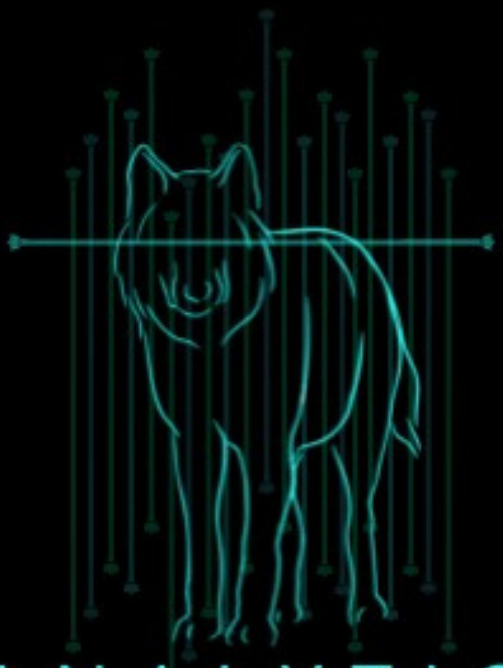


ONG - **Soy Henry**

Week #1

WORK PROPOSAL

Milestones week #1 - Index

Understanding of the current situation

The report required by our client is on the subject of Migration Flows (FM). MFs refer to the number of migrants arriving in or departing from an area over the course of a specified period of time. These zones can be countries, regions, cities or continents. The term arose to replace other terms with a more derogatory connotation, such as "mass" or "wave" of immigrants.

The phenomena of migration and human mobility are an innate characteristic of the human being, of an ancient nature and have affected almost without exception all societies in the world. Migration is a complex topic, and as such it can be distorted to alarming degrees by misinformation and politicization.

The understanding of FM from the perspective that Data Science provides us will be the characteristic criterion of our approach, without falling into ideological biases or opinions of a personal nature.

We believe that exploring and exposing MF in a clear, precise and scientific way is enormously useful for society in general, companies, policy makers and all the agents that are part of this world, that are directly or indirectly affected. for the phenomenon.

In our preliminary research we have been able to determine that the characteristics of FM vary depending on the places, times, country policies, demographic, sociopolitical, environmental and economic aspects, among others. We will explore all these fields in detail, increasing our understanding and reserving the prerogative to modify the focus of the report as our knowledge and expertise on the phenomenon under study deepens and increases.

Objectives: job and group specific

In the first stage we will explore all the data and we will try to explain which are the determining features that affect the MF, we will understand them in depth and we will determine which are the most important. We will carry out flow analysis by zones and continents, to broadly understand the scale, direction and frequency of the movements.

Subsequently, we will sift said data at the country level, and we will make an exhaustive study of their characteristic features, classifying them into various categories according to their net amount of migration.

Subsequently, we will carry out deep analytical tasks and apply Machine Learning (ML) Models trying to answer, among others, the following questions:

- *What makes an emigrant leave a country?*
- *What are the predominant characteristics of the countries most chosen by immigrants?*
- *What consequences does FM generate in both types of countries?*
- *How can we measure that effect?*
- *What are the characteristic features of countries with net positive and negative FM?*
- *What makes a country have net FM close to zero?*

As a result of all this work, we will generate a detailed report of the discoveries that our team of analysts will make, with the respective conclusions and insights provided by the data.

We will create a unique database in the world that will contain all this information and will be updated automatically.

In addition, we will create a technological interface that allows remote access not only to the database, but also to metrics, statistics and specific indicators that we will develop according to our future discoveries.

Reach and out of reach

Initially, the scope of the project is extended, upon reaching the agreed deadline on 2/28, to deliver the following products to the client:

- Access to the private repository with all the information on the process.
- Detailed report of the theme.
- +5 KPIs that we consider decisive, with their explanation and information on use and interpretation.
- Interactive Dashboard with metrics, statistics and KPIs.
- Trained ML learning model, with access to the code and step-by-step instructions for the process.
- Report meeting with the entire team to present the deliverables, a step by step presentation of the entire process and a presentation by our analysts.

This project does not include:

- The rigorous forecast of future FM flows.
- Modifications in the requirements made beyond the second week of work.
- Provide information to refugees, asylum seekers and those displaced by political and armed conflicts regarding courses of action in their situation.

Repository

The official Analytic Hound repository for this project has already been created and will be delivered to the client once all the documentation is duly pushed and ready to be consumed.

Proposed solution

TECH STACK

We will use an architecture that contemplates the tools that are seen in the following diagram. We reserve the right to modify the stack in the future, if necessary for unforeseen operational issues.

WORK METHODOLOGY

The work methodology is supported by the SCRUM software development framework. Our members are trained in the methodology and we will use several of its tools to effectively coordinate our actions.
We will mainly focus on:

Sprint Planning Meeting: We will use this tool at the beginning of each weekly Sprint with the entire team. We will inspect the Product Backlog (the tasks, requirements and functionalities that the project requires that week) and select the specific items that we will work on during the weekly Sprint.

Daily Scrum: Daily, in the morning prior to the meeting with our external advisor Gonzalo Posse, the team will meet in this space. For no more than 15 minutes, the project will be monitored in which compliance with the assumed tasks will be controlled. In addition, the daily work objectives will be agreed and possible problems that have directly limited or prevented the fulfillment of the objectives will be analyzed.

Sprint Review: At the end of the weekly Sprint, after meeting with our PO Jonathan Deiloff, the team will come together to retrospectively evaluate performance. The goal is to reflect on the performance of the Sprint and identify opportunities for improvement for the next Weekly Sprint.

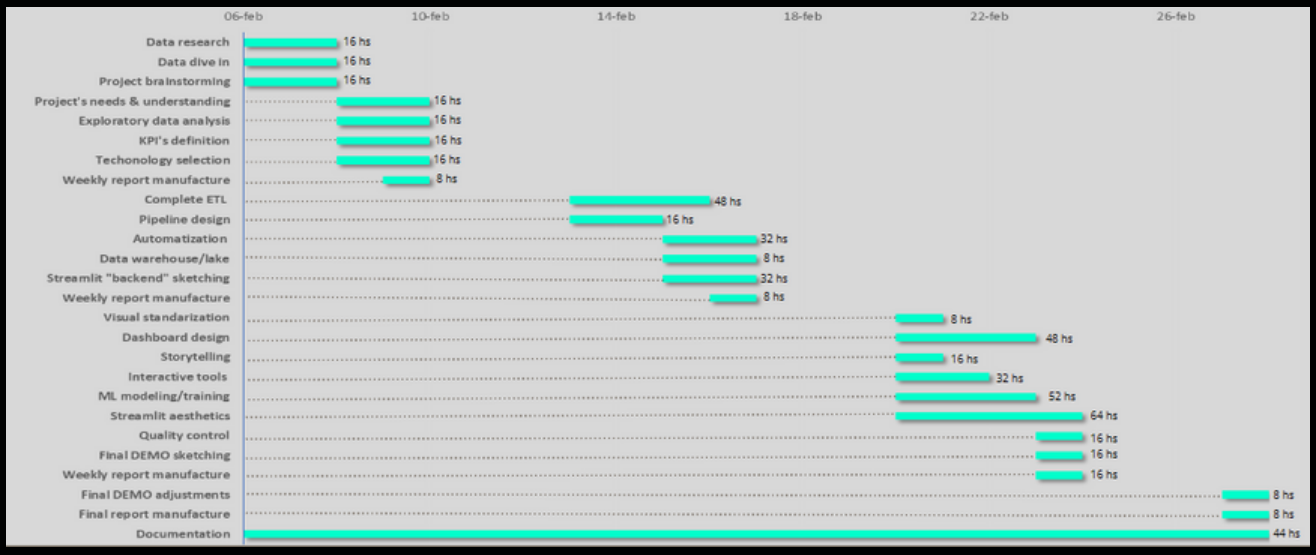
Proposed solution

ROLES AND RESPONSIBILITIES

Our team has diverse professional profiles and the whole spectrum in the Data industry.

- Functional Analyst - Mauro Pini
- Data Engineer - Lucas Rodriguez
- Data Engineer - Eugenia Ball
- Data Analyst - Belén Zapata
- Data Scientist - Alan Mysler

GENERAL SCHEDULE AND ESTIMATE OF EFFORTS



KPI's to use

According to the preliminary internal report of the Analytics team, the initial working hypotheses for the formulation of the KPIs yield various results. As the workflow progresses, we can add KPIs not initially considered and/or eliminate any of those proposed in this initial stage of project development.

The methodology for determining the KPIs follows the concept of the recognized SMART objectives. The acronym stands for "Specific", "Measurable", "Attainable", "Relevant" and "Time-bound".

NOTE: for KPIs measured in percentage, we prefer the decimal notation instead of writing the percentage in base 100. Ex: 0.2 = 20%. This decision by our Analytics team will facilitate the subsequent handling of the information.

- KPI_1 = Increase in annual net migration, measured in quantity (difference between the number of people who immigrate to a country and the number of people who emigrate from that country between two periods).

$$KPI\_1 = MN(n) - MN(n-1)$$

MN(n) - > Net migration for the period (n)
MN(n-1) - > Net migration for the period (n)

Eg: KPI_1 = 150,000 implies that the objective to be achieved is that there is a net MF of 150,000 entrants to the country per year.

KPI's to use

- KPI_2 = Increase in annual net migration, measured as a percentage (percentage change with respect to the MN with respect to the previous period).

$$KPI\_2 = \frac{MN(n) - MN(n-1)}{MN(n-1)}$$

Ex: KPI_2 = 0.12 implies that the objective to be achieved is that there is a percentage increase of 12% per year of the net MF.

- KPI_3 = Annual percentage increase in skilled immigration. This KPI measures the annual percentage change of immigrants that a country receives, who have higher education levels than high school.

$$KPI\_3 = \frac{MC(n) - MC(n-1)}{MC(n-1)}$$

Ex: KPI_3 = 0.15 implies that the objective to be achieved is that there is a percentage increase of 15% in the number of qualified immigrants.

KPI's to use

- KPI_4 = Increase in the index of conditions desired by migrants by X%. This index is subject to further specification, since it is a basket of the features that we find most relevant when it comes to explaining MF reception in the countries that receive the most MN.

$$KPI\_4 = \sum_{i=1}^{m} \left(\frac{Fi(n) - Fi(n-1)}{Fi(n-1)}\right) / \mathrm{m}$$

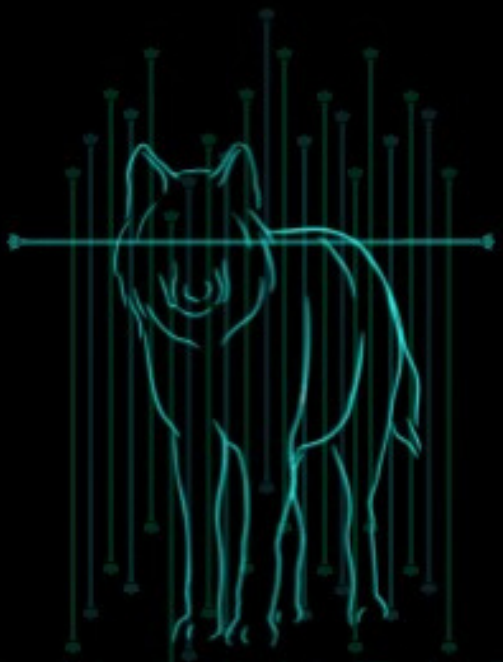
This KPI provides us with the weighted average of the percentage variations between period n and period n-1, for the "m" indicators chosen. It provides an effective estimate of the degree of desirability of the country to receive FM.

Eg: KPI_4 = 0.12 implies that the objective to be achieved is that there is an annual percentage increase of 12% of the indicator. This would occur, for example, in a scenario with 3 selected Features where their annual percentage variation values are: F1 =0.05 , F2=0.15 and F3=0.16.

- KPI_5 = Decrease in the index of conditions unwanted by migrants by X%. This index is subject to further specification since it is a basket of the features that we find most relevant to explain the low levels of MF reception. The logic is similar to that of KPI_4, except that this basket measures the features that most explain a negative FM in a country.

$$KPI\_5 = \sum_{i=1}^{m} \left(\frac{Fi(n) - Fi(n-1)}{Fi(n-1)}\right) / \mathrm{m}$$

In this case, a decreasing KPI over time is desirable because it indicates that the adverse conditions are less. Ideally, KPI_5 will take negative signs. Intuitively, it can be thought that a slowdown in the inflation rate, a decrease in unemployment and insecurity indices are desirable, issues that would give us a negative KPI_5.



ANALYTIC

HOUND