Cluster Analysis of School Performance in Mexico City

Leon Palafox[1][0000−0003−3448−5133] and Daniel Villegas[2]

[1] Universidad Panamericana, Facultad de Ingenieria, Mexico lpalafox@up.edu.mx
[2] Universidad Panamericana, Facultad de Ingenieria, Mexico
dan.villegas.a@gmail.com

Abstract. This paper will briefly discuss the current education problem in Mexico City and some factors that play a role in its development.

Keywords: Education · Machine Learning · Clusterization

1 Introduction

As many others around the world, Mexico's education system presents a problem that afflicts the performance of it's students as well as its overall quality. However, the educational problem obeys to particular functions of each country, state and locality.

Many factors could be accounted to have an influence in the evaluation of a school inside its district, including students performance, the amount of students evaluated, the municipality in which it is located, the overall budget it possess to invest in quality education and hiring of teaches as well as socioeconomic factors such as the cost of living in the area where it is located, etc.

Having a way to visualise and identify this problem becomes a major point in trying to mitigate an education deficit amongst regional government and policy makers. This becomes a complex sociocultural problem that involves many dimensions of the schools of a region as well as the timeframe in which the study is made with many dynamic factors playing a key role in which schools are afflicted and can help map out and identify them.

In this work we test different Machine Learning techniques to try to identify what are the primary factors in the performance of schools in Mexico City, the applications of this works may help in a future to create new public policy in the education and further studies that try to model a solution for this problem.

It becomes important as well to consider the framework from which we will evaluate the performance, so as to set a standard which can be later replicated in studies outside of Mexico City.

By analyzing the performance of students, as well as their economic situations, we can create more efficient strategies to combat educational backwardness.

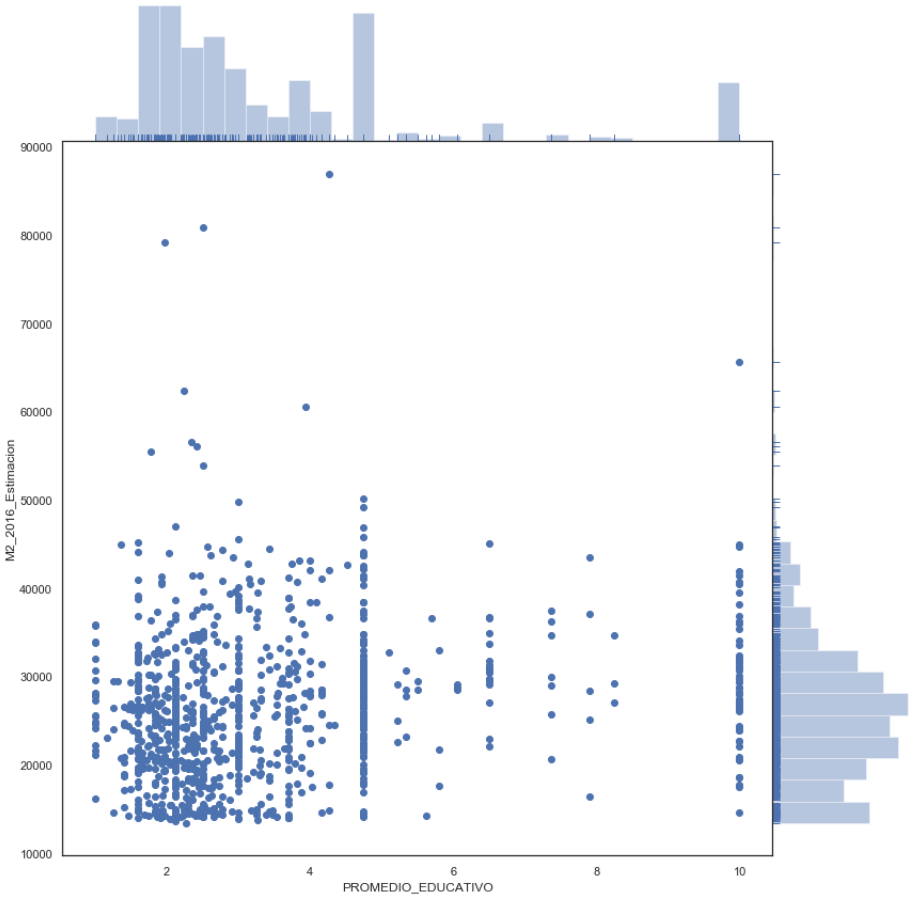

Fig. 1. Education mean against cost of living per square meter in Mexico City's neighborhoods.

1.1 Sec.1

2 Methodology

We extracted extensive data from different sources of information,then we tested multiple Machine Learning algorithms to try and find patterns in the information.

2.1 Methodology

Like we mentioned before, the geographical coverage of this work is limited to Mexico City using the open data policies to study the performance of several school levels in an international exam which measures both written language and math abilities of students.

2.2 Data

The different data sources we used are referenced in the Figure **??**.

- Inventario Nacional de Viviendas 2016: National inventory of living quarters
- Censo de Poblacin y Vivienda: national census in Mexico
- SCINCE: System to query the census information
- DENUE: National directory of economic activities
- Softec: Private database with information regarding the cost of the quare meter.
- Plan Nacional para la Evaluacin de los Aprendizajes (PLANEA): National framework exam withinformation of school performance in written language and maths.

3 Algorithm Description

Information about schools performance, as well as geographical data was scrapped from the PLANEA web page and database accessed publicly. Then, as the clusterization method selected an evaluation for K-means cluster was estimated.

4 Results

References

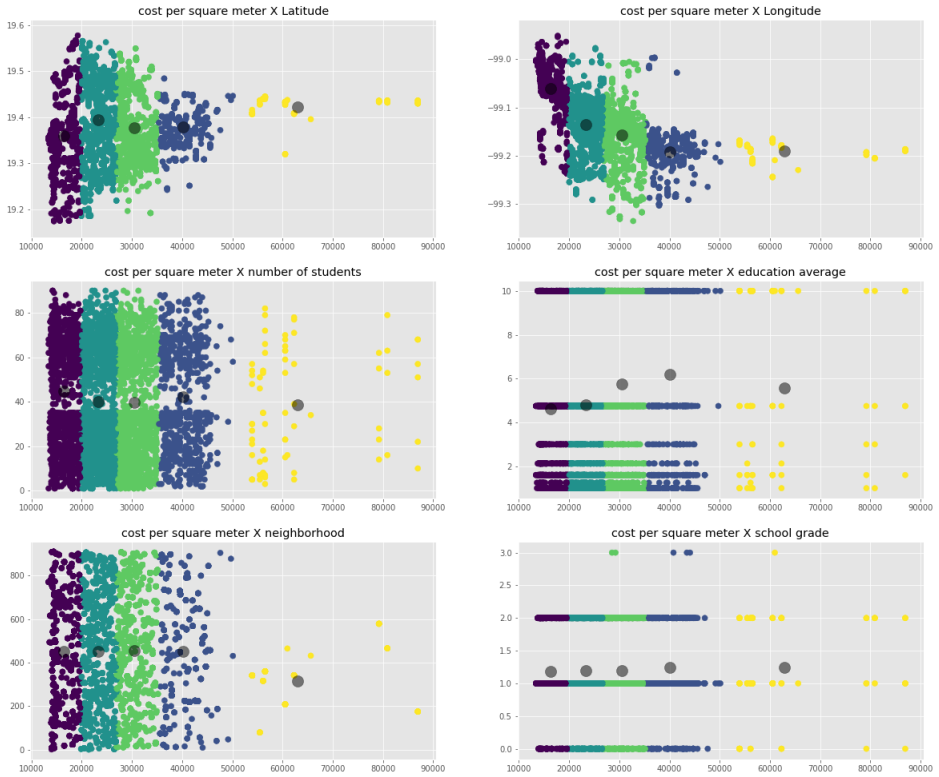

Fig. 2. Cross plotting of cost per square meter against other features after clusterization.

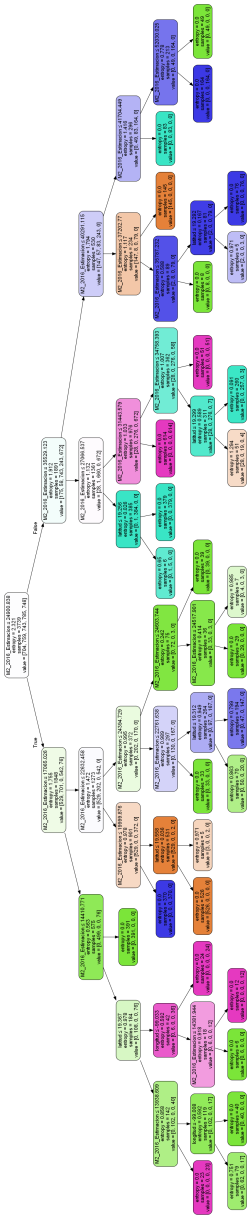

Fig. 3. Desition Tree created from the cluster output.

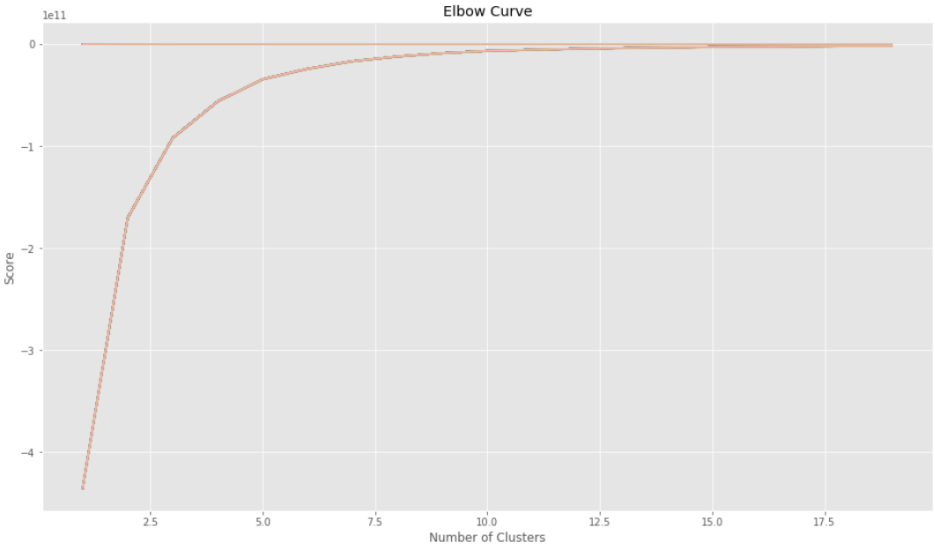

Fig. 4. Elbow Curve generated for the scrapped data

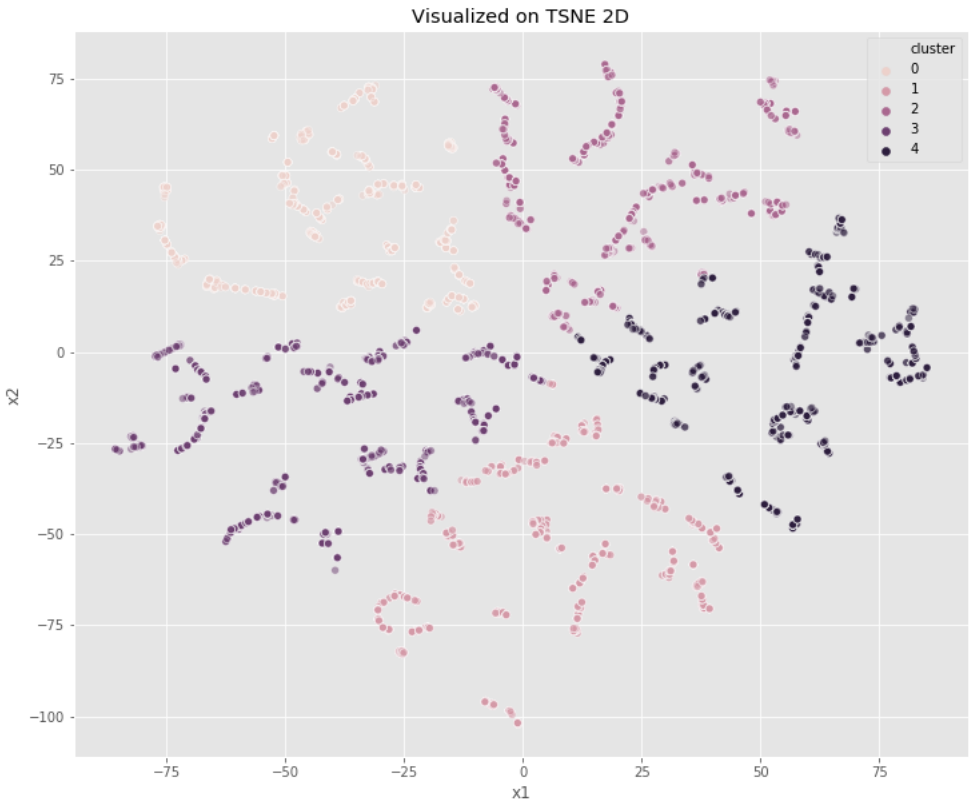

Fig. 5. TSNE algorithm applied to cluster results for visualization in 2D.