COLUMBIA
SIPA
School of International and Public Affairs

RIO
PREFEITURA

BIG DATA EDUCATION
An Exploratory Study of Rio de Janeiro City Public School System

Alberto Zeraik
Bruno Bondarovsky
Eduardo Padua
Fernando Ivo Cavalcante
Luiz Eduardo Ricon
Victor Zajdhaft

LÍDERES CARIOCAS

January, 2015

OBJECTIVE:

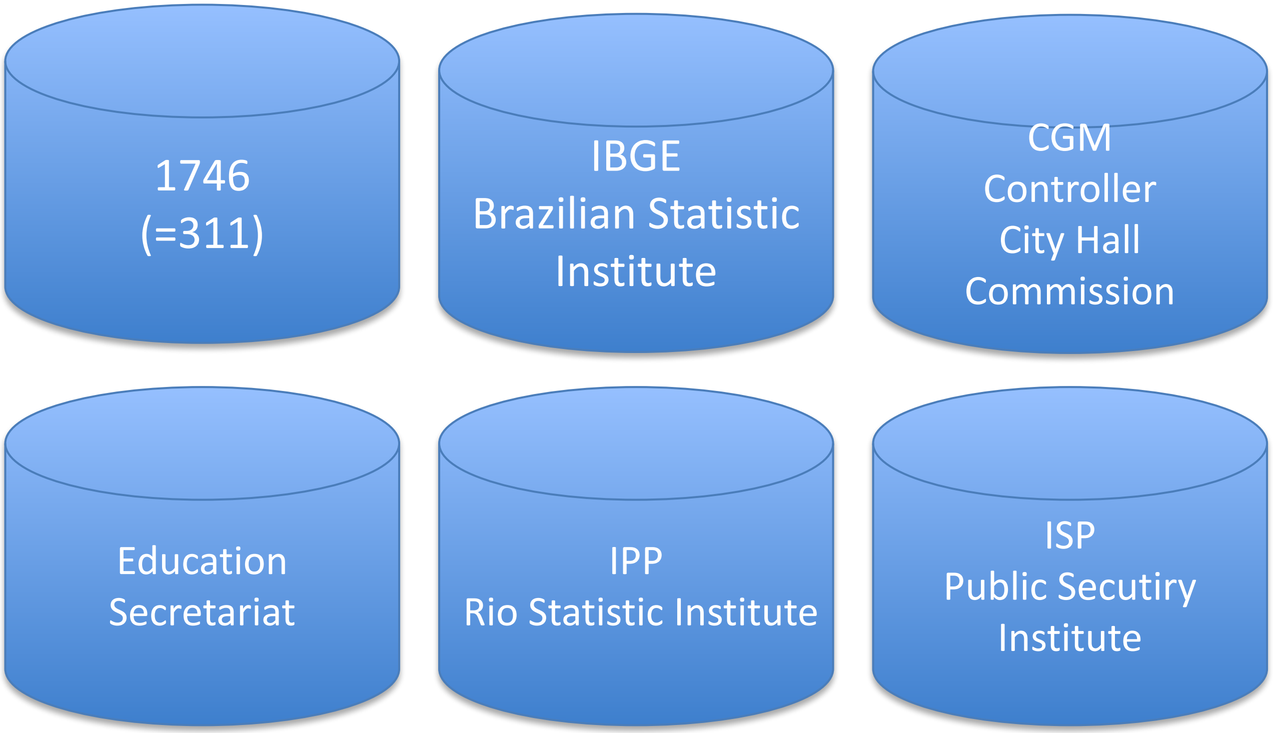
Propose a methodology based on the big data technology in order to identify behaviors or conditions that might impact on school and student performance.

The goal of this proposal is to offer policy makers and public education managers a new tool to support their decision making process in order to achieve higher educational outcomes.

METHODOLOGY

- Database Identification
- Indicators definition
- Data Analysis:
    - Multiple Linear Regression of the indicators for each school, based on performance
    - Multiple Linear Regression of the greater impact indicators
    - Comparision between the highest impact indicators in groups of the 20 best and the 20 worst schools in performance

DATABASES



1746
(=311)

IBGE
Brazilian Statistic Institute

CGM
Controller
City Hall
Commission

Education
Secretariat

IPP
Rio Statistic Institute

ISP
Public Secutiry
Institute

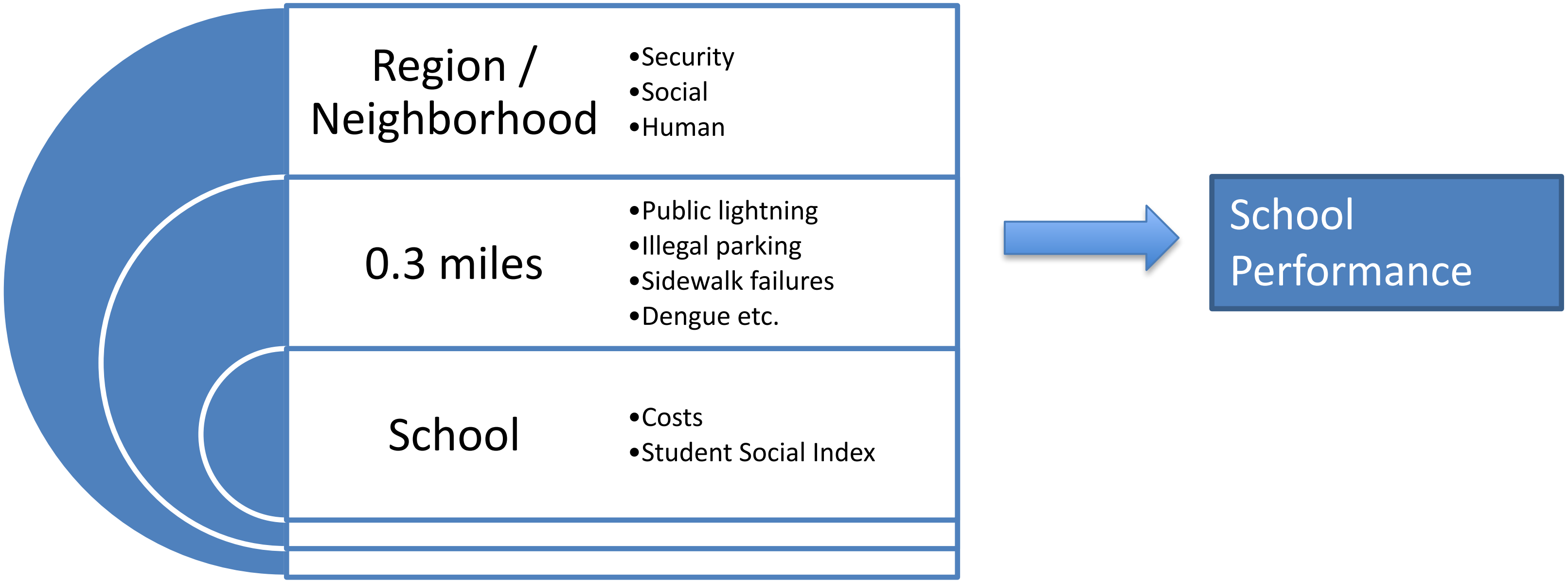
INDICATORS

INDICATORS	SOURCE	RECORDS
IDH, IDH-L, IDH-E, IDH-R (per Neighborhood)	IBGE - Brasilian Statistic Institute	130 x 4
IDS (per Neighborhood)	IPP - City Data Institute	226
Variance and Average student-school distance	Education Secretariat	24000 x 2
Social Economic Condition of students (per School)	QEDU Website	885
Average Distance age-school grades (per School)	QEDU Website	885
Annual student cost (per School)	CGM	1500
Dengue fever (500m radius from school)	1746 (=311)	940
Potholes fix requests (500m radius from school)	1746 (=311)	940
Public Illumination requests (500m radius from school)	1746 (=311)	940
Illegal parking requests (500m radius from school)	1746 (=311)	940
Security assistance requests per school	1746 (=311)	940
Public security occurrences per neighborhood	Public Security Institute	152
IDEB - Brazilian Education Performance index	Education Secretariat	1450
IDE-RIO - RIO's Education Performance index	Education Secretariat	1450

COLUMBIA SIPA School of International and Public Affairs

RIO PREFEITURA

DATA ANALYSIS



Region / Neighborhood

- Security
- Social
- Human

0.3 miles

- Public lightning
- Illegal parking
- Sidewalk failures
- Dengue etc.

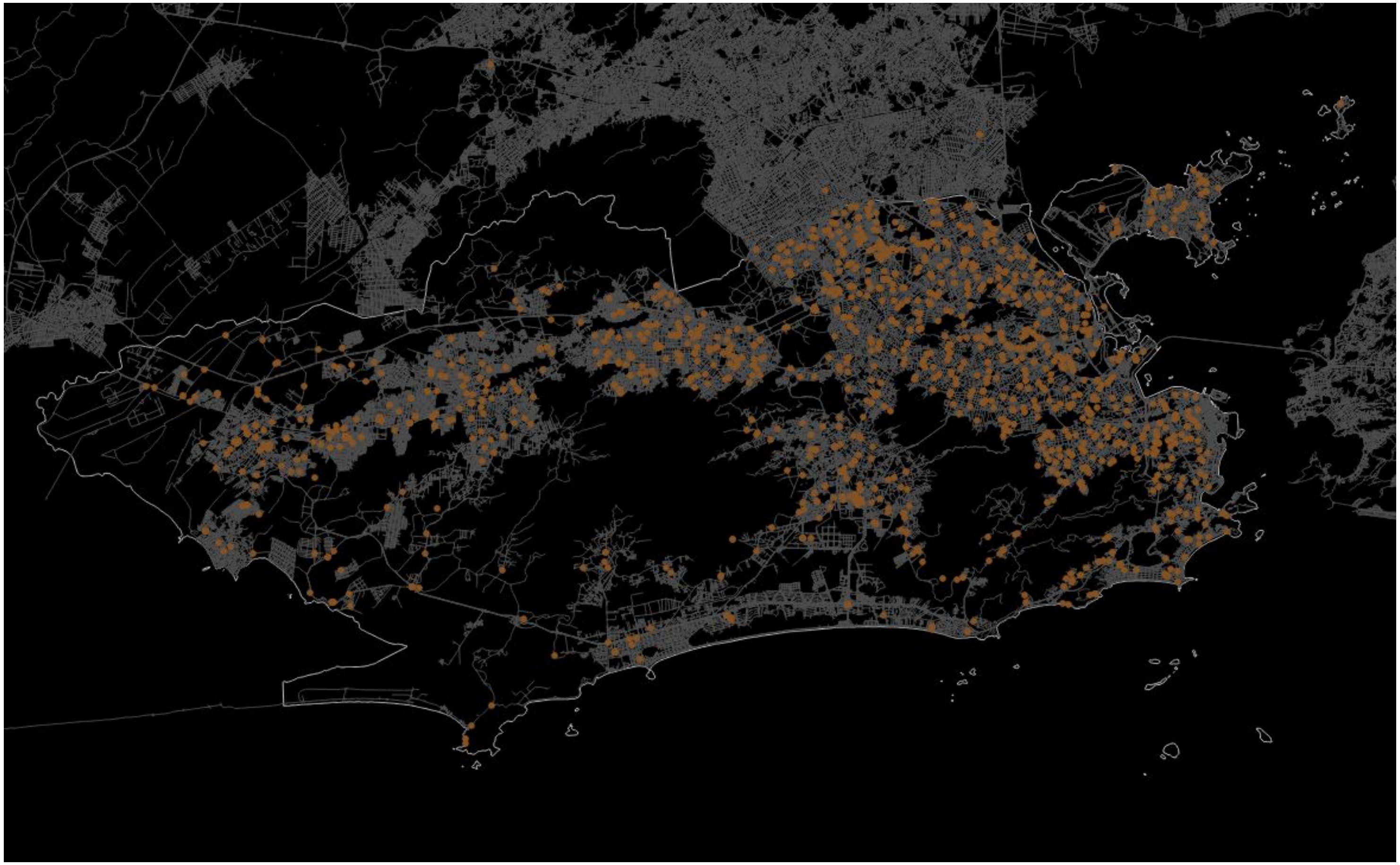
School

- Costs
- Student Social Index

School Performance

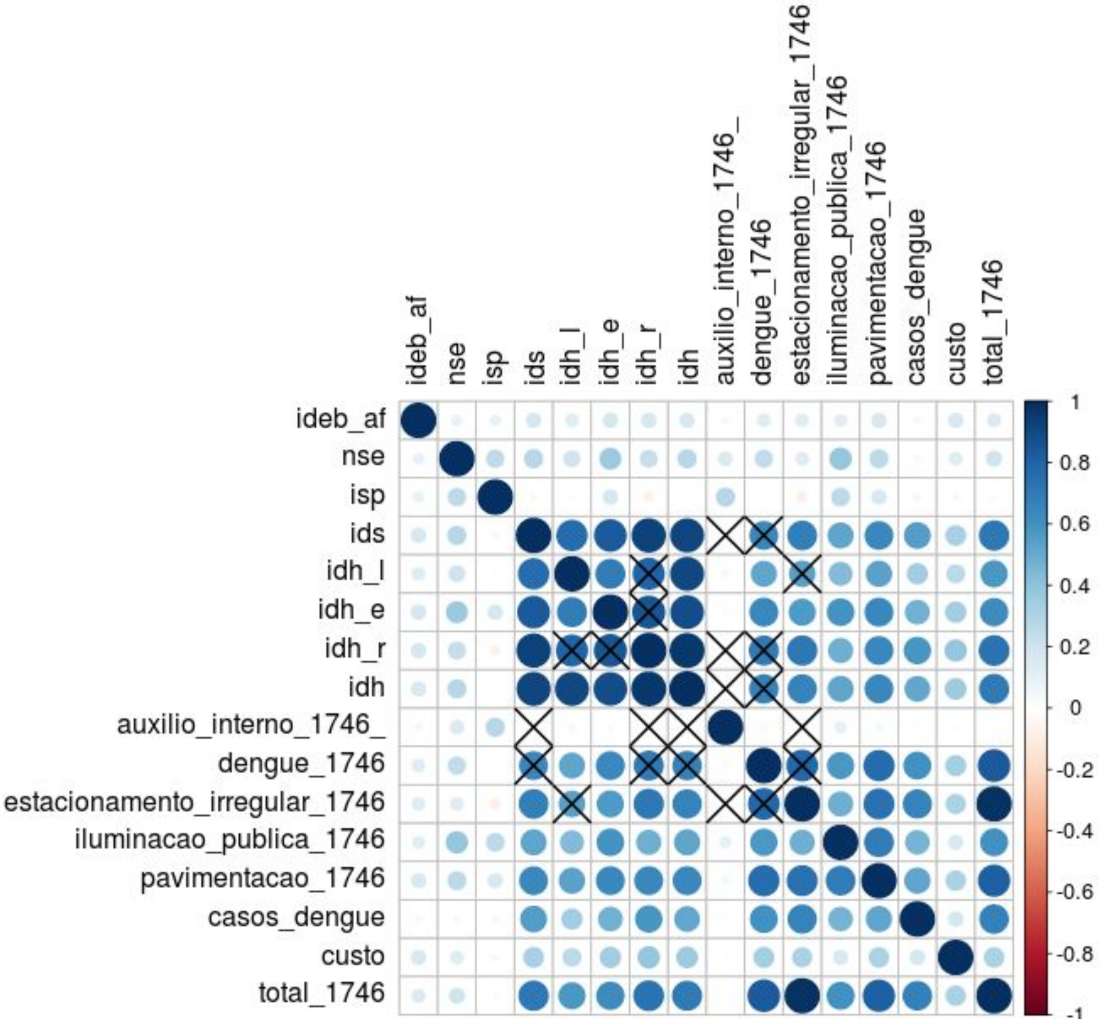
ANALYSIS RESULTS

- Schools geografically distributed on Rio de Janeiro's territory



ANALYSIS RESULTS

- Multiple Lin
on School perf

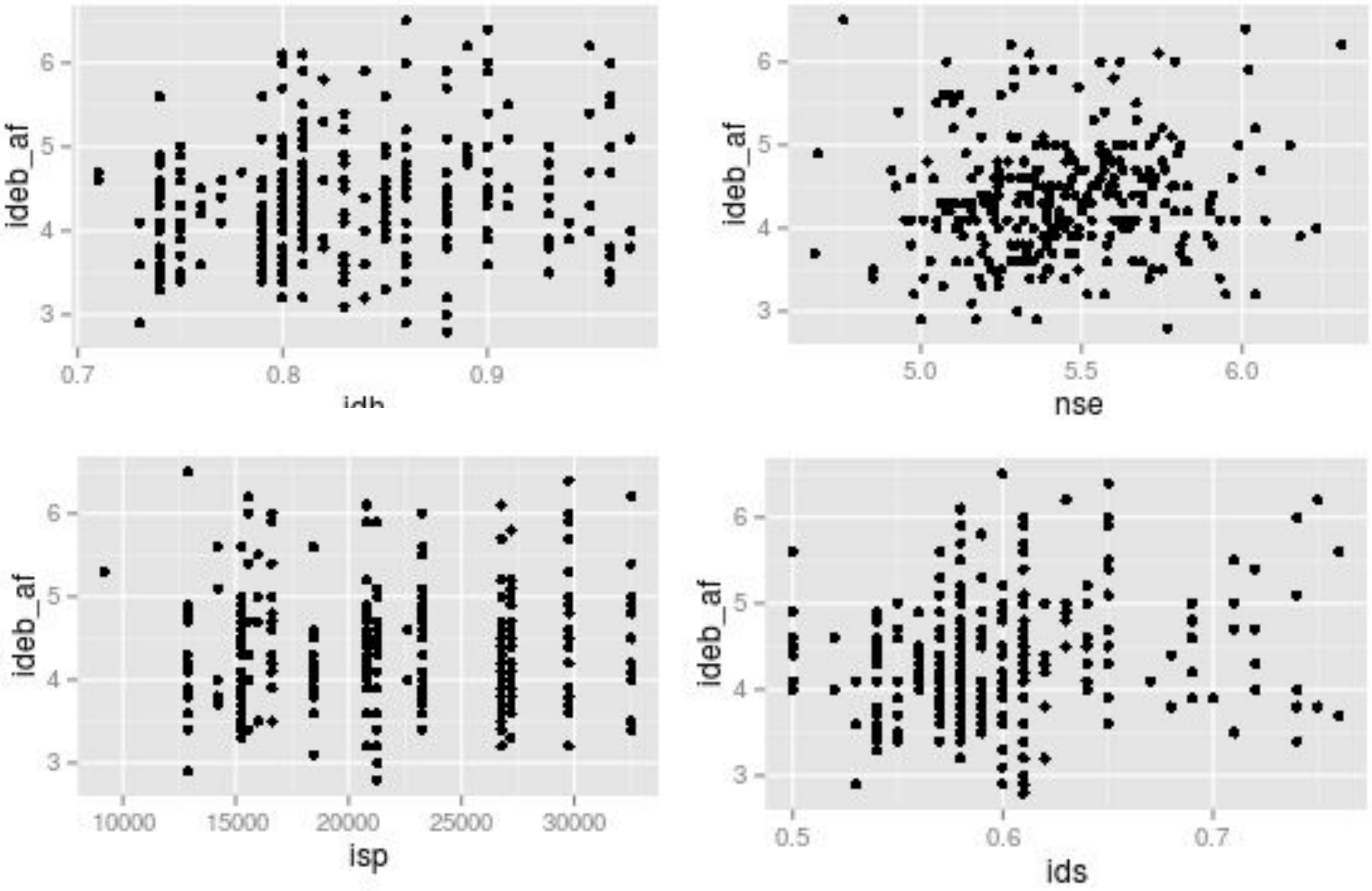


COLUMBIA
SIPA
School of International and Public Affairs

RIO
PREFEITURA

ANALYSIS RESULTS

- Multiple Linear Regression based on School performance



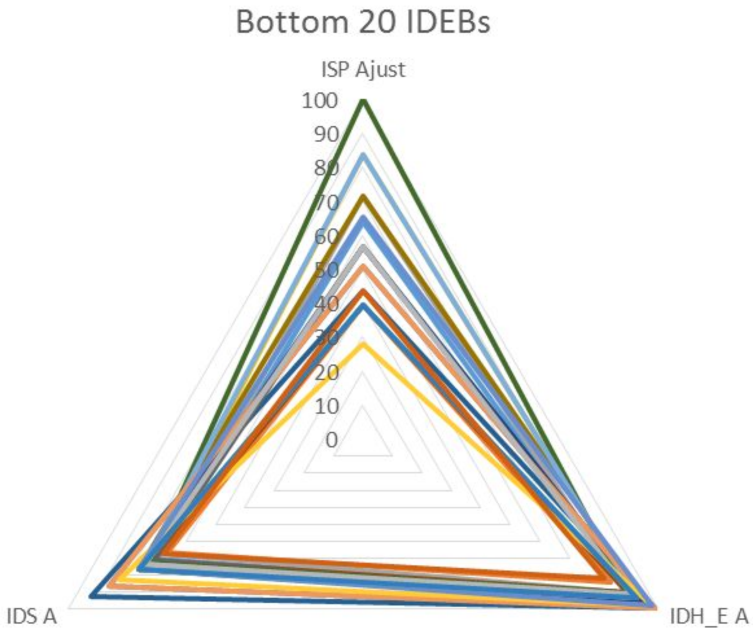
COLUMBIA
SIPA
School of International
and Public Affairs

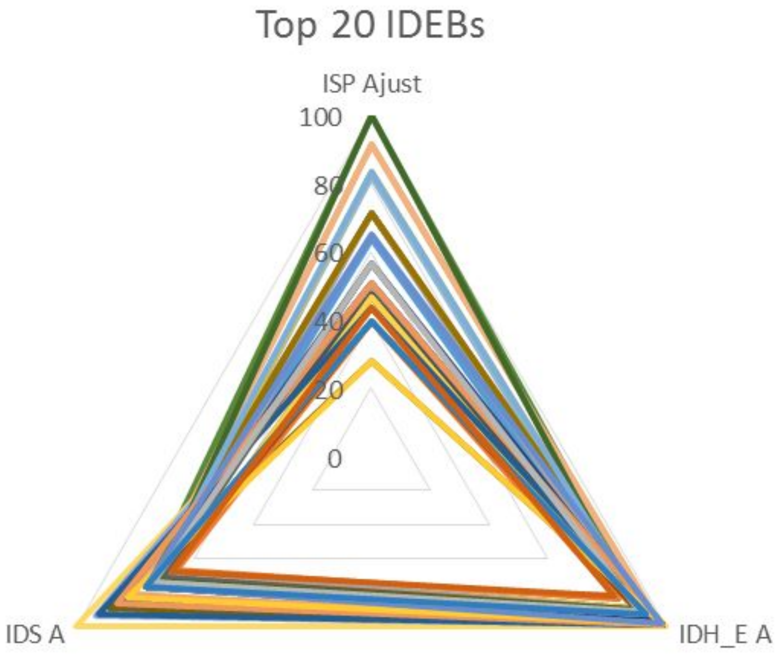
RIO
PREFEITURA

ANALYSIS RESULTS

- Variables Overview comparison: Top/Bottom 20

  Non significant variables

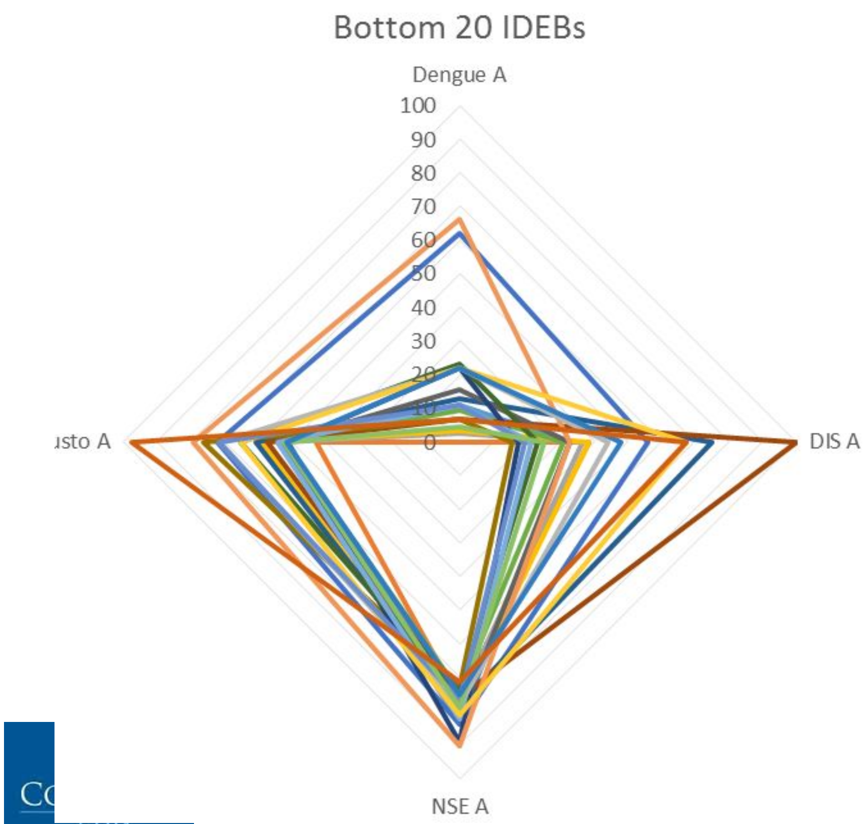


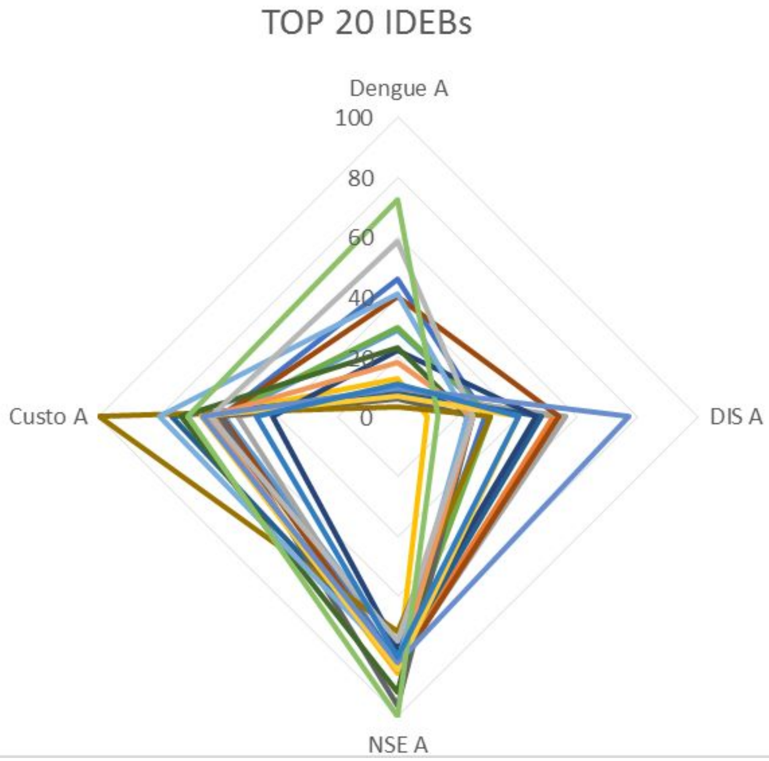


ANALYSIS RESULTS

- Variables Overview comparison: Top/Bottom 20

  Potencially significant variables





CONCLUSIONS

The sample used and the applied methodology showed no clear correlation between the data analyzed and school performance. That suggest that outside school environment factors do not have strong interference in student performance.

Recommendations:

- Use inner school indicators: assets, climate, principal performance, teachers, parents and community engagement
- Focus environment indicators on near school surrounding
- Avoid aggregated indicators
- Use a more rich and complete sample

LEARNED LESSONS

The experience showed us it is possible to reach better results if we go deeper crossing others indicators.

In a few weeks, we put together 6 City Hall areas, the BIG DATA PENSA Group, hundred of thousands of records crossed through 6 different databases.