# APIs and Language Processing with Python for Twitter

**Objectives**

The aim of this talk is to give a fun and interesting introduction to interacting with APIs in Python, specifically the Twitter REST API. Additionally, using Twitter data this talk also aims to show introductory language processing via neural networks. I'd like to introduce the libraries Tweepy, Carmen, and TextBlob to give a comprehensive introduction to these topics.

**Description**

Have you wanted to know in real time what everyone's saying about anything—ranging from the Amazon fires to the latest iPhone release? Fear not, Twitter is the key. Learn about engaging with the Twitter API and using language processing as a tool to discover interesting patterns about your data! In this talk, I'll cover the concept of the Twitter REST API, along with a new geolocating tool developed for Python, and lastly, a simple way to analyse tweet sentiments.

**Abstract**

When approaching APIs and language processing, the amount of information is overwhelming. You can get lost in article after article and come out with no real knowledge about the topic.

With this talk, I would like to change that at least by a small margin. I'd like to provide a simple, easy to follow, comprehensive yet engaging introduction into an exciting field with countless applications from sentiment analysis, to universal translators.

**Outline**

1.  Introduction and Background (10 mins)

In this section I will give a brief background about me and the problem I intended to solve along with its application

*About Me:* I am a second year student at Ecole Polytechnique within the Bachelor program following a double major in Mathematics and Computer Science. Over the summer, I worked as a Data Science research intern which led to some really cool projects!

*The Problem:*
1) Research that used my code: Aiming to answer if social movements on twitter led to any significant political changes [with Gilets Jaunes and #MeToo]
2) What I did: Create a Twitter Penetration Rate calculator that given a time period and region will give back the concentration of Twitter usage.

2.  API Streaming (5 mins)

Here, I will show how basic real time streaming works with Tweepy and all the possible parameters we could use. Potentially with a live example.

- GitHub link: https://github.com/sonalsannigrahi/TwitterLive/blob/master/livestream.py
- Will cover: language, geobox, trackword, user, hashtag

3.  Carmen and Geolocation (7 mins)

Once you are streaming tweets, an interesting data set to collect is where the tweets are coming from.

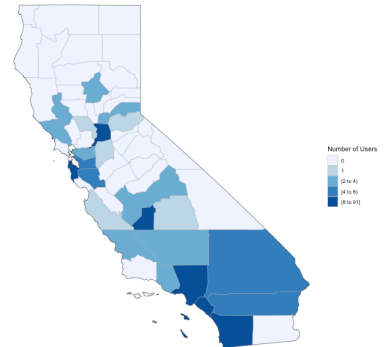*Problem:* Twitter doesn't always give tweet/user location as by default it is set to Inactive.

*Solution:* Carmen: A new python library

- Dictionary Search and Mapping with Carmen
- Example of Geolocated Data
- Some visualisations in R

US Twitter Usage at the County Level between start and end
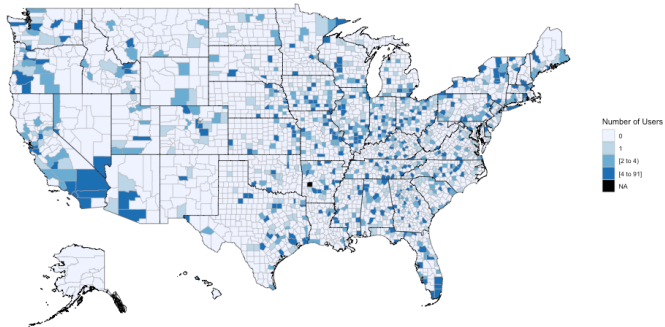


US Twitter Usage at the County Level (California) between start and end



- Accuracy and Reliability of Carmen

4. Basic Language Processing (8 mins)

Here, I will discuss how to create a simple Tweet sentiment analyser.
- The Task: Natural Language Processing to determine a Positive (+1), Neutral (0), Negative(-1) sentiment
- Simple Naïve Bayes Classifier
- TextBlob for visualisations

5. Q&A (5 minutes)

Extra time yielded to any questions the audience might have.