

NO + DROP - OUT = APP

A Data Science Web App to Understand Scholar Desertion in Colombia

Highlights



% of students in afternoon sessions

MOST ADVERSE FEATURES



% of students affected by armed conflict



children and teen fertility rates

Alejandra Castelblanco • Carlos Coy • Daniel Jiménez • Alberto Lagos • Diego Mora • Luis J. Parra

Background



Scholar desertion is an important phenomenon that affects boys, girls and teenagers around the globe. Not attending school severely impacts the individual's possibilities to improve his or her wellbeing.

In Colombia, for every 100 children that start school, only 44 graduate.

Our Proposal: An Interactive Web App

A tool to improve resource allocation and public policy making.



EXPLORE DATA

FIND RELEVANT FEATURES

COMPARE & RANK Municipalities

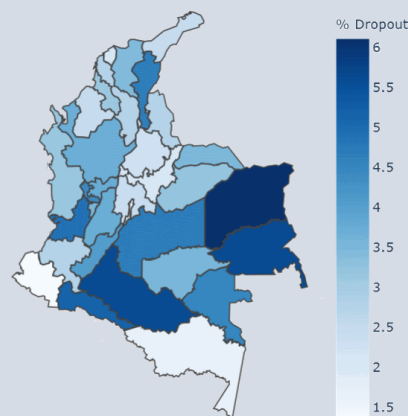
Data processing

Data from multiple dimensions at the municipality level was collected:

- Economical
- Health
- Criminality
- Sociodemographic
- Schools
- Teachers

We combined data from multiple institutions into one single database available for analysis via AWS, PostgreSQL and Python.

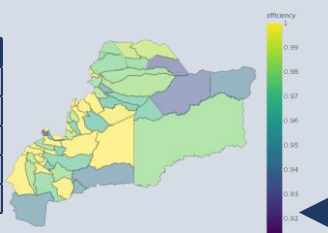
1. Exploring Desertion Across Colombia



Scholar desertion occurs differently across distinct regions. The **Orinoquia** and **Amazon** regions have the highest desertions rates. With the application, it is possible to explore spatial trends at the departmental and municipal level.

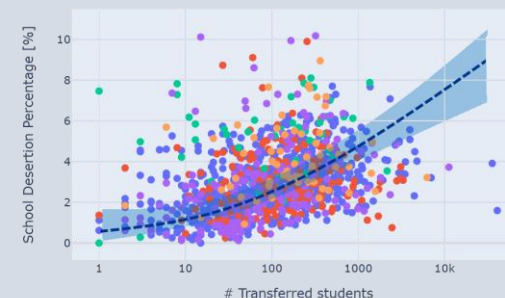
4. Benchmarking Desertion

Municipality	Efficiency
Villavicencio	100.0%
Cubarral	100.0%
Tame	97.7%
S. C. Guaraoa	96.8%
El Calvario	91.3%



Using EDA, we can rank and compare the performance of municipalities regarding desertion. Efficiency rules take a set of inputs (e.g. #teachers and #schools) that produce a desertion rate. The users can vary their definitions of efficiency and compare municipality performance.

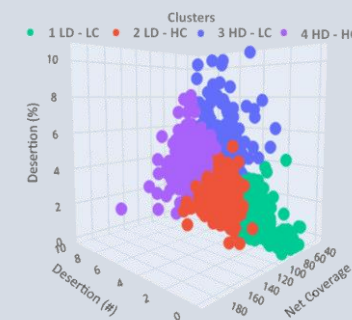
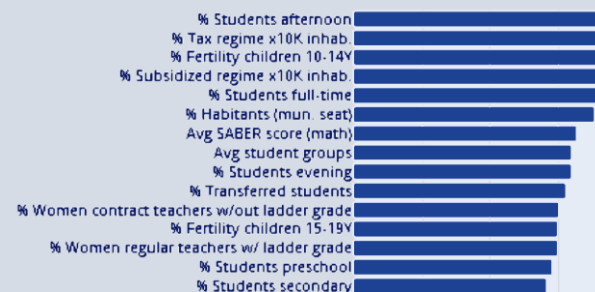
2. Exploring Desertion and its Relationships



With the application, it is possible to analyze relationships between variables and to explore their behavior in time. The goal is to get acquainted with the problem.

3. Finding the Most Important Features

We carried out a K-means clustering for the response variables: desertion and coverage. Then, using Decision trees, XG-Boost and Random Forest classifiers, we identified relevant features related to desertion.



[Link to the App: http://13.59.29.168:5000/](http://13.59.29.168:5000/)