1. Explain the use case
2. Collect data. Explain why you choose this data, from these data sources
3. Clean data and execute exploratory data analysis (+vizualisation)
4. Choose the database type (compare several types and explain why)
5. Create an entity-relationship diagram (at least 4 entities)
6. Create a database (database, tables)
7. Add data to the database
8. Create 5 scripts showing the insights
9. Prepare the 10 pages report
10. Prepare presentation

## **Requirements & Deliverables**

Each student should upload all the project materials to Github.

You should deliver:

* Planning of your project in Trello/Jira,
* Code in Python for data collection and cleaning,
* ER model,
* data sources and metadata,
* database script,
* report (10 pages)
* slides.

The link to the Github repository and the report in pdf format should be sent on Friday, 03/06/2022til 10 PM.

## **Presentation**

The presentation time limit is 30 minutes. You will have 5 minutes for Q&A.

The slides of your presentation must include the content listed below and a demo of your project:

* Title of the project + Name
* Description of your Project (Planning, ER, Database Schema, Queries, Methodology)
* Challenges
* Process
* Highlights
* Main results
* Demo

Possible Data sources:

<https://opendata.paris.fr/pages/home/>

<https://data.gov.uk/>

<https://data.iledefrance.fr/pages/home-open-data/>

|  | Data Analytics |
| --- | --- |

Project Title

Name Surname

June, 2022

**Table of content**

**Introduction**

Business Use Case

Goal

Plan

**Data and data sources**

**Data collection**

**Data cleaning and Exploratory data analysis**

**Entities. ERD**