

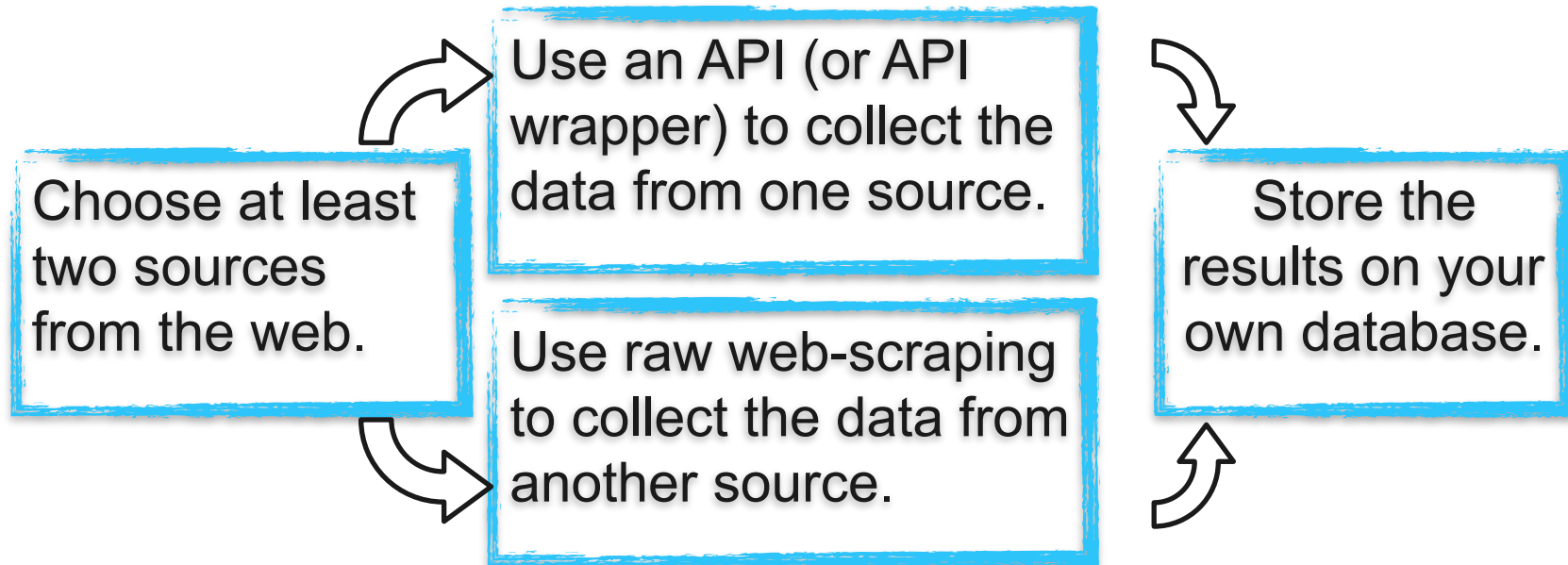


Project 2 - Data gathering processes

Web Scraping and Databases

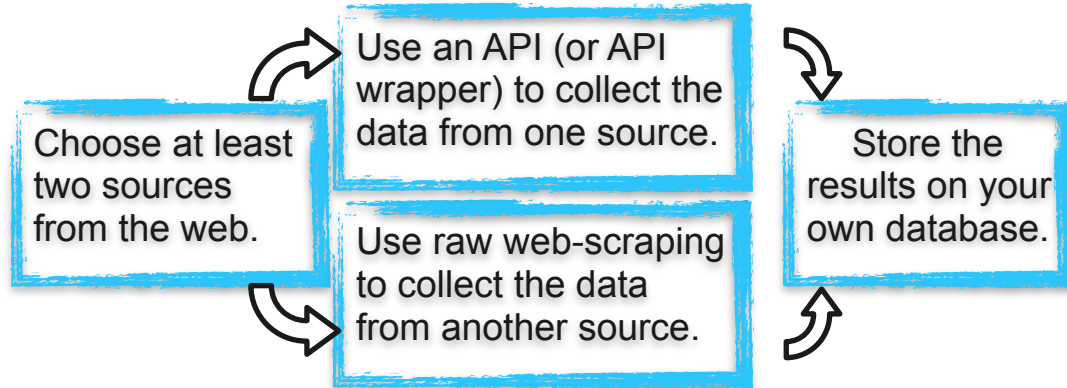
MARCH, 19 2020 | SAO

Main Objectives



Main Objectives

The main objective of this project is to train your skills on obtaining data from different sources and safely storing it in a database. This will also help your develop skills in designing good process workflows.



Specific Objectives

1. You'll create one code to generate structured data from an **API** (or API wrapper);
2. You'll create another code to generate structured data from **raw web-scraping** (using **beautiful-soup** or any related tool you prefer);



Specific Objectives

3. You'll design and create your database with the structured data obtained;
4. Don't forget to document the process.



Hints

- Keep in mind that you'll use this database later in the course. So it's best to choose something you are familiar with, something you are interested in.
- Specifically, you'll use the data you've chosen for the visualization project. So try to find something you have interest in seeing some graphics and statistics from.
- It would be lovely if you could relate both sources. That is, if you could later use both of your sources in the same context.

Deliverables

- The url of both of your Jupyter Notebook files on your GitHub (the API and the raw web-scraping codes).
- The dump of your database.
- [EXTRA] Draw the Entity-Relationship-Diagram of your database ([ERD](#))

Deadline

- 25/03/2020

