Validating your choice for the final project

To validate your choice, start by reading the document *project.pdf* carefully. Then, complete this template and submit the completed template as a pdf file (assignment "Validation of your choice for the final project").

Please specify the version of this document (V1 for initial proposal): V1

Please specify the changes made to the previous version to take my comments into account (if this is not the initial version of your proposal):

Before submitting, please check the following (tick the box for every point that you have checked). **Please make sure that you have checked everything in the check box below and that you have completely answered the three questions below.** Uncomplete proposals may not be approved, and I will ask you to complete the proposal and check all points below before I can approve. As a reminder, the project is individual.

1. You have specified who the decision maker is in this project (section 1 below).	☒
2. You have specified what problem(s) you plan to address / what decisions you plan to support (section 1 below).	☒
3. Your data are real and you combine different data.	☒
4. You have specified (section 2 below) what data you are going to use (tables/worksheets *and* attributes of the tables/columns of the worksheets), **and** you have provided the exact source of the different data (typically, exact URLs from which the data can be directly accessed).	☒
5. You have specified (section 3 below) how you will use Tableau (and any other tool or language) to support / make the decision(s).	☒
6. In your data, you have measures of interest to analyze, and these include measures that you can have an influence on as a decision maker.	☒
7. In your data, you have the information needed to analyze measures at different levels of detail (hierarchies).	☒
8. You have geographic data.	☒
9. You have temporal data (dates).	☒
10. You have data at sufficiently detailed level to deal with the problem addressed in your project. (For example, Statista may be an interesting data source, but it is typically not sufficient, as it provides aggregated data.)	☒
11. In your data, you have enough attributes (dimensions and measures, in Tableau vocabulary) to make the project interesting and the answer to the problem not self-evident.	☒
12. If you are using anonymized data, you will still be able to make enough interesting and relevant analyses with these data.	☒

1. Who is the decision maker in this project, and what problem(s) do you plan to address / what decisions do you plan to support?

The decision maker in this project is the mayor of Paris. He or she wants to understand what influences the prices of homes. The reason for that prices are a proxy for demand and he or she is planning to make less popular neighborhoods more appealing with targeted investments in infrastructure, cultural activities, renovations, etc.

2. What data will you use, from what sources? Please be specific (tables/worksheets, attributes/columns, exact sources of data).

1. Primary source: Airbnb listings in Paris
    a. Source :
        i. Inside Airbnb (https://insideairbnb.com/)
        ii. URL of data: http://data.insideairbnb.com/france/ile-de-france/paris/2023-03-13/visualisations/listings.csv
    b. Important columns:
        i. Name
        ii. Description
        iii. Neighbourhood
        iv. Latitude
        v. Longitude
        vi. Property
        vii. Room type
        viii. Accommodates (number of ppl staying)
        ix. Price

2. Primary source: Events in Paris
    a. Source:
        i. OpenData Paris
        ii. https://opendata.paris.fr/api/explore/v2.1/catalog/datasets/que-faire-a-paris-/exports/csv?lang=fr&timezone=Europe%2FBerlin&use_labels=true&delimiter=%3B
    b. Important columns:
        i. Name
        ii. Code postal
        iii. Longitude
        iv. Latitude
3. Secondary source: Population evolution by commune (arrondissement)
    a. Source:
        i. INSEE
        ii. https://www.insee.fr/fr/statistiques/3698339
    b. Important columns:
        i. Commune
        ii. Year
        iii. Population
4. Secondary source (optional): Social housing by département
    a. Source:
        i. OpenData Paris
        ii. https://opendata.paris.fr/explore/dataset/logements-sociaux-finances-a-paris/information/?disjunctive.code_postal&disjunctive.ville&disjunctive.bs&disjunctive.mode_real&disjunctive.nature_programme
    b. Important columns:
        i. Total number of flats per location
        ii. Departement
        iii. Year of financing
5. Secondary source: Welfare by département in 2017
    a. Source:
        i. INSEE
        ii. https://statistiques-locales.insee.fr/
    b. Important columns
        i. Commune
        ii. Niveau de vie winsorisé

6. Additional sources will be added if needed (more factors that influence the prices of homes)

3. How will you use Tableau (and any other tool or language, as appropriate) to support / make the decision(s)?

1. **Data Preparation:** To prepare the data I will use **Python** to clean and prepare the data. This will entail to create or remove columns and to clean fields that are not very pleasant.
2. **Data Visualization**: Then I will load the different tables into **Tableau** where I will create all my visualizations. For my topic the geographical visualizations of Tableau are particularly interesting. I will create Stories composed of multiple dashboards. The goal is that the decision maker (the mayor of Paris) is as well informed about the topic as possible. I will create maps, time series graphs,and other visualizations in Tableau.