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 FAO (Food and Agriculture Organization of the UN) and their goal is to "to achieve food security for all" (https://www.fao.org/about/en/). Their budget for 2018-19 was around 1 billion US $. In order to help countries, they first need to know where the world supply of food comes from.

I am analyzing the Food production per country and see how it evolves over time. Combined with data about the GDP and Greenhouse gas emissions, I will be able to make conclusions about what countries produce which kind of food, how this changed over time and what the environmental impact of it is. This information is crucial for the

FAO to better understand where food comes from and what its environmental impact is. Once they know that, they can make more informed decisions about which country to support or what technologies to foster where for example.

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

Primary source: Agricultural Production

- Source : Food and Agriculture Organization
  - https://www.fao.org/faostat/en/#data/QCL
- Columns :
  - Year : 1961-2021
  - Countries : All FAO-countries (around 160)
  - Item produced (different levels of granularity)

Secondary source: Agriculture emissions

- Source: Climate Watch (https://www.climatewatchdata.org/data-explorer/historical-emissions)
- Columns:
  - Country
  - Year (from 1990-2020)
  - All GHG emmisions for that year and country of the agricultural sector

Secondary source: GDP

- Source: World Bank (https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?end=2021&start=1960)
- Columns:
  - Country
  - Year (from 1990-2020)
  - GDP

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, to change their names, to convert units.
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 visualizations. I will be able to combine multiple sources of data to create enticing visualizations.