IRON
HACK

Project 3 - Data gathering processes

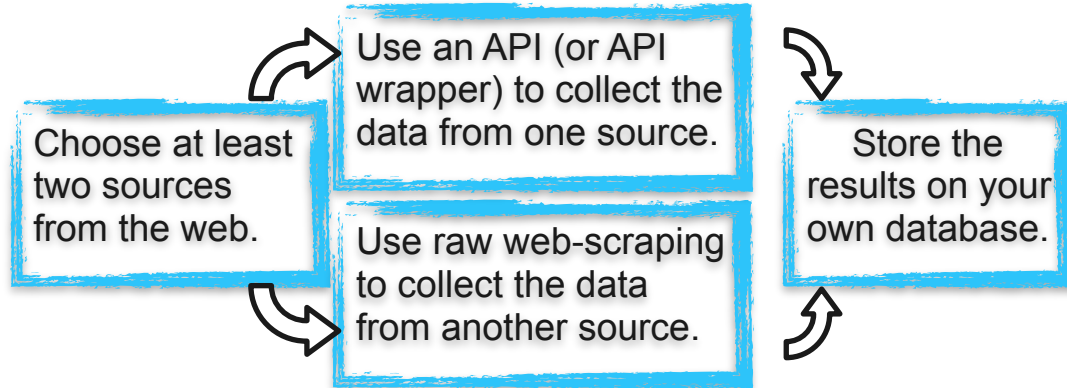Web Scraping and Databases

MARCH, 17 2020 | SAO

Main Objectives

Choose at least two sources from the web.

Use an API (or API wrapper) to collect the data from one source.

Use raw web-scraping to collect the data from another source.

Store the results on your own database.
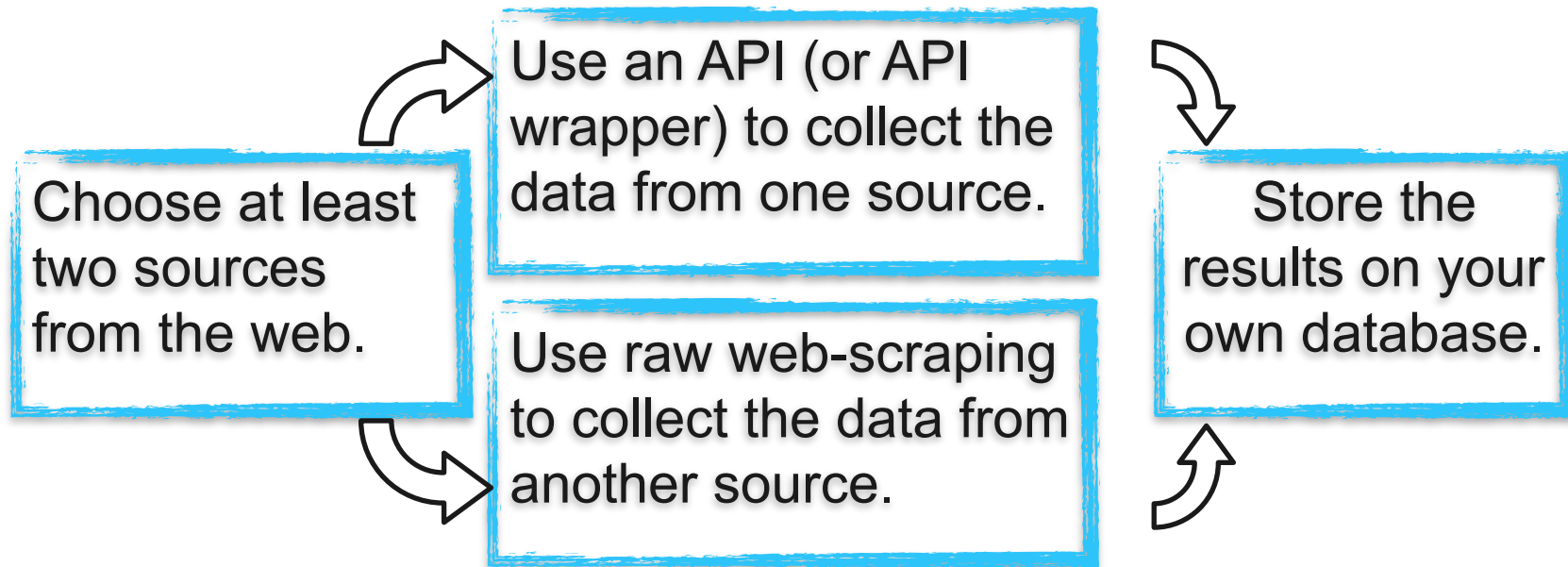
Main Objectives

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.

```
Choose at least
two sources
from the web.
  ↗
Use an API (or API
wrapper) to collect the
data from one source.
  ↘
Use raw web-scraping
to collect the data
from another source.
  ↗
Store the
results on your
own database.
  ↘
```

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