

December 6, 2023

# 1 EVALUACIÓN TÉCNICA

## 1.1 Introducción

Este documento resume el proceso y los resultados de la evaluación técnica para el cargo de Analista en la Unidad de Inteligencia de Datos del Centro de Estudios del Ministerio de Educación. El propósito de esta evaluación es analizar la desvinculación de estudiantes en establecimientos particulares subvencionados y entender las dinámicas detrás de estos fenómenos.

## 1.2 Fundamentos y Metodología

La evaluación se basa en el análisis de dos conjuntos de datos proporcionados: Rendimiento 2022 y Matrícula Oficial 2023. Las siguientes bibliotecas y herramientas de Python fueron utilizadas para procesar y analizar los datos:

- Pandas: para la manipulación y limpieza de los datos.
- Numpy: para operaciones numéricas.
- Matplotlib y Seaborn: para la visualización de datos.

El código adjunto en el archivo IPYNB/PY y script.py contiene todos los pasos de la carga, limpieza, transformación y análisis de los datos.

Se entiende como alumno desvinculados de establecimientos particulares subvencionados alumnos que no renuevan su matrícula entre el año 2022 y 2023 en el algún establecimiento particular subvencionado.

En el [codigo](#) se utiliza la siguiente condición:

```
df_no_promovidos = df[
    ~(
        (df.set_index(['COD_ENSE_x', 'COD_GRADO_x']).index.isin(CODE_4TO_MEDIO)) &
        ((df['SIT_FIN'] == 'P') |
         (df['SIT_FIN_R'] == 'P'))
    )
]

disengagement_condition = (
    (df_no_promovidos['RBD_x'].notna() & df_no_promovidos['RBD_y'].isna()) |
    (df_no_promovidos['COD_REG_RBD_x'].notna() & df_no_promovidos['COD_REG_RBD_y'].isna())
)
```

Se filtra y quita los “4tos medios” o grados similares que se encuentra promovidos, para tener la información más limpia y no abultar la desvinculación.

### 1.3 Procesamiento de Datos y Cálculo de Desvinculación

Se crea script donde se procesan los datos.

Se agregan las siguientes columnas:

- `count_disengaged`: El número de estudiantes desvinculados durante el periodo 2022-2023.
- `count_matricula`: La cantidad de estudiantes matriculados en el año 2023.
- `matricula_teorica`: La suma de `count_matricula` y `count_disengaged`, reflejando la cantidad total de estudiantes que estaban matriculados o se desvincularon.
- `tasa_disengaged`: Tasa de desvinculación, es la proporción de `count_disengaged` sobre `matricula_teorica`.

```
[6]: from script import read_data_statics
stats = read_data_statics()
```

### 1.4 Análisis de Datos

El análisis realizado identificó las tasas de desvinculación de estudiantes por región y por cada uno de los establecimientos. Los resultados se presentan en tablas y gráficos correspondientes.

```
[7]: #Stats for rbd-establishment
stats.get("RBD").get("stats")
```

```
[7]:
```

|      | RBD   | NOM_RBD_x  | COD_REG_RBD_x | \ |
|------|-------|--|---------------|---|
| 0    | 45    | COLEGIO ITALIANO SANTA ANA                         | 15            |   |
| 1    | 50    | COLEGIO ADVENTISTA DE ARICA                        | 15            |   |
| 2    | 52    | LICEO BICENTENARIO DOMINGO SANTA MARIA             | 15            |   |
| 3    | 56    | COLEGIO SAN MARCOS DE ARICA                        | 15            |   |
| 4    | 60    | COLEGIO ALEMAN DE ARICA                            | 15            |   |
| ...  | ...   | ...  | ...           |   |
| 5501 | 42154 | ESCUELA HOSPITALARIA ENTRE VALLES                  | 4             |   |
| 5502 | 42158 | INSTITUTO APRENDER ALERCE                          | 10            |   |
| 5503 | 42161 | ESCUELA ESPECIAL PARA NIÑOS CON AUTISMO NUESTRA... | 5             |   |
| 5504 | 42166 | CEIA INAPEWMA PUERTO LIMACHE                       | 5             |   |
| 5505 | 42168 | COLEGIO PAIHUEN LA SERENA                          | 4             |   |

|      | NOM_REG_RBD_A_x | count_disengaged | count_matricula | matricula_teorica | \ |
|------|-----------------|------------------|-----------------|-------------------|---|
| 0    | AYP             | 100              | 940             | 1040              |   |
| 1    | AYP             | 28               | 547             | 575               |   |
| 2    | AYP             | 451              | 4088            | 4539              |   |
| 3    | AYP             | 41               | 1027            | 1068              |   |
| 4    | AYP             | 45               | 1043            | 1088              |   |
| ...  | ...             | ...              | ...             | ...               |   |
| 5501 | COQ             | 0                | 1               | 1                 |   |

|      |       |   |     |     |
|------|-------|---|-----|-----|
| 5502 | LAGOS | 0 | 253 | 253 |
| 5503 | VALPO | 0 | 35  | 35  |
| 5504 | VALPO | 0 | 36  | 36  |
| 5505 | COQ   | 0 | 179 | 179 |

|      |                 |
|------|-----------------|
|      | tasa_disengaged |
| 0    | 0.09615385      |
| 1    | 0.04869565      |
| 2    | 0.09936109      |
| 3    | 0.03838951      |
| 4    | 0.04136029      |
| ...  | ...             |
| 5501 | 0               |
| 5502 | 0               |
| 5503 | 0               |
| 5504 | 0               |
| 5505 | 0               |

[5506 rows x 8 columns]

```
[8]: #Stats for regions
stats.get("REG").get("stats")
```

```
[8]: COD_REG_RBD NOM_REG_RBD_A_x count_disengaged NOM_REG_RBD_A_y \
0          1          TPCA          5594          TPCA
1          2          ANTOF          3431          ANTOF
2          3          ATCMA          1418          ATCMA
3          4          COQ          8167          COQ
4          5          VALPO          21944         VALPO
5          6          LGBO          7593          LGBO
6          7          MAULE          8225          MAULE
7          8          BBIO          12249         BBIO
8          9          ARAUC          13156         ARAUC
9          10         LAGOS          9982          LAGOS
10         11         AYSEN          1356          AYSEN
11         12         MAG          946           MAG
12         13         RM          79195          RM
13         14         RIOS          4895          RIOS
14         15         AYP          3668          AYP
15         16         NUBLE          3351         NUBLE
```

|   |                 |                   |                 |
|---|-----------------|-------------------|-----------------|
|   | count_matricula | matricula_teorica | tasa_disengaged |
| 0 | 57410           | 63004             | 0.08878801      |
| 1 | 48774           | 52205             | 0.06572167      |
| 2 | 21386           | 22804             | 0.06218207      |
| 3 | 101977          | 110144            | 0.07414839      |
| 4 | 216562          | 238506            | 0.09200607      |

|    |        |        |            |
|----|--------|--------|------------|
| 5  | 89104  | 96697  | 0.07852364 |
| 6  | 106635 | 114860 | 0.07160892 |
| 7  | 170624 | 182873 | 0.06698091 |
| 8  | 126219 | 139375 | 0.09439283 |
| 9  | 92863  | 102845 | 0.09705868 |
| 10 | 13523  | 14879  | 0.09113516 |
| 11 | 12191  | 13137  | 0.07201035 |
| 12 | 828119 | 907314 | 0.08728511 |
| 13 | 41723  | 46618  | 0.10500236 |
| 14 | 33771  | 37439  | 0.09797270 |
| 15 | 53412  | 56763  | 0.05903493 |

### 1.4.1 Grafico por Región

```
[9]: import matplotlib.pyplot as plt
import matplotlib.ticker as ticker
import seaborn as sns

# Supongamos que este es tu DataFrame
df = stats.get("REG").get("stats")

# Configuración para un gráfico más grande
fig, ax1 = plt.subplots(figsize=(10, 6))

# Graficar 'matricula_teorica'
sns.barplot(x='NOM_REG_RBD_A_x', y='matricula_teorica', data=df, color='blue',
            label='Matrícula Teórica', ax=ax1)

# Graficar 'matricula'
sns.barplot(x='NOM_REG_RBD_A_x', y='count_matricula', data=df, color='green',
            label='Matrícula', ax=ax1)

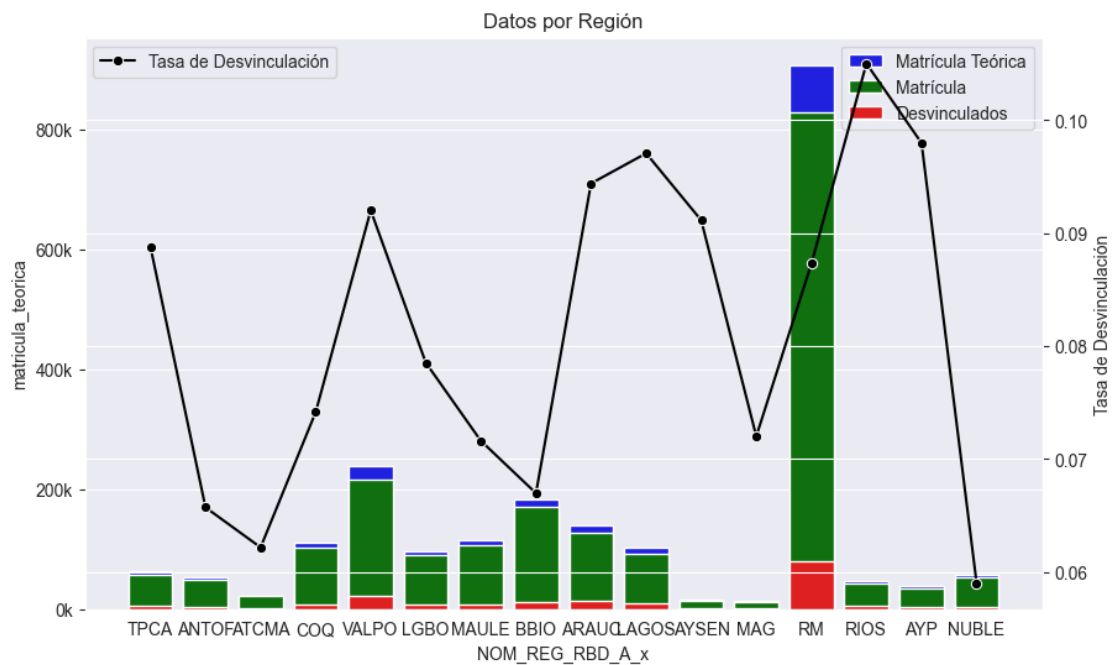
# Graficar 'desvinculados'
sns.barplot(x='NOM_REG_RBD_A_x', y='count_disengaged', data=df, color='red',
            label='Desvinculados', ax=ax1)
ax1.yaxis.set_major_formatter(ticker.FuncFormatter(lambda y, pos: '{:.0f}k'.
            format(y/1000)))

# Crear un segundo eje Y para la tasa de desvinculación
ax2 = plt.gca().twinx()
sns.lineplot(x='NOM_REG_RBD_A_x', y='tasa_disengaged', data=df, ax=ax2,
            color='black', label='Tasa de Desvinculación', marker='o')

# Añadir leyenda
plt.legend(loc='upper left')
```

```
# Títulos y etiquetas
plt.title('Datos por Región')
plt.xlabel('Región')
plt.ylabel('Cantidad')
ax2.set_ylabel('Tasa de Desvinculación')

plt.show()
```



## 1.5 Totales Nacionales

```
[10]: totales = stats.get("REG").get("totales")
      totales.transpose()
```

```
[10]:      count_matricula  count_disengaged  matricula_teorica  tasa_disengaged
      totales      2014293      185170      2199463      0.08418873
```

Teniendo en cuenta la desvinculación, en este contexto, se refiere a los estudiantes que no continuaron su educación en algún establecimiento particular subvencionado en el año siguiente.

Los datos proporcionados muestran una panorámica de la desvinculación de estudiantes en distintas regiones de establecimientos particulares subvencionados. El análisis se centra en el número de estudiantes desvinculados, la matrícula real, la matrícula teórica y la tasa de desvinculación en cada región.

### 1.5.1 Desvinculación por Región

El número de estudiantes desvinculados (`count_disengaged`) y la tasa de desvinculación (`tasa_disengaged`) varían considerablemente entre las regiones, lo que sugiere diferencias en la retención de estudiantes en los establecimientos particulares subvencionados. Algunos puntos destacados son:

- La Región de los Ríos (RIOS), Arica y Parinacota (AYP), los Lagos (LAGOS), la Araucanía (ARAUC), Valparaíso (VALPO) y Aysén (AYSEN) destacan con las tasas más altas, alcanzando un 10.50%, 9.79%, 9.70%, 9.44%, 9.20%, 9.11%. Este indicador sugiere que casi uno de cada diez estudiantes se desvincula, lo que podría señalar desafíos únicos en retención que ameritan una atención particular.
- La Región de Ñuble (NUBLE) muestra la tasa más baja de desvinculación con un 5.90%, lo que sugiere una mayor efectividad en mantener a los estudiantes dentro del sistema educativo de esa área. Este resultado podría ser reflejo de políticas educativas exitosas o de un contexto socioeconómico favorable que facilita la continuidad escolar.
- La tasa promedio de desvinculación a través de todas las regiones se sitúa en un término medio, con regiones como Antofagasta (ANTOF) y Atacama (ATCMA) mostrando tasas de desvinculación por debajo del promedio, lo que puede evidenciar un entorno educativo más estable o una menor incidencia de factores que propician la desvinculación.

### 1.5.2 Desvinculación por Establecimiento

- Desvinculación por Establecimiento: La cantidad de estudiantes desvinculados varía significativamente entre los establecimientos, lo que podría indicar diferencias en la capacidad de retención de cada colegio. Por ejemplo, el Liceo Bicentenario Domingo Santa María (Arica) tiene un número alto de desvinculados (739), lo que podría requerir una investigación más profunda para entender las causas.
- Tasa de Desvinculación: La tasa de desvinculación también varía, con el Colegio Italiano Santa Ana (Arica) mostrando una tasa relativamente alta (0.200680), mientras que el Colegio San Marcos de Arica (Arica) tiene una tasa más baja (0.055198). Esto puede reflejar la eficacia de las políticas de retención de estudiantes o el impacto de factores externos como el cambio socioeconómico en la región.

## 1.6 Conclusiones

Finalmente, analizando los datos disponibles, podemos concluir lo siguiente: Las regiones con altas tasas de desvinculación, como los Ríos (RIOS) y la Araucanía (ARAUC), podrían beneficiarse significativamente de programas de intervención diseñados para abordar los factores particulares que inciden en estas cifras. Estos programas deberían ser personalizados, basándose en un entendimiento cabal de las causas locales de desvinculación, para que sean eficaces en mejorar la retención de estudiantes.

Por otro lado, regiones como Ñuble, Atacama y Antofagasta, que exhiben tasas más bajas de desvinculación, proporcionan una excelente oportunidad para el análisis y la identificación de prácticas y políticas exitosas. La comprensión de las estrategias que han resultado efectivas en estas áreas puede ofrecer patrones replicables que, una vez adaptados, podrían aplicarse en otras regiones para fortalecer la permanencia estudiantil.

En aquellos establecimientos con tasas altas de desvinculación, es prioritario identificar las causas subyacentes específicas que contribuyen a este fenómeno. Esto permitirá el desarrollo de programas de apoyo dirigidos, estrategias de intervención temprana y, si se requiere, la mejora de la calidad educativa y del entorno escolar.

Consolidar como política educativa la adopción de enfoques basados en la evidencia es fundamental. Entender los éxitos de los establecimientos con bajas tasas de desvinculación permitirá construir puentes de colaboración con aquellos que necesitan implementar y mejorar sus estrategias de retención estudiantil. Este enfoque colaborativo, lejos de ser un simple intercambio, puede ser una fuente de innovación y un medio para compartir recursos valiosos. Más aún, fomenta una cultura de mejora continua, la cual es esencial para abordar la desvinculación de manera efectiva y sostenible.

## **1.7 Referencias**