VIVA CHALLENGE

Project 1: VIVA - Lima

Many children have the dream of building a better country and becoming good citizens. They have all the potential to achieve it, but we have to help them to improve their soft skills and that is what the NGO VIVA does. They work with school directors and teachers, being the last ones the main users of VIVA's methodology. This methodology is measured by 3 main *indicators*: **Self-esteem**, **Sense of belonging** and **Conflict management**. The main objective of the project is to use data analytics to measure the teacher's impact on children to help VIVA's decision making related to its methodology. In order to do that, we will identify the children who have an improvement in the 3 indicators and then we will find out what the characteristics of their teachers are, and we will do the same with those children who had less or no improvement. Based on those characteristics, we will be able to propose a set of actions with the aim of improving the results of VIVA's methodology, such as encouraging the traits of the teachers that had successful results with the students.

Open questions:
- How to deal with interactions between variables that could explain the indicators? interactions between categorical variables vs. continuous variables, categorical variables vs. categorical variables, etc.
- How to deal with variables that could be related to bias & fairness in the models, like age, gender, etc, should they be eliminated from the dataset?
- How to identify / handle correlation vs causation?
- We plan to propose some improvements in the methodology based on the results from the data analysis. Is there a way to evaluate if the improvements we propose in the methodology work at a lower cost than a RCT?
- RCT Background: VIVA performed a RCT in 2018, with the objective of evaluating if using the methodology led to better *indicators* compared to not using the methodology. For this, a group of 60 schools was randomly partitioned into a treatment and control group. This experiment led to significant improvements in the students indicators, which brought evidence regarding the effectiveness of the VIVA methodology. These kinds of interventional experiments haven't been performed since then, though.
    - We can think of the project we're currently working on as a study to find important relationships between teachers characteristics and students indicators from observational data only.
- In the case that a good teacher has a bad group of students then we can wrongly associate the teacher possibly positive characteristics to a poor performance in the students' indicators. How to identify and handle this situation?
    - Another aspect of this question is related to the fact that the indicator results don't only depend on the teachers but also on the students, so a match between their characteristics may be the responsible for these results. Is this getting closer to a recommender systems approach? Is that a better way to handle the problem? Or should we just add the students aggregated variables

as extra "columns" in our dataset of teachers? There's also some text data from the students' comments.

Project 2: VIVA Webinars

Given the current Peru emergency state situation due to the COVID-19 global pandemic, VIVA has started to offer live Zoom webinars to teachers, school directors and student's parents from all over the country, with an unexpected average attendance of 400 participants. New needs arose: measuring their audience (number of attendants, participants permanence during the webinar, region, etc.) through the data provided by Zoom, and increasing the number of attendees across the country. Zoom also provides the log of the attendants' comments during the webinars, where they express their needs and current situation among other things. The main objective of this project is use this information to measure the participation and interactions of the attendants during the webinar and understand better the opinions and needs of teachers and directors through their comments. This will help VIVA to prepare the webinars according to the current needs of teachers and directors. Furthermore, this will also provide information to VIVA about which schools to work with more closely next, following their original methodology described in the first project.

Open questions:
- What are important metrics that we could use to measure the participants' engagement? We are currently considering the time they spend at the webinar, how many comments they make, the results from the feedback forms and if they are recurrent participants across webinars.
- Could there be more ways to link this project with the first one, which is the "business" core of VIVA?
- Is there a way to get more information from the attendees that don't interact during the webinar?
- Are there good guidelines or strategies for designing good feedback forms that people may be willing to fill?
- How can we make better use of the comments data from the webinars?