# A Look at Health Care in Sogamoso, Boyacá

PQRSDF Sogamoso Sector Salud

**Team 222**: Aura Ramírez, David Vergara, Fabio Calderón, Katherin Del Risco, Laura Quiza, Sandra Rivera.

DS 4A COLOMBIA

## Highlights

- Text analysis of the PQRSDF.
- Predictive model an predictive tool for classify the type of PQRSDF.
- Dashboard for visualization and analysis of relevant information from the Sogamoso Health System.

Dashboard Link

## Background

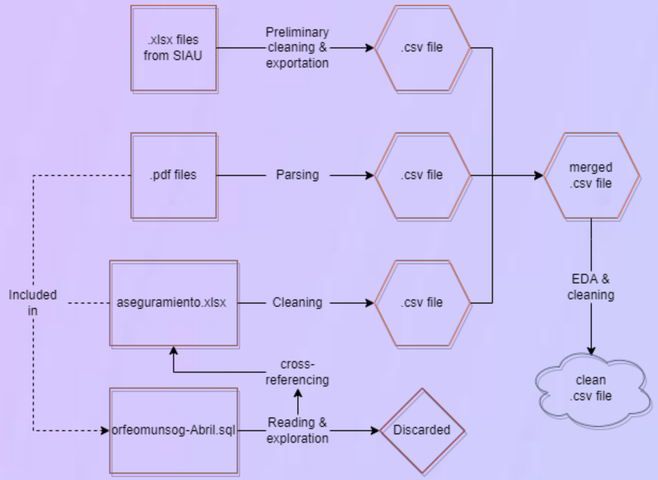
Sogamoso is a city located in Boyacá, Colombia, with around 132,985 inhabitants. The Municipal Health Secretariat, which oversees the Health System of the city, has detected that some of the EAPB (Benefit Plan Administration Entities) do not guarantee proper attention to their affiliates, and thus users constantly fill PQRSDF (Petitions, complaints, claims, suggestions, denunciations, and compliments).

## Objective

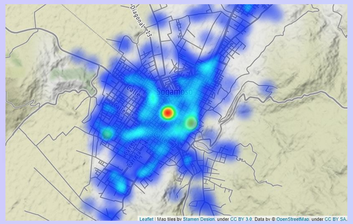
Detect and explore the aspects of the Health System that most affect the population of the city and that therefore require greater attention from the institutions. Also, to model and predict in order to classify the type of PQRSDF.

## Data

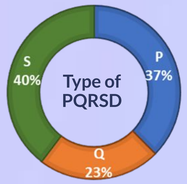
We received 4 datasets from the Municipal Government and we cleaned each one. Hence, after deciding that only three of the four datasets were worthy of our attention and having performed some more data cleaning, we joined them. This led to a `pandas` dataframe (in Python) with 4324 entries.

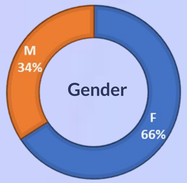


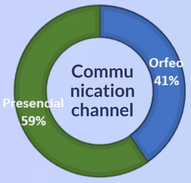

## Visualizations and Insights



Using the `googlemaps` library in Python we connected Google Maps' API for georeferencing to get coordinates for those PQRSDF having a non-NaN address associated with them. We can observe that not all PQRSDF were filled in Sogamoso, but rather some requests come from other cities/towns in the region (e.g. Bogotá, Tunja, Yopal).

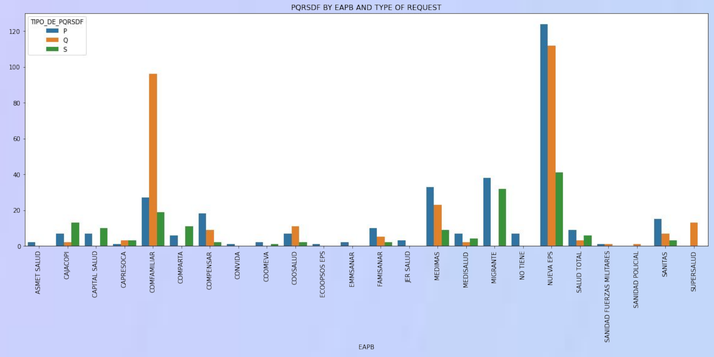



### Distribution of PQRSDF for each EAPB

It can be seen in the figure below that Comfamiliar has a lot more complaints (Q) than petitions (P), and Nueva EPS has a similar number for both of them. These two EAPB, together with Medimas, seem to be the ones that fail the most at providing good health care services. Although the reason for this might just be that these are precisely the EAPBs with the highest number of users.



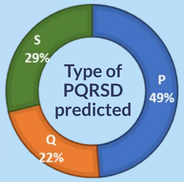

### Text Analysis

The text on the PQRSDF allows to identify the principal reasons for the requests. Also, it can be filtered by gender. We detected the most common words in the requests: 'Busca', 'inconformidad', 'manifiesta', 'informacion', 'solicita', and 'medicamentos'.


## Model

We decided to try out a classification model, meaning that we wanted to take any PQRSDF text and predict its type. For this purpose we ran a logistic regression model and a random forest.

After several attempts with our models, including stratification, cross validation and hyperparameter tuning, we chose the logistic regression model as our best model, with an accuracy of 67%. Although its accuracy is close to the one of the random forest, the logistic model is more consistent with new unseen data, so we used it on our original data set to predict the missing values in Type of PQRSDF.




**Finally, with the logistic regression model and the most optimal vector conversion parameters we created a tool that allows us to enter new PQS and predict their type.**

## Conclusions

- The Municipal Health Secretariat (SMS) needs to improve its way of imputing the information of the PQRSDF filled by the users, so the relevant data can be recorded for all requests.
- As with any data science project, the accuracy of the models and predictive tools strongly depends on the integrity of the dataset provided. So, even though our models did perform good with a dataset having so many NaN values, if the SMS had provided a more solid dataset, a higher accuracy might be reached.