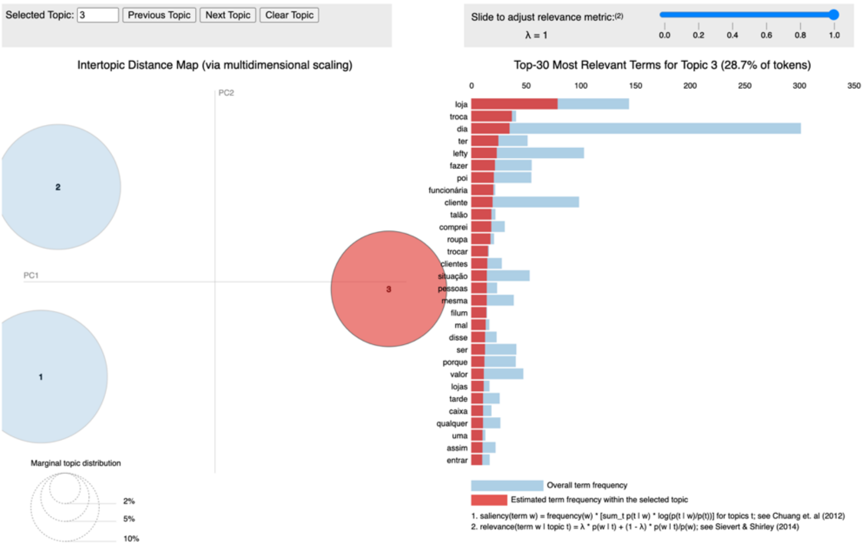

Figure 16 - Topic Modelling: Cluster 3

5 Evaluation

5.1 Evaluate results

Every failure is an opportunity to learn and become better. Therefore we believe that the ability of a company to deal with challenging situations says more about them rather than when receiving compliments. Lefties have a reputation for bad service and our main goal was to understand the biggest weaknesses and also understand how they deal with them. After analyzing the Lefties page on the national complaint portal and performing several techniques on the complaints, we came to the following conclusions.

Firstly we managed to web scrap a total of 177 complaints which formed our database. Out of this, 62% were unresolved, which already shows a point of improvement for the brand. It is also important to mention that we can split the complaints into two groups: pre-pandemic physical store complaints and also a post-pandemic scenario where the complaints were referring to gaps in the delivery services. The first ones are related directly with Lefties and in the second one, they are intermediate.

We then proceeded to get further insights on the complaints by performing some text mining techniques on the data such as normalization, steaming, tokenization, and removing stop words and special characters. This will provide a clearer text input to analyze as it summarizes the simplest meaning of present words.

After the clean data, we have performed different modeling techniques, and here is a summary of the main conclusions. By using n-grams and word frequency, we understood that the main topics and queries mentioned in the complaints are about people who have made a purchase but have failed to receive it. Online is also a key keyword that we highlight. Additionally, we also identify many complaints regarding customer support.

From Sentiment analysis, we can see that the distribution is skewed towards the negative sentiments which are expected as they are complaints. Another important insight is that the most negative complaints seem to have been resolved by Lefties.

Finally, we have divided the complaints into three main clusters: 1 for the online shoppers, 2 for the return difficulties, and 3 for physical store issues. We believe this segmentation is very valuable to Lefties as they can address the different topics using different strategies.

In-store complaints are normal for every business and are also the ones that Lefties can control better. Nevertheless, they are not having a high response rate or trying to offer solutions. We suggest an approach where they acknowledge the situation and apologize. Then can make it up by offering for example a discount.

Complaints from the online store and consequent delivery are sometimes not Lefties' fault. Nevertheless, the transport company is representing them and they should still acknowledge the issue and reply to the complaints.

Their customer support and return policy is also a big highlighted issue that they could work on. Also, customer support includes addressing the clients and answering their complaints. Overall, Lefties has a lot of room to improve. All companies make mistakes, but dealing with them is one step closer to success.