

# Best Schools in Colombia

Diana Barbosa, Yeison Cuitiva and Carmelo Hurtado - UNIANDES

Marcelo Burbano - PRISMAPAR



Fig. 1. Search of best schools: <http://mejorescolegioscolombia.com> ó <http://bestschoolscolombia.com>

**Abstract**— Several publications recognized in Colombia, such as the magazine Dinero and Semana, are publishing school searchers in Colombia according to the results of the PRUEBAS SABER 11°. However, for parents it is not enough in the process of searching for schools to their children, since they need other criteria such as: student process, physical facilities such as laboratories, level of inclusion of students, quality of teachers, among others.

The objective of this implementation is to develop a best school search engine based on a ranking that considers new relevant criteria and that allows parents to perform specific searches according to their particular interests to find the best school for their children.

For the definition of the ranking, the following sources of information were consulted: ICFES: <http://www.icfesinteractivo.gov.co/>, Ministry of Education: <https://www.mineducacion.gov.co>, Colombian Open Data: [www.datos.gov.co](http://www.datos.gov.co) and Synthetic Index of Educational Quality: <http://superate20.edu.co/isce/> and education experts from the Ministry of Education, the District Secretary of Education and Prismapar.

**Index Terms**— Schools, ranking, visual analytics, parents, data quality, childrens

## 1 INTRODUCTION

The implementation of the best school search engine, which includes private as well as public schools in the municipalities of Colombia, is born because there is no tool in Colombia to evaluate schools with a comprehensive vision that includes relevant criteria such as: the importance of the student, the quality of teachers, environments and learning methodologies.

The definition of the ranking starts with an "Ideal Ranking" by the Prismapar expert, then the search is made of the information required

in the ideal ranking, based on the information obtained, the first version is defined, then around 3 are made modifications on the formula for the ranking, product of finding more information or discarding another one because of its quality.

The search engine for best schools has three components, the first related to the distribution of schools according to the result of the ranking, the second allows identifying the characteristics of the ranking and the last, compares the result of the ranking with that obtained by the magazine Dinero.

Parents can make use of the search engine of better schools through the web pages: <http://mejorescolegioscolombia.com> or <http://bestschoolscolombia.com>, there they find the schools ordered by the ranking and can search from filters of calendar, type, number of students and the variation to the ranking according to the needs.

## 2 EXPOSITION

The Project was developed in several stages, the first one was definition meetings with the expert to define the requirement for the visualization, after that, the necessary information was searched for and then the new formulas for the ranking were defined. From this,

- Diana Milena Barbosa Rojas - Master student in information engineering at the Universidad de los Andes. E-mail: [dm.barbosa@uniandes.edu.co](mailto:dm.barbosa@uniandes.edu.co).
- Carmelo Hurtado Orozco - Master student in information engineering at the Universidad de los Andes. E-mail: [c.hurtadoo@uniandes.edu.co](mailto:c.hurtadoo@uniandes.edu.co).
- Yeison Javier Cuitiva - Master student in information engineering at the Universidad de los Andes. E-mail: [yj.cuitiva10@uniandes.edu.co](mailto:yj.cuitiva10@uniandes.edu.co).

work was started based on the Tamara framework to determine the What-Why-How of the visualization and finally it was built in D3.js. Finally some insights were found based on the tasks defined with the expert.

#### Search and data collection:

The first step in the process was the search and collection of data corresponding to the schools and their results in each of the tests Saber 3, 5, 9 and 11. These were downloaded from the page <http://www2.icfesinteractivo.gov.co/> where is the detail of the results of these tests. There it was necessary to download a file for each department of the country and for the period of application of the tests (2018-I 0 2018-II) and then consolidate the information in a single dataset. In addition to this, it was necessary to find the data corresponding to the synthetic index of quality of education in the link <http://superate20.edu.co/isce/> to add this index to the dataset by school. It was also necessary to do data cleaning processes on some attributes that had inconsistent or missing information.

#### Description of the ranking:

The ranking is developed through 3 formulas, the first focused on the PRUEBAS SABER, the second based on the synthetic index of educational quality (ISCE), which evaluates aspects such as: progress, efficiency, performance and school environment, and the third formula related to the inclusion of the school from the students enrolled and those who present the SABER 11. The following explains the detail of each of the formulas used in the calculation of the ranking:

##### 1. First: PRUEBAS SABER

The mathematical formula applied is the following:

$$(Average\ Global\ Score\ 3^{\circ} / 500) + (Average\ Global\ Score\ 5^{\circ} / 500) + (Average\ Global\ Score\ 9^{\circ} / 500) + (Average\ Global\ Score\ 11^{\circ} / 500)$$

With a maximum score of 500 points, according to the rating of each of the PRUEBAS SABER, the percentage obtained by the school in the 3°, 5°, 9° and 11° grades is calculated from the maximum possible value.

##### 2. Second: Synthetic Index of Education Quality

The mathematical formula applied is the following:

$$(ISCE\_Primary + ISCE\_Secondary + ISCE\_Middle) / 30$$

With a maximum score of 30 points, given that the ISCE index is measured over 10 points, the percentage obtained by the school in primary, secondary and middle school is calculated from the maximum possible value.

##### 3. Third: Level of Inclusion

$$(Evaluated\ in\ grade\ 11 * Average\ Global\ Score\ 11) + ((Enrolled - Assessed) * (Average\ Global\ Score\ 11 - deviation\ below\ in\ the\ SABER\ 11\ tests))$$

Considering that the evaluated ones correspond to the average of the score of the PRUEBAS SABER 11 and those not evaluated are below the average of the score of the PRUEBAS SABER 11, the average of the score of global of the students enrolled in the school is calculated.

Finally, to define the ranking, the three results of the formulas are added together and the average is obtained to define the final value of the new ranking.

### 3 WHAT, WHY AND HOW

Based on the Tamara Framework, the what, why and how was defined as follows:

#### 3.1 What

Attributes and data types:

- ranking: Indicates the position of the school in the ranking generated. Type of attribute: quantitative, sequential ordering
- dane\_code: Indicates the unique DANE code assigned to the school. Type of attribute: Ordinal, sequential ordering
- City: Indicates the city where the school is located. Type of attribute: categorical.
- Department: Indicates the department where the school is located. Type of attribute: categorical.
- School: Name of the school. Type of attribute: categorical.
- Sector: Sector to which the school belongs (Official or Private). Type of attribute: categorical.
- Calendar: Calendar to which the school belongs (A or B). Type of attribute: categorical.
- Matriculados: Number of students enrolled in grade 11 for a given period (2018-I or 2018-II). Type of attribute: quantitative, sequential ordering
- Publicados: Number of students who took the SABER 11 exam for a given period (2018-I or 2018-II). Type of attribute: quantitative, sequential ordering
- Puesto\_saber\_11: Position that the school occupies in the tests Saber 11 2018 (Including first and second period). Type of attribute: quantitative, sequential ordering
- Promedio\_puntaje\_global\_saber\_11: average score obtained by the school in the tests Saber 11. Type of attribute: quantitative, sequential ordering
- Desviación\_general: standard deviation obtained by the school in the tests Saber 11. Type of attribute: quantitative, sequential ordering
- Isce\_media: score obtained in the synthetic index of educational quality for the courses corresponding to secondary education. Type of attribute: quantitative, sequential ordering
- Isce\_secundaria: score obtained in the synthetic index of educational quality for the courses corresponding to secondary education. Type of attribute: quantitative, sequential ordering
- Isce\_primaria: score obtained in the synthetic index of educational quality for the courses corresponding to primary education. Type of attribute: quantitative, sequential ordering
- Saber\_9\_general: average score obtained by the school in the tests Saber 9. Type of attribute: quantitative, sequential ordering
- Saber\_5\_general: average score obtained by the school in the tests Saber 5. Type of attribute: quantitative, sequential ordering
- Saber\_3\_general: average score obtained by the school in the tests know 3. Type of attribute: quantitative, sequential ordering
- Formula\_1: Value obtained for formula 1  $(Promedio\_puntaje\_global\_saber\_11 + Saber\_9\_general + Saber\_5\_general + Saber\_3\_general) / 4$ . Type of attribute: quantitative, sequential ordering
- Formula\_2: Value obtained for formula 2  $(Isce\_media + Isce\_secundaria + Isce\_primaria) / 3$ . Type of attribute: quantitative, sequential ordering
- Formula\_3: Value obtained for formula 3  $(Promedio\_puntaje\_global\_saber\_11 * Publicados) + ((Matriculados - Publicados) * (Promedio\_puntaje\_global\_saber\_11 - Desviación\_general)) / Matriculados$ . Type of attribute: quantitative, sequential ordering
- Total: Total score obtained by the school in the new ranking  $(Formula\_1 + Formula\_2 + Formula\_3) / 3$ . Type of attribute: quantitative, sequential ordering

- **Ranking dinero:** Average score obtained by each school in the ranking of the Dinero magazine. Type of attribute: quantitative, sequential ordering

The availability of this Dataset is static.

### 3.2 Why

The following tasks were defined:

Main tasks:

1. Summarize the distribution of the best schools in Colombia, from the generation of a new ranking (Summarize - Distribution).
2. Compare the result for each of the formulas of the new ranking of schools according to the filter made by the user. (Compare – Features)
3. Compare the result of the new ranking with that obtained by the money magazine based on the scores obtained by each in the SABER 11 tests (Compare - Distribution).

Secondary tasks:

1. Calculate the total score for each defined formula of the new ranking by School (Derive - Features).
2. Derive the position of each school in the new ranking (Derive - Features).
3. Determine the difference between students enrolled / registered by each school and those evaluated to identify if this influences the results obtained (Derive - Trends).

### 3.3 How

1. Summarize the distribution of the best schools in Colombia, from the generation of a new ranking (Summarize - Distribution).

Mark: Line  
Channels: hue - parts of the ranking  
Idiom: Horizont Bar Chart  
Encode: Express  
Manipulate: Select  
Reduce: Filter

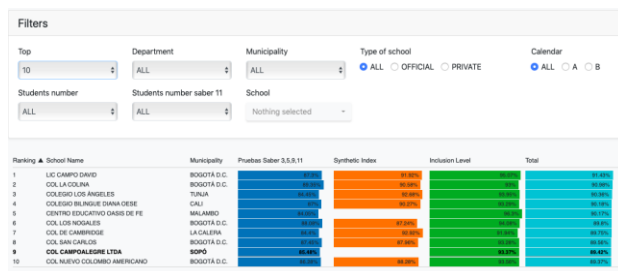


Fig. 1. Visualización para la tarea 1.

2. Compare the result for each of the formulas of the new ranking of schools according to the filter made by the user. (Compare – Features)

Mark: Line  
Idiom: Parallel coordinates  
Encode: Express  
Map: Hue, Luminance.  
Manipulate: Change  
Reduce: Filter  
Facet: SuperImpose  
Channel: Vertical position, Color Hue

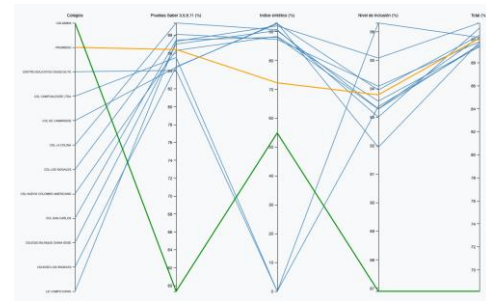


Fig. 2. Visualización para la tarea 2.

3. Compare the result of the new ranking with that obtained by the money magazine based on the scores obtained by each in the SABER 11 tests (Compare - Distribution).

Mark: Line  
Channel: hue - ranking / revista Dinero  
Idiom: Group Bar Chart  
Encode: Separate Order and Align  
Manipulate: Select – Shows detail of the name and ranking of the School  
Reduce: Filter

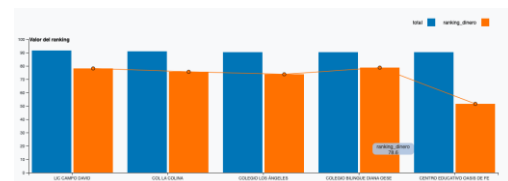


Fig. 3. Visualización para la tarea 3.

### 3.4 Insights

For each of the tasks, the following insights were found:

Task 1: Summarize the distribution of the best schools in Colombia, from the generation of a new ranking (Summarize - Distribution).

- The best performing schools in the new index are mostly private schools that are located in Bogotá, Tunja, Cali, La Calera, Sopo, Floridablanca, Envigado, Chia, Ibagué and Pamplona.
- There are only 2 official colleges in the top 100, 10 colleges in the top 500 and 47 in the top 1000.
- There is a higher number of Calendar B colleges in the top 50.

Task 2: Compare the result for each of the formulas of the new ranking of schools according to the filter made by the user. (Compare – Features)

- In the three new formulas (Evidence Saber, Synthetic Index and inclusion level) the performance of the best schools is significantly higher than the average of the country.
- The best schools have a higher level of inclusion and better test results than others, for the synthetic index the majority is between 70% and 90%, finding that the difference is made in the inclusion level and the results of the tests.

Task 3: Compare the result of the new ranking with that obtained by the money magazine based on the scores obtained by each in the SABER 11 tests (Compare - Distribution).

- The ranking is higher for all schools than that obtained in the magazine Dinero, but the order does change considerably.

#### **4 CONCLUSION**

We presented a new and simple search engine to select the best schools in Colombia, and deal with the problem of many parents, an inherent problem throughout the country.

Colombia's best schools search engine allows users enter data according to your need; department, municipality, type of school, calendar, number of students, number of students that presented the Knowledge tests and name of the schools are the input parameters available in our application.

For the generation of the ranking, the search engine for the best schools in Colombia takes into account the most important criteria of an educational environment, result of the tests know of the primary and baccalaureate, the synthetic index of quality and the level of inclusion.

We compare the ranking of the best schools in Colombia with the ranking generated by the magazine Dinero and the results show that our ranking assigns higher scores to the schools, additionally, it allows the generation of a dynamic ranking according to specific criteria

The results are impressive, we will be able to satisfy the needs of many families from Colombia and facilitate such a complex task with a few clicks, guaranteeing an exact election and based on official information of the educational entities of the country, these are ICFES, Ministry of Education, Government of Colombia and Education Bogotá.

#### **ACKNOWLEDGMENTS**

The authors wish to thank John Alexis Guerra Gomez.