Perinatal and neonatal mortality analysis in Colombia

Problem definition and project overview

As of 2011, The World Health Organization (WHO for its acronym) estimates state that about 4 million neonates die every year, and nearly 41% of all under-five child deaths are among newborn infants, babies in their first 28 days of life or the neonatal period [1]. Accordingly, perinatal and neonatal mortality is now a part of the 2030 agenda for Sustainable Development of the United Nations and is also one of the topics of interest to the Colombian Ministry of Health [2].

Our aim is to carry out an analysis of different databases related to perinatal and neonatal death, as well as demographic, economic, social and geographic data for Colombia, that can help us understand which features may influence and contribute to the said death rate in different locations in the country by looking for possible correlations that may constitute the first step to the search for solutions to this global health issue.

Impact of a possible solution to the problem.

Characterization of the social, economic, demographic and geological variables that can be correlated and may have an impact on the newborn and fetal death rates is of capital importance for the localization of the most affected regions, as well as for the identification of causal relationships between said factors and, upon inclusion of larger databases, identification and prediction of new factors, possibly distribution of genetic disorders and others. Such an analysis may eventually lead to the implementation of more specialized and well-designed social programs as well as health campaigns, among other humanitarian initiatives.

Methods

One of the bases of this project is the data visualization which allows for the identification, comparison, summarization and explanation of the perinatal and neonatal death rates in Colombia, with the objective to find trends and behavior that can support the programs of prevention and reduction of said rate in this country.
Additionally, this project provides the analytical and visual representation tools to continue the control over this topic as new data is sourced by DANE, and support in a good way the future of the country on this matter.
Below are some of the graphics that are planned to be carried out throughout the project. These charts are designed to help recognize patterns that can contribute to the analysis of this problem:

- Maps that relate geolocation with outcomes of the pregnancies (fetal, non-fetal and neonatal data) at the departmental level in Colombia. Through this visualization, the areas and populations that may be most affected by this event will be identified.
- Charts that allow the exploration of possible correlations between the pregnancy outcome and the descriptive variables that have been classified as variables of interest.
- Interactive graphics that allow visualizing the behavior of pregnancy results in Colombia over the years.

Models

In order to find possible causes for the perinatal and neonatal death rates in Colombia, we will use the following set of techniques:

- Descriptive analysis: Descriptive analysis of the data will be used throughout the project with a view to identifying statistical values to achieve the proposed objectives.
- Georeferenced maps: Through this visualization tool and the departmental descriptive statistics, the aim is to identify and analyze different possible areas where the perinatal and neonatal death rate is high.
- Correlations: Based on the previous methods, we intend to explore possible correlations between variables that may influentiate the outcome of the pregnancy.
- Clusters analysis: With the help of clustering algorithms, it is intended to find groups of characteristics, with a view to predicting the outcome of the pregnancy.
- Decision trees: Based on its branched structure and its conditional control sentences, it will be possible to analyze the different options and consequences, with a view to predictive trends in pregnancy outcomes that depend on temporal, geographic and demographic data.
- Random forest: In case the results of the decision trees do not meet the expectations in the prediction of the pregnancy result based on different categorical variables over time, it is possible to iteratively perform said algorithm with different parameters within the same data, seeking to reduce possible noisy models, so as to generate more robust predictions.

Exploratory Data Analysis (EDA)

Data Cleaning. In order to concatenate the 3 datasets, some column names were changed and values standardized across datasets. It was observed that the dataset contained death reports with information from people of all ages, so it was reduced to the data that effectively contribute to the objective analysis, that is, data associated with over-one-year-old people was discarded. Afterwards, missing information rate was computed and it was possible to identify columns with huge amount of missing or unspecified data. Figure 1 shows a table sorting the columns with the highest missing-value rate, the ones with a score of 70% or more were chosen to be analysed individually, in order to determine which ones can safely be discarded.

	Porcentaje faltantes
OCUPACION	99.978045
C_MCM1	99.828022
C_ANT32	99.758498
MAN_MUER	99.502360
CODOCUR	99.487724
C_ANT12	99.396246
C_ANT22	99.348677
C_DIR12	99.235245
C_PAT2	98.455853
C_ANT3	91.982875
C_PAT1	89.974020
FECHA_NACM	80.965275
C_ANT2	80.723773
GRU_ED1	73.182334
ULTCURPAD	71.528413
C_ANT1	69.830583

Figure 1. Missing-values rate on columns
-

A simple exploratory analysis was carried out for each variable, aimed to allow a heuristic determination of the information content. In addition, some columns were imputed based on other variables, namely, the missing values of the field related to the number of alive successful pregnancy deliveries was imputed based on the total of pregnancies and the number of born-dead children, and in the other way around to calculate the missing values of born-dead field. Furthermore, some categorical values were relabeled, since they were given as non-self-explanatory integer values, or unified to build new features; e. g. department and municipality were used to create the zip codes and individual day, month and year to create the complete date, and some other transformations.

EDA. Initially, some easily understood variables were plotted as discriminated distributions (births, fetal deaths and newborn deaths). The plots shown in figure 2 were obtained as a first approximation to the EDA by considering only a subset of the data for the sake of efficiency. Since all of those variables do not give enough information by itself, they were discriminated by our variable of interest (pregnancy outcome).

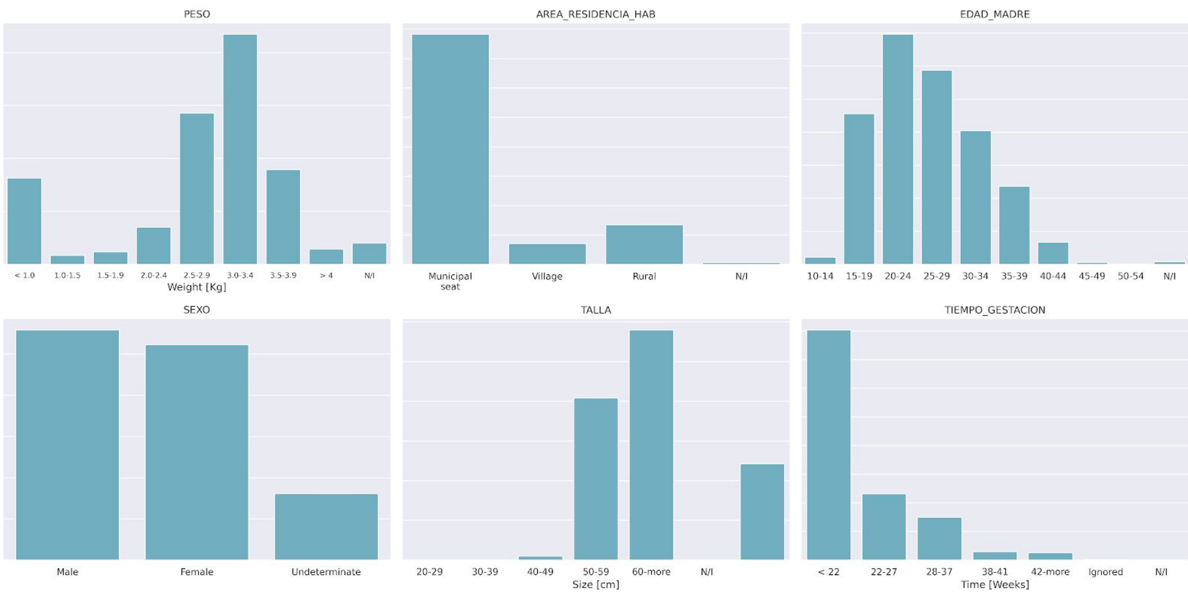

Figure 2. Distribution of some of the variables of interest.

The discriminated distribution is presented in figure 3. Here, every category was normalized independently from the others so as to allow an easier visualization of the relative differences in the distributions; consequently, the scale is not the same for all the distributions. These plots show already some important features of the dataset in a more detailed way; particularly, it can be seen that the weight distribution for newborns is a normal-like, and most of the fetal deaths are concentrated below 1.0 kg, which is an expected conclusion since fetal deaths don't make it to the end of the normal nine-month pregnancy period and therefore they do not reach the total of their development. Although we have considerably less data for the non-fetal deaths, we can see that the weight distribution for this group differs from that for the newborns, in that its mean is displaced to the left and has a peak below 1 kg. Further on, it can be noted that post-birth deaths with its seemingly bimodal distribution looks already very much like a different population than normal births, with a more normal-like distribution. Weight can be thus proposed to be a good factor indicating newborn viability.

Although distributions for age of the mother are similar across categories, it can be seen that pregnancy outcome is also sensitive to this variable since the probability of a fetal death increments with mother's age, as can be seen from the respective plot, where the distribution for fetal deaths is more spread out. Some general features are that the peak is at 20-25 years old, and it is a worrying situation the number of mothers in the age range 10-14, and even 15-19 years old.

Some general characteristics of mother's age distribution:

- The number of pregnancies in the age range of 10-14 is higher than wanted, as this kind of pregnancies are treated as a public health issue.
- The peak of all distributions is found in the ages of 20-24 years old.
- In the age range 35-49, it would be interesting to see if this observation holds true for the complete data set. Nonetheless, this suggests a greater risk as mothers age, which matches public health information available.

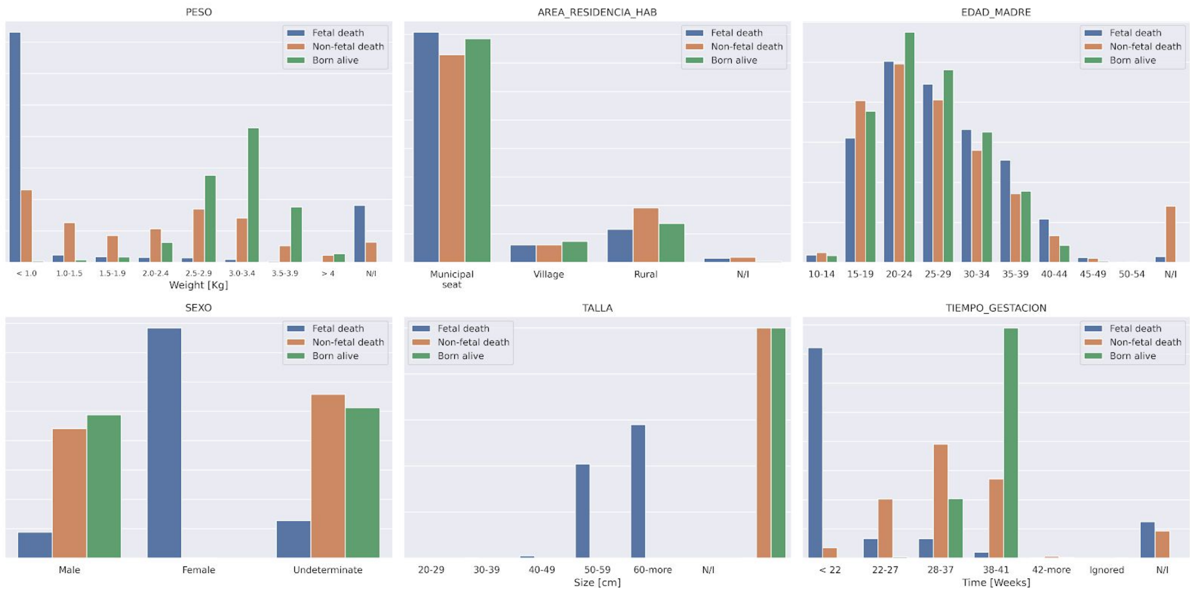

Figure 3. Distribution of some of the variables of interest, discriminated by final pregnancy outcome.

The sex and pregnancy time plots can be understood together as follows: most of the fetal deaths occur before 22 weeks, which may affect the sex determination procedure since at this stage the fetus is not yet fully developed; that way can the sex-undetermined peak in the sex plot (which is due to fetal deaths only) be understood. Additionally there does not seem to be a relationship between gender and the probability that the infant will be born alive or suffer a fetal or non-fetal death.

Finally, there does not seem to be substantial differences among distributions in residence area, that is, the fact that some mother lives in a city or in a rural location does not seem to correlate substantially with the outcome of the pregnancy.
From the graph of gestation time, it can be seen that the infant is more likely to suffer a non-fetal death if the infant is born in the 28-37 week period; the effect of this is however apparently weaker as compared to others, such as weight or mother's age, so it needs to be further explored to confirm this claim.
From the baby length graph, we can see that we only have available data for non-dead newborns.

Some more plots can show further insights on this data and on Colombian population in general. In figure 4 the father and mother's age distribution are shown, and it can clearly be seen some slight differences. Specifically, mother's distribution peaks at an early age (20-25) and rapidly drops, while the fathers' peaks at 25-30 and is more spread to the right, showing that male parents tend to be older than females. Although this is a known stereotype of relationships in our country, it would be interesting to dig deeper into this matter and check what is happening to mothers in ages between 10-19 and verify the possible pregnancy outcome in this specific scenario

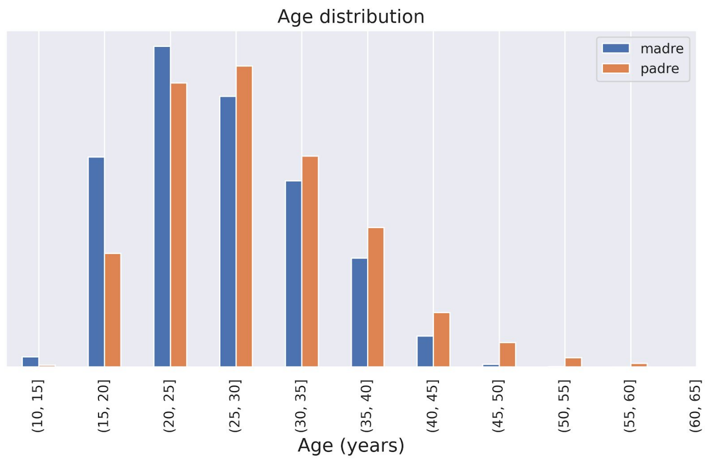

Figure 4. Age distribution of both male and female parents.

The distribution of ages of post-birth deaths is shown in figure 5. Left plot shows the raw numbers, as categorized by different ages. Here we can see that most of these deaths occur between 1-6 days, and 1-5 months. However, the time-normalized distribution (right plot), which can be heuristically interpreted as a time-dependent death probability, shows that the first hour of live is critical to determine the survival of newborns. This last point is already an important insight from the data, for it shows where the focus should be set: babies between 1-6 days, and 1-5 months, and specially less than 1 hour old newborns and the particular conditions that lead to the so-mentioned behaviour.

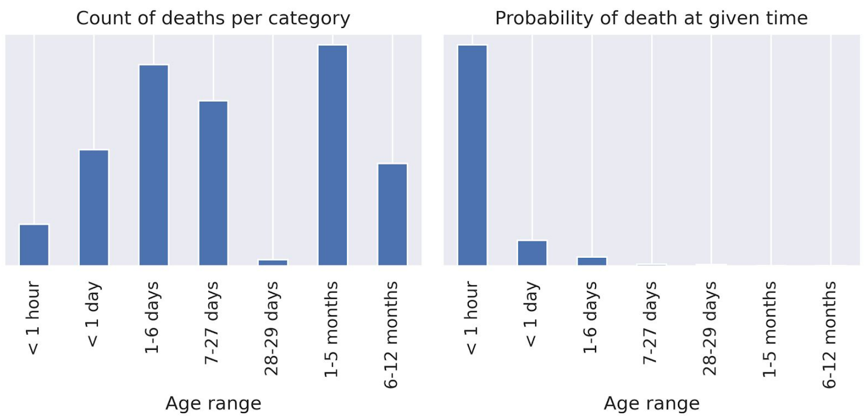

Figure 5. Distribution of death ages for post-birth deaths.
a. Absolute numbers (no normalization).
b. Time normalization, accounting for the different time sizes of the bins. This can be interpreted as a time-dependent

death-probability.

Lastly, pregnancy outcome counts were plotted accounting for the different racial and cultural identification of the person in question (figure 6). Particularly, this plot shows differences between indigenous, afro and “raizal” populations, and population that do not identify themselves with any of these, among others. It can be seen that the probability of post-birth death is higher for these three aforementioned populations, as compared to the latter. This might be due to cultural and/or behavioural differences among these populations; however, we must not discard the possibility of the presence of systematic discrimination against these populations, as well as to their territories. This can further be observed in choropleths and other forms of geographical visualization techniques.

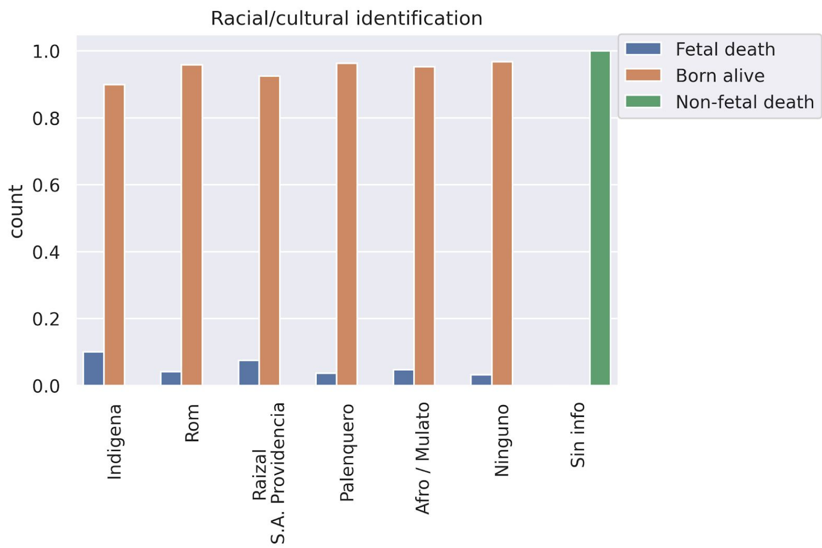

Figure 6. Pregnancy outcome as discriminated by racial/cultural identification. The scale is not the same for every distribution, so as to observe differences in distributions among categories.

The other variables can be further studied in this or any other way by formulating some questions to the dataset. These are the questions we have come up with (and have not answered) so far:

- Influence of multiplicity of pregnancy on outcome.
- Compare mother’s and father’s educative levels. Are these distributions any different? Look for combinations of these two, maybe data can already show inequity?
- Influence of number of pregnancies, successful pregnancies and deaths for a given mother on outcome. Check distribution, geographical distribution, correlation with race, culture, age, among others.

Time series:

In order to understand better how this phenomena is behaving during the time, more information had to be collected; fortunately, lots of data was found, divided by years, in the same web page where the original data was gotten. It was necessary to mix them all and then adjust these new dataset to be compatible with the one that we had generated.

Information from 2008 to 2018 could be added and plotted, as a result, it could be possible to visualize that children born alive is actually a stationary phenomena. As figure 7 shows, it has very high peaks and low valleys in almost the same periods, which is pretty interesting; further investigation can be done in order to identify the reason why this is happening and maybe correlate this with some other variables.

Our interest variable was plotted separated, so if the non-alive children plot is analyzed, figure 8, it is possible to see that in the period from 2010 to 2016 the amount of values were increasing but then it starts to decrease. Finally, the non-alive children was plotted but in the form of rates, identified as orange in figure 9, and roughly it show it’s stationarity and it’s maximum value not greater than 10%.

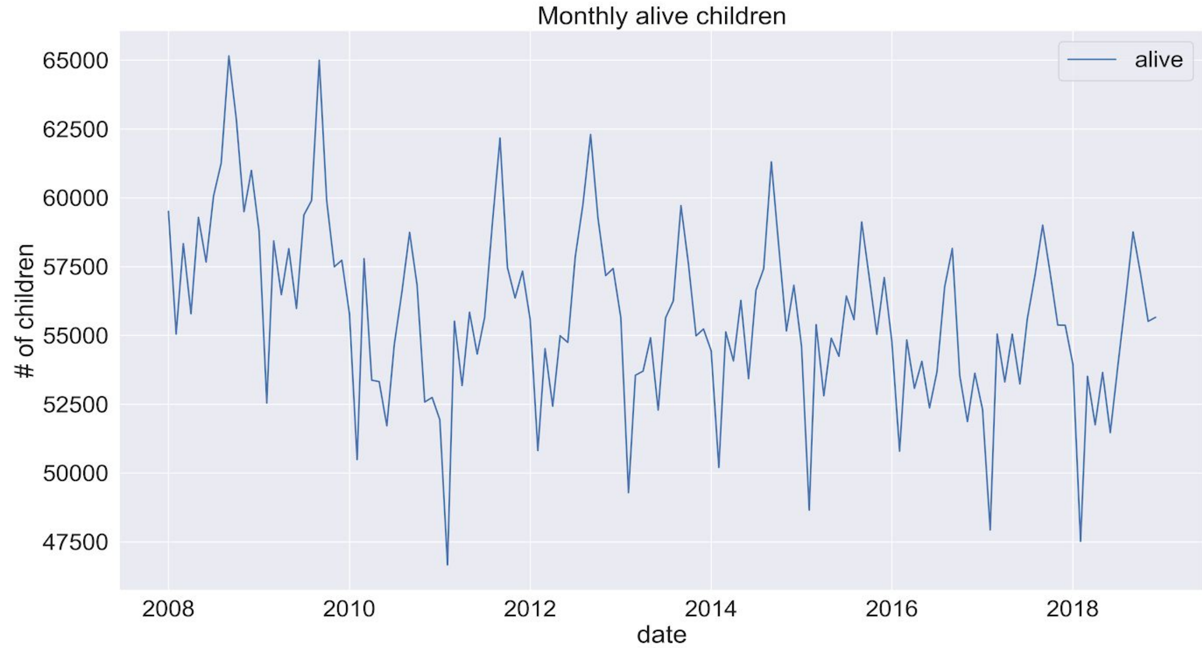

Figure 7. Monthly alive children

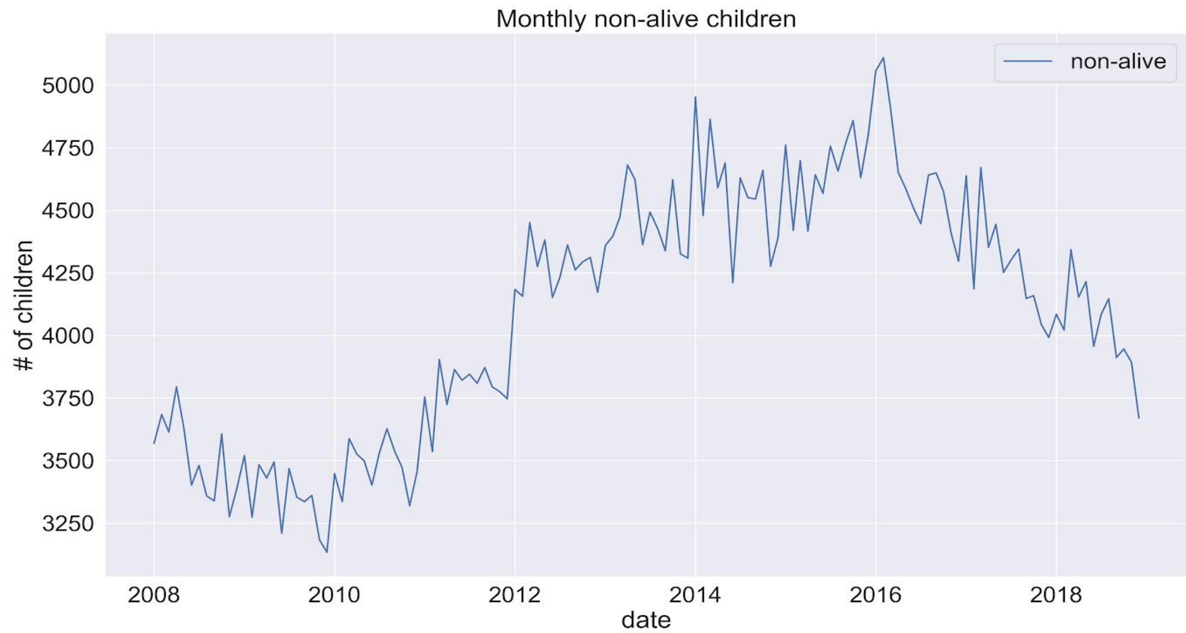

Figure 8. Monthly non-alive children

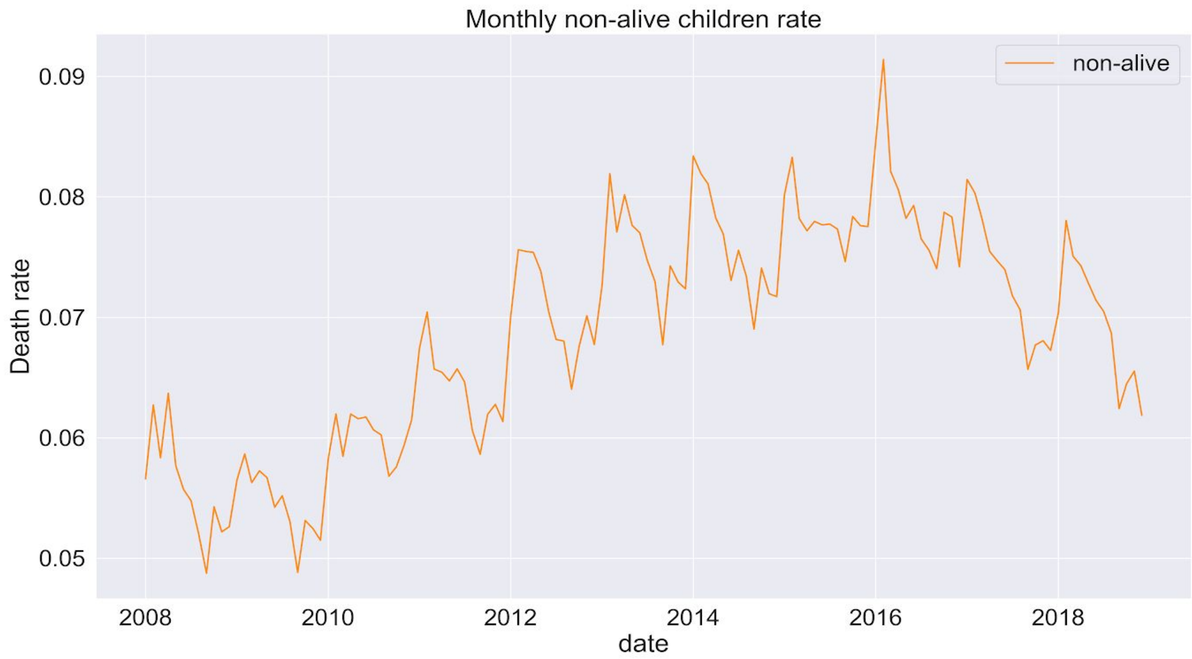

Figure 9. Monthly children death rate.

Mortality rates by department:

One of the best ways to visually analyze the data, even without using any model, is to observe the mortality rates by departments in a geopolitical fashion. This could allow us to observe and identify strange behaviours in some departments, which can be useful for subsequent analysis.

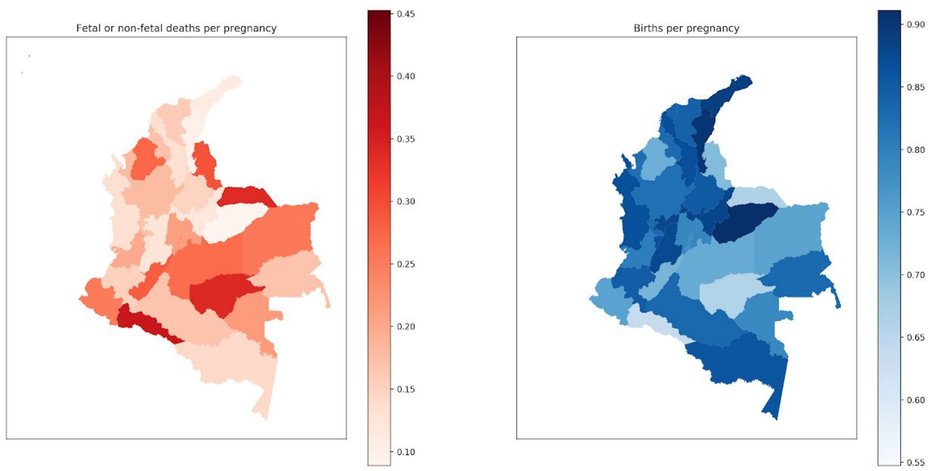

Figure 10.a) On the left, the heat map representing the total deaths (fetal or non fetal) per pregnancy. b) On the right, the heat map representing the successful births per pregnancy.

In the first place, in figure 10 it is possible to observe the rates of total deaths (fetal and non fatal) and births per department. At a glance it is possible to identify that Cauca, Putumayo, Arauca and almost all the Orinoquia region are problematic regions when we talk about deaths in the first year of live. In the same way, when we look at the image of births per pregnancy we look that successful pregnancies are more possible in the Andean region, but is important to look carefully what happen in the periferia because many of those regions do not have many inhabitants so the rates could change dramatically for specific problems and we want to keep our attention in the bigger picture.

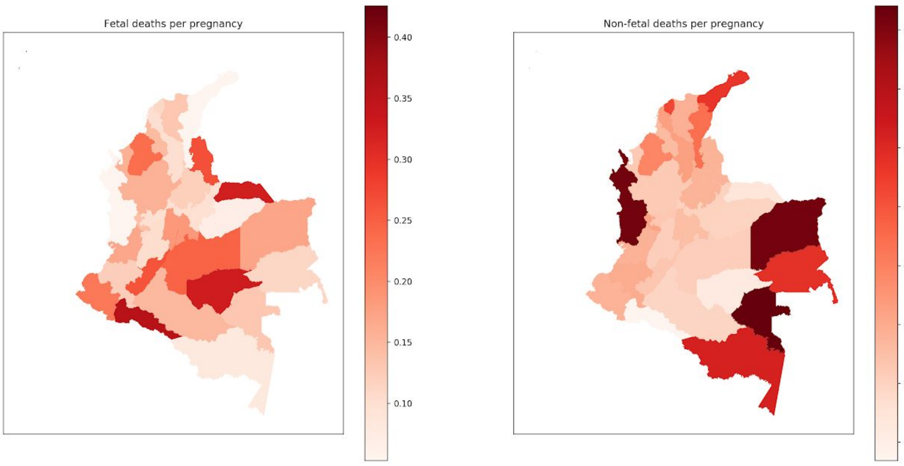

Figure 11.a) On the left, the heat map representing fetal deaths per pregnancy. b) On the right, the heat map representing the non fetal deaths per pregnancy.

Another important observation is about how deaths occur in each department, in poor, rural departments the fetal deaths could be undervalued because people do not have the capability to go to hospitals, but non-fetal deaths are different because people must report the baby's death since they are already a citizen. In figure 11 is possible to observe that poor, rural departments in the periferia have much bigger rates when we talk about non fetal deaths but the rate of fetal deaths is almost zero in those departments, on the other hand the central departments of the Andean region have bigger rates for fetal deaths but much lower rates for non fetal deaths. One possible explanation is that in the central region the access to hospitals is better so people could go even for small problems in pregnancy, this on one side increase the possibility of having a healthy baby and on the other side It also increase chance of losing baby while being in the hospital, leading to miscarriage being reported more frequently.

Correlated Variables:

The variables were splitted into continuous and categorical, in order to find relationship between them. The correlation methodology was applied to continuous variables, as a correlation matrix presented in figure 12

Here, we find that some of these variables have a correlation between them, for instance, the number of pregnancy controls (NUM_CONSUL) is highly correlated with father’s age (EDAD_PADRE) and size of baby (TALLA), also, the variable number of pregnancies

(NUMERO_EMBARAZOS) has a high correlation with number of both dead and alive children (N_HIJOSV and N_HIJOSM).

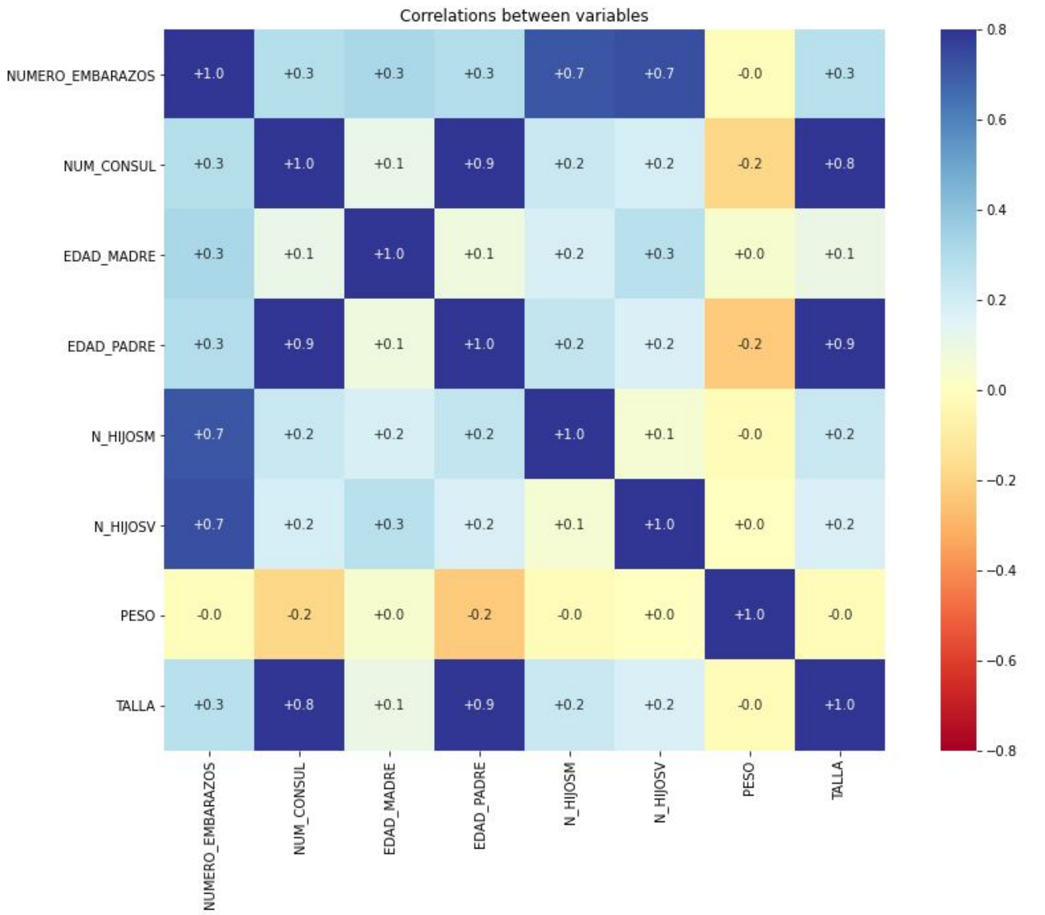

Figure 12. Correlation matrix performed on continuous variables on the dataset.

Subsequently, the exploration of models that could predict the outcome of pregnancy begins with these variables. To do this, a new column called “NACIDO_VIVO_INV” is used where the value 1 represents the subjects who died and 0 the subjects who lived. When training a model (using the natural log) with the variables “NUMERO_EMBARAZOS”, “EDAD_MADRE” and “PESO” which correspond to the number of pregnancies of the mother, age of the mother and weight of the baby respectively, we obtain the results shown in figure 13. From this we can see some interesting coefficients, the R2 is however very low, which indicates that the model does not fit well with our desired result, being the reason is why further work must be conducted over the categorical variables so as to include these in the model.

The regression model:

The model for this project is the logistic regression

- $$Y = B_0 + B_1X_1 + B_2X_2 + ... + B_8X_8 + \xi$$

```
FORMULA="NACIDO_VIVO_INV~NUMERO_EMBARAZOS+PESO+SEXO_CAT+SEG_SOCIAL_CAT+SITIO_EVENTO_CAT+TIEMPO_GESTACION_CAT+TIPO_PARTO_CAT+AREA"
modelo_full_output=smf.ols(FORMULA, data=df_cat_num_models).fit()
print(modelo_full_output.summary())
print(modelo_full_output.aic)
```

where the dependent variable (1: if a baby died and 0:if the baby live) that is explained for two continuous variables and six categorical variables. The model is:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:        NACIDO_VIVO_INV   R-squared:                       0.772
Model:                            OLS   Adj. R-squared:                  0.772
Method:                 Least Squares   F-statistic:                 1.976e+05
Date:                Sat, 04 Jul 2020   Prob (F-statistic):               0.00
Time:                        02:53:18   Log-Likelihood:             1.0938e+06
No. Observations:             1579068   AIC:                        -2.188e+06
Df Residuals:                 1579040   BIC:                        -2.187e+06
Df Model:                          27
Covariance Type:            nonrobust
===============================================================================================
                                  coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------------------
Intercept                       0.5765      0.001    519.923      0.000       0.574       0.579
SEXO_CAT[T.2]                   0.0055      0.000     27.659      0.000       0.005       0.006
SEXO_CAT[T.3]                   0.4698      0.001    442.209      0.000       0.468       0.472
SEXO_CAT[T.9]                   0.0582      0.030      1.919      0.055      -0.001       0.118
SEG_SOCIAL_CAT[T.2]             0.0066      0.000     16.468      0.000       0.006       0.007
SEG_SOCIAL_CAT[T.3]             0.0045      0.000     21.541      0.000       0.004       0.005
SEG_SOCIAL_CAT[T.4]             0.0023      0.001      3.649      0.000       0.001       0.004
SEG_SOCIAL_CAT[T.5]             0.0006      0.003      0.221      0.825      -0.005       0.006
SEG_SOCIAL_CAT[T.9]            -0.0538      0.001    -38.503      0.000      -0.057      -0.051
SITIO_EVENTO_CAT[T.2]          -0.1323      0.001   -103.483      0.000      -0.135      -0.130
SITIO_EVENTO_CAT[T.3]           0.1730      0.001    137.910      0.000       0.171       0.175
SITIO_EVENTO_CAT[T.4]           0.4731      0.006     78.043      0.000       0.461       0.485
SITIO_EVENTO_CAT[T.5]           0.3938      0.006     66.261      0.000       0.382       0.405
SITIO_EVENTO_CAT[T.6]           0.2248      0.015     14.852      0.000       0.195       0.254
SITIO_EVENTO_CAT[T.9]           0.1169      0.002     68.447      0.000       0.114       0.120
TIEMPO_GESTACION_CAT[T.2]      -0.0099      0.001     -6.795      0.000      -0.013      -0.007
TIEMPO_GESTACION_CAT[T.3]      -0.4825      0.001   -439.492      0.000      -0.485      -0.480
TIEMPO_GESTACION_CAT[T.4]      -0.5014      0.001   -457.402      0.000      -0.504      -0.499
TIEMPO_GESTACION_CAT[T.5]      -0.4807      0.002   -247.089      0.000      -0.485      -0.477
TIEMPO_GESTACION_CAT[T.9]      -0.1374      0.001   -149.526      0.000      -0.139      -0.136
TIPO_PARTO_CAT[T.2]            -0.0053      0.000    -26.133      0.000      -0.006      -0.005
TIPO_PARTO_CAT[T.3]             0.0057      0.001      7.459      0.000       0.004       0.007
TIPO_PARTO_CAT[T.9]             0.0307      0.001     29.648      0.000       0.029       0.033
AREA[T.2]                       0.0056      0.001      3.997      0.000       0.003       0.008
AREA[T.3]                      -0.0122      0.001     -8.693      0.000      -0.015      -0.009
AREA[T.9]                      -0.3229      0.005    -59.545      0.000      -0.334      -0.312
NUMERO_EMBARAZOS                0.0029   8.54e-06    333.762      0.000       0.003       0.003
PESO                           -0.0132   7.63e-05   -173.049      0.000      -0.013      -0.013
==============================================================================
Omnibus:                  1539195.849   Durbin-Watson:                   1.999
Prob(Omnibus):                  0.000   Jarque-Bera (JB):         103216441.853
Skew:                           4.710   Prob(JB):                         0.00
Kurtosis:                      41.471   Cond. No.                     4.02e+03
==============================================================================
```

Figure 13. The regression model.

Interpretation of coefficients:

- Sexo T.2 = Male; Sexo T.3 = Indeterminate; Sexo T.9 = Not information; if the baby is an indeterminate or can't define the sex in the moment of evaluation has in average

0.4698 option of died that a baby that has a define sex.
- Seg_social T.2 = Not have insurance services; Seg_social T.3 = subsidy; Seg_social T.4 = exceptional; Seg_social T.5 = especial; Seg_social T.9 = Not information; mother has no have a insurance service has in average more probability that a baby died.
- Sitio_Evento T.3 = the event place is the house; Sitio_Evento T.4 = the event place is public; Sitio_Evento T.5 = the event place is other; Sitio_Evento T.6 = the event place is the job; If the event is in a public street in average has more probability that a baby died.
- Tiempo_Gestacional T.2 = if the pregnancy time is between 22-27 weeks; Tiempo_Gestacional T.3 = if the pregnancy time is between 28-37 weeks; Tiempo_Gestacional T.4 = if the pregnancy time is between 38-41 weeks; Tiempo_Gestacional T.5 = if the pregnancy time is unknown; Tiempo_Gestacional T.9 = Not information; if the mother has a few week of pregnancy time has in average more probability that a baby died.
- Tipo_Parto T.2 = if the born type is cesarean; Tipo_Parto T.3 = if the born type is instrumental; Tipo_Parto T.9 = if the born type is unknown; if born type is for cesarean has less in average probability to die.
- AreaT.2 = the location is small village ; AreaT.3 = the location is rural or farm; AreaT.6 = the location is unknown;
- Número de Embarazos: number of pregnancies of the mother;
- Peso: Weight of the baby.

Although this model could obtain a relatively high R2 score of 0.77, it is still in analysis and development in order to confirm all the statements listed above and just to make sure we are not having data leakage issues.

Interface

The final front-end product will feature ideally three pages: an overview page, an analytics page and a recommendations page. The overview page, figure 14, will have some summarized statistics and maybe some context about Colombia's current situation in relation to this topic. Also, some extra information to allow users to understand all the content. The Analytics page will contain maps showing perinatal and neonatal death rates by department (Colombia's geographical subdivisions), or even to some more geographical detail (it should be at least by department). These maps will have, as well as historical information, all of these plots allowing users to interactively explore among different periods, departments and pregnancy outcomes. The recommendations page will also have some visualizations of the historical data, and graphs showing possible correlations between the pregnancy outcome and the most relevant descriptive variables, as well as some results gotten from the models created.

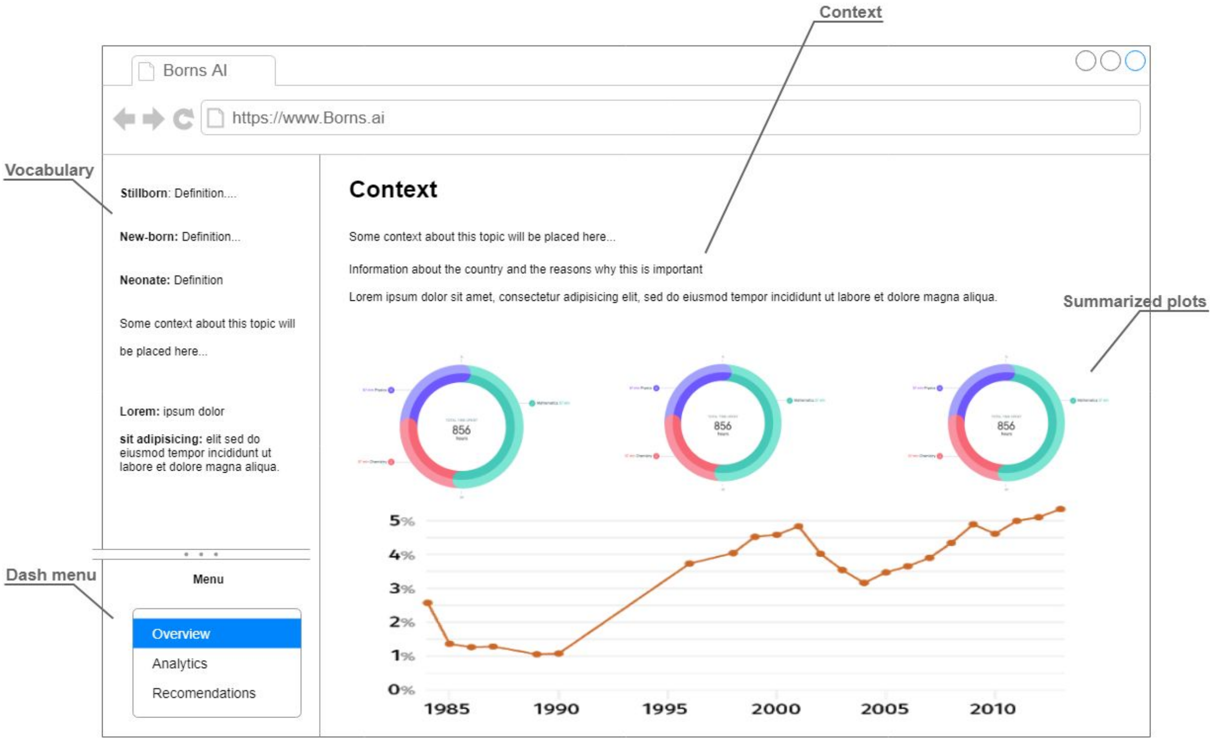

Figure 14. Overview page.

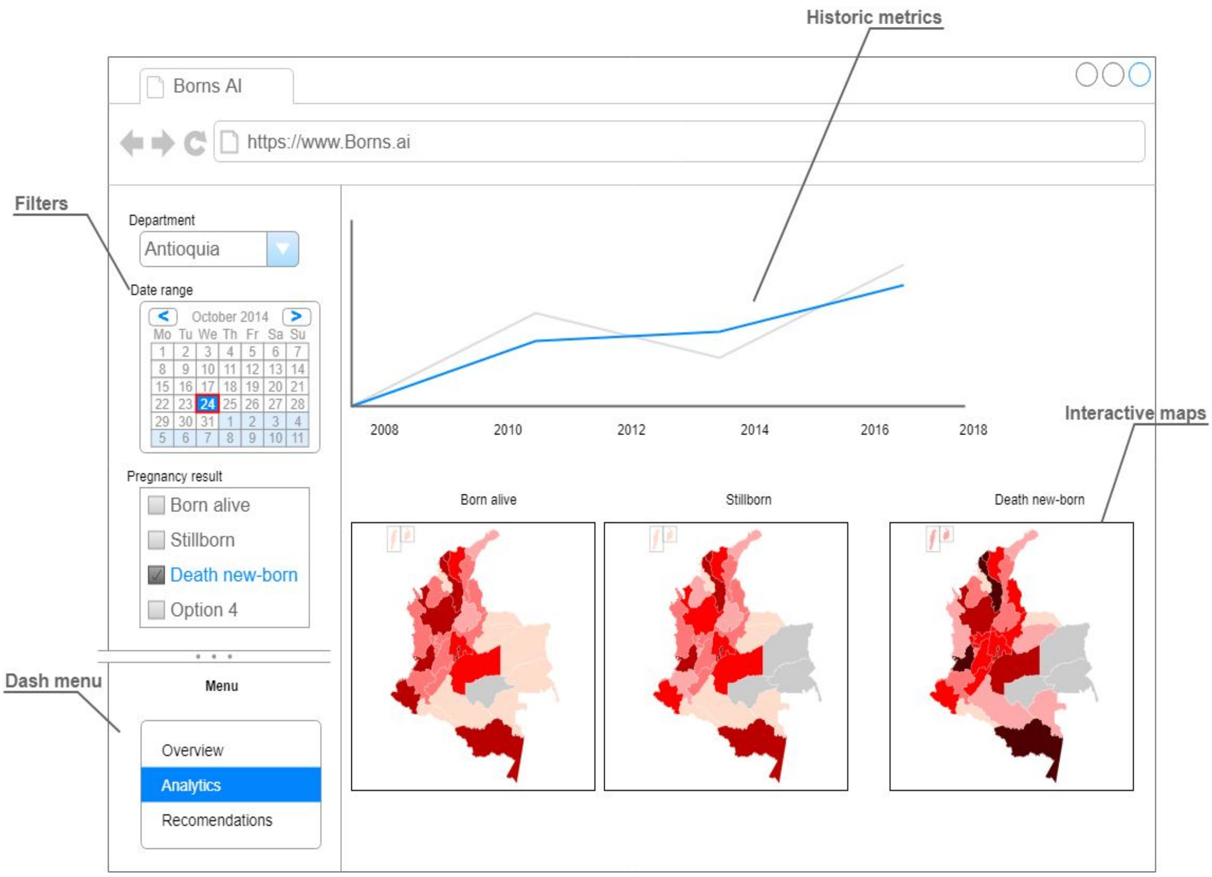

Figure 15. Analytics page.

Final notes.

As of week 7, basic EDA seems to be completed for a subset of the 2018 data. In order to further extend this analysis to the full dataset including data from 2008 and on, an AWS-RDS instance was created and all of this data was uploaded to it. With the advances presented so far, next week will consist mainly on extending this same analysis for the full dataset, some confirmatory data analysis -particularly for the hypothesis introduced in the EDA section-, new -better- modelling, and the development of a fully functional frontend application able to account for the results obtained so far, namely the EDA as well as geographical and correlation analysis, which would successfully give an end to the first version of this project.

Appendix.

a. Dataset description.

An entire dataset of the Colombian population was sourced from the National Administrative Department of Statistics site (DANE for its acronym in Spanish) with geographical, geospatial and demographic features as well as perinatal and neonatal information [3].The unified information gathered has more than 1'855.962 data points about births, as well as fetal and neonatal deaths, described by about 60 or more columns. This could potentially be complemented by the resource distributions through the country and region characteristics, such as water quality, electricity, health facilities and education access, as well as poverty indicators and some others, sourced from the Humanitarian Data Exchange (HDX).

Since we have several possible databases with lots of features that are not directly related to newborn infants and death rates it is challenging to dig into such an amount of information and find relationships that eventually can lead us to the most insightful ones. Moreover, It will be really time-consuming to clean up, summarize and visualize all of that amount of data in order to show it in an organized and user-friendly way.
This dataset includes in its majority a set of categorical variables, which are counted as ordered natural numbers starting from 1 (1,2,3,...).
From this information we can highlight the following fields:

Common fields on all of the datasets (Fetal death, newborn death, alive newborn):

- **Location:** Department, City/Municipality of death or born. Categorical variables with Colombian standardized encoding called Divipola (for its acronym in Spanish of División Político-Administrativa)
- **Habitual residence**: Country, Department, City. Categorical values. Divipola encoding.
- **Gender**: Categorical variable (Male, Female, Unknown).
- **Born type**: Categorical variable (Natural, Cesarean, Instrumental, Ignored, No data)

- **Pregnancy type**: Categorical variable (Normal, Twin, Triple, Multiple, Ignored)**Pregnancy time:** Categorical variable (<22 weeks, <27 weeks, …)
- **Mother's age:** Categorical variable. Encoded in ranges of 4. (10-14,15-19,...)
- **Amount of children alive:** Discrete variable
- **Amount of children not alive:** Discrete variable

Fetal/newborn death

- **Death cause identification method:** Categorical value. (Necropsy, Medical history, Lab. tests, familiar interviews, no data)
- **Death cause:** Categorical variable. Encoded with the International Statistical Classification of Diseases and Related Health Problems, a medical classification list by the World Health Organization (CIE-10)
- **Diagnostic:** Direct, indirect, Precedent medical history, others. Categorical variable. Encoded with CIE-10
- **Probable death:** Categorical variable. (Natural, Violent, In studies)

Alive newborn

- **Weight**: Categorical variable. Encoded on ranges of 0.5 pounds (<1.000, <1.499,...)
- **Height:** Categorical variable. Encoded on ranges of 10 centimeters (<20 , <29, …)

References.

[1] Newborn death and illness, World Health Organization, 2011.
https://www.who.int/pmnch/media/press_materials/fs/fs_newborndealth_illness/en/
[2] Resolution adopted by the General Assembly on 25 September 2015. United Nations, 2015, p. 16 https://undocs.org/A/RES/70/1
[3] COLOMBIA - Estadísticas Vitales. Posted: 24 Jan, 2020. Source; Dane.
http://microdatos.dane.gov.co/index.php/catalog/652/get_microdata