

Social Data Acquisition with Reddit API

Layla Bouzoubaa

7/11/23

Hello 🖐️

🎓 Doctoral student, IS

💊 NLP, social computing, drugs

🕒 Passionate about 🐱 👫 🍷 🍣 🦪 🍵

Outline

- Introduce API
- Setup Reddit API
- Reddit API with examples
- Play time

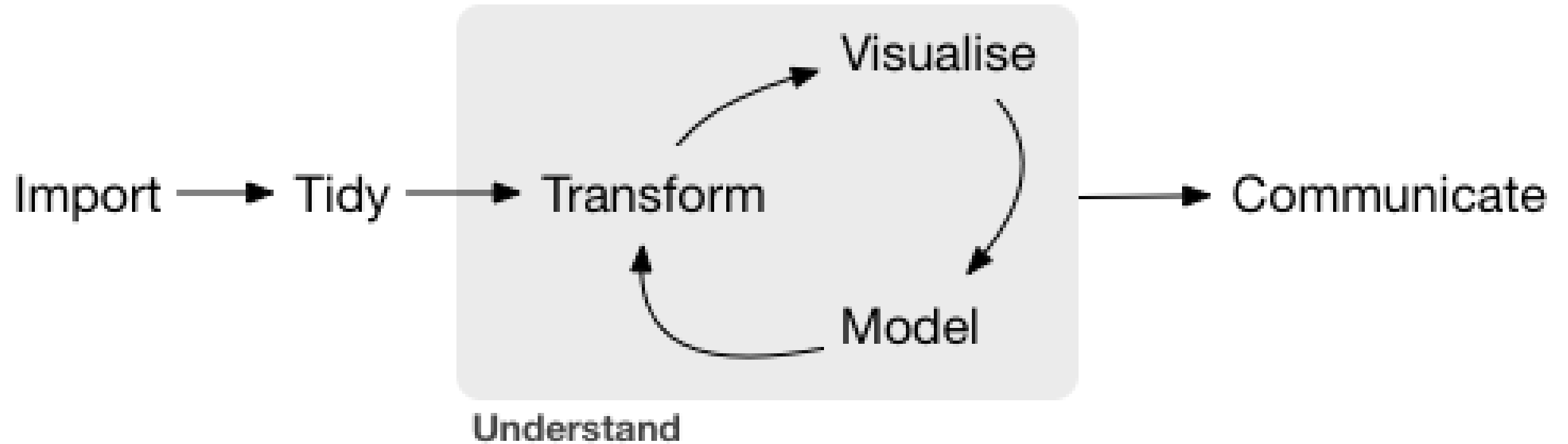
A meme featuring Woody and Buzz Lightyear from the movie Toy Story. Buzz is in the foreground, wearing his green and purple space suit, with his right arm raised and fingers spread. Woody stands behind him, looking on with a concerned expression. The background is a simple indoor setting with a door and a window.

DATA

LOTS AND LOTS OF DATA

makeameme.org

Typical Data Science Pipeline



Program

Reddit

- The “front page of the internet”
- Reddit has an API that allows developers to access and extract data from the platform.



What is an API?

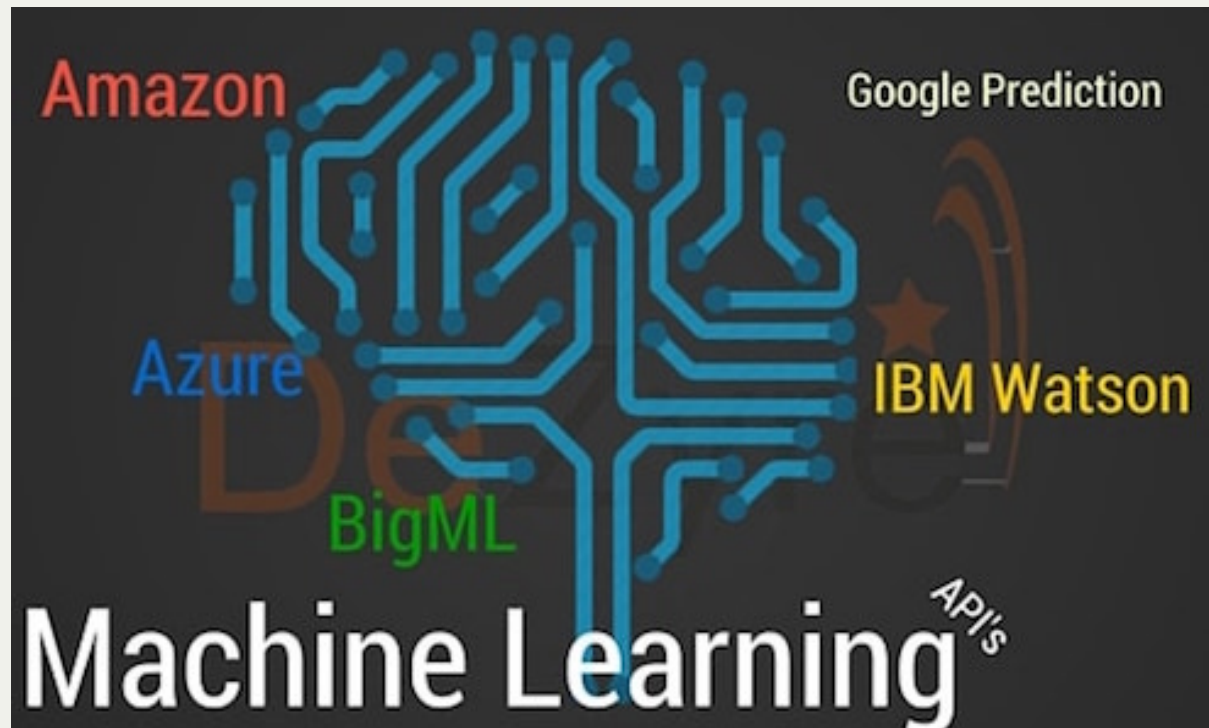
Application Program Interface

- Allows a user to programmatically pull data from a source given that source provides one
 - *ex. The NYT, Twitter, Facebook, Google, US Census*



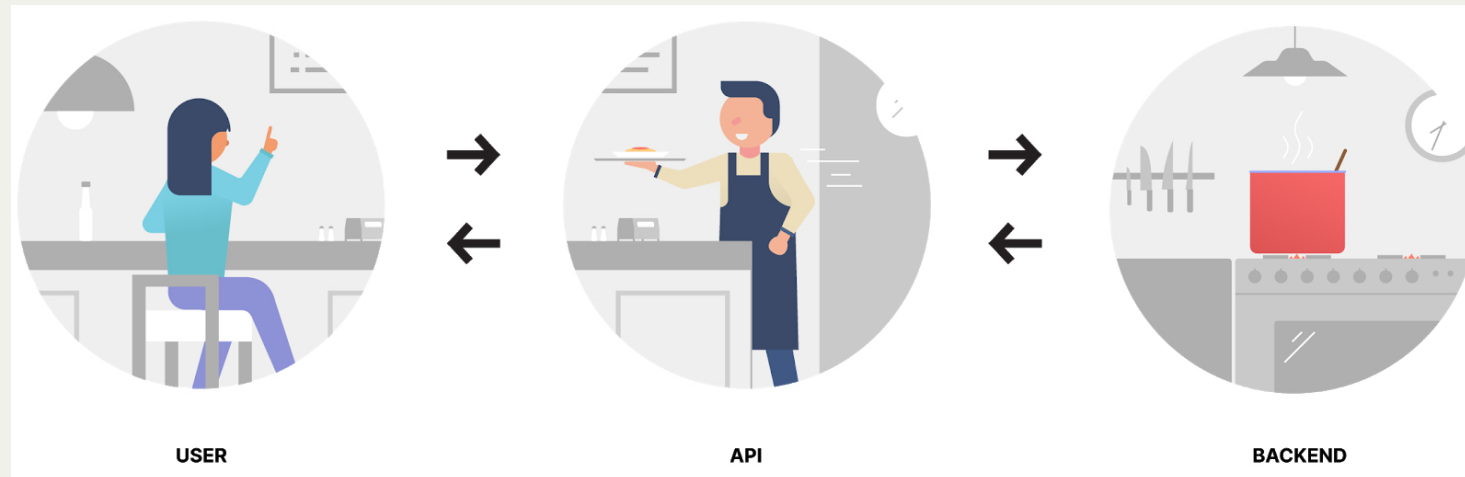
How Does it Work?

- Like a Graphical User Interface (GUI) allows you to interact with your code, an API lets your code interact with other code
- Set of commands, functions, objects, and protocols



Not Only for Getting Data

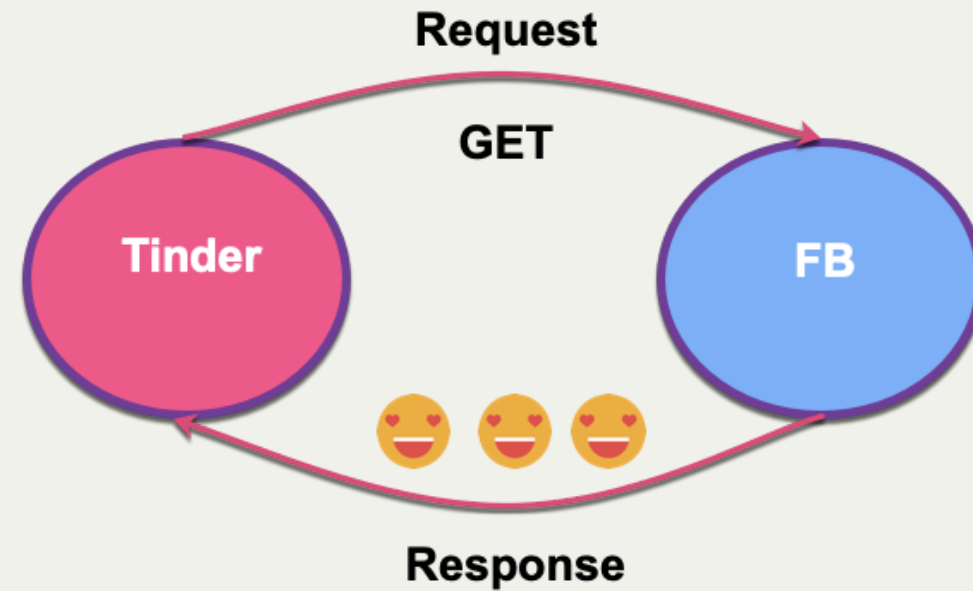
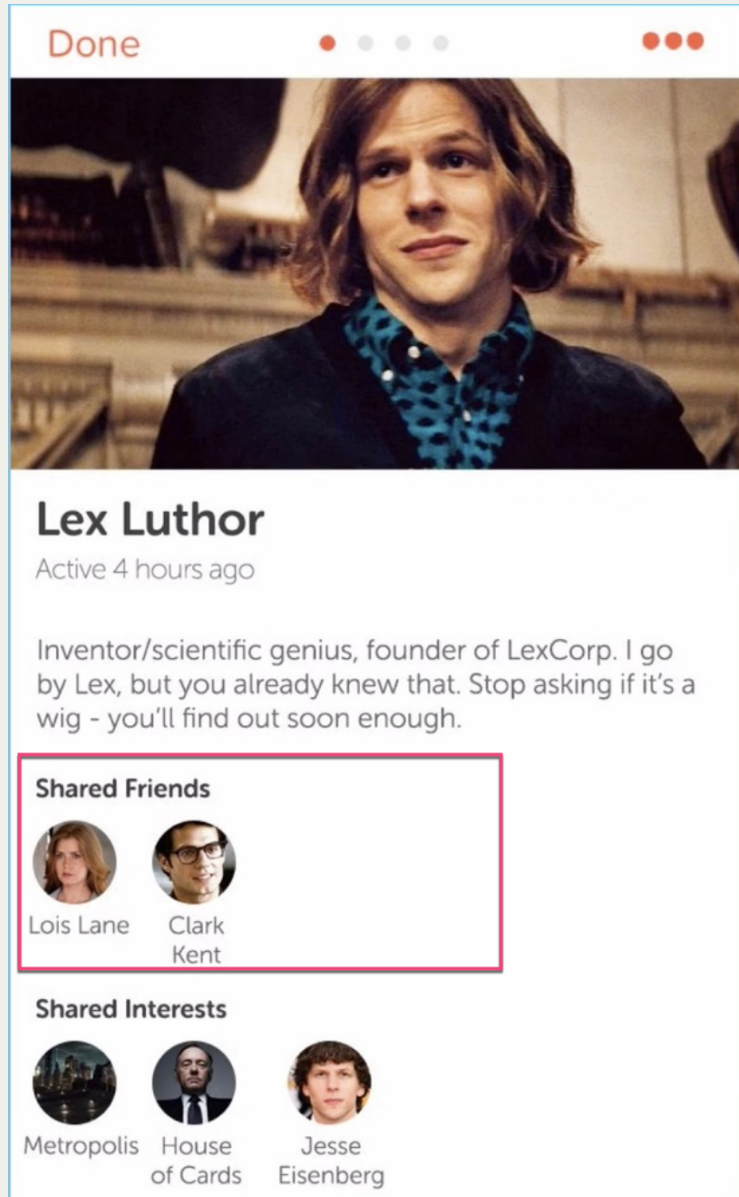
- Build APIs for your ML models so others can use them
- Write (post) to a database
 - and update (put) and delete (delete)
- Functions
 - more control over what is exposed to end user



The API

- **GET**: used to retrieve data from the server
- **POST**: send data to the server to create a new resource
- **PUT**: update an existing resource on the server
- **DELETE**: delete an existing resource on the server

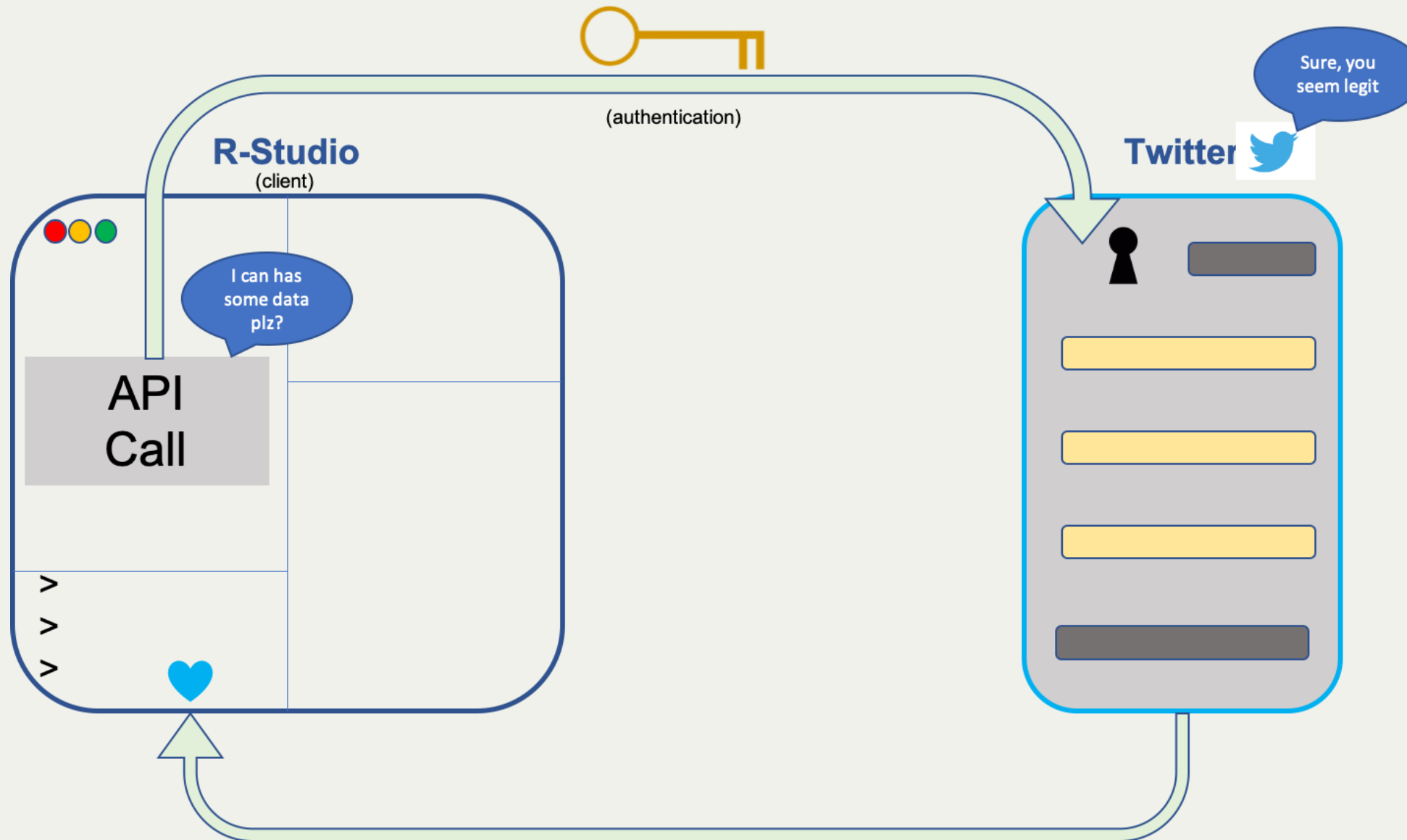
There is this App...



It's that easy?

- 1. You must first agree to the terms of use and obtain an API Key.
 - 1. This authenticates you to the server that holds the data
 - 2. Sometimes it's simple and sometimes it's a process...be prepared
- 2. Store this key in a local `.py` file
 - 1. DO NOT SHARE / PUBLISH YOUR API KEY!! 💀
- 3. Run your API call
 - 2. DO NOT SHARE / PUBLISH YOUR API KEY!! 💀

Keeper of the Keys



Understanding Reddit API

- Reddit API is a RESTful API that uses HTTP requests to communicate with the server.
- API endpoints are used to perform specific tasks such as getting information about a subreddit, submitting posts or comments, and searching for content.
- To access the Reddit API, we need to authenticate with our API key.

The Problem

Which Friends characters are discussed more?



The One with the Steps

Source: [reddit.com/r/friends_tv_show](https://www.reddit.com/r/friends_tv_show)

1. Obtain Reddit API Key
2. Install required libraries
 1. PRAW
 2. requests
3. Authenticate with Reddit
4. Do data science

```
pip install praw  
pip install requests
```


The One with the Notebook

See the notebook [here](#) to follow along

Resources

- What is an API?
- Reddit API Documentation

