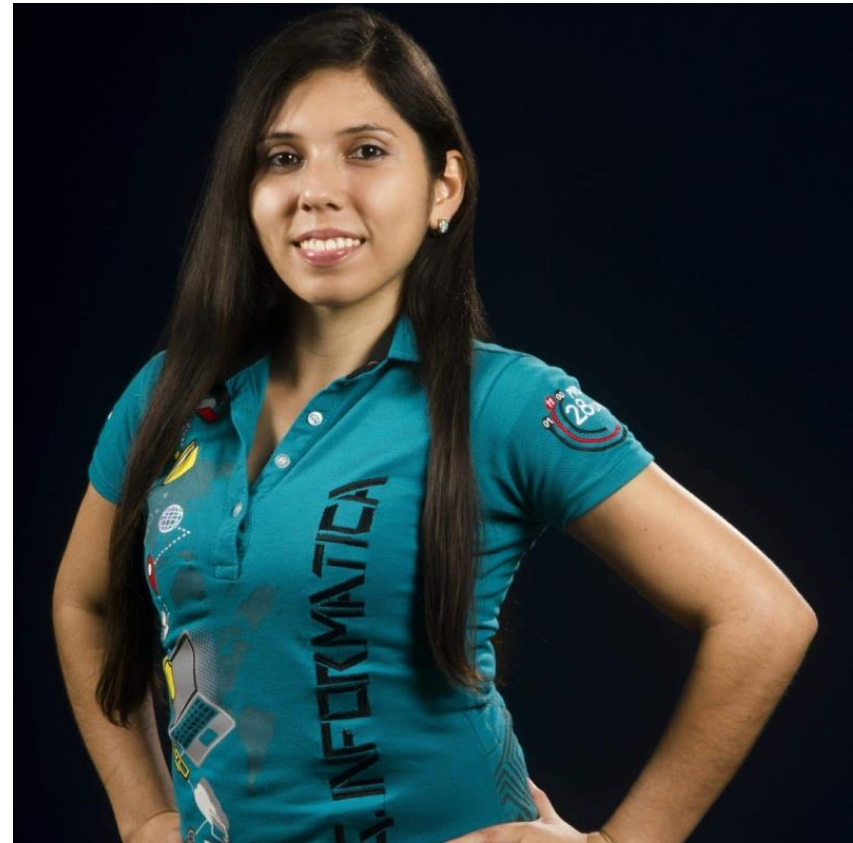


**Hello Hello**

**#PyConCo2020**

# Angelica Landazabal

- Computer Science Engineering and a technology lover. (Karateka too)
- I like the development of web applications, marketing and, of course, data analysis ;)
- It motivates me to continue learning about various IT topics, personal / professional growth and helping others so they are also motivated to learn.
- "Keep going" is my motto :)



✉ [angelica.informatik@gmail.com](mailto:angelica.informatik@gmail.com)

🐙 [ALandazabal](#)

🐦 [@ALandazabal15](#)

django girls  
Colombia





**SciPy**  
LATAM • 2019

7ª CONFERENCIA  
LATINOAMERICANA DE  
PYTHON CIENTÍFICO



OCT 8 - 10, 2019 • BOGOTÁ, COLOMBIA • [CONF.SCIPLYA.ORG](http://CONF.SCIPLYA.ORG)



**Lindsey Heagy**



**Travis Oliphant**



**Damián Avila**



<https://www.youtube.com/channel/UClb88lwUvIFikmhTzVGsVGA/featured>

# **"Applying sentiment analysis to your social network"**

What is Sentiment Analysis?

# Objectives:

- Understand What is Natural Language Processing y which are its applications.
- Extract data from a Social Network.
- Cleaning the data extracted.
- Apply the Sentiment Analysis technique to 2 sets of data (Spanish - English).
- Show data processed.

# What will we need?



colab



<https://github.com/ALandazabal/PyConCo2020-workshop>



# Natural Language Processing (NLP )

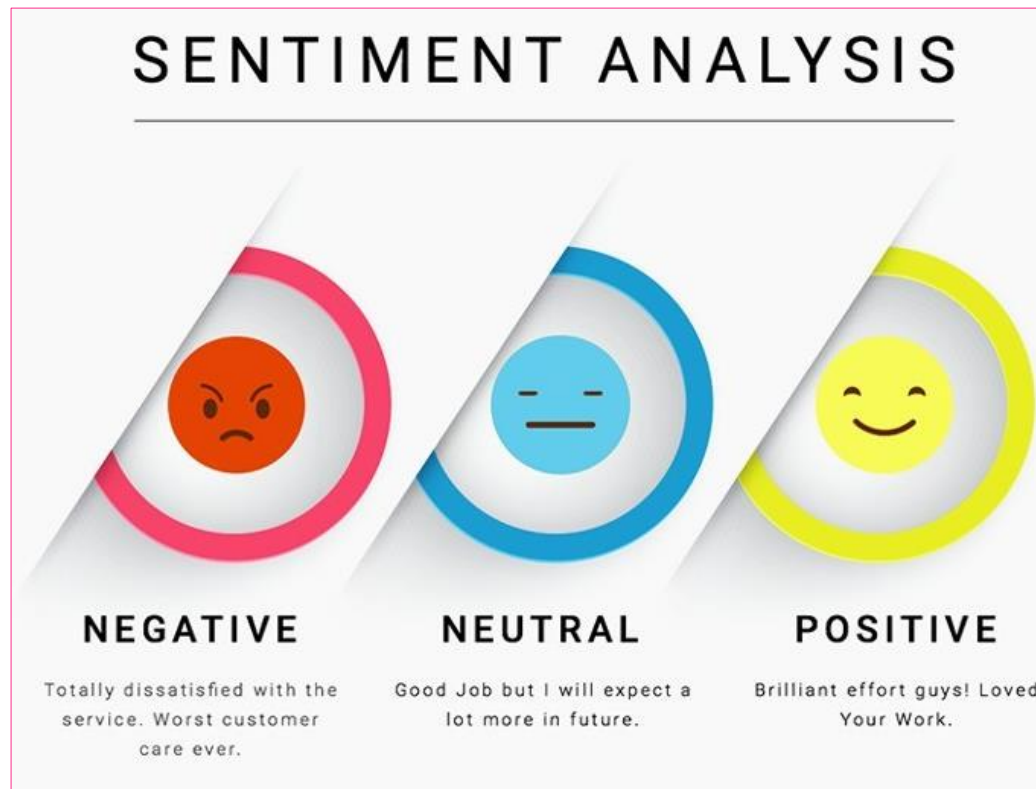
Is the study that it try to understand, analyze, interpret and obtain mean of human language with a computer to will be used in data science, AI and linguistics.

The main idea is perceive of a set of words the intention of its creator.

# NLP applications:

- **Translation:** It is the most used application because it allows you to understand more quickly the text written in another language.
- **Automatic summarization:** the idea is get a reduced version of a text created by extraction or abstraction.
- **Chatbots:** A system capable of having a coherent conversation with a human about a specific topic.

- **Sentiment Analysis:** Identify subjective information in texts as judgments or opinions used for companies and famous people that want to be aware of their reputation on the internet.

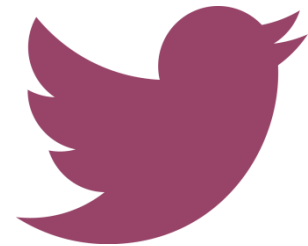


# Steps for Sentiment Analysis:

# Step 1: Extract Data

- **Social Network:** Twitter

Twitter is what is happening in the world and what people are talking about at the moment. You can access Twitter on the Web or from your mobile device.





# Step 1: Extract Data

- **API Twitter:** Twitter provides companies, developers and users programmatic access to data through its API (application programming interface).

You should be a twitter developer account and create an app to request the access tokens.

**Note:** We only access public information

# Step 1: Extract Data

Twitter Developer Labs is live: [Learn more >](#)

## Use cases

### Advertise

---

Programmatically create and manage Twitter ad campaigns.

[Learn more >](#)

### Publish and curate

---

Tell great stories with Twitter content.

[Learn more >](#)

### Analyze

---

Evaluate Twitter data to inform business decisions.

[Learn more >](#)

### Engage

---

Create connections with conversational experiences.

[Learn more >](#)



<https://developer.twitter.com/en.html>

# Step 1: Extract Data

- **Tweepy:** An easy-to-use Python library for accessing the Twitter API.

**Install with :** `pip install tweepy`

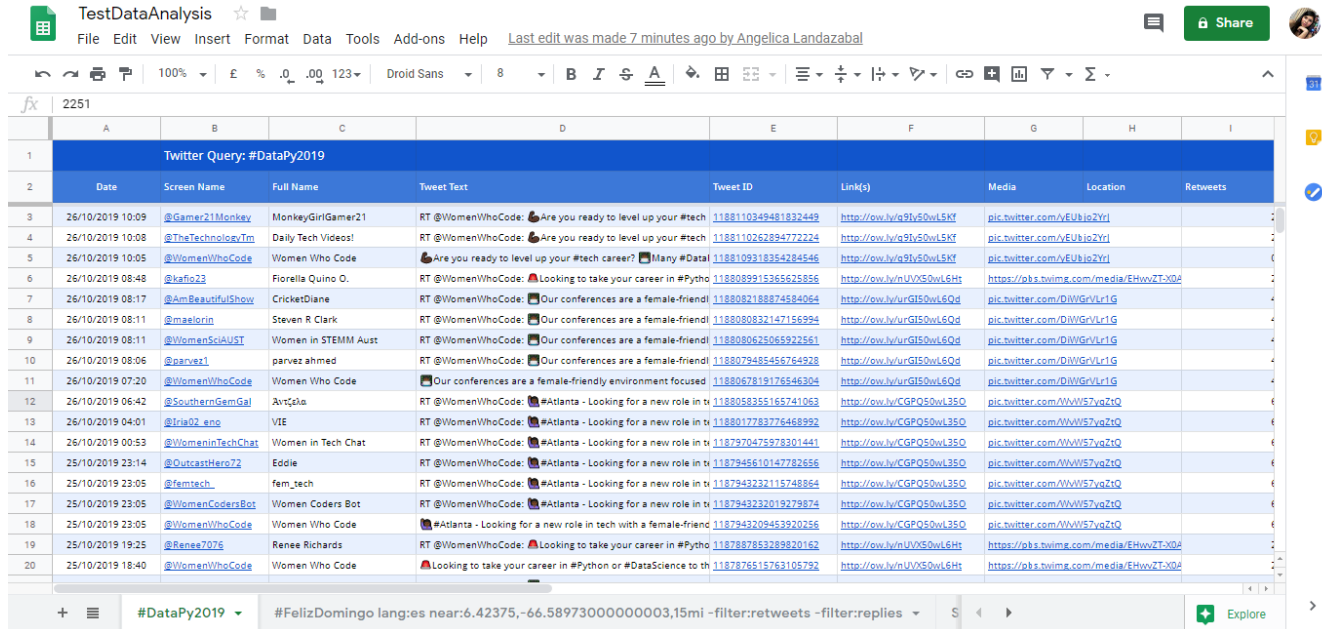
- **xlsxwriter:** Is a Python module for creating Excel XLSX files.

**Install with :** `pip install XlsxWriter`



# Step 1: Extract Data

- **Twitter Archiver:** is an Add-ons in Drive Spreadsheet help us to extract data of twitter with the parameters given.



	A	B	C	D	E	F	G	H	I
1	Twitter Query: #DataPy2019								
2	Date	Screen Name	Full Name	Tweet Text	Tweet ID	Link(s)	Media	Location	Retweets
3	26/10/2019 10:09	@Gamer21Monkey	MonkeyGirGamer21	RT @WomenWhoCode: 🐼 Are you ready to level up your #tech	1188110349481832449	<a href="http://ow.ly/q9ly50wl5MY">http://ow.ly/q9ly50wl5MY</a>	<a href="pic.twitter.com/yEUBja2YrI">pic.twitter.com/yEUBja2YrI</a>		
4	26/10/2019 10:08	@TheTechnologyTm	Daily Tech Videos!	RT @WomenWhoCode: 🐼 Are you ready to level up your #tech	1188110262894722224	<a href="http://ow.ly/q9ly50wl5MY">http://ow.ly/q9ly50wl5MY</a>	<a href="pic.twitter.com/yEUBja2YrI">pic.twitter.com/yEUBja2YrI</a>		
5	26/10/2019 10:05	@WomenWhoCode	Women Who Code	🐼 Are you ready to level up your #tech career? 🐼 Many #Data	1188109318354284546	<a href="http://ow.ly/q9ly50wl5MY">http://ow.ly/q9ly50wl5MY</a>	<a href="pic.twitter.com/yEUBja2YrI">pic.twitter.com/yEUBja2YrI</a>		
6	26/10/2019 08:48	@kaflo23	Fiorella Quino O.	RT @WomenWhoCode: 🐼 Looking to take your career in #Pytho	1188089915365625856	<a href="http://ow.ly/nUvXS0wl6Ht">http://ow.ly/nUvXS0wl6Ht</a>	<a href="https://pbs.twimg.com/media/EHwvZT-X04">https://pbs.twimg.com/media/EHwvZT-X04</a>		
7	26/10/2019 08:17	@AmBeautifulShow	CricketDiane	RT @WomenWhoCode: 🐼 Our conferences are a female-friendl	1188082188874584064	<a href="http://ow.ly/urGI50wl6Qd">http://ow.ly/urGI50wl6Qd</a>	<a href="pic.twitter.com/DlWGrVlr1G">pic.twitter.com/DlWGrVlr1G</a>		
8	26/10/2019 08:11	@maslorin	Steven R. Clark	RT @WomenWhoCode: 🐼 Our conferences are a female-friendl	1188080832147156394	<a href="http://ow.ly/urGI50wl6Qd">http://ow.ly/urGI50wl6Qd</a>	<a href="pic.twitter.com/DlWGrVlr1G">pic.twitter.com/DlWGrVlr1G</a>		
9	26/10/2019 08:11	@WomenSciAUST	Women in STEMM Aust	RT @WomenWhoCode: 🐼 Our conferences are a female-friendl	1188080635065922561	<a href="http://ow.ly/urGI50wl6Qd">http://ow.ly/urGI50wl6Qd</a>	<a href="pic.twitter.com/DlWGrVlr1G">pic.twitter.com/DlWGrVlr1G</a>		
10	26/10/2019 08:06	@parvez1	parvez ahmed	RT @WomenWhoCode: 🐼 Our conferences are a female-friendl	1188079485456764928	<a href="http://ow.ly/urGI50wl6Qd">http://ow.ly/urGI50wl6Qd</a>	<a href="pic.twitter.com/DlWGrVlr1G">pic.twitter.com/DlWGrVlr1G</a>		
11	26/10/2019 07:20	@WomenWhoCode	Women Who Code	🐼 Our conferences are a female-friendly environment focused	1188067819176546304	<a href="http://ow.ly/urGI50wl6Qd">http://ow.ly/urGI50wl6Qd</a>	<a href="pic.twitter.com/DlWGrVlr1G">pic.twitter.com/DlWGrVlr1G</a>		
12	26/10/2019 06:42	@SouthernGemGal	Avtçila	RT @WomenWhoCode: 🐼 #Atlanta - Looking for a new role in ti	1188058355165741063	<a href="http://ow.ly/CGPQ50wl35Q">http://ow.ly/CGPQ50wl35Q</a>	<a href="pic.twitter.com/WVW57vaZtQ">pic.twitter.com/WVW57vaZtQ</a>		
13	26/10/2019 04:01	@Fris92_ene	VIE	RT @WomenWhoCode: 🐼 #Atlanta - Looking for a new role in ti	1188017783776468992	<a href="http://ow.ly/CGPQ50wl35Q">http://ow.ly/CGPQ50wl35Q</a>	<a href="pic.twitter.com/WVW57vaZtQ">pic.twitter.com/WVW57vaZtQ</a>		
14	26/10/2019 00:53	@WomenInTechChat	Women in Tech Chat	RT @WomenWhoCode: 🐼 #Atlanta - Looking for a new role in ti	1187970475978301441	<a href="http://ow.ly/CGPQ50wl35Q">http://ow.ly/CGPQ50wl35Q</a>	<a href="pic.twitter.com/WVW57vaZtQ">pic.twitter.com/WVW57vaZtQ</a>		
15	25/10/2019 23:14	@OutcastHero72	Eddie	RT @WomenWhoCode: 🐼 #Atlanta - Looking for a new role in ti	1187945610147782656	<a href="http://ow.ly/CGPQ50wl35Q">http://ow.ly/CGPQ50wl35Q</a>	<a href="pic.twitter.com/WVW57vaZtQ">pic.twitter.com/WVW57vaZtQ</a>		
16	25/10/2019 23:05	@femtech	fem_tech	RT @WomenWhoCode: 🐼 #Atlanta - Looking for a new role in ti	1187943232115748864	<a href="http://ow.ly/CGPQ50wl35Q">http://ow.ly/CGPQ50wl35Q</a>	<a href="pic.twitter.com/WVW57vaZtQ">pic.twitter.com/WVW57vaZtQ</a>		
17	25/10/2019 23:05	@WomenCodersBot	Women Coders Bot	RT @WomenWhoCode: 🐼 #Atlanta - Looking for a new role in ti	1187943232019278874	<a href="http://ow.ly/CGPQ50wl35Q">http://ow.ly/CGPQ50wl35Q</a>	<a href="pic.twitter.com/WVW57vaZtQ">pic.twitter.com/WVW57vaZtQ</a>		
18	25/10/2019 23:05	@WomenWhoCode	Women Who Code	🐼 #Atlanta - Looking for a new role in tech with a female-friend	1187943209453920256	<a href="http://ow.ly/CGPQ50wl35Q">http://ow.ly/CGPQ50wl35Q</a>	<a href="pic.twitter.com/WVW57vaZtQ">pic.twitter.com/WVW57vaZtQ</a>		
19	25/10/2019 19:25	@Renee7076	Renee Richards	RT @WomenWhoCode: 🐼 Looking to take your career in #Pytho	1187887853289820162	<a href="http://ow.ly/nUvXS0wl6Ht">http://ow.ly/nUvXS0wl6Ht</a>	<a href="https://pbs.twimg.com/media/EHwvZT-X04">https://pbs.twimg.com/media/EHwvZT-X04</a>		
20	25/10/2019 18:40	@WomenWhoCode	Women Who Code	🐼 Looking to take your career in #Python or #DataScience to th	1187878615763105792	<a href="http://ow.ly/nUvXS0wl6Ht">http://ow.ly/nUvXS0wl6Ht</a>	<a href="https://pbs.twimg.com/media/EHwvZT-X04">https://pbs.twimg.com/media/EHwvZT-X04</a>		



<https://docs.google.com/spreadsheets/d/1tUilV5Y-yWivYHnLw4zWh7DEKDCC7DNVbtqPXEnZU/edit?usp=sharing>

# Step 1: Extract Data

- **Facebook-Scraper:** Scrape Facebook public pages without an API key. Inspired by twitter-scraper.

**Install with :** `pip install facebook-scraper`





# Step 2: Load Data

- **Pandas:** Powerful data structures for data analysis, time series, and statistics.

**Install with :** `pip install pandas`

- **NLTK:** The Natural Language Toolkit (NLTK) is a Python package for natural language processing.

**Install with :** `pip install nltk`



# Step 3: Clean Data

- **String:** contains some constants, utility function, and classes for string manipulation.  
**Install with :** `pip install strings`
- **Re:** This module provides regular expression matching operations similar to those found in Perl.  
**Install with :** `pip install re`
- **Emoji:** The entire set of Emoji codes as defined by the unicode consortium is supported in addition to a bunch of aliases.  
**Install with :** `pip install emoji`



# Objectives:

- Understand What is Natural Language Processing y which are its applications. ✓
- Extract data from a Social Network. ✓
- Cleaning the data extracted. ✓
- Apply the Sentiment Analysis technique to 2 sets of data (Spanish - English).
- Show data processed.

# Step 4: Process Data

- **Naive Bayes:** It is a classification and prediction technique that constructs models to determine what size is the possibility in which a set of data originates certain results..

$$P(\text{Clase} \mid \text{Datos}) = \frac{P(\text{Datos} \mid \text{Clase}) * P(\text{Clase})}{P(\text{Datos})}$$

- **Bag of word:** It is a mathematical representation of the testing data set. This has all the unique words of a document and the occurs frequency of each one.



# Step 4: Process Data

- **Textblob:** It provides a consistent API for diving into common NLP tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, and more.

## Sentiment Analysis returns two values:

**Polarity:** float which lies in the range of  $[-1,1]$  where 1 means positive statement and -1 means a negative statement.

**Subjectivity:** Subjective sentences generally refer to personal opinion. Subjectivity is also a float which lies in the range of  $[0,1]$ .

**Install with:** `pip install textblob`





# Step 5: ShowData

- **Matplotlib:** is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms.  
**Install with:** `pip install matplotlib`
- **WordCloud:** This tool will be quite handy for exploring text data and making your report more lively.  
**Install with:** `pip install wordcloud`



# Objectives:

- Understand What is Natural Language Processing y which are its applications. ✓
- Extract data from a Social Network. ✓
- Cleaning the data extracted. ✓
- Apply the Sentiment Analysis technique to 2 sets of data (Spanish - English). ✓
- Show data processed. ✓

# Use cases of Sentimental Analysis

- Brand monitoring
- Competitive research
- Flame detection and customer service prioritization
- Product analysis
- Market research and insights into industry trends
- Workforce analytics/employee engagement monitoring

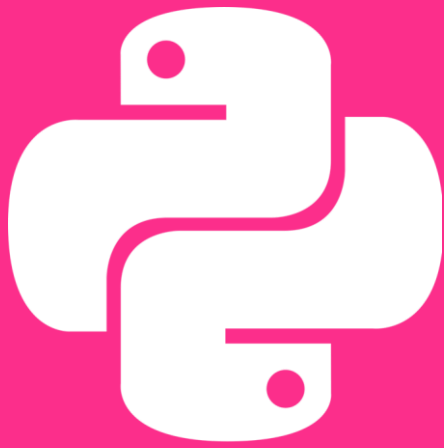
# Sources

- **The Definitive Guide to Natural Language Processing :** <https://monkeylearn.com/blog/definitive-guide-natural-language-processing/>
- **API Twitter:** <https://developer.twitter.com/>
- **Jupyter Notebook:** <https://jupyter.org/>
- **Excel files with python:** <https://xlsxwriter.readthedocs.io/>
- **Tweepy Library:** <https://www.tweepy.org/>
- **Pandas:** <https://pandas.pydata.org/>
- **NLTK:** <https://www.nltk.org/>
- **Re:** <https://docs.python.org/3/library/re.html>

# Sources

- **String module:** <https://www.journaldev.com/23788/python-string-module>
- **TextBlob:** <https://textblob.readthedocs.io/en/dev/>
- **Matplotlib:** <https://matplotlib.org/>
- **WordCloud:**  
<https://www.datacamp.com/community/tutorials/wordcloud-python>
- **Sentiment Analysis: Types, Tools and Use Cases:**  
<https://www.altexsoft.com/blog/business/sentiment-analysis-types-tools-and-use-cases/>





**“Las personas mienten,  
los datos no”  
PlatziConf 2019**



**angelica.informatik@gmail.com**



**ALandazabal**



**@ALandazabal15**