**Introduction**

In this project, executed in a bootcamp in Ironhack, my team and I performed an exploratory data analysis on different kind of diets (Vegan, vegetarian and omnivore). Analysing with diet is the most or the least ecofriendly word wide as well in each specific country.

The data is from 2019 and shows the total foodprint per country.

We gathered the data from Food and Agriculture Organization of United Nations (FAOSTAT), and Kaggle.

FAOSTAT: <http://www.fao.org/faostat/en/#data>

Kaggle: <https://www.kaggle.com/selfvivek/environment-impact-of-food-production>

Please contact me in case you would like to have more information <https://www.linkedin.com/in/imanollaconcha/>

**Data analysis**:

Here’s a list of the major computations done to explore, clean and process the data:

* Database import, getting the info and displaying them
* Joingin all the datasets from FAO
* Creating the realtion between FAO datasets and Kaggle datasents in order to join them
* Joining the FAO datasents and Kaggle datasets.
* Creating all the columns that we will use in tableau
* Database filtering to get only Spanish data
* New database import
* Databases merging
* Target variable definition
* Null values identification (6, 7, 8, 9, 66, 77, 88, etc.)
* Dealing with null values (substitution by Mean or Random Forest Method depending on null percentage over total individuals)
* Categorical variables treatment (One Hot Encoding or Label encoding depending on the nature of the variable)
* Correlation study
* X and Y definition
* Dealing with imbalanced data (SMOTE)
* Train test split
* Logistic Regression
* Confusion matrix
* Precision, recall and f1 scores

**Introduction**:

In this project, executed in collaboration with Ironhack, my team and I performed an exploratory data analysis focusing on people's perception about climate change. Specifically, we were interested in analysing the factors that influence citizens who do not believe in climate change. We gathered the data from the European Social Survey (ESS). This survey has the aim to measure behaviour patterns of the European population.

Please contact me in case you would like to receive the raw dataset (I couldn't upload it due to the size): <https://www.linkedin.com/in/npferrari/>

**Data analysis**:

Here’s a list of the major computations done to explore, clean and process the data:

* Database import
* Columns renaming
* Database filtering to get only Spanish data
* New database import
* Databases merging
* Target variable definition
* Null values identification (6, 7, 8, 9, 66, 77, 88, etc.)
* Dealing with null values (substitution by Mean or Random Forest Method depending on null percentage over total individuals)
* Categorical variables treatment (One Hot Encoding or Label encoding depending on the nature of the variable)
* Correlation study
* X and Y definition
* Dealing with imbalanced data (SMOTE)
* Train test split
* Logistic Regression
* Confusion matrix
* Precision, recall and f1 scores

**Results**:

We have found a correlation between 38 different variables and people who believe or not in climate change. Out of this variables, the most insightful one related to political interest, human values, education and demographics. Check the presentation to have a full overview about our findings!

**Authors**:

Nicolò Paolo Ferrari Jorge Palacios Marta Coll

**Content of this repo**:

* Source of the data: ESS (European Social Survey)
* Google Colab containing all the codes performed during the analysis
* Presentation (PDF format)
* Visualisations done in Tableau (see presentation)