

Educación Secundaria Argentina

Factores Condicionantes al Desempeño Académico

Matias Grouman, Buenos Aires 2021

Matias Grouman

- Licenciado en Administración de Empresas (UADE)
- Experiencia en Industria Retail (Kimberly Clark)
- Investigación Independiente



“El analfabeto del futuro será quien no sepa cómo aprender”

Alvin Toffler (Sociólogo, Nueva York 1928 - 2016)

MUNDO Francia Medida

Francia prohíbe el uso de celulares en salas de clases y recreos

Los estudiantes menores de 15 años no podrán usar teléfonos móviles, tablets y relojes inteligentes en las escuelas. La disposición de casilleros para guardar los aparatos, así como las sanciones, dependerán de cada institución.

Valentina Jofré 5 SEP 2018 04:30 AM



“La datificación es quizá la fuerza más poderosa del nuevo mundo educativo digital”

Ben Williamson (Investigador, Reino Unido)

Objetivos

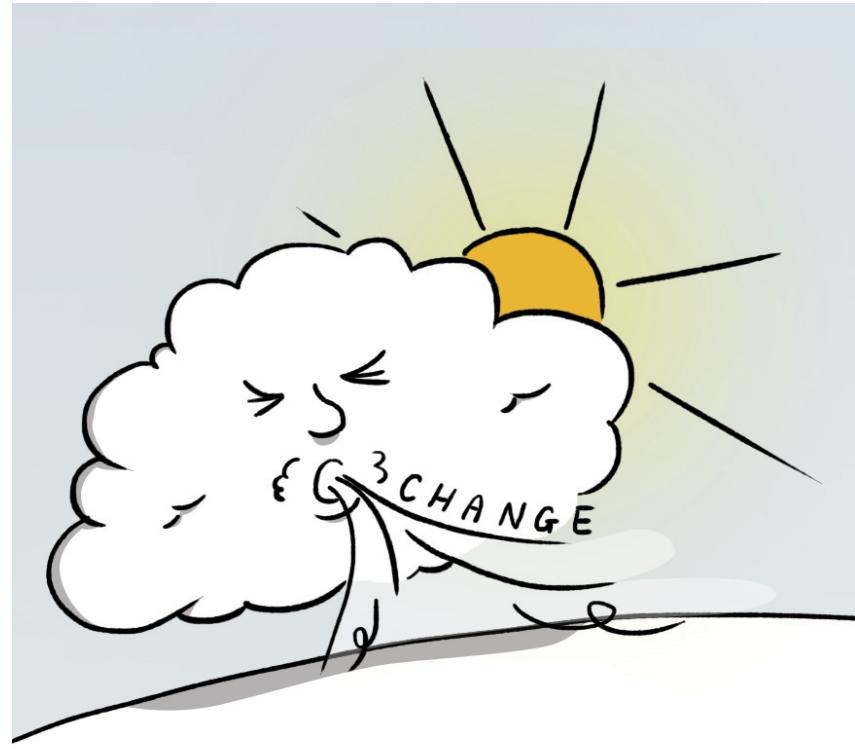
- **Identificar factores relevantes que hacen un desempeño mejor en los alumnos del último año de secundaria**
- Estudiar la relación entre el nivel socioeconómico y el desempeño académico
- Relevar la percepción de las tecnologías
- Determinar la relevancia del nivel educativo de los padres en el desempeño académico
- Brindar campos de estudio en próximas encuestas

Agenda

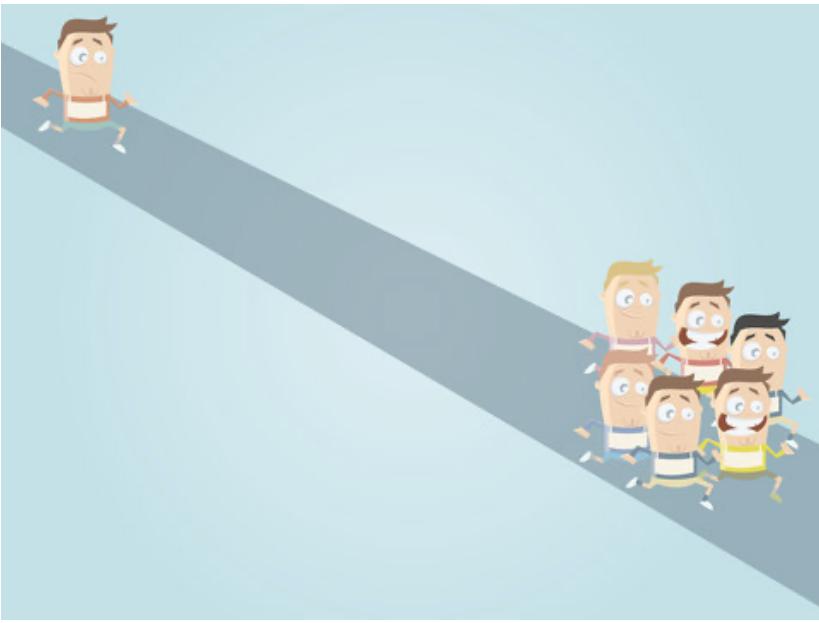
- Motivo del estudio
- Trabajos Realizados
- ¿Qué es Aprender? - Datos disponibles
- Exploración y Transformación
- Modelo aplicado
- Conclusiones
- Preguntas

Motivo del Estudio

¿Por qué aplicar Ciencia de Datos en la Educación?



globalización
futuro
cambios



educación
docentes
tecnologías



lógica
comprensión
matemáticas

Trabajos Realizados

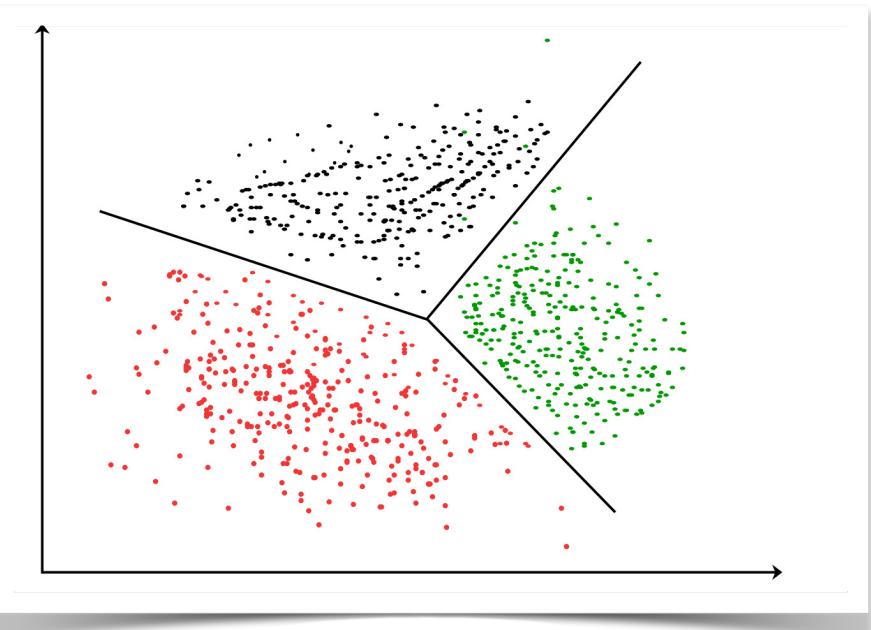
Aplicaciones de Data Mining en Educación

Exploración



- Exploración Aprender 2016 (Templado I.)
- Aprender 2016 Primaria y Secundaria (Llach & Cornejo)
- Investigaciones Sociales (Fundación UCA)

Clustering



- Segmentación de alumnos y docentes por conocimientos tecnológicos (Clemens, Varela & Co.)
- Personalización de contenido (Romero & Ventura)

Predictión



- Predicción del Desempeño y Sugerencia sobre estudios terciarios (Baradwaj & Mal)

Aprender Argentina

Datos disponibles

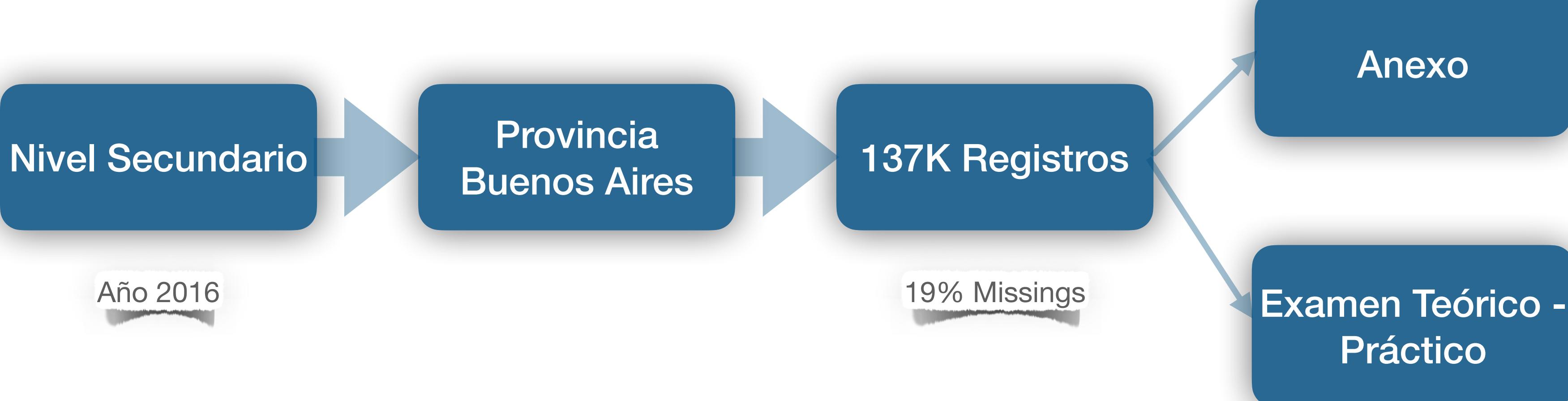


Aprender

PISA

Encuesta censal que busca medir el desempeño de los alumnos en Matemática, Lengua y Ciencias.

El análisis involucra a los alumnos del último año de Primaria y Secundaria, de alcance Nacional.



- Socio demográfico
- Socio económico
- Acceso a TIC
- Trayectoria escolar
- Clima escolar

Aprender Argentina

Datos disponibles



Cuestionario Anexo

- Respondido por el alumno
- 163 variables
- Preguntas de opción múltiple
- Variables ordinales



Selección de Buenos Aires

- Provincias con distintas realidades
- Menor reducción posible del dataset
- Factores externos controlados

Exploración y Transformación

Preparación de los Datos

1. Datos Geográficos:

Unión de base con datos del Instituto Geográfico Nacional, creación variable Subcategory

2. Categorización de Preguntas:

Se categorizaron las preguntas del cuestionario anexo según notas metodológicas

3. Imputación de Faltantes:

Agrupación de respuestas “no sabe”, “no contesta”, etc.

4. Tratamientos Previos:

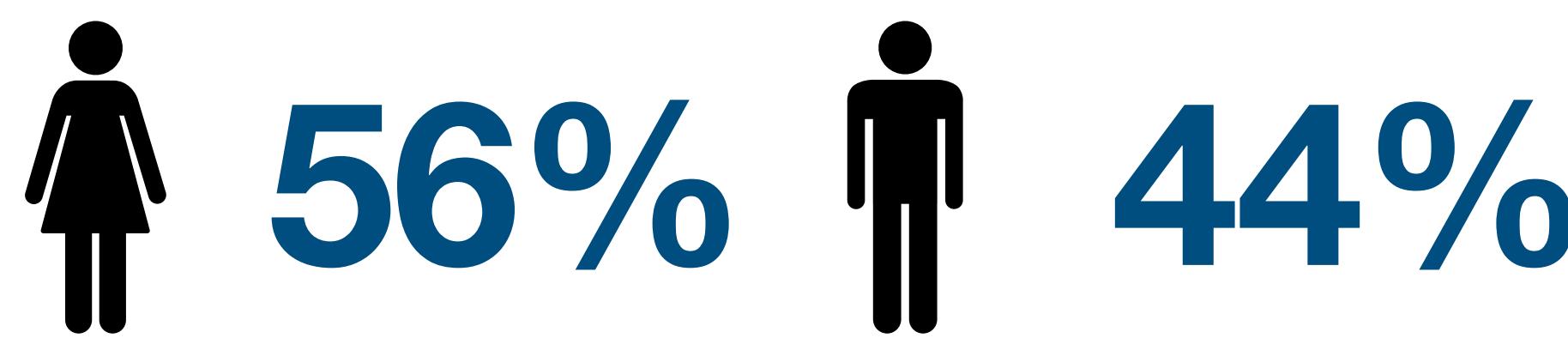
El dataset cuenta con un trabajo de missings donde 50% o más fue descartado y sino imputado con regresiones.

5. Descomposición del Nivel Socioeconómico:

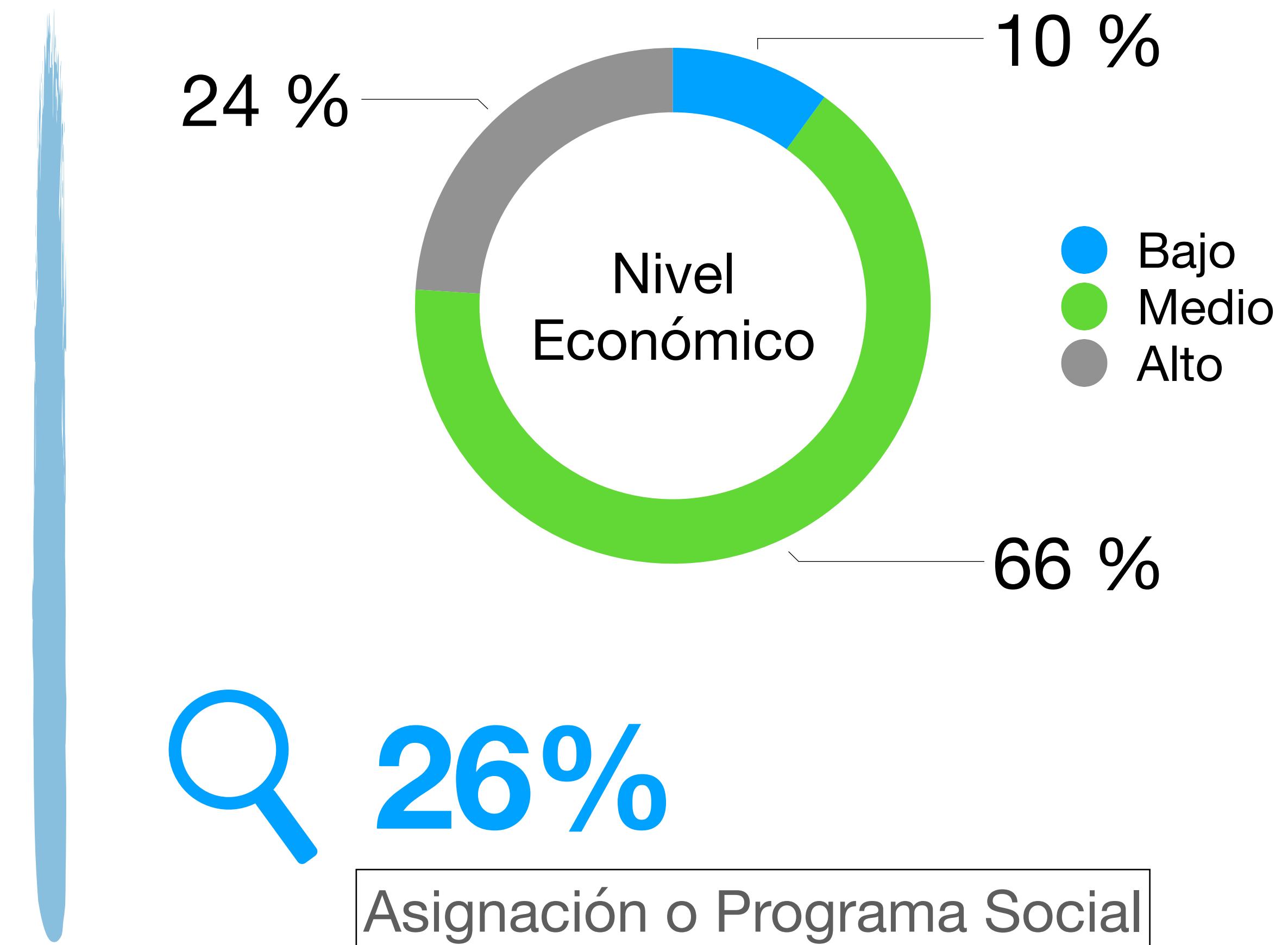
Se construye en base al cuestionario anexo siendo entrenado con bases del INDEC y validado con correlaciones de Spearman.

Exploración y Transformación

¿Quiénes componen el dataset?

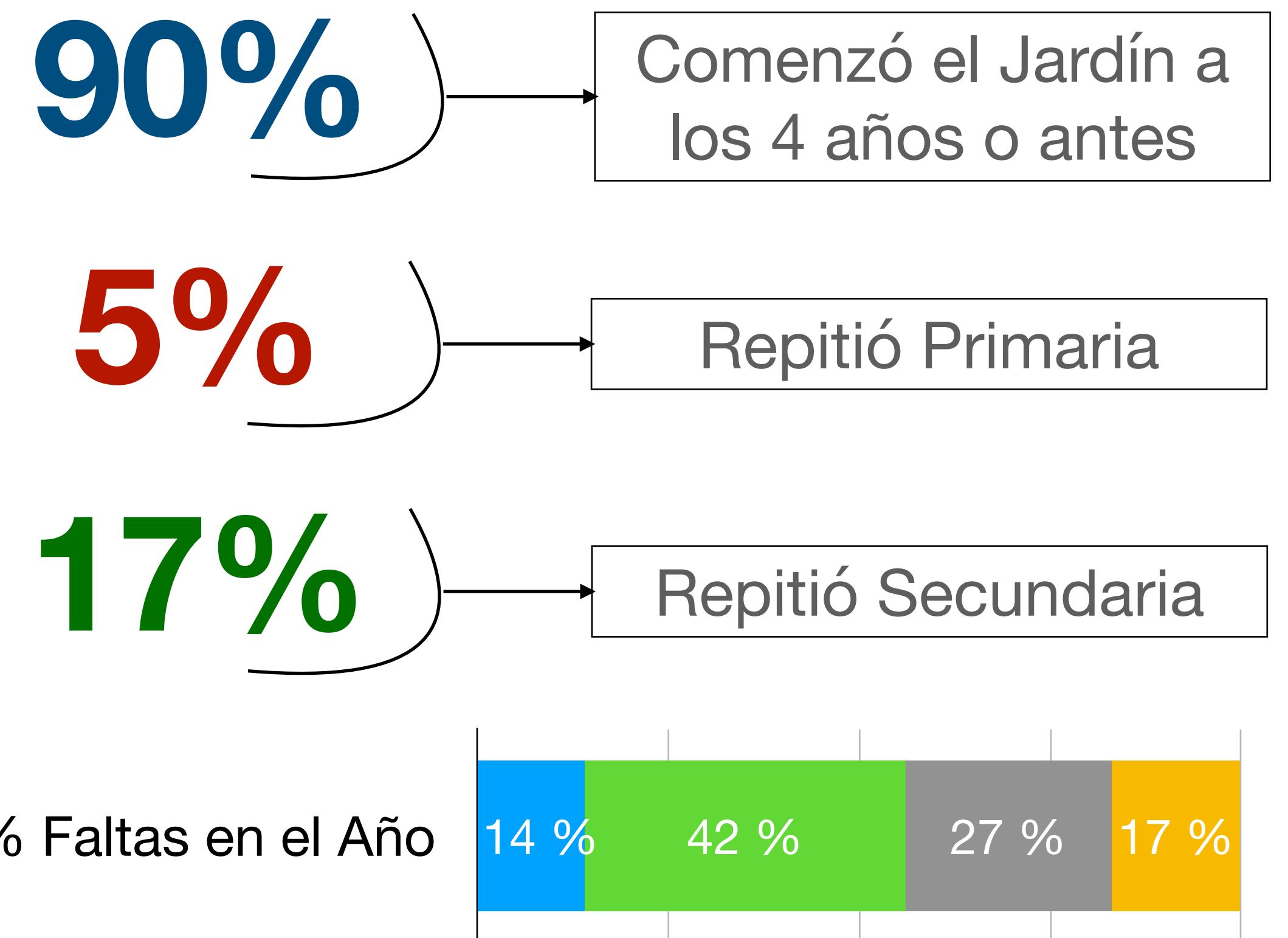


87% → 17 a 18 años

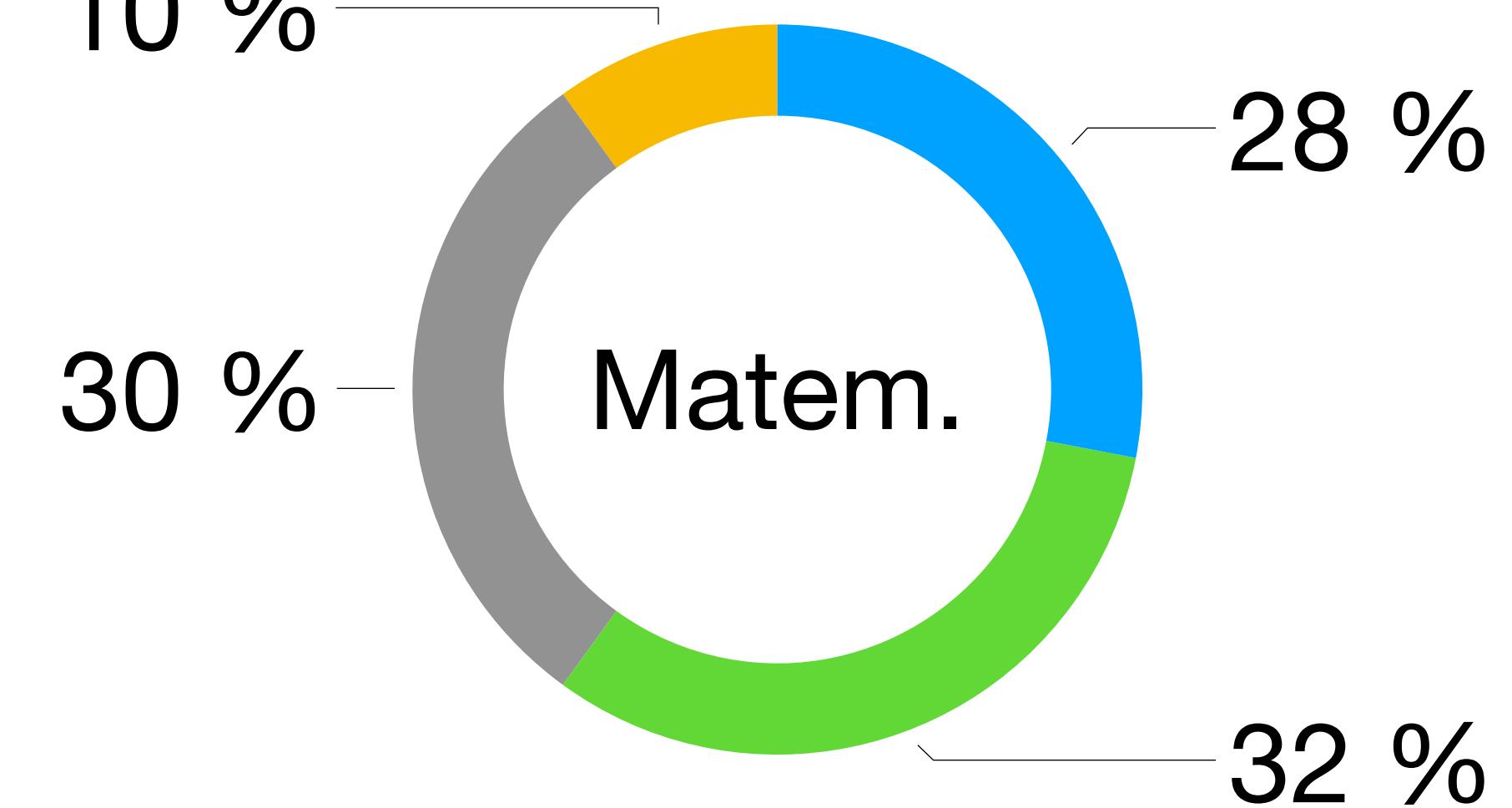
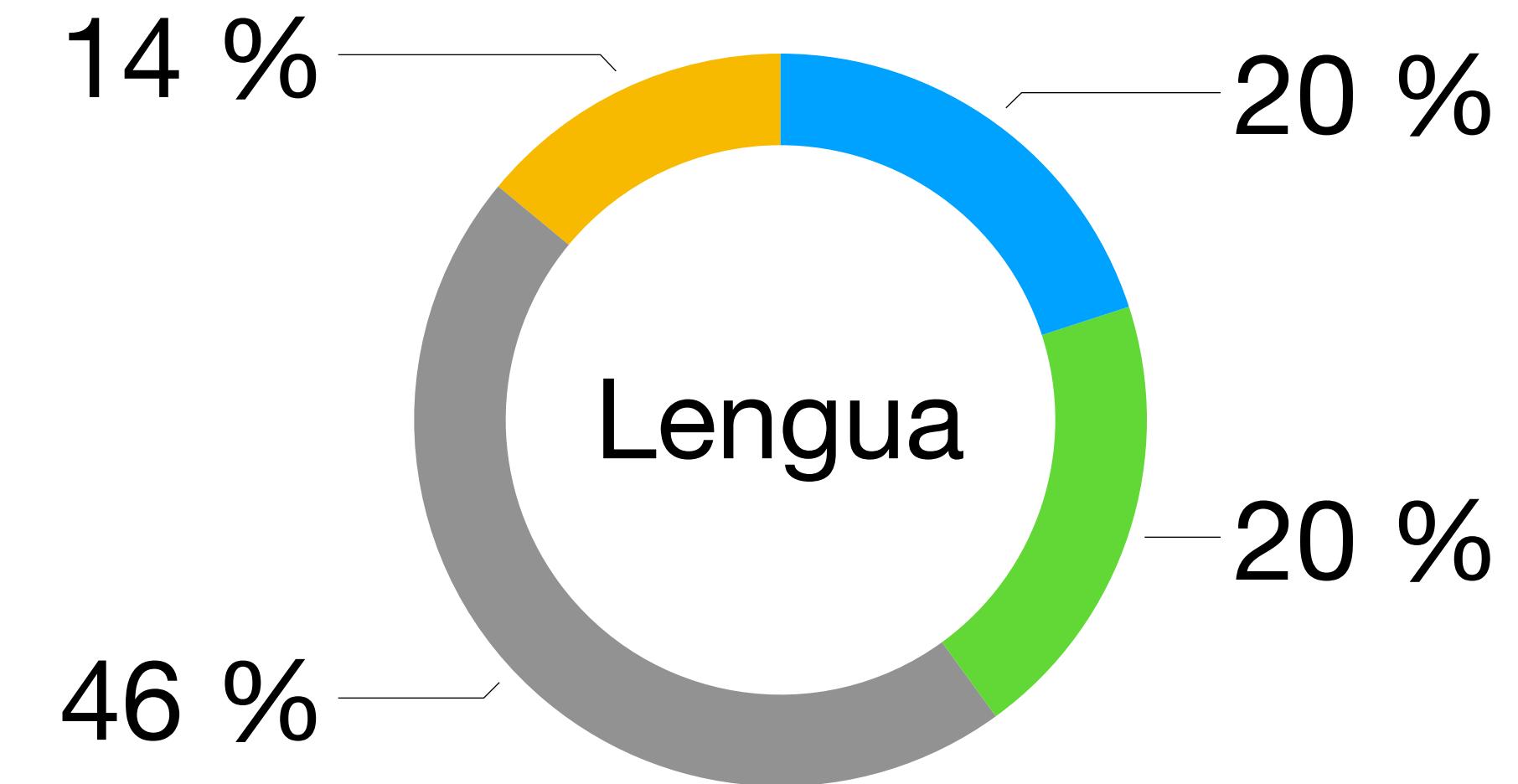


Exploración y Transformación

Primeros datos sobre su educación



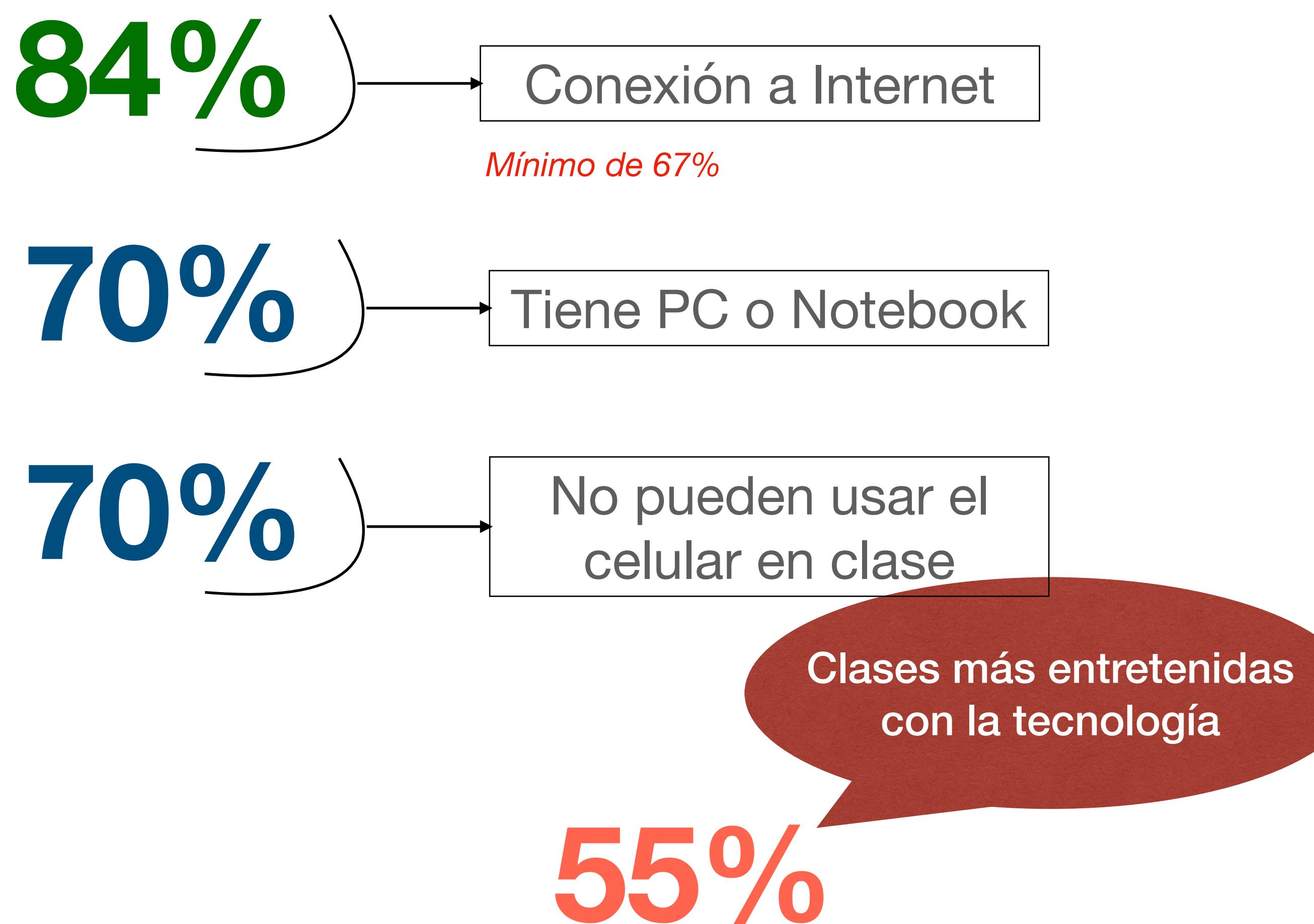
- Debajo de básico
- Básico
- Satisfactorio
- Avanzado



Menos de 8 De 8 a 17 De 18 a 24 Más de 24

Exploración y Transformación

Percepción de la Tecnología



Actividades más usadas

- Mensajes
- Ver videos
- Redes sociales
- Buscar información para la Escuela

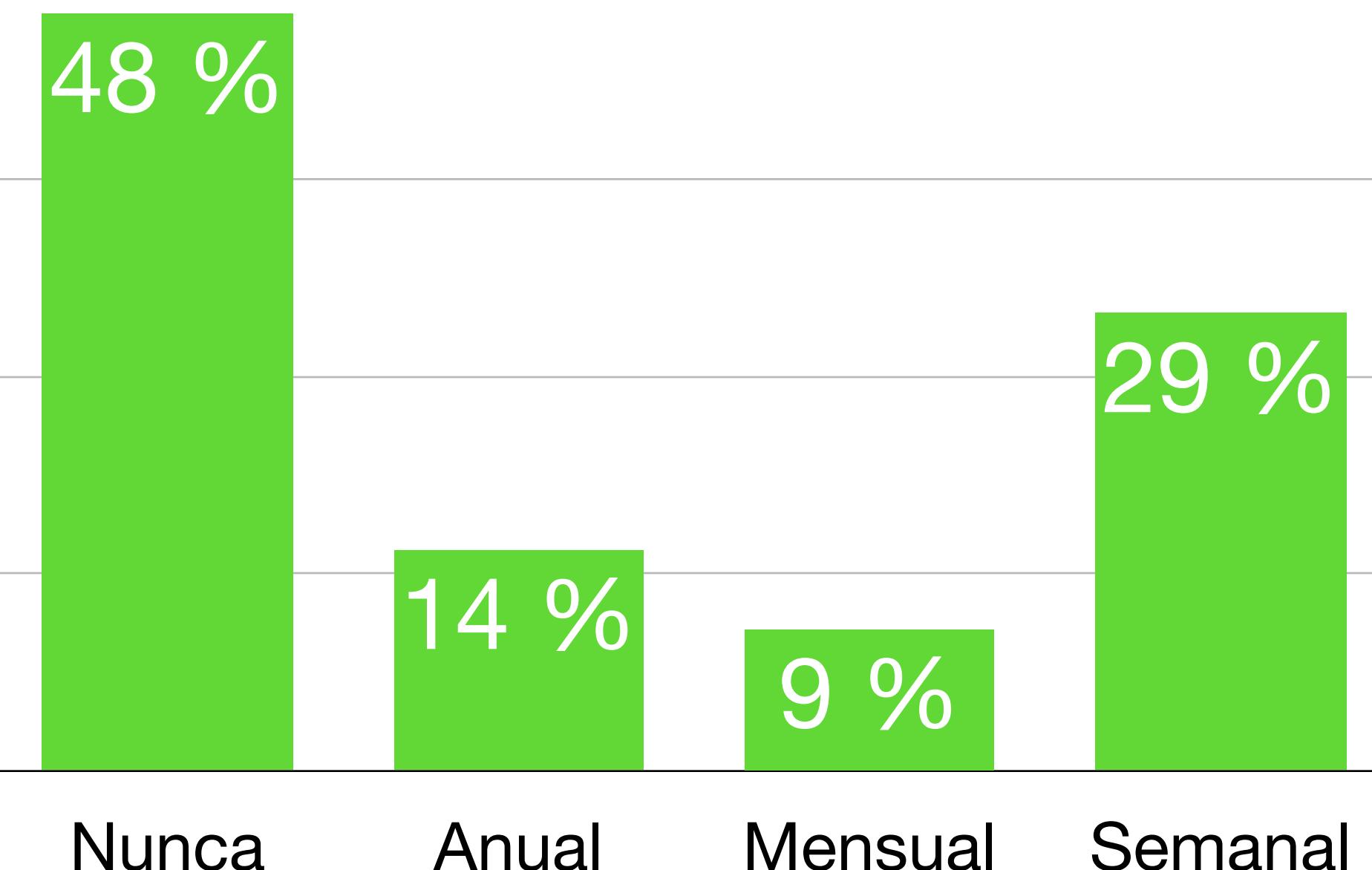
Actividades menos usadas

- Leer artículos o libros
- Estudiar para la Escuela

Exploración y Transformación

Uso de la tecnología en el aula

Frecuencia de uso de la computadora
en Informática



Altos niveles de correlación con el sentimiento
del alumno en el aula al usar tecnologías

Actividades más usadas

- Buscar información en internet
- Producir textos y documentos
- Trabajar en colaboración
- Leer textos y realizar cálculos

Actividades menos usadas

- Usar simulaciones o videojuegos educativos
- Programar

Modelo Aplicado

Parámetros de decisión



Puntaje por Materia

Variable continua - Valores entre 0 y 1000
Distribución Normal
Sin valores outliers

1	Por debajo de básico
2	Básico
3	Satisfactorio
4	Avanzado

Modelo Aplicado

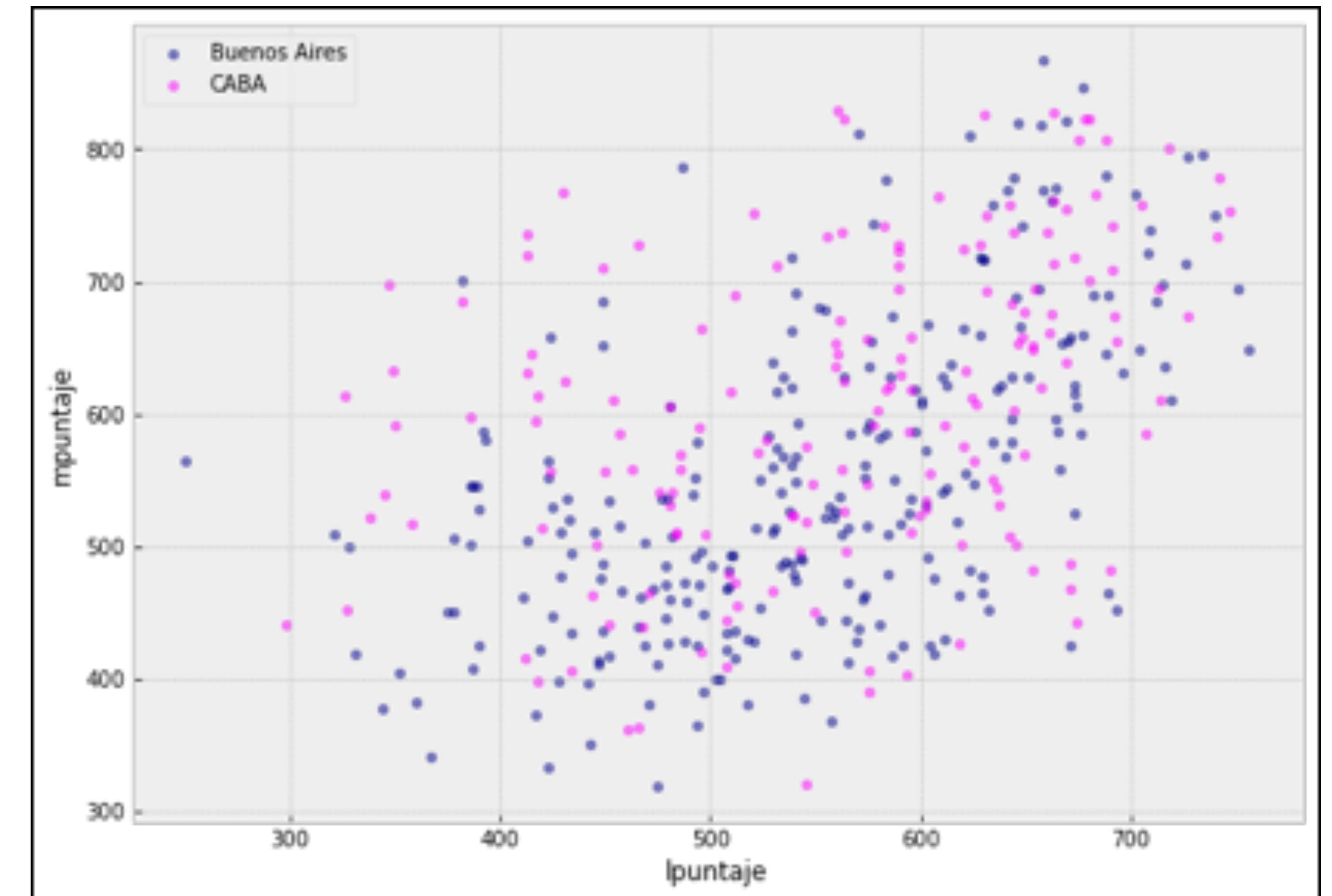
Investigaciones sociales



Datos Estacionales

- Mínima expresión = Municipio
- Non Stationary Data
- Muestra seleccionada
- Muestra pequeña por Mun.
- Target sin agrupación

Distribución de desempeño por región:



Modelo Aplicado

Feature Importance - Correlación de Kendall

- Mejor performance con variables ordinales
- Mide la coincidencia entre variables

Variable	Lengua
Sector	21,5% / 22,8%
Tenes Smartphone	20,7% / 21,5%
Tiene materias previas de años anteriores	17,5% / 18,7%
Usa PC para leer artículos	18,6% / 19,5%

Variable	Matemática
Tenes Smartphone	26,6% / 28,8%
Nivel Socioec.	26,1% / 28,2%
Sector	25,8% / 28,0%
Tiene materias previas de años anteriores	19,5% / 21,2%

Primer resultado para Puntaje y segundo para Desempeño

Modelo Aplicado

Feature Importance - Random Forest

- No requiere distribución normal
- Contribución de cada variable en la predicción

Parámetros

- **25% testeo**
- **Accuracy 47%**

Indicadores

- **Mean Decrease Impurity:** Cuanto aporta agregar una variable
- **Permutation importance:** disminución del score cuando una variable cambia

Lengua	Matemática
Nivel educativo de los padres	Nivel educativo de los padres
Cantidad de personas que viven en el hogar	Cantidad de personas que viven en el hogar
Cantidad de habitaciones en el hogar	Cantidad de habitaciones en el hogar
Sector	Sector
Tiene Smartphone	Tiene Smartphone

Modelo Aplicado

Feature Importance - XGBoost

- Robustez y alta escalabilidad
- Agrega features que mejoren la performance

Lengua

Peso de las variables	Ganancia
Municipio	Sector
Nivel educativo de los padres	Repitó primaria
Cantidad de habitaciones en el hogar	Tiene Smartphone
Auto concepto Aprender	Auto concepto Aprender

Parámetros

- 25% testeo
- Accuracy 48%

Matemática

Peso de las variables	Ganancia
Municipio	Sector
Nivel educativo de los padres	Repitó primaria
Cantidad de habitaciones en el hogar	Nivel socio económico
Auto concepto Aprender	Auto concepto Aprender

Modelo Aplicado

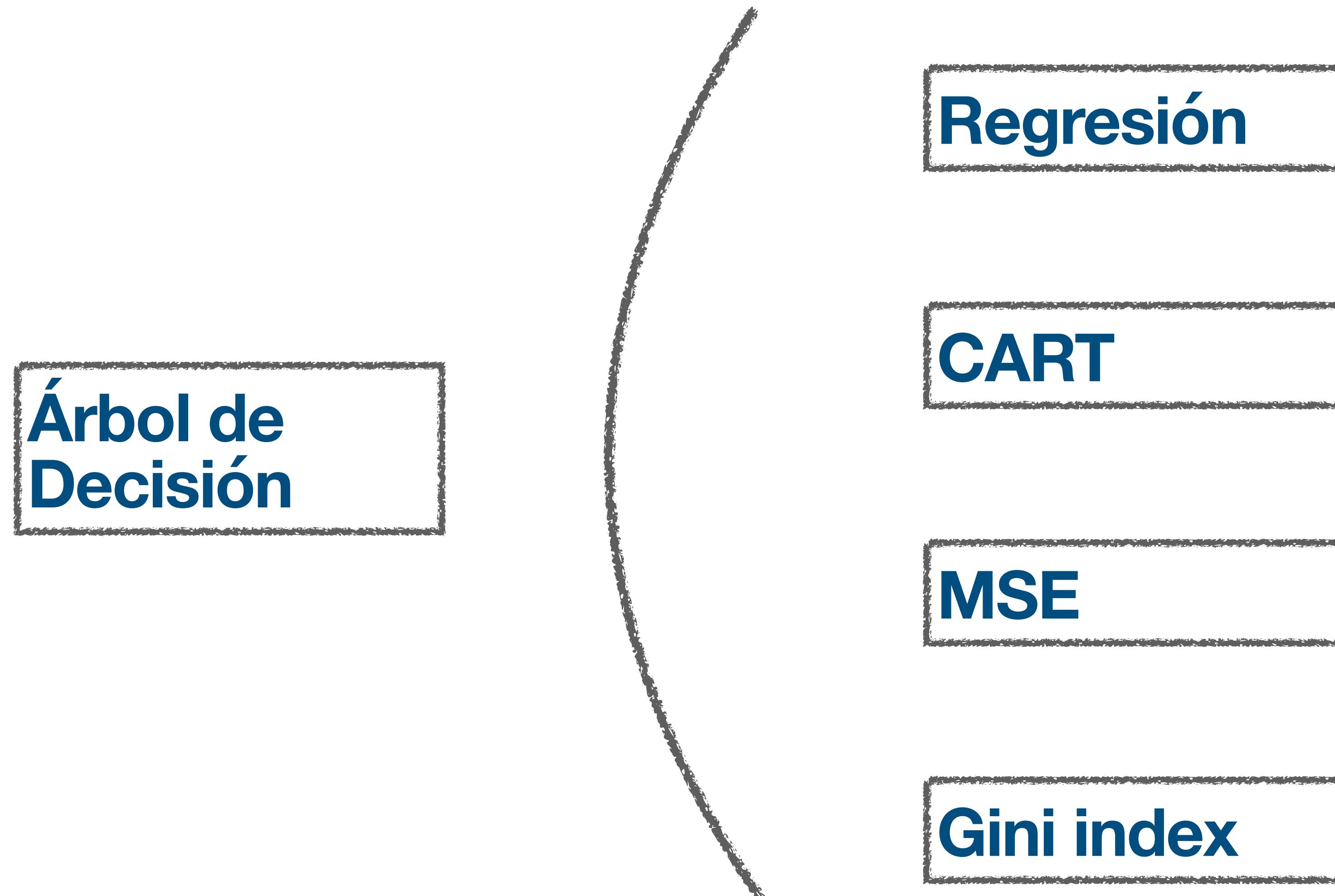
Predicción



Primeros pasos

- A. Aplicación del modelo por materia (Matemática y Lengua)
- B. Creación de tabla para eliminar variables sin uso
- C. Ejecución del modelo

Modelo Aplicado



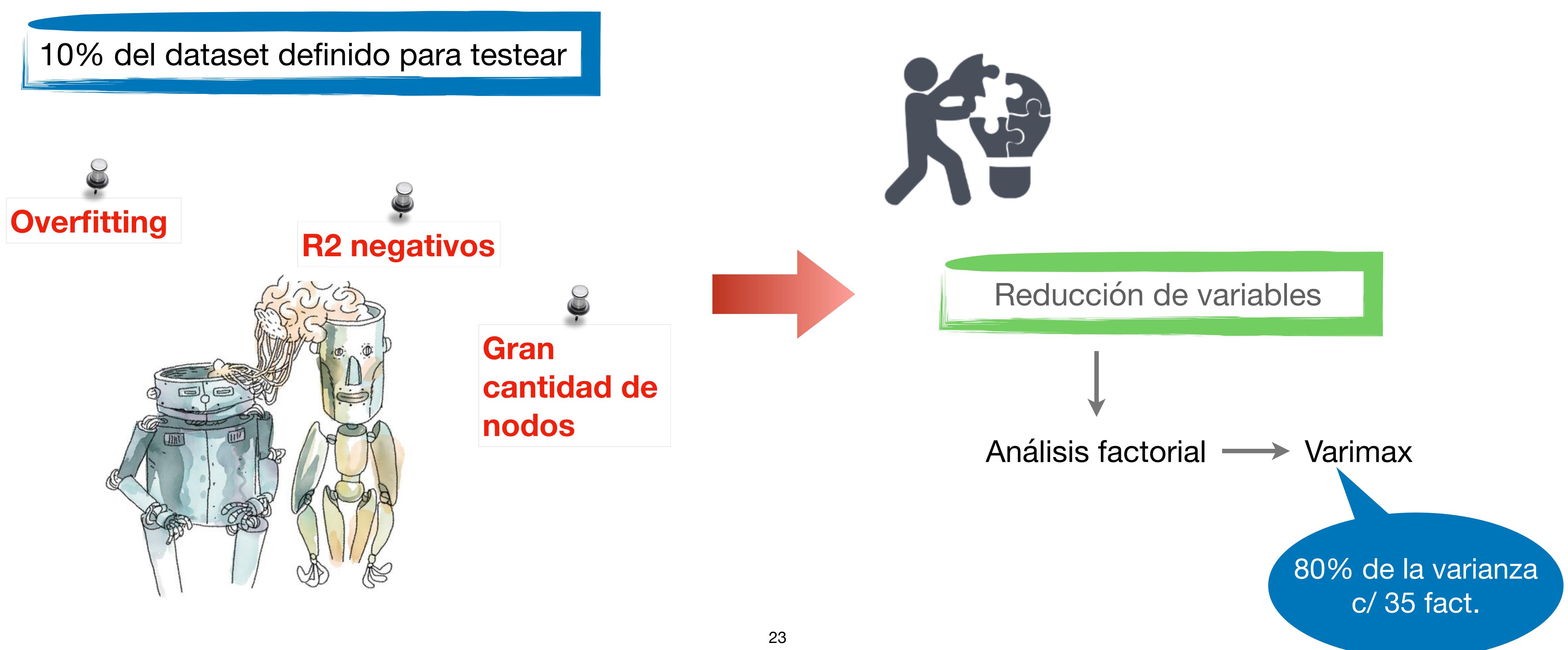
*Classification and
Regression Trees*

*Mean Squared Error,
busca minimizar la
pérdida*

*Impureza de cada
feature*

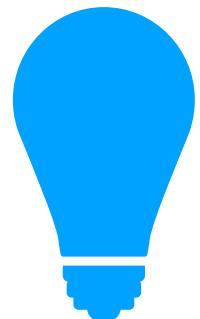
Modelo Aplicado

Ejecución sin reducción de variables

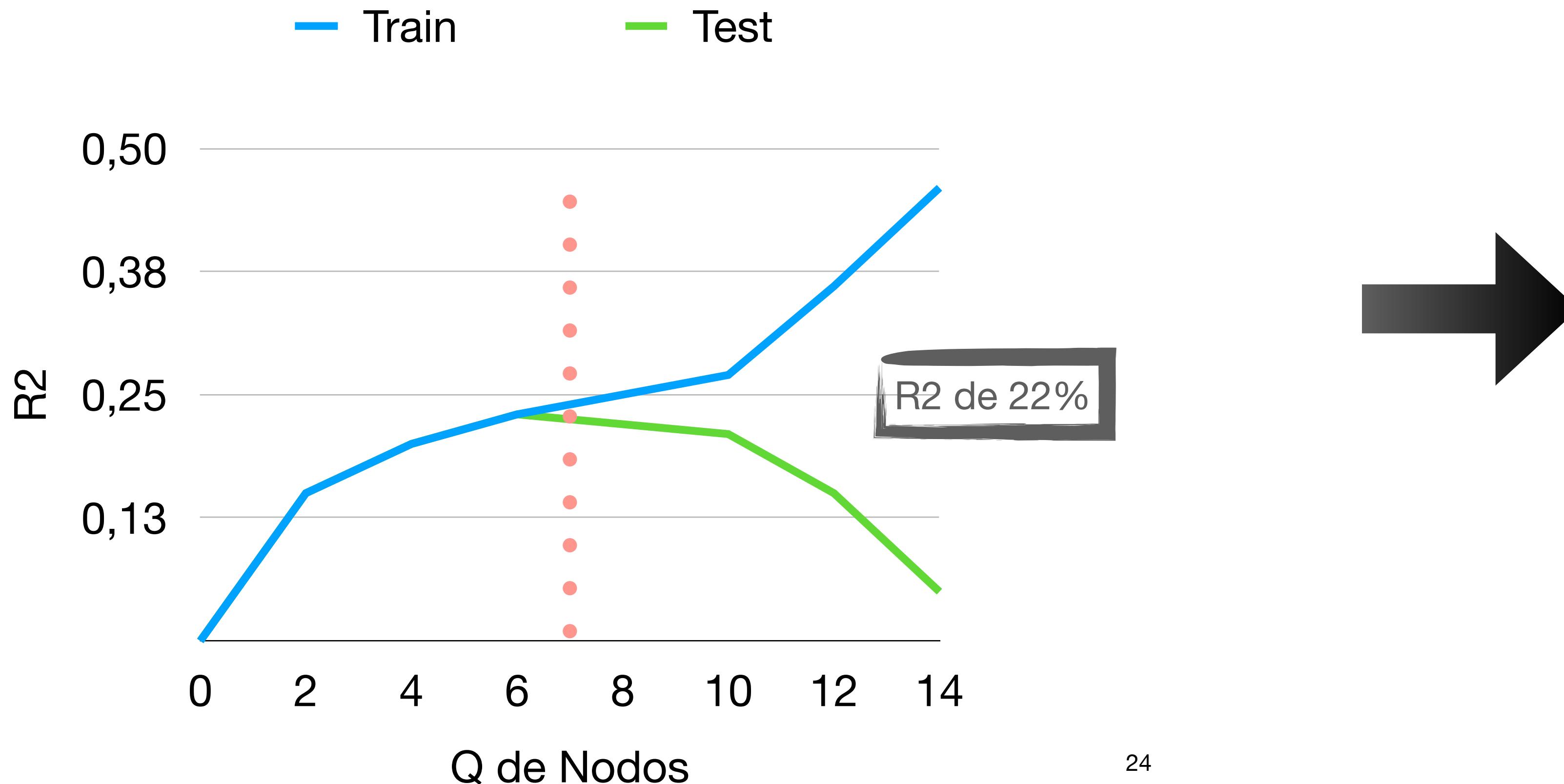


Modelo Aplicado

Ejecución con reducción de variables - Lengua



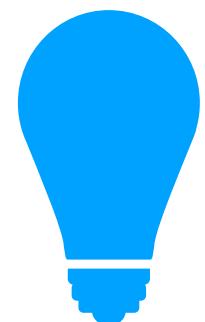
Búsqueda de parámetros para evitar overfitting



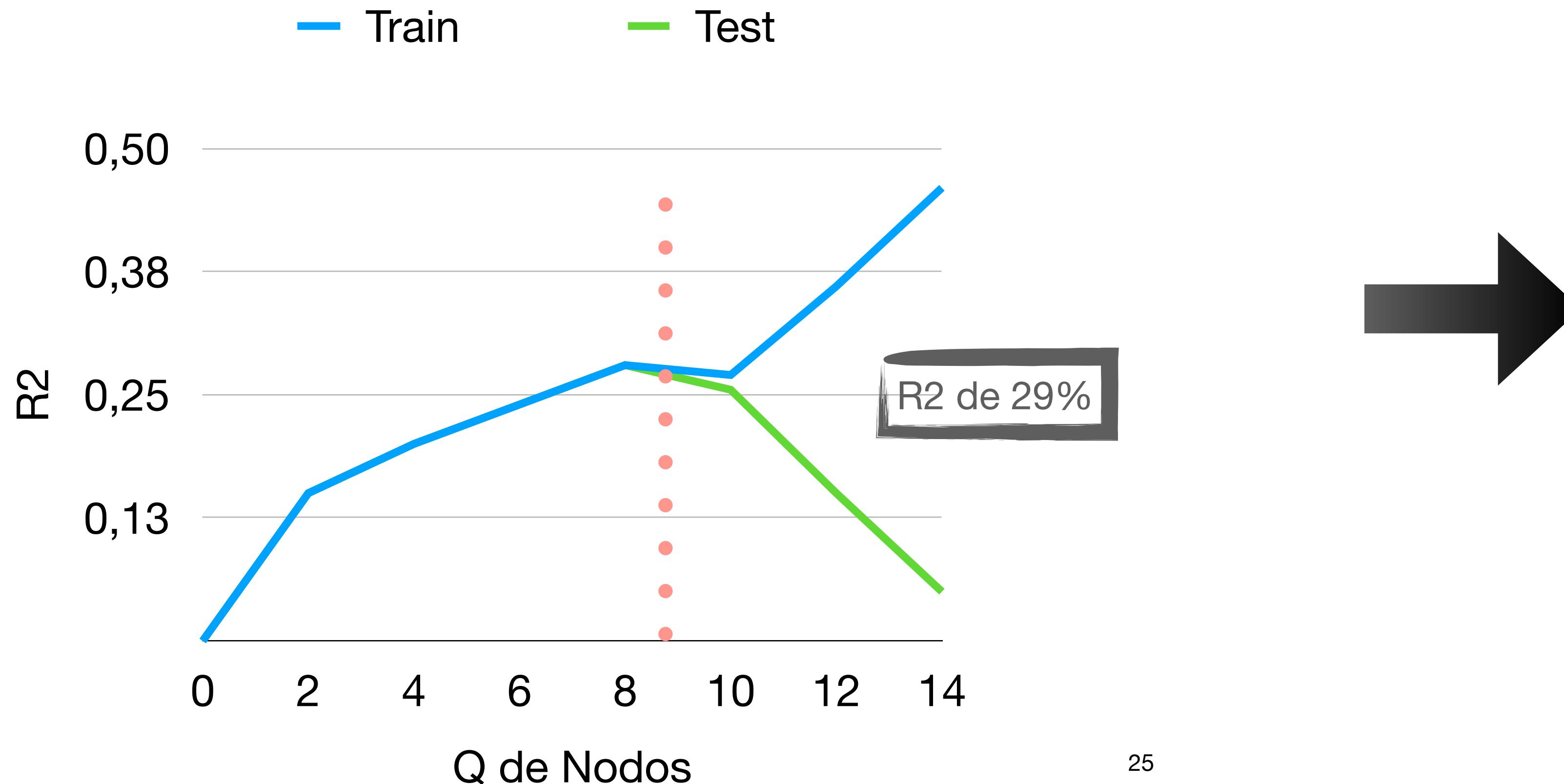
Factores Relevantes
Auto concepto Aprender
Trayectoria escolar
Sector
Tiene Smartphone

Modelo Aplicado

Ejecución con reducción de variables - Matem.



Búsqueda de parámetros para evitar overfitting

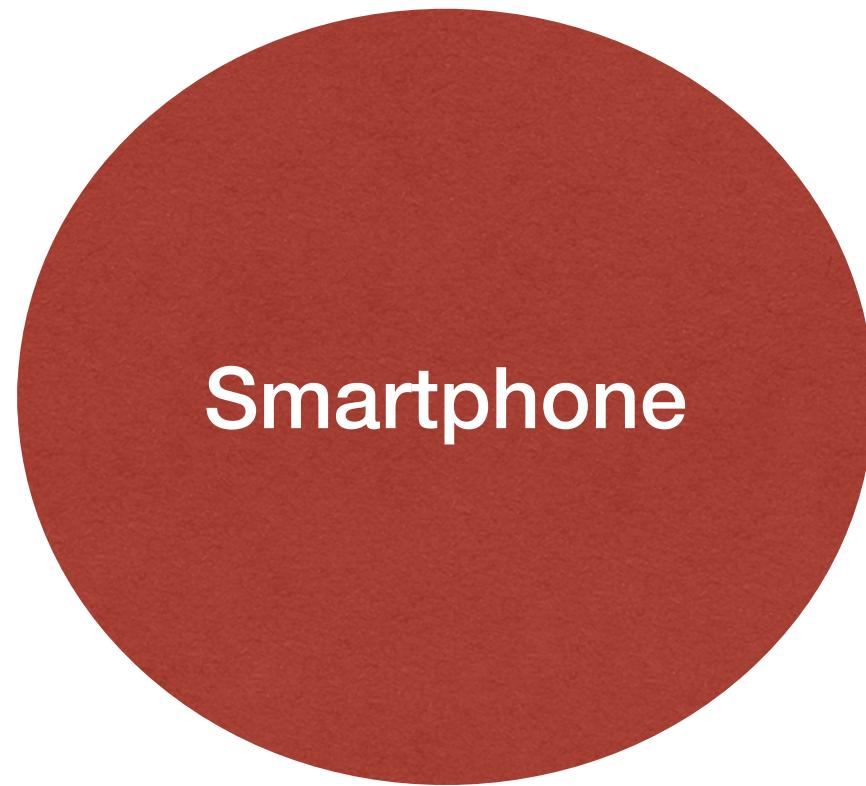
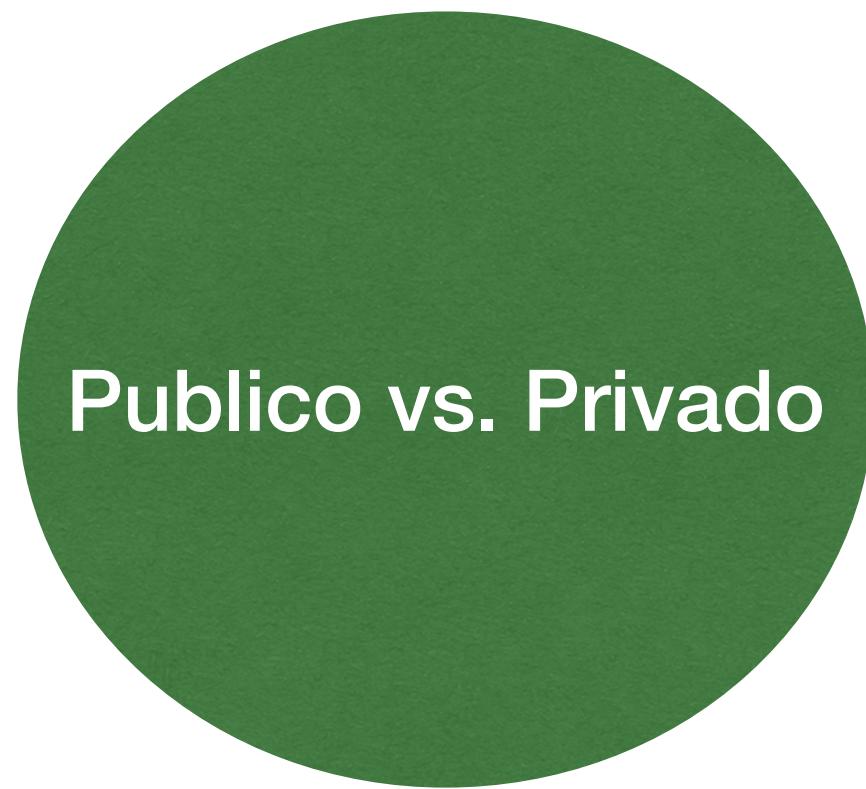


Factores Relevantes
Sector
Tiene Smartphone
Acceso TIC en el hogar
Auto concepto escuela

Conclusiones

- Acceso a Internet de Buenos Aires 84% (Superior al promedio País)
- Los alumnos utilizan la tecnología para tareas escolares →
Lectura de artículos, búsqueda de información y trabajo en equipo
- La tecnología no tiene relación significativa al desempeño, pero sí al
la percepción de las clases
- Nivel socioeconómico con baja relevancia estadística →
Nivel educativo de los padres con
relevancia en algunos modelos
- No hay robustez para un modelo predictivo, pero sí para explicar
qué ajustes tendrán impactos positivos y cuáles no.
- **¿Dónde debemos enfocarnos?**

Conclusiones



Espacio abierto a preguntas

Muchas Gracias