

Análisis sobre la insuficiencia del desarrollo educacional con relación al rendimiento académico y otras variables

Integrantes:

- Rodolfo Cruz
- Nicolás San Martín
- Benjamín Santander
- Daniela Victorero

Github: <https://github.com/dvictorerol/proyecto-data-science>

Motivación:

Deseamos saber si existe una relación entre la situación socioeconómica del individuo y su capital cognitivo, para tomar medidas que permitan mejorar y equiparar el desarrollo cognitivo del individuo, así como su acceso considerando el contexto de reformas educacionales y pandemia.

En primer lugar, se define capital cognitivo como “la suma de los saberes adquiridos a consecuencia del capital cultural de origen más las oportunidades educacionales formales” (Rosas, 2013, pág. 55).

En segundo lugar, según Pierre Bourdieu, se define capital cultural como “al conjunto de bienes culturales disponibles, transmitidos de generación en generación, los que explican las diferencias de los resultados escolares y en la ubicación relativa en la pirámide social.” (Rosas, 2013, pág. 15)

Antecedentes:

1. Libro: Rosas, & Santa Cruz, C. (2013). *Dime en qué colegio estudiaste y te diré que CI tienes : radiografía al desigual acceso al capital cognitivo en Chile*. Eds. UC.

Audiencia:

Investigadores del área educacional, así como padres de familia que estén interesados en el desarrollo educacional del país.

Objetivos:

1. El objetivo de la investigación es responder a cómo se encuentra el nivel de desarrollo educacional del país, para poder ofrecer proyecciones y/o propuestas sobre el problema de la insuficiencia educacional a nivel nacional.
2. Los beneficios y/o accionables del proyecto, van relacionados, en primer parte, a una toma de conciencia respecto a la situación educativa del país, para así, en segunda parte, lograr presentar argumentos claves en base a datos oficiales y open source, para los fines que se estimen convenientes.

Datos:

Los datos a utilizar en nuestro proyecto son provenientes de bases de datos abiertas.

Se adjuntan links a las bases de datos

1. Mineduc, <https://centroestudios.mineduc.cl/datos-abiertos/>
2. DEMRE, <https://investigador.demre.cl/>
3. INE, <https://stat.ine.cl/?lang=es>
4. Agencia de Calidad de la Educación, <https://informacionestadistica.agenciaeducacion.cl/#/bases>

Descripción de los datos:

Para el caso de los datos del Mineduc, que serán los datos fundamentales, generalmente se presentan en formato .csv.

El tipo de los datos generalmente son de tipo int64 y object.

Por Ejemplo:

Las siguientes imágenes representan como se encuentran estructurados los datos en la Agencia de Calidad de la Educación:

Características generales			
Nombre archivo	ldps8b2019_rbd		
Descripción	Tabla con puntajes indicadores de desarrollo personal y social a nivel de colegio del proceso SIMCE 2019		
Número de variables	18		
Número de observaciones	5.961		

Volver al índice

Variables			
Nombre	Descripción	Valores posibles	Tipo
rbd	Rol base de datos del establecimiento		Numérico
agno	Año de la evaluación		Numérico
grado	Grado en el que se rinde la evaluación	Ver tabla grado	Cadena
nom_rbd	Nombre del establecimiento		Cadena
cod_reg_rbd	Código de región del establecimiento año 2019		Numérica
nom_reg_rbd	Nombre de la región del establecimiento año 2019		Cadena
cod_pro_rbd	Código de provincia del establecimiento año 2019		Numérica
nom_pro_rbd	Nombre del departamento provincial del establecimiento año 2019		Cadena
cod_com_rbd	Código de comuna del establecimiento año 2019		Numérica
nom_com_rbd	Nombre de la comunas del establecimiento año 2019		Cadena
nom_deprov_rbd	Nombre Deprov del establecimiento año 2019		Cadena
cod_depe2	Código de dependencia 3 categorías	Ver tabla dependencia 2	Numérica
cod_grupo	Código de grupo socioeconómico	Ver tabla GSE	Numérica
cod_rural_rbd	Código de ruralidad del establecimiento	Ver tabla Ruralidad	Numérica
ind_am	Puntaje Indicador Autoestima académica y motivación escolar del colegio	0-100	Numérico
ind_cc	Puntaje Indicador Clima de Convivencia escolar del colegio	0-100	Numérico
ind_pf	Puntaje Indicador Participación y formación ciudadana del colegio	0-100	Numérico
ind_hv	Puntaje Indicador Hábitos de vida saludable del colegio	0-100	Numérico

Dependencia 2	
Valor	Descripción
1	Municipal
2	Particular subvencionado
3	Particular pagado

GSE	
Valor	Descripción
1	Bajo
2	Medio bajo
3	Medio
4	Medio alto
5	Alto

Ruralidad	
Valor	Descripción
1	Urbano
2	Rural

Grado	
Valor	Descripción
4b	4° básico
6b	6° básico
8b	8° básico
2m	II medio

Podemos observar que el formato entregado en los archivos de la Agencia de Calidad de la Educación describe las variables de estudio involucradas y su clasificación cuando corresponda.

Para el caso de los datos entregados por el Mineduc también se presenta una estructura de los datos similar, es decir, se entregan las variables de todos los Dataset respectivos:

Nombre	Tipo	Descripción	Valores
GD_ALU	Numérico	Indica si el alumno es atendido en un grupo diferencial ⁸	0: No está en grupo diferencial 1: Si está en grupo diferencial
COD_REG_ALU	Numérico	Código de la región de residencia del estudiante.	0: Sin Información
COD_COM_ALU	Numérico	Código oficial ⁹ comuna de residencia (autodeclarado y voluntario)	0: Sin Información
NOM_COM_ALU	Cadena	Nombre de la comuna de residencia	
COD_RAMA	Numérico	Código de rama (sólo media técnico-profesional y artística)	Ver Anexo V
COD_SEC	Numérico	Código de sector económico (sólo media técnico-profesional y artística)	Ver Anexo VI
COD_ESPE	Numérico	Código de especialidad (sólo media técnico-profesional y artística)	Ver Anexo VI
FEC_ING_ALU	Numérico	Fecha de ingreso del alumno al establecimiento	
PROM_GRAL	Numérico	Promedio general anual ¹⁰	
ASISTENCIA	Numérico	Porcentaje anual de asistencia	
SIT_FIN	Cadena	Situación de promoción al cierre del año escolar:	P: Promovido ¹¹ R: Reprobado Y: Retirado en blanco: Sin información
SIT_FIN_R	Cadena	Situación de promoción al cierre del año escolar, con indicador de traslado:	P: Promovido R: Reprobado Y: Retirado T: Traslado ¹² en blanco: Sin información
COD_MEN	Cadena	Mencion	Ver anexo X

Además, indica si estas variables se encuentran en todos los Datasets de distintos años, algo que es útil para poder realizar una automatización en la obtención y limpieza de los datos.

[illegible]

En general los datos serán obtenidos vía descarga, aunque es posible optimizar su obtención.

Preguntas de investigación:

1. ¿Es la educación una inversión al largo plazo?
2. ¿Existe una relación entre el nivel socioeconómico del estudiante y capital cognitivo?
3. ¿Influye la ubicación en el rendimiento académico?
4. ¿Cuáles son las regiones con mayor rendimiento académico?
5. ¿Es acaso la inversión al estudiante un factor que determine el éxito académico?

Diseño Tentativo:

- Para el análisis de los datos esperamos utilizar la librería Seaborn para poder obtener un resumen de los datos.
- A su vez se utilizarán las librerías Sklearn, este se utilizará para realizar regresiones lineales, polinomiales u múltiples si es necesario para los datos, Numpy para poder utilizar variables numéricas como Arrays.
- También se hará uso de pandas para la creación de los DataFrames.
- Estadísticamente se hará uso de diversos gráficos que ofrece seaborn, para poder analizar la correlación de los datos, medidas de dispersión, así como todo el análisis de datos