

**Deserción Escolar en Chile**

Integrantes: Nicolas Diaz, David Leal, Boris Bravo y Pia Soto

Profesor: Francia Berna y Kate Epple

Fecha: Julio - 2023

Índice

1. Introducción

2. Entendimiento del problema de negocio

2.1. Definición de la problemática

* + 1. Problemática de negocio
    2. Objetivo, alcance y preguntas a responder

2.2. Contexto del negocio

2.2.1 Recursos de datos

2.2.2 Supuestos

2.3. Plan del proyecto

3. Comprensión de los datos (Data Understanding)

3.1. Origen de los datos

3.2. Descripción de los datos

3.2.1 Estadísticos descriptivos

3.2.2 Tabla de contingencia

3.3. Exploración preliminar de los datos

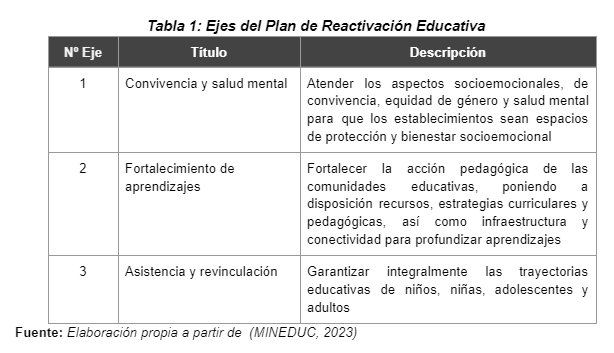
4. Bibliografía

5. Anexos

1.Introducción

Pocas semanas después de identificar el primer caso de Coronavirus en Chile, en marzo de 2020 las autoridades del momento decretaron el cierre de colegios como parte de la estrategia de cuidado de la población nacional ante, la en ese momento, inminente crecida de casos e impacto de la pandemia. Años después, comprendimos que la emergencia tuvo impactos en la vida, familias y la sociedad de distintos países del mundo. Fuera del ámbito de salud, el sistema educativo sufrió un serio golpe, principalmente por la prolongación de las medidas de no presencialidad en las escuelas, donde Chile fue uno de los países cuyos estudiantes estuvieron más tiempo fuera de las aulas (UNESCO, UNICEF y CEPAL, 2022).

El impacto de la emergencia en el sistema educativo tiene diversas dimensiones. El Ministerio de Educación de Chile, generó un plan de reactivación donde encontramos tres ejes prioritarios a solucionar los impactos de la pandemia:



Como aparece en Tabla 1, uno de los problemas que el Ministerio de Educación identificó son los serios niveles de asistencia y la necesidad de Re vinculación de las trayectorias educativas. El Ministerio, establecimientos educacionales y otras instituciones de la sociedad civil conocen del problema de la deserción escolar y están implementando estrategias para reducirla. Pero este no es un problema nuevo, encontrando en la literatura especializada y en la prensa (Cordano, 2018) información sobre la importancia y el crecimiento de este problema, donde ya antes del inicio de la pandemia del COVID 19 encontramos declaraciones sobre el tamaño del problema en la región.

En Latinoamérica este es un problema grave, porque en países como Argentina y Uruguay las cifras de deserción llegan hasta 30% en secundaria. En naciones de Centroamérica hablar de este tema es complicado, porque todavía falta mucho por avanzar en cobertura” (Cordano, 2018). Aunque las declaraciones afirman que las cifras son menores en Chile que en otros países vecinos, en educación media la deserción llega al 10% con el factor de pobreza como un elemento destacable en 8 de cada 10 casos.

La deserción educativa es un fenómeno preocupante, por lo tanto, existe una extensa bibliografía que analiza sus causas y efectos en la población. El MINEDUC afirma en su plan de Reactivación que este es un problema que es necesario comprenderlo desde un enfoque más amplio:

“La pandemia ha producido consecuencias en múltiples ámbitos en el sistema educativo, que han repercutido fuertemente en los estudiantes, sus familias, los establecimientos educacionales y sostenedores, siendo una de las dimensiones más afectadas la asistencia. Los datos del Ministerio de Educación muestran que 1.208.124 estudiantes (38% de la matrícula) presentaron inasistencia grave (menor al 85%) en 2022, lo que significa un alza de 10 puntos porcentuales respecto a los 878.361 estudiantes en esta categoría en el año 2019. A su vez, la inasistencia grave se considera un predictor de la exclusión educativa, por lo que las cifras del 2022 podrían anticipar mayor interrupción de las trayectorias educativas en el sistema educativo (BID, 2022; UNICEF y UNESCO, 2022).”

Deserción, aumento del porcentaje de inasistencia grave y sumado a la trayectoria nacional de no retención de personas en el sistema educativo, nos permite identificar que este es un tema relevante que necesita integrar nuevas estrategias de prevención, identificación y alerta que posibilite mejores soluciones de alto impacto en el sistema educativo, las familias y personas para poder terminar sus trayectorias educacionales.

No todo son malas noticias. El MINEDUC entrega una enorme cantidad de datos a la sociedad y las metodologías de minería de datos y machine learning posicionan una oportunidad de usar la información disponible para estructurar nuevas soluciones enfocadas en solucionar el problema y contribuir a proteger a los actuales estudiantes del sistema y reintegrar a quienes están momentáneamente fuera de las instituciones educativas para nivelar las oportunidades del sociales para los niños, niñas, jóvenes y adolescentes de nuestro país.

El presente informe marca el punto de inicio de nuestra investigación que busca encontrar esos factores que inciden en el aumento en la tasa de incidencia de la deserción escolar en Chile para entregar a los tomadores de decisiones herramientas basadas en datos para una alerta temprana tomar medidas sobre los estudiantes.

2.Entendimiento del problema

Durante los últimos años la deserción escolar se ha vuelto un indicador a medir con múltiples factores de incidencia en el estudiantado a nivel nacional.

El año 2022 un informe liberado por el Ministerio de Educación arrojó que en Chile 50.529 estudiantes desertaron del sistema escolar durante el 2022, lo que representa un 1.7% de la matrícula del país.

Lo que buscamos con el presente trabajo es encontrar esos factores que inciden en el aumento en la tasa de incidencia de la deserción escolar en Chile para entregar a los tomadores de decisiones herramientas basadas en datos para una alerta temprana tomar medidas sobre los estudiantes.

2.1. Definición de la problemática

Los altos índices de abandono del sistema educacional en los últimos años en Chile no sólo se atribuyen a la pandemia, las cifras superan los 40.000 estudiantes en 2021, llegando a 50.529 niños y adolescentes desescolarizados en 2022; es decir, 24% más que el año 2019. A esto se suma una tasa cercana al 40% en lo que respecta a ausentismo (menos de un 85% de asistencia, según criterios del Ministerio). Escenario que no sólo impacta en los aprendizajes de los estudiantes del sistema escolar chileno, sino también en las posibilidades de inserción en la Educación Superior y mundo laboral de parte de estos estudiantes.

En el marco normativo nacional, la Ley General de Educación, define la educación como un proceso permanente que abarca las distintas etapas de la vida de las personas, y que tiene por finalidad que los y las estudiantes alcancen su desarrollo espiritual, ético, moral, afectivo, intelectual, artístico y físico, mediante la transmisión y el cultivo de valores, conocimientos y destrezas; para conducir su vida en forma plena, convivir y participar en forma responsable, solidaria, democrática y activa en la comunidad, y para trabajar y contribuir al desarrollo del país.

2.1.1 Problemática de negocio

Tomando en cuenta que el abandono escolar es un proceso que involucra múltiples factores y que por lo mismo no es repentino, detectar los síntomas de riesgo es clave:

Factores académicos del alumno, tales como motivación o aprendizaje, Los factores antes expuestos, en los alumnos generan instancias de deserción, a los alumnos de bajo rendimiento les afecta la conducta y el aprendizaje en la escuela, lo que se traduce en bajo rendimiento y riesgo de fracaso y de deserción.

Factores académicos del profesor, prácticas pedagógicas tales como: la aplicación poco variada de estrategias didácticas y un manejo deficiente de éstas. Asimismo, prácticas pedagógicas inadecuadas y descontextualizadas de parte de los profesores, genera en los alumnos bajas expectativas y falta de proyección a través de la educación, no comprometiéndose con la recepción de contenidos en el aula.

Factor económico, tiene carencias o necesidades materiales, genera situaciones de vulnerabilidad, revelando una situación de pobreza en la familia, indicando ingresos económicos deficientes, lo que motivaría su ingreso al mundo laboral, restándole tiempo y educación a los estudios posibilitando una deserción escolar,

Factor familiar, muestra gran incidencia en los logros educacionales, se evidencia que el bajo nivel educacional de los padres dificulta que el hogar pueda constituirse en instancia de apoyo a las actividades escolares, constituyéndose en un factor que incide directamente, Por otra parte surge el factor de la disgregación familiar, donde el adolescente vive con familiares o terceros, alejado de sus padres; o pueden vivir sólo con el padre o la madre por el abandono de uno de ellos; esto influye en el rendimiento escolar y consecutivamente en el abandono escolar.

Factor sociocultural, encontramos 3 puntos importantes que hacen relevancia a:

* Entorno comunitario: hace referencia al medio social más inmediato que rodea, específicamente en contexto vulnerable, en el que está inserto e interactúa el estudiante. En concordancia con lo antepuesto, los entornos comunitarios empobrecidos donde existen niveles de violencia, inseguridad y abandono de niños y adolescentes tienen un impacto educativo negativo, más aún si la escuela ignora esta situación y no establece relaciones con los distintos sectores de la comunidad.
* Drogas: aparece el consumo de drogas, o alcohol en los adolescentes. Al respecto, los docentes declaran que las drogas pueden producir cambios negativos en los alumnos, sobre todo en el proceso de enseñanza - aprendizaje y las conductas que éstos realizan, desencadenado ausentismo escolar.
* Grupo de pares: hace referencia al grupo de amigos y de iguales con que el alumno comparte cotidianamente. Cuando el grupo de pares interviene negativamente en el alumno puede provocar que éstos no actúen por sí solos y que se dejen llevar por los demás un grupo de personas no bien adaptadas al colegio y con bajo rendimiento, a la larga se perjudicarán buscando la aceptación del grupo. En este sentido, el grupo de pares o amigos pueden ser factores determinantes en la vida de los alumnos respecto de su rendimiento y en la permanencia de los adolescentes en la escuela y liceos.

2.1.2 Contexto del negocio

Mediante el análisis de datos se busca obtener información relevante para detectar de forma anticipada situaciones de riesgo. Se encuentran involucrados instituciones educacionales, docentes, alumnos, situación social de familias, datos relacionados a la región, antecedentes escolares, brechas y desigualdades entre los servicios educativos, incidencia de la pobreza en los hogares y trabajo infantil.

2.2.1 Recursos de datos

Para el presente trabajo comenzamos a analizar las bases de datos que entrega el sitio https://datosabiertos.mineduc.cl donde tomamos algunas variables necesarias en nuestro análisis.

Es posible, que ante la necesidad en presentaciones posteriores presentemos otras bases de datos e incluso de otros sitios donde encontramos información relevante como el índice de vulnerabilidad escolar (IVE) de JUNAEB o algunos datos de CASEN.

La siguiente tabla está creada solamente con datos de MINEDUC

***Tabla 2: Variables de Investigación***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **N.º** | **Variable** | **Tipo** | **Descripción** | **Base de datos** |
| 1 | ano | integer | Año del proceso | Asistencia declarada mensual por año; Matrícula por estudiante; Rendimiento académico por estudiante |
| 2 | mes\_escolar | integer | Mes de observación de asistencia | Asistencia declarada mensual por año |
| 3 | rbd | integer | Rol Base de Datos del establecimiento | Asistencia declarada mensual por año; Matrícula por estudiante; Rendimiento académico por estudiante |
| 4 | dgv\_rbd | integer | Dígito verificador del RBD | Asistencia declarada mensual por año; Matrícula por estudiante; Rendimiento académico por estudiante |
| 5 | nom\_rbd | string | Nombre del establecimiento | Asistencia declarada mensual por año; Matrícula por estudiante; Rendimiento académico por estudiante |
| 6 | cod\_reg\_rbd | integer | Código de región en que se ubica el establecimiento (Anexo I) | Asistencia declarada mensual por año; Matrícula por estudiante; Rendimiento académico por estudiante |
| 7 | nom\_reg\_rbg\_a | string | Nombre abreviado de la región en que se ubica el establecimiento (Anexo I) | Asistencia declarada mensual por año; Matrícula por estudiante; Rendimiento académico por estudiante |
| 8 | cod\_pro\_rbd | integer | Código de la provincia en que se ubica el establecimiento (Anexo I) | Asistencia declarada mensual por año; Matrícula por estudiante; Rendimiento académico por estudiante |
| 9 | cod\_com\_rbd | integer | Código oficial de comuna en que se ubica el establecimiento | Asistencia declarada mensual por año; Matrícula por estudiante; Rendimiento académico por estudiante |
| 10 | nom\_com\_rbd | string | Nombre de la comuna | Asistencia declarada mensual por año; Matrícula por estudiante; Rendimiento académico por estudiante |
| 11 | cod\_deprov\_rbd | integer | Código del Departamento Provincial en que se ubica el establecimiento | Asistencia declarada mensual por año; Matrícula por estudiante; Rendimiento académico por estudiante |
| 12 | nom\_deprov\_rbd | string | Nombre del Departamento Provincial en que se ubica el establecimiento | Asistencia declarada mensual por año; Matrícula por estudiante; Rendimiento académico por estudiante |
| 13 | rural\_rbd | integer | Área geográfica en que se ubica el establecimiento  0: Urbano  1: Rural | Asistencia declarada mensual por año; Matrícula por estudiante; Rendimiento académico por estudiante |
| 14 | cod\_depe | integer | Dependencia administrativa:  1: Corporación Municipal  2: Municipal DAEM  3: Particular Subvencionado  4: Particular Pagado  5: Corporación de Administración delegada  6: Servicio Local de Educación | Asistencia declarada mensual por año; Matrícula por estudiante; Rendimiento académico por estudiante |
| 15 | cod\_depe2 | integer | Código de Dependencia del Establecimiento (agrupado):  1: Municipal  2: Particular Subvencionado  3: Particular Pagado (o no subvencionado)  4: Corporación de Administración delegada (DL 3166)  5: Servicio Local de Educación | Asistencia declarada mensual por año; Matrícula por estudiante; Rendimiento académico por estudiante |
| 16 | cod\_ense | integer | Tipo de enseñanza (Anexo II) | Asistencia declarada mensual por año; Matrícula por estudiante; Rendimiento académico por estudiante |
| 17 | cod\_ense2 | integer | Niveles de Enseñanza agrupados (Anexo III) | Asistencia declarada mensual por año; Matrícula por estudiante; Rendimiento académico por estudiante |
| 18 | cod\_grado | integer | Código de grado (Anexo IV en versiones 2011-218 y 2019) | Asistencia declarada mensual por año; Matrícula por estudiante; Rendimiento académico por estudiante |
| 19 | let\_cur | string | Letra del curso | Asistencia declarada mensual por año; Matrícula por estudiante; Rendimiento académico por estudiante |
| 20 | mrun | integer | Identificador único del estudiante (máscara del RUN que comparte el MINEDUC) | Asistencia declarada mensual por año; Matrícula por estudiante; Rendimiento académico por estudiante |
| 21 | gen\_alum | integer | Sexo del estudiante  0: Sin Información  1: Hombre  2: Mujer | Asistencia declarada mensual por año; Matrícula por estudiante; Rendimiento académico por estudiante |
| 22 | fec\_nac\_alumno | date | Fecha de nacimiento del estudiante (AAAAMM)  190001: Sin información  180001: Sin información | Asistencia declarada mensual por año; Matrícula por estudiante; Rendimiento académico por estudiante |
| 23 | cod\_com\_alum | integer | Código oficial de la comuna de residencia (auto declarado y voluntario) | Asistencia declarada mensual por año; Matrícula por estudiante; Rendimiento académico por estudiante |
| 24 | nom\_com\_alum | string | Nombre de la comuna de residencia | Asistencia declarada mensual por año; Matrícula por estudiante; Rendimiento académico por estudiante |
| 25 | dias\_asistidos | integer | Número total de días asistidos. Es posible encontrar registros con asistencia 0 debido a que se encuentran retirados del establecimiento, pero se mantienen en el sistema | Asistencia declarada mensual por año |
| 26 | dias\_trabajados | integer | Número total de días trabajados | Asistencia declarada mensual por año |
| 27 | asis\_promedio | integer | Proporción de días asistidos respecto al  total, de días trabajados en el mes | Asistencia declarada mensual por año |
| 28 | let\_rbd | string | Letra del establecimiento | Matrícula por estudiante; Rendimiento académico por estudiante |
| 29 | num\_rbd | integer | Número del establecimiento | Matrícula por estudiante; Rendimiento académico por estudiante |
| 30 | estado\_estab | integer | Estado del establecimiento:  1: Funcionando  2: En receso  3: Cerrado  4: Autorizado sin matrícula | Matrícula por estudiante; Rendimiento académico por estudiante |
| 31 | cod\_ense3 | integer | Niveles de enseñanza con educación especial reasignado (Anexo 3) | Matrícula por estudiante |
| 32 | cod\_grado2 | integer | Código de grado recodificado para equivalencias con Educación Especial y Educación de adultos (Anexo VI) | Matrícula por estudiante |
| 33 | cod\_jor | integer | Jornada en la que asiste a clases:  2: Tarde  3: Mañana y tarde  4: Vespertina / Nocturna  99: Sin Información | Matrícula por estudiante; Rendimiento académico por estudiante |
| 34 | cod\_tip\_cur | integer | Índice de tipo de curso  0: Curso Simple  1 – 4: Curso Combinado  99: Sin Información | Matrícula por estudiante; Rendimiento académico por estudiante |
| 35 | cod\_des\_cur | integer | Descripción del curso (solo aplica para enseñanza media TP)  0: No aplica  1: Sólo Liceo  2: Dual  3: Otro | Matrícula por estudiante; Rendimiento académico por estudiante |
| 36 | edad\_alum | integer | Edad al 30 de junio del correspondiente año escolar  En caso de edades atípicas, se debe a problemas en el ingreso de la información en el Sistema de Información General de Estudiantes (SIGE) del nivel de enseñanza correspondiente. A contar del año 2016, los casos con edades MISSING corresponden a casos no validados con Registro Civil que originalmente tenían 99 años o más; con matrícula en los niveles Parvularia, Básica niños y jóvenes, Media HC niños y jóvenes o Media TP niños y jóvenes que originalmente tenían 25 años o más, o con matrícula en niveles distintos a Parvularios y Educación Especial y que originalmente tenían 0 años. | Matrícula por estudiante; Rendimiento académico por estudiante |
| 37 | gd\_alu | integer | Indica si el alumno es atendido en un grupo diferencial  Sólo estudiantes matriculados en Básica niños y jóvenes  0: No está en grupo diferencial  1: Si está en grupo diferencial | Matrícula por estudiante; Rendimiento académico por estudiante |
| 38 | cod\_reg\_alu | integer | Código de la región de residencia del estudiante  0: Sin Información | Matrícula por estudiante; Rendimiento académico por estudiante |
| 39 | cod\_sec | integer | Código de sector económico (sólo media técnico-profesional y artística) | Matrícula por estudiante; Rendimiento académico por estudiante |
| 40 | cod\_espe | integer | Código de especialidad (sólo media técnico-profesional y artística) | Matrícula por estudiante; Rendimiento académico por estudiante |
| 41 | cod\_rama | integer | Código de rama (sólo media técnico-profesional y artística) | Matrícula por estudiante; Rendimiento académico por estudiante |
| 42 | cod\_men | integer | Código de la mención de la especialidad Técnico Profesional | Matrícula por estudiante; Rendimiento académico por estudiante |
| 43 | ens | integer | Recodificación del Código de Enseñanza en 13 niveles | Matrícula por estudiante |
| 44 | int\_alu | integer | Indicador de alumno integrado  . : Sin Información  0: No  1: Sí  Estudiante que presenta Necesidades Educativas Especiales de carácter permanente o transitorio, según lo que establece el Artículo 2 del Decreto 170/2010 del Ministerio de Educación, disponible para descarga en http://www.leychile.cl/Navegar?idNorma=1012570&buscar=necesidades+educativas+especiales | Rendimiento académico por estudiante |
| 45 | fec\_int\_alu | date | Fecha de ingreso del alumno al establecimiento | Rendimiento académico por estudiante |
| 46 | prom\_gral | integer | Promedio general anual  No necesariamente corresponde al promedio aritmético de los promedios por subsector. Para estudiantes retirados se informa promedio “0,0”. | Rendimiento académico por estudiante |
| 47 | asistencia | integer | Porcentaje anual de asistencia | Rendimiento académico por estudiante |
| 48 | sit\_fin | string | Situación de promoción al cierre del año escolar:  P: Promovido  R: Reprobado  Y: Retirado  en blanco: Sin información  Promovido (Aprobado): estudiante que rinde satisfactoriamente sus evaluaciones, de acuerdo con la legislación vigente durante un año lectivo.  Reprobado: estudiante que no rinde satisfactoriamente sus evaluaciones, de acuerdo con la legislación vigente durante un año lectivo.  Retirado: estudiante que, por retiro formal o no formal, no está en condiciones de ser evaluado | Rendimiento académico por estudiante |
| 49 | sit\_fin\_r | string | Situación de promoción al cierre del año escolar, con indicador de traslado:  P: Promovido  R: Reprobado  Y: Retirado  T: Trasladado  en blanco: Sin información  Trasladado: estudiante que se cambió de establecimiento (o a otro curso, en el mismo establecimiento) durante el año escolar | Rendimiento académico por estudiante |

*Fuente: Elaboración propia a partir de bases de datos abiertos de Mineduc*

2.2.2 Supuestos

El Ministerio de Educación comparte distintas bases de datos que nos ayudan en el desarrollo de esta investigación al tener información por colegio, mes y año de distintas métricas de los colegios.

Para el desarrollo de seguimiento de los estudiantes en sus trayectorias educativas tenemos el MRUN que es una máscara del RUN de cada estudiante, pero anonimizado por el ministerio. Ahí nuestro primer supuesto: el número del MRUN es el mismo de un estudiante en distintas bases de datos y no existe un cambio de MRUN de una misma persona de un mes a otro o desde la base de datos de asistencia a la de calificaciones, por nombrar algunos ejemplos.

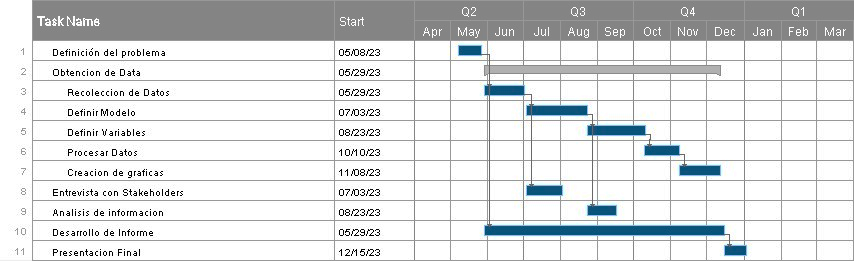
El segundo supuesto es que podemos hacer seguimiento de las trayectorias educativas porque la información es fidedigna y está actualizada por cada estudiante. Por ejemplo, si un establecimiento miente con la información de asistencia de uno o más estudiantes, esto ensucia la muestra.

Un tercer supuesto, es que podemos hacer seguimiento a las trayectorias educativas de los estudiantes con machine learning y otras herramientas que aprenderemos en este diplomado. Por lo menos, encontramos en la literatura (Smith & Gutiérrez, 2002; Kotsiantis, Pierrakeas & Pintelas, 2003) que nos hacen pensar que esto es posible usando las mismas bases de datos que nosotros queremos trabajar.

Un cuarto supuesto, es que los factores registrados por datos pueden ser factores de deserción escolar o síntomas de la misma. La revisión bibliográfica, nos muestra que existen dos tipos de razones y las más afectadas son las externas al colegio al considerarse más relevantes en la interrupción de las trayectorias educativas (Contreras, 2021; Rodríguez, Villanueva, Dombrovskaia& Valenzuela, 2023; Opazo & Sepúlveda, 2009; Espinoza, Castillo, González & Loyola, 2012; Santos, 2009; CESC, 2007). Por eso las iniciativas para estudiantes de educación primaria y secundaria (como Fundación Súmate del Hogar de Cristo) o para estudiantes en educación terciaria (como Fundación Portas) se enfocan en acompañamiento y apoyo social, comunidades y redes.

El programa SAT de Peñalolén de la primera década del milenio también tenía ese enfoque. Nuestro supuesto es que, aunque la literatura no encuentre una relación directa y exclusiva, estos si nos pueden proveer de síntomas que nos permitan descubrir algunos patrones comunes de asistencia, cambio en las calificaciones u otro que nos permitan reconocer a un estudiante desertor antes de que deje el sistema escolar.

El último supuesto, es que podemos seguir buscando datos que al cruzarlos nos entreguen más información. Por ejemplo, tenemos las postulaciones al sistema de admisión (SAE) incluso con la latitud y longitud de los estudiantes lo que posiblemente nos permitiría descubrir si los desertores intentan reinsertarse al sistema escolar o si son vecinos de otras personas con trayectorias educativas truncadas.

2.3. Plan del proyecto

3.Comprensión de los datos (Data Understanding)

3.1. Origen de los datos

Analizamos las bases de datos que entrega el sitio https://datosabiertos.mineduc.cl donde tomamos algunas variables necesarias en nuestro análisis, tales como

* Matriculas Anual de los periodos 2018 al 2022
* Rendimiento Anual de los periodos 2018 al 2022

En el primer informe identificamos tres grupos de bases de datos: asistencia mensual por estudiante, matrícula anual por estudiante y rendimiento anual por estudiante. Al analizar los datos nos percatamos de la ausencia de base de datos de asistencia mensual por estudiante de los años 2020 y 2021. Por esa razón, no tomamos ese grupo completo de datos por dos razones: queremos investigar más para saber qué hacer en ese caso en particular y los datos de estadística descriptiva de esa base lo podemos resolver con el porcentaje de asistencia anual en la base de rendimiento anual por estudiante. Esperamos a futuro este grupo de datos nos permite descubrir algunas tendencias de asistencia para la generación de un sistema de alerta temprana.

|  |  |
| --- | --- |
| **Limpieza de datos y transformación** | |
|  |
| **BBDD Matrícula utilizamos** | **BBDD Matrícula no utilizamos** |  |
| ['AGNO', 'COD\_DEPE2', 'RURAL\_RBD', 'ESTADO\_ESTAB', 'COD\_ENSE', 'COD\_ENSE2', 'COD\_GRADO', 'MRUN', 'GEN\_ALU'] | ['RBD', 'DGV\_RBD', 'NOM\_RBD', 'COD\_REG\_RBD', 'NOM\_REG\_RBD\_A', 'COD\_PRO\_RBD', 'COD\_COM\_RBD', 'NOM\_COM\_RBD', 'COD\_DEPROV\_RBD', 'NOM\_DEPROV\_RBD', 'COD\_DEPE', 'ESTADO\_ESTAB', 'COD\_ENSE', 'COD\_ENSE3', 'COD\_GRADO2', 'LET\_CUR', 'COD\_JOR', 'COD\_TIP\_CUR', 'COD\_DES\_CUR', 'FEC\_NAC\_ALU', 'EDAD\_ALU', 'COD\_REG\_ALU', 'COD\_COM\_ALU', 'NOM\_COM\_ALU', 'COD\_SEC', 'COD\_ESPE', 'COD\_RAMA', 'ENS'] |  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
| *\* Estos por dos razones: la información podía ser de mejor acceso por otra variable similar o porque no incluimos esa variable en el análisis actual, aunque creemos que eso puede estar presente a futuro, por ejemplo, con los datos del establecimiento, región y comuna.* | |  |
|  |
|  |

|  |  |
| --- | --- |
| **Limpieza de datos y transformación** | |
|  |
| **BBDD Rendimiento utilizamos** | **BBDD Rendimiento no utilizamos** |  |
| ['RURAL\_RBD', 'COD\_ENSE2', 'COD\_GRADO', 'MRUN', 'GEN\_ALU', 'PROM\_GRAL', 'ASISTENCIA'] | ['AGNO', 'RBD', 'DGV\_RBD', 'NOM\_RBD', 'COD\_REG\_RBD', 'NOM\_REG\_RBD\_A', 'COD\_PRO\_RBD', 'COD\_COM\_RBD', 'NOM\_COM\_RBD', 'COD\_DEPROV\_RBD', 'NOM\_DEPROV\_RBD', 'COD\_DEPE', 'COD\_DEPE2', 'ESTADO\_ESTAB', 'COD\_ENSE','LET\_CUR', 'COD\_JOR', 'COD\_TIP\_CUR', 'COD\_DES\_CUR', 'FEC\_NAC\_ALU', 'EDAD\_ALU', 'COD\_REG\_ALU', 'COD\_COM\_ALU', 'NOM\_COM\_ALU', 'COD\_RAMA', 'COD\_SEC', 'COD\_ESPE', 'SIT\_FIN', 'SIT\_FIN\_R', 'COD\_MEN'] |  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
| *\* Hasta el momento el uso está asociado a información relevante en la comprensión de matrícula, que es el corazón de nuestra métrica de estudiantes regulares / estudiantes desertores.* | |  |
|  |
|  |

Respecto a la limpieza de datos, la calidad de los mismos permitió un trabajo rápido. Solamente nos vimos en la obligación de configurar el separador de los csv importados desde coma (,) a punto y coma (;) y cambiar el formato del promedio de notas a float, porque en la base del ministerio estaba con coma (,) y confundía a Python el tipo de valor de float a objeto.

3.2. Descripción de los datos

A continuación, compartiremos nuestro análisis de datos en tablas y gráficos que muestran la situación de estudiantes desertores y de sistema regular en el periodo 2018 al 2022.

3.2.1 Estadísticos descriptivos

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **PROM\_GRAL\_REGULAR** | | | | |
|  |  |  |  |  |
| **Estadísticos Descriptivos** | **2018** | **2019** | **2020** | **2021** |
| **count** | 2.769.803 | 2.809.892 | 2.729.690 | 2.757.522 |
| **mean** | 5,459 | 5,496 | 5,882 | 5,854 |
| **std** | 1,528 | 1,560 | 1,271 | 1,261 |
| **min** | 0 | 0 | 0 | 0 |
| **25%** | 5.3 | 5.3 | 5.5 | 5.5 |
| **50%** | 5.8 | 5.9 | 6.2 | 6.2 |
| **75%** | 6.3 | 6.3 | 6.7 | 6.6 |
| **max** | 7 | 7 | 7 | 7 |
|  |  |  |  |  |
| **PROM\_GRAL\_DESERTORES** | | | | |
|  |  |  |  |  |
| **Estadísticos Descriptivos** | **2018** | **2019** | **2020** | **2021** |
| **count** | 87.363 | 85.342 | 46.649 | 70.087 |
| **mean** | 3,283 | 3,382 | 4,337 | 4,110 |
| **std** | 2,556 | 2,669 | 2,438 | 2,432 |
| **min** | 0 | 0 | 0 | 0 |
| **25%** | 0 | 0 | 2.5 | 2.3 |
| **50%** | 4.5 | 4.6 | 5.2 | 5 |
| **75%** | 5.4 | 5.7 | 6.3 | 6.1 |
| **max** | 7 | 7 | 7 | 7 |

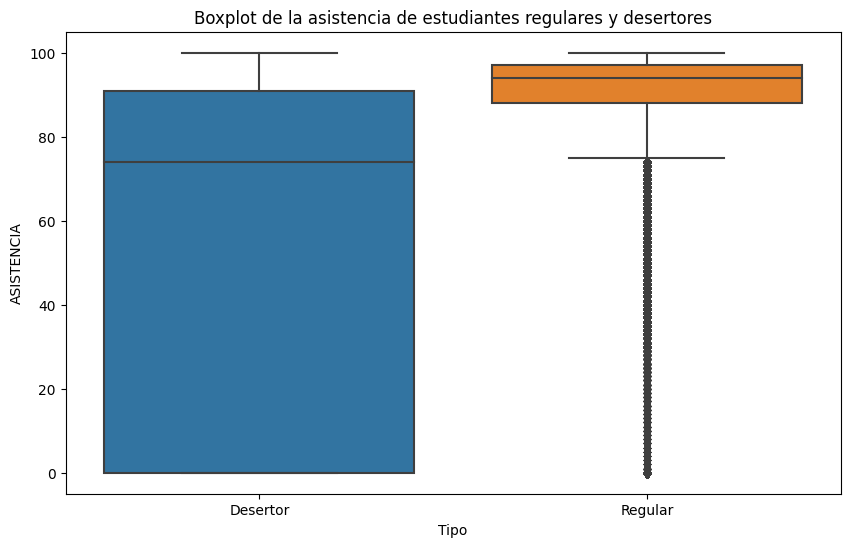
3.2.2 Tabla de contingencia

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **ESTADISTICAS DE ASISTENCIA** | | | | | | | |
|  |  |  |  |  |  |  |  |  |  |
|  |  | **Regular** | **Desertor** | **Regular** | **Desertor** | **Regular** | **Desertor** | **Regular** | **Desertor** |
|  |  |  |  |  |  |  |  |  |  |
| **Estadísticos Descriptivos** |  | **2018** | | **2019** | | **2020** | | **2021** | |
| **count** |  | 2.769.803 | 87.363 | 2.809.892 | 85.342 | 2.729.690 | 46.649 | 2.757.522 | 70.087 |
| **mean** |  | 86,62 | 52,63 | 85,47 | 52,47 | 90,80 | 71,12 | 86,34 | 63,76 |
| **std** |  | 23,48 | 41,78 | 23,84 | 41,97 | 18,09 | 38,47 | 21,66 | 39,43 |
| **min** |  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **25%** |  | 88 | 0 | 86 | 0 | 85 | 61 | 85 | 16 |
| **50%** |  | 94 | 74 | 93 | 74 | 97 | 85 | 91 | 85 |
| **75%** |  | 97 | 91 | 97 | 91 | 100 | 100 | 99 | 94 |
| **max** |  | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  | **ESTADISTICAS PROMEDIO GENERAL** | | | | | | | |
|  |  |  |  |  |  |  |  |  |  |
|  |  | **Regular** | **Desertor** | **Regular** | **Desertor** | **Regular** | **Desertor** | **Regular** | **Desertor** |
|  |  |  |  |  |  |  |  |  |  |
| **Estadísticos Descriptivos** |  | **2018** | | **2019** | | **2020** | | **2021** | |
| **count** |  | 2.769.803 | 87.363 | 2.809.892 | 85.342 | 2.729.690 | 46.649 | 2.757.522 | 70.087 |
| **mean** |  | 5,46 | 3,28 | 5,50 | 3,38 | 5,88 | 4,34 | 5,85 | 4,11 |
| **std** |  | 1,53 | 2,56 | 1,56 | 2,67 | 1,27 | 2,44 | 1,26 | 2,43 |
| **min** |  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **25%** |  | 5,3 | 0 | 5,3 | 0 | 5,5 | 2,5 | 5,5 | 2,3 |
| **50%** |  | 5,8 | 4,5 | 5,9 | 4,6 | 6,2 | 5,2 | 6,2 | 5 |
| **75%** |  | 6,3 | 5,4 | 6,3 | 5,7 | 6,7 | 6,3 | 6,6 | 6,1 |
| **max** |  | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 |

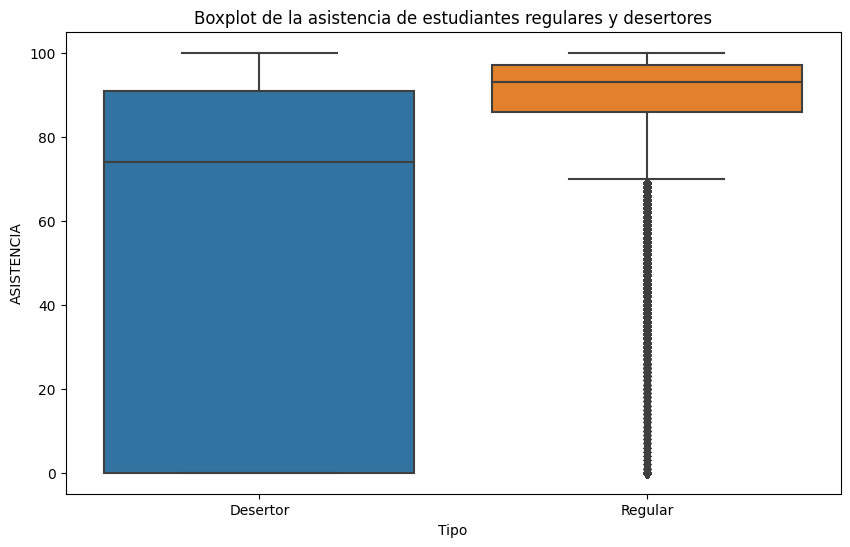
3.3. Exploración preliminar de los datos

**Gráficos de asistencia de estudiantes regulares y desertores 2018-2021**

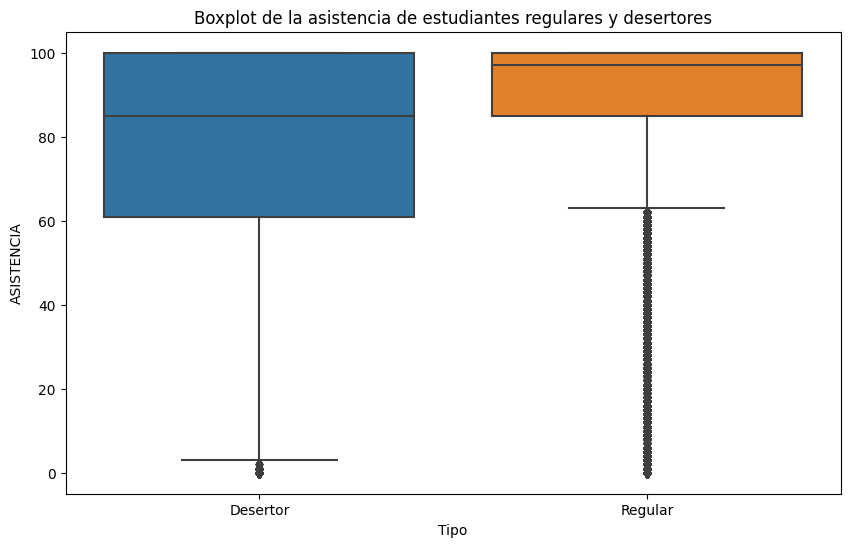
**Asistencia de estudiantes regulares y desertores 2018**



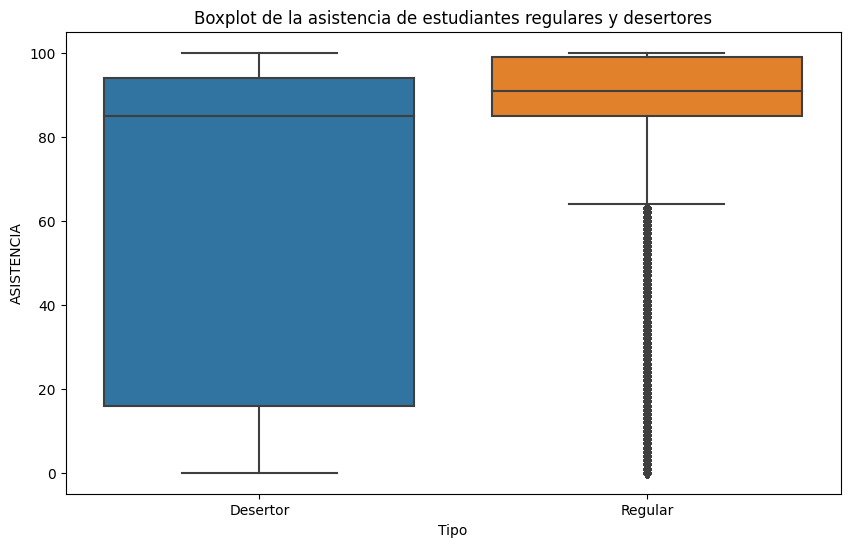
**Asistencia de estudiantes regulares y desertores 2019**



Asistencia de estudiantes regulares y desertores 2020



**Asistencia de estudiantes regulares y desertores 2021**

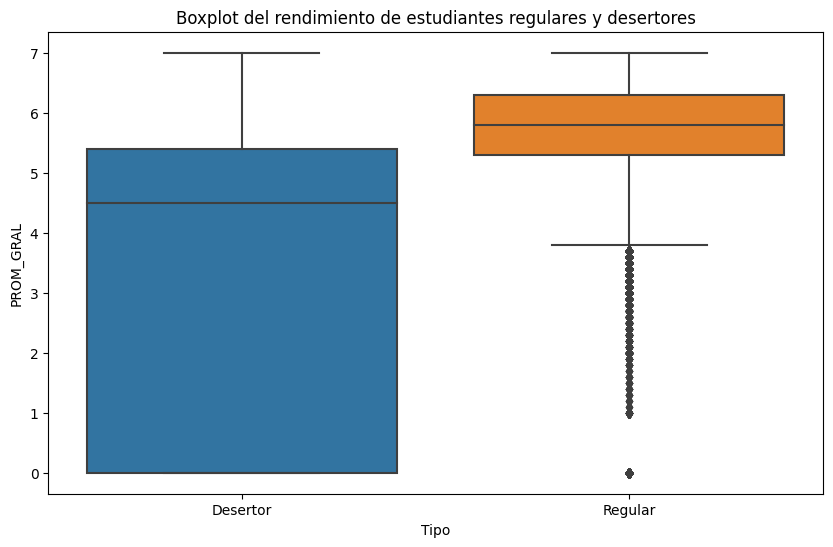


En los gráficos Boxplot de asistencia de alumnos regulares y desertores por año, se aprecia que los estudiantes regulares tienen una asistencia mayor que los estudiantes desertores. Las cajas de los Boxplot de alumnos regulares se concentran sobre el 80% en cambio las cajas de alumnos desertores parten cerca del 20% con excepción del año 2020 donde aumenta la asistencia incluso de los alumnos desertores. Esto se puede inferir de la baja cantidad de datos de asistencia que se posee para ese año declarado con alerta sanitaria a partir de la pandemia por brote de coronavirus a nivel mundial.

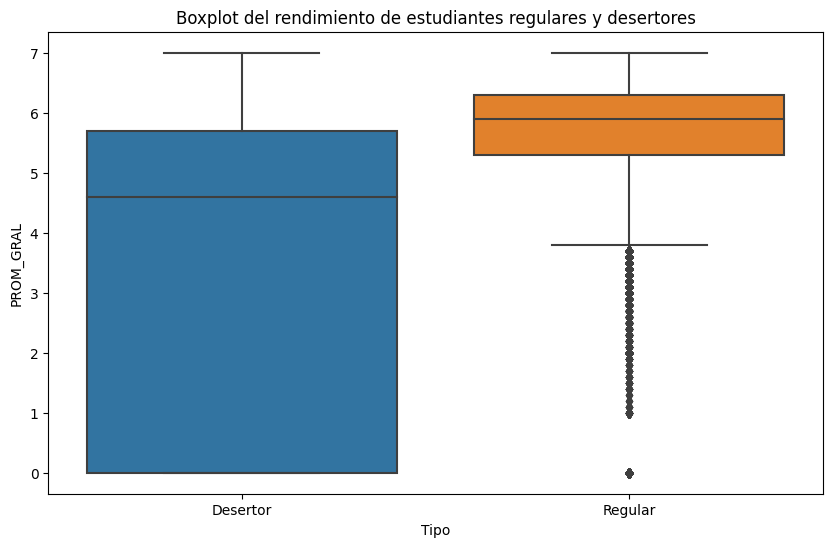
Por lo tanto, podemos concluir que la mediana de los alumnos desertores tiene asistencia más baja en relación con la misma medida de los alumnos regulares.

**Gráficos de rendimiento de estudiantes regulares y desertores 2018-2021**

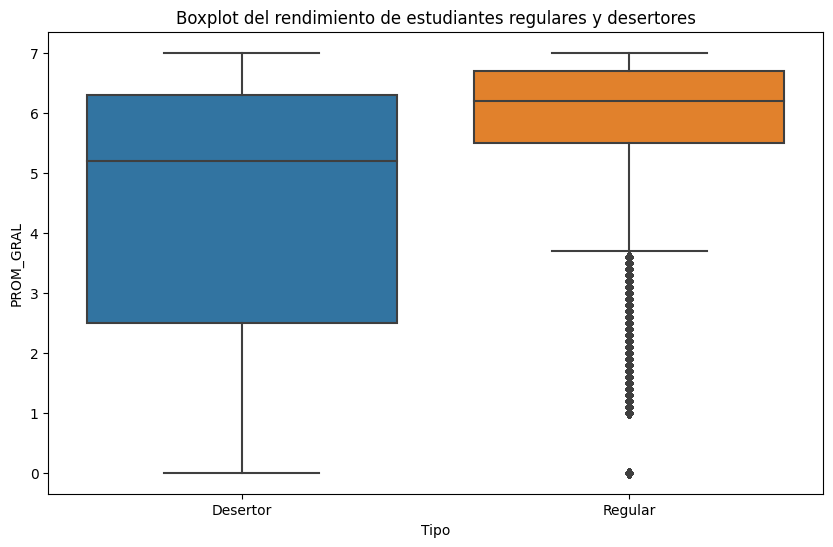
**Rendimiento de estudiantes regulares y desertores 2018**



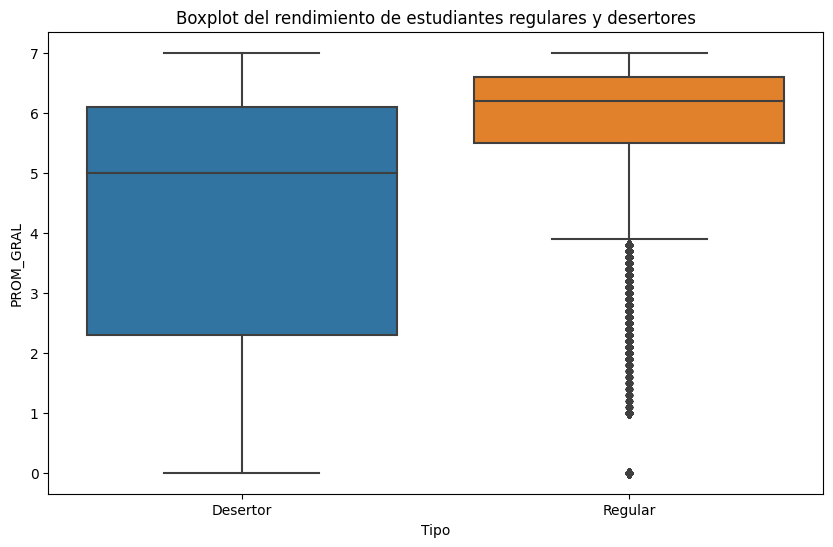
**Rendimiento de estudiantes regulares y desertores 2019**



**Rendimiento de estudiantes regulares y desertores 2020**



**Rendimiento de estudiantes regulares y desertores 2021**



En los gráficos Boxplot de rendimiento de alumnos regulares y desertores por año, se aprecia que los estudiantes regulares concentran un mejor rendimiento en notas que los estudiantes desertores.

Por lo tanto, podemos concluir que los alumnos desertores tienen un rendimiento más bajo en relación a los alumnos regulares.

Por otro lado, en los alumnos regulares se aprecia un leve aumento en la mediana del rendimiento pasando de 5,8 y 5,9 para los años 2018 y 2019 respectivamente a 6,2 en los años 2020 y 2021.

Al analizar todos los gráficos podemos notar que la mediana de los estudiantes desertores en todos los indicadores analizados está bajo el primer cuartil de los estudiantes regulares. Por lo tanto, el rango intercuartílico, demuestra una mayor concentración y homogeneidad de los estudiantes no desertores o regulares.

4.Bibliografia

* Cao, L., & Zhang, C. (2007). The evolution of KDD: Towards domain-driven data mining. International Journal of Pattern Recognition and Artificial Intelligence, 21(04), 677-692.
* Centro de Estudios de Seguridad Ciudadana (CESC). (2007). Sistema de Alerta Temprana de la deserción escolar en Peñalolén (SAT), Chile. Recuperado de http://cesc.uchile.cl/buenaspracticasenprevencion/bbp\_docs/03\_sistema\_alerta\_temprana\_chile.pdf (Visitado el 15 de mayo, 2023)
* Contreras, C. (2021). Determinación de variables predictivas de deserción inicial para generar un sistema de alerta temprana. Análisis sobre una muestra de estudiantes beneficiarios de la beca de nivelación académica en una universidad pública en Chile. Calidad en la Educación, (54), 12-45.
* Cordano F., M. (2018, abril 15). La deserción escolar en Chile es un problema que se comenta poco, pero existe. Recuperado de http://www.economiaynegocios.cl/noticias/noticias.asp?id=460171 (Visitado el 26 de mayo, 2023)
* Espinoza, Ó., Castillo, D., González, L. E., & Loyola, J. (2012). Factores familiares asociados a la deserción escolar en Chile. Revista de Ciencias Sociales (Ve), 18(1), 136-150.
* Kotsiantis, S. B., Pierrakeas, C. J., & Pintelas, P. E. (2003). Preventing student dropout in distance learning using machine learning techniques. In Knowledge-Based Intelligent Information and
* Engineering Systems: 7th International Conference, KES 2003, Oxford, UK, September 2003. Proceedings, Part II 7 (pp. 267-274). Springer Berlin Heidelberg.
* Opazo, C., & Sepúlveda, L. (2009). Deserción escolar en Chile:¿ volver la mirada hacia el sistema escolar?. REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación, 7(4), 120-135.
* Pyo, S., Uysal, M., & Chang, H. (2002). Knowledge discovery in database for tourist destinations. Journal of Travel Research, 40(4), 374-384.
* MINEDUC (2023). Plan de reactivación educativa 2023. Recuperado de https://reactivacioneducativa.mineduc.cl/wp-content/uploads/sites/127/2023/04/Plan-Reactivacion-05.04.2023-mas-tablas.pdf (Visitado el 26 de mayo, 2023)
* UNESCO, CEPAL, UNICEF. (2022). La encrucijada de la educación en América Latina y el Caribe Informe regional de monitoreo ODS4-Educación 2030 (Vol. 16). Organización de las Naciones Unidas para la Educación, a Ciencia y la Cultura. Recuperado de https://repositorio.cepal.org/bitstream/handle/11362/48153/1/S2200834\_es.pdf (Visitado el 26 de mayo, 2023)
* Rodríguez, P., Villanueva, A., Dombrovskaia, L., & Valenzuela, J. P. (2023). A methodology to design, develop, and evaluate machine learning models for predicting dropout in school systems: the case of Chile. Education and Information Technologies, 1-47.
* Santos, H. (2009). Dinámica de la deserción escolar en Chile. Santiago de Chile: Centro de Políticas Comparadas de Educación, CPCE.
* Smith Uldall, J., & Gutiérrez Rojas, C. (2022). UNA APLICACIÓN DE APRENDIZAJE AUTOMÁTICO (MACHINE LEARNING) EN POLÍTICAS PÚBLICAS. PREDICCIÓN DE ALERTA TEMPRANA DE DESERCIÓN ESCOLAR EN EL SISTEMA DE EDUCACIÓN PÚBLICA DE CHILE. Multidisciplinary business review, 15(1), 20-35.

5.Anexos

* **Anexo I:**

<https://docs.google.com/spreadsheets/d/1-LfMlVWOPK2EZtm90WP_iDWaBnSOk5E4/edit#gid=207276751>

* **Anexo II:** https://docs.google.com/spreadsheets/d/1LfMlVWOPK2EZtm90WP\_iDWaBnSOk5E4/edit#gid=1065370180