**Project Journal**

**This project was developed by Alberto Cruz and Itandehui Rivera. The problem that was addressed was the lack of housing in Ireland. They worked with 3 datasets in which the statistics of population growth, housing creation and homeless people were recorded.**

**02/12/2024-09/12/2024**

**Choose topic**

In the first week, everyone worked separately collecting information on topics of interest and from which we could find datasets with the necessary characteristics.

Some of the possible topics and the first problem we went through was cancer and and pollution but due that we couldn’t find dataset in the format we needed and with enough number of records were discarted

**Alberto Cruz and Itandehui Rivera**

**Objectives**Once the topic was chosen, homelessness in Ireland, the next thing was to identify the objectives of the project, what did we want to show? With the datasets available on the website, it was decided to make a contrast between population growth, new housing projects and homeless people.

**Itandehui Rivera**

**Finding References**

Once the objectives of the project were defined, the next thing was to obtain references, research and previous data that could explain the situation. Research was carried out in newspapers and different websites and the most relevant information was chosen to include in the project.

**Alberto Cruz**

**09/12/2024-16/12/2024**

**Project Report (first steps)**

With the selected investigations and reports, a summary and analysis of the information was made to write the abstract and the introduction of the report.

**Itandehui Rivera**

**Coding**

Another problem we faced was choosing the type of file we would use and how we would approach it, streaming, web scraped, etc.

We decided to work with JSON files since it is what we have the most experience with.

In this part we retrieve 2 dataset (population, housing expectations) and store the data in sqlit3

The data was also cleaned by identifying empty columns, converting data and everything necessary to be able to work with the information.

**Alberto Cruz (household formation) and Itandehui Rivera (Population expectations)**

**16/12/2024 -23/12/2024**

**Coding**

Working each one in one dataset we process and analize the data showing with diferrent grafics the informaion of most interest

Here we face some problems such as when we wanted to graph the columns they were not of the necessary data type or that there were boxes without values.

In the population dataset the numbers could not be recognized since they had a comma (,) so it was necessary to replace this symbol to later convert to float

Although it took us time in the end, the solution was simple, go to the preprocessing part and change the types of values ​​or fill in boxes with averages

**Alberto Cruz (household formation) and Itandehui Rivera (Population expectations)**

**Report**

Advance in the report according with advances in the code

**Alberto Cruz**

**02/01/2025- 05/01/2015**

**Coding**   
work on the third dataset The only problem we face here is that we did not find one with enough records, so a new one was created out of 3 datasets.

**Itandehui Rivera**

**Report**

Finish the report

**Alberto Cruz**

**Video**

Each member record a video explaining their parto the code

**References and Cover**

Give format to the references and put the cover in the report

**Itandehui Rivera**