

# Análisis del rendimiento académico y las incidencias externas en Argentina

clusterAI

Mauro Dario Corigliano, Santiago Manuel Sosa, Matias Grosso Basto

## INTRODUCCIÓN

A partir de la encuesta realizada en el 2018 por el gobierno de la nación denominada "Aprender", que consistía en evaluar el nivel académico en Matemática y Lengua, y, en paralelo, una serie de preguntas que daban dimensión del entorno en el cual desarrollaban sus estudios, se planteó mediante este análisis la búsqueda de una posible relación entre variables propias del entorno del encuestado, que influyen en el respectivo rendimiento en matemáticas.

Según datos oficiales, participaron de esta encuesta el 94% de las escuelas primarias del país en 19.600 establecimientos educativos.

## OBJETIVO

El objetivo es evaluar las incidencias de distintas variables culturales, económicas y socioambientales en el desempeño en la asignatura matemática de alumnos de sexto grado de los colegios de Argentina.

## DATASET

El dataset elegido posee aproximadamente 580.000 filas y 124 columnas. Las filas representan las distintas muestras (samples:alumnos) que realizaron dichas encuestas, y las columnas muestran los resultados obtenidos en los puntos a evaluar para cada una de las etiquetas (labels). De las 124, columnas para este estudio, se consideraron como las punto de interés las siguientes:

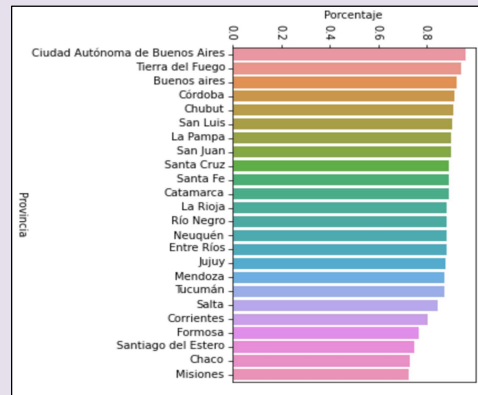
Nacionalidad	Provincia	Edad	Sexo
Nota Matemática	Nota Lengua	Ámbito (Púb o priv)	Sector (Rural o urb)
Asistirá al Secundario	Tiene celular	Repetió	Educación de la madre
Educación del padre	Cantidad de libros que hay en la casa	Tiempo de viaje	Viaje Secundario

## REFERENCIAS

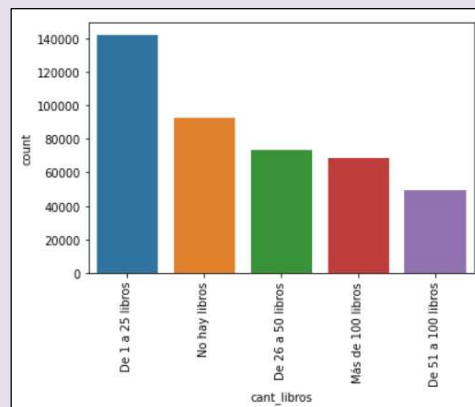
1. Randal S. Olson [olsonran@upenn.edu](mailto:olsonran@upenn.edu) and Jason H. Moore [jhmoore@upenn.edu](mailto:jhmoore@upenn.edu) Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA, USA
2. <https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html>
3. <https://stackoverflow.com/questions/41667397/interactive-boxplot-with-pandas-and-jupyter-notebook>

## ANÁLISIS EXPLORATORIO DE DATOS

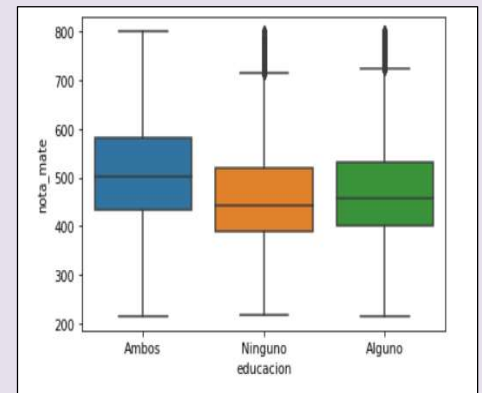
Evaluamos qué porcentaje de padres completaron al menos el primario. En el gráfico vemos que la variable supera el 80% en casi todas las provincias, y en aquellas donde no lo hace, se encuentra cercano al 70%.



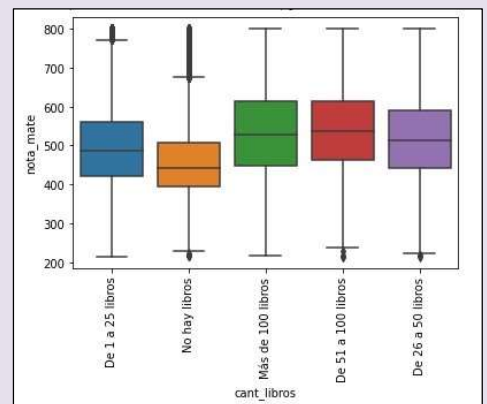
Finalmente, analizamos la incidencia de la disponibilidad y acceso a libros por parte de los encuestados en su rendimiento en matemáticas. Para esto, se analizó la cantidad de libros con los que cuentan los encuestados en sus hogares :



Y finalmente evaluamos la variación del desempeño en matemáticas en base a la formación primaria de los padres de los encuestados :



De esta forma, vemos que la media del desempeño en matemática es considerablemente menor para los alumnos que no tienen libros en sus casas y un poco mayor en los alumnos que tienen hasta 25 libros. Es decir, que la disponibilidad de libros en los hogares, incide en el desempeño de los estudiantes.



## RESULTADOS

Como hemos podido sustentar en el presente informe, los resultados indican que existe una tendencia a obtener mejores notas en matemática cuando se forma parte de un hogar con padres con estudios iniciales concluidos y además, el apoyo con libros que puedan encontrarse en los ámbitos donde crecen los encuestados, incide en su desempeño de forma notable.

## CONCLUSIONES

Como conclusión encontramos que este análisis de datos podría continuar con un modelo de predicción de desempeño en matemáticas en Machine Learning a partir de ciertos inputs, o un modelo no supervisado de clustering que nos permita separar variables automáticamente y agrupar según similitudes. Ambos casos fueron evaluados por el equipo, pero el score obtenido en el entrenamiento no fue satisfactorio, por lo cual no se incluyeron dichas conclusiones. Tal como mencionamos en el apartado de materiales y métodos, el dataset no posee información suficiente (features relevantes) para poder predecir el desempeño en matemática de un alumno (variable categórica) mediante variables de su entorno social (por ejemplo, educación de los padres, tiempo de viaje a la escuela, etc).