

# **ESCUELA POLITÉCNICA NACIONAL**

**FACULTAD DE SISTEMAS**

**UNIDAD DE TITULACIÓN**

**DESARROLLO DE UN MODELO PARA PREDECIR EL  
RENDIMIENTO ACADÉMICO DE ESTUDIANTES DE LA EPN EN  
BASE A SU NIVEL DE ACCESO A TICS Y FACTORES  
SOCIOECONÓMICOS.**

**TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DE GRADO DE  
MAGISTER EN SISTEMAS DE INFORMACIÓN MENCION INTELIGENCIA DE  
NEGOCIOS Y ANALÍTICA DE DATOS MASIVOS**

**ANDRÉS FERNANDO REINOSO QUIJO**

andres.reinoso01@epn.edu.ec

**Director: Dr. Edison Loza**

edison.loza@epn.edu.ec

**Codirector: Dra. Lorena Recalde**

lorena.recalde@epn.edu.ec

**Quito, enero 2023**

## **APROBACIÓN DEL DIRECTOR**

Como director del trabajo de titulación: DESARROLLO DE UN MODELO PARA PREDECIR EL RENDIMIENTO ACADÉMICO DE ESTUDIANTES DE LA EPN EN BASE A SU NIVEL DE ACCESO A TICS Y FACTORES SOCIOECONÓMICOS desarrollado por Andrés Fernando Reinoso Quijo, estudiante de la Maestría en Sistemas de Información Mención en Inteligencia de Negocios y Analítica de Datos Masivos, habiendo supervisado la realización de este trabajo y realizado las correcciones correspondientes, doy por aprobada la redacción final del documento escrito para que prosiga con los trámites correspondientes a la sustentación de la Defensa Oral.

---

**Dr. Edison Loza**  
**DIRECTOR**

## **APROBACIÓN DEL CODIRECTOR**

Como codirectora del trabajo de titulación: DESARROLLO DE UN MODELO PARA PREDECIR EL RENDIMIENTO ACADÉMICO DE ESTUDIANTES DE LA EPN EN BASE A SU NIVEL DE ACCESO A TICS Y FACTORES SOCIOECONÓMICOS desarrollado por Andrés Fernando Reinoso Quijo, estudiante de la Maestría en Sistemas de Información Mención en Inteligencia de Negocios y Analítica de Datos Masivos, habiendo supervisado la realización de este trabajo y realizado las correcciones correspondientes, doy por aprobada la redacción final del documento escrito para que prosiga con los trámites correspondientes a la sustentación de la Defensa Oral.

---

**Dra. Lorena Recalde**  
**CO - DIRECTORA**

## **DECLARACIÓN DE AUTORÍA**

Yo, Andrés Fernando Reinoso Quijo, declaro bajo juramento que el trabajo aquí descrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y que he consultado las referencias bibliográficas que se incluyen en este documento.

La Escuela Politécnica Nacional puede hacer uso de los derechos correspondientes a este trabajo, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.

---

**Ing. Andrés Fernando Reinoso Quijo**

## **DEDICATORIA**

Dedico este trabajo a quienes con sus sacrificios, enseñanzas, aventuras y amor marcaron mi existencia, y ahora; con su recuerdo me siguen impulsando y motivando a seguir adelante. Mis abuelos Andrés & Rosario, mi hermano Paul.

## **AGRADECIMIENTO**

A mis tutores y docentes cuyas enseñanzas dejaron huella en mi mente, mostrándome el horizonte profesional, despertando curiosidad en mi carrera y orientándome con su experiencia.

Mi profundo agradecimiento a Elizabeth, mi esposa, amiga, colega y compañera. Otra meta alcanzada en nuestra vida.

# ÍNDICE DE CONTENIDO

<b>1. INTRODUCCIÓN</b> .....	<b>6</b>
1.1. OBJETIVO GENERAL .....	7
1.2. OBJETIVOS ESPECÍFICOS.....	7
1.3. ALCANCE.....	7
1.4. MARCO TEÓRICO.....	8
1.4.1. APRENDIZAJE DE MÁQUINA .....	8
1.4.2. RANDOM FOREST REGRESSION .....	8
1.4.3. SUPPORT VECTOR MACHINE (SVM) .....	9
1.4.4. LINEAR REGRESSION (LR).....	10
1.4.5. ARTIFICIAL NEURAL NETWORKS (ANN).....	11
1.4.6. DECISION TREES (DT).....	11
1.4.7. TRABAJOS RELACIONADOS .....	12
<b>2. METODOLOGÍA</b> .....	<b>14</b>
2.1. INVESTIGACIÓN ACCIÓN (IA) .....	14
2.2. CROSS-INDUSTRY STANDARD PROCESS FOR DATA MINING CRISP-DM	15
2.2.1 COMPRENSIÓN DEL NEGOCIO .....	16
2.2.1.1 DETERMINAR LOS OBJETIVOS DEL NEGOCIO .....	16
2.2.1.2 CRITERIOS DE ÉXITO DEL NEGOCIO .....	17
2.2.1.3 EVALUACIÓN DE LA SITUACIÓN.....	17
2.2.1.4 INVENTARIO DE RECURSOS.....	17
2.2.1.5 REQUISITOS, SUPUESTOS Y RESTRICCIONES.....	17
2.2.1.6 TERMINOLOGÍA .....	18
2.2.1.7 COSTES Y BENEFICIOS .....	18
2.2.1.8 CRITERIOS DE ÉXITO DE MINERÍA DE DATOS .....	19
2.2.2 COMPRENSIÓN DE LOS DATOS.....	19
2.2.2.1 RECOLECCIÓN DE DATOS INICIALES .....	19
2.2.2.2 DESCRIPCIÓN DE LOS DATOS.....	19
2.2.2.3 EXPLORACIÓN DEL CONJUNTO DE DATOS .....	20
2.2.2.4 VERIFICAR LA CALIDAD DE LOS DATOS .....	26
2.2.3 PREPARACIÓN DE LOS DATOS.....	26

2.2.3.1	SELECCIÓN DE DATOS .....	26
2.2.3.2	LIMPIEZA DE DATOS .....	27
2.2.3.3	CONSTRUIR DATOS .....	28
2.2.3.4	INTEGRAR DATOS.....	28
2.2.4	MODELADO.....	31
2.2.4.1	SELECCIÓN DE TÉCNICA DE MODELADO .....	31
2.2.4.2	CONSTRUIR UN MODELO.....	32
3.	RESULTADOS Y DISCUSIÓN.....	34
3.1.	DISCUSIÓN .....	36
4.	CONCLUSIONES Y RECOMENDACIONES.....	38
4.1.	CONCLUSIONES .....	38
4.2.	RECOMENDACIONES.....	39
	ANEXOS .....	43



## LISTA DE FIGURAS

Figura 1. Estructura algoritmo Random Forest Regressor .....	9
Figura 2. Clasificador en un espacio vectorial .....	9
Figura 3. HiperPlano SVM [9] .....	10
Figura 4. Estimación LR.....	11
Figura 5. Posibles topologías ANN .....	11
Figura 6. Esquema Decision Tree.....	12
Figura 7. Ciclo de Investigación Acción aplicado al presente caso de estudio .....	15
Figura 8. Metodología CRISP – DM [21].....	15
Figura 9. Modelo de proceso CRISP-DM.....	16
Figura 10. Archivo inicial de datos EPN .....	18
Figura 11. Cabecera df1 .....	20
Figura 12. Cabecera df2 .....	24
Figura 13. Cabecera df3 .....	25
Figura 14. Representación gráfica de nulos y ceros .....	31

## LISTA DE TABLAS

Tabla 1. Tipos de datos DataFrame factores socioeconómicos .....	20
Tabla 2. Cantidad de datos no nulos df1 factores socioeconómicos .....	21
Tabla 3. Cantidad de datos no nulos df2 factores académicos .....	24
Tabla 4. Cantidad de datos no nulos df3 Encuesta factores socioeconómicos .....	25
Tabla 5. Resumen de valores nulos.....	28
Tabla 6. Atributos correlacionados variable objetivo .....	32
Tabla 7. Atributos más representativos, base de pruebas con todos los algoritmos.....	34
Tabla 8. Resultados de Entrenamiento, Validación y Prueba.....	35
Tabla 9. Resumen de resultados de ejecución.....	36

## LISTA DE ANEXOS

Anexo I – Glosario de términos .....	44
Anexo II – Descripción de datos .....	44
Anexo III – Proceso de limpieza de datos .....	44
Anexo IV– Integración de datos .....	44

## RESUMEN

Esta investigación tiene como objetivo determinar en qué relación los factores socioeconómicos y su nivel de acceso a las Tecnologías de la Información y la comunicación (TICs) afecta el rendimiento académico de estudiantes universitarios.

Para este estudio se hace uso de registros académicos, socioeconómicos y nivel de acceso a las TICs obtenidos de los registros de una de las universidades más grandes y reconocidas del país.

En todas sus fases se hace uso de metodologías de investigación formales con el objetivo de comprender el entorno educativo con los datos objeto de análisis, pasando por las fases de preparación y limpieza datos, para posteriormente modelar varios algoritmos de aprendizaje automático de máquina.

Con los resultados obtenidos se realiza una evaluación para determinar cuáles son los factores más influyentes en el rendimiento académico de los estudiantes. De donde se ha llegado a la determinación que los factores: año de inicio de estudios, edad de inicio de estudios, carrera que cursa el estudiante, facultad, créditos por materia registrada, porcentaje de créditos acumulados, número de miembros en el entorno familiar, ingresos económicos del padre, ingresos económicos de la madre, afiliación al seguro social del padre, afiliación al seguro social de la madre, propiedades que cuenta el estudiante, si el estudiante tiene vehículo, código Ipf del estudiante, si el computador del estudiante es para uso personal, influyen directamente en su calificación final y rendimiento académico.

Los resultados de este estudio constituyen información importante para autoridades de control académico en la tarea de mejorar las condiciones de estudiantes, brindándoles todos los medios para el éxito.

**Palabras clave:** Rendimiento Académico, Inteligencia artificial, Aprendizaje de Máquina, CRISP DM

## ***ABSTRACT***

The aim of this research was to establish the academic performance of university students by identifying the relationship between socioeconomic factors and the level of access to the Information and Communication Technologies (ICTs).

The study took place in a public university, one of the largest and recognized higher education institutions in Ecuador, South America. Academic records, socioeconomic background and level access to ICT was considered for the research.

The study consider Action Research and CRISP DM as a framework; undertaking the steps of data analysis, data preparation, cleaning phases, and finally consider several machine learning algorithms.

The study revealed a solid dataset with 15 variables that reflect its correlation and affect academic performance. year of beginning of studies, age of beginning of studies, career that the student is studying, faculty, credits for registered subject, percentage of credits accumulated, number of members in the family environment, father's salary, mother's salary, father's social security affiliation, mother's social security affiliation, properties owned by the student, if the student has a vehicle, the student's IPF code, if the student's computer is for personal use, have a direct influence on the final grade and academic performance.

The result of this research represent an important information for academic control authorities in the higher education; in the task of improving the conditions of students, providing them with all the needs for success.

### **Keywords:**

Academic Performance, Artificial Intelligence, Machine Learning, CRISP DM

## 1. INTRODUCCIÓN

El nivel de acceso a tecnología en la sociedad es un factor que ha sido objeto de análisis en campos como el productivo, social, económico, comercial, entre otros [1]. Sin embargo, debido al escenario de hacinamiento que se estableció como eje para evitar la propagación de la COVID-19 en especial en Latinoamérica [2], las TICs han tomado un rol fundamental en el desarrollo de actividades cotidianas. La formación académica no fue la excepción y tuvo que adaptarse para que ésta pueda ser realizada en su totalidad a través de medios electrónicos en línea. Dicha reforma ha puesto de manifiesto que el aislamiento social puede generar desigualdades entre sectores de la sociedad, incluso llegando a afectar el rendimiento académico de estudiantes [3].

Por ello, en búsqueda de la democratización de la educación superior se han generado iniciativas en diversas ramas de la educación que, entre otras, hacen uso de técnicas y herramientas informáticas que permiten identificar la brecha de desigualdad. Este es el caso del uso de minería de datos soportada por aprendizaje automático, que generalmente utiliza un conjunto de datos de entrenamiento etiquetados (preclasificados), los cuales se procesan para realizar predicciones sobre los mismos [4].

La presente investigación tiene como objetivo desarrollar un modelo predictivo que permita identificar el rendimiento académico de estudiantes de la Escuela Politécnica Nacional (EPN) en base a su nivel de acceso a las TICs y factores socioeconómicos de manera conjunta.

Las predicciones resultantes brindarán información a las autoridades universitarias para que puedan tomar decisiones informadas, lo cual proporcionará una adecuada integración académica de los estudiantes manteniendo un rendimiento académico dentro del estándar.

## **1.1. Objetivo General**

Generar un modelo de datos que permita predecir el rendimiento académico de estudiantes de pregrado de la EPN en base a su nivel de acceso a TICs y factores socioeconómicos, usando algoritmos de aprendizaje de máquina supervisados para aportar insumos para la toma estratégica de decisiones por las autoridades de la EPN.

## **1.2. Objetivos específicos**

- Realizar una revisión de literatura de indicadores de factores socioeconómicos y de acceso a TICs que puedan incidir en el rendimiento académico de estudiantes en el entorno latinoamericano.
- Analizar y comprender la situación actual de la EPN respecto del rendimiento académico a través de los datos proporcionados en la encuesta de enero del 2020 que hace relación a los factores socioeconómicos y nivel de acceso de TICs aplicada a estudiantes de la EPN.
- Probar y evaluar el desempeño de diferentes algoritmos de aprendizaje de máquina y aprendizaje profundo para determinar el algoritmo que brinde el mejor rendimiento y menor margen de error con los datos recolectados.
- Generar un documento científico presentando los resultados de la investigación a las autoridades académicas de la EPN para que puedan ser considerados en la toma estratégica de sus decisiones.

## **1.3. Alcance**

El alcance del este proyecto implica el uso de metodologías de investigación formal con la implementación de la metodología Cross Industry Standard Process for Data Mining (CRISP – DM) y algoritmos de aprendizaje de máquina para identificar los factores que inciden en el rendimiento académico de estudiantes. Este proyecto no incluye mecanismos de recolección automática de datos ni la implementación del algoritmo en cualquier sistema que permita su interacción con el usuario final.

## **1.4. Marco Teórico**

En esta sección se detallan las definiciones expuestas por varios autores relacionados a las diferentes temáticas que se abordan en este proyecto, tales como: Aprendizaje de Máquina, Random Forest Regression, Support Vector Machine, Linear Regression, Artificial Neural Networks, Decision Trees.

### **1.4.1. Aprendizaje de Máquina**

El aprendizaje de máquina es una rama interdisciplinaria entre la estadística e informática que permite generar algoritmos robustos a la hora de identificar una variedad de patrones en el conjunto de datos que son analizados [5]. Dentro del presente proyecto de investigación se aplicó diferentes algoritmos al mismo set de datos con el objetivo de evaluar su rendimiento y porcentaje de error, asegurándonos de esta manera de utilizar el más preciso para la predicción deseada.

### **1.4.2. Random Forest Regression**

Random forest es un algoritmo no paramétrico y, a diferencia de los enfoques de regresión tradicionales, se basa en modelos CART®. El objetivo de CART es comprender (aprender) la relación entre una variable dependiente ( $y$ ) y un conjunto de variables predictoras ( $X$ ). El algoritmo de aprendizaje emplea la partición recursiva en la que se seleccionan divisiones en las variables  $X$  para crear agrupaciones homogéneas de  $y$ . La partición recursiva continúa hasta que el subconjunto de  $y$  en cada nodo tiene el mismo valor o la división adicional no agrega ninguna mejora adicional [6]. La estructura del algoritmo random forest con  $N_{Tree} = 3$  [7] se representa en la Figura 1.



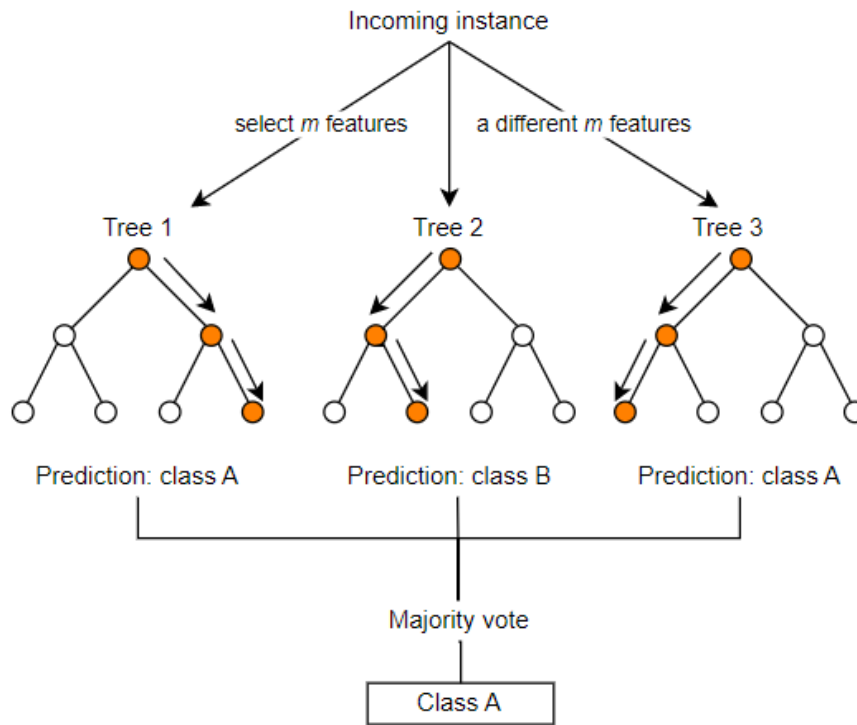


Figura 1. Estructura algoritmo Random Forest Regressor

### 1.4.3. Support Vector Machine (SVM)

SVM trabaja en tareas de clasificación binaria, está relacionada y contiene elementos de estadística aplicada no paramétrica, redes neuronales y aprendizaje automático [8].

Para entender su funcionamiento, podemos representar un espacio vectorial formado por dos clases con puntos  $x_1$  y  $x_2$ , que estas dos clases están situadas en zonas distintas del espacio vectorial, por lo que pueden separarse o clasificarse mediante un hiperplano separador  $w^T x = 0$ . [9], como se muestra en la Figura 2.

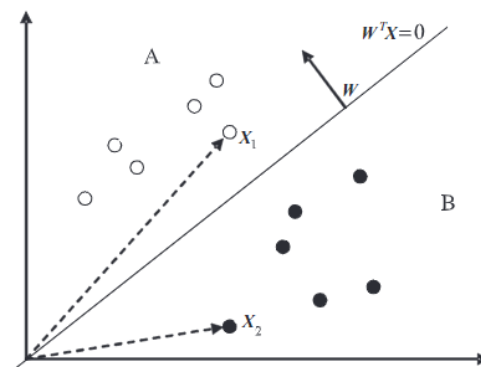


Figura 2. Clasificador en un espacio vectorial

La expresión del hiperplano separador es entonces  $w^T x = 0$ . Si se quiere saber cuál es la clase de un punto  $x_1$ , hay que determinar si el ángulo entre el vector  $w$  asociado al plano y la muestra  $x_1$  es mayor o menor que  $90^\circ$ . Si es inferior a  $90^\circ$ , la muestra se clasifica como perteneciente a la clase A. De lo contrario, la clasificación será B. Para ello, se calcula el producto escalar entre los vectores. Si es positivo, significa que el ángulo es menor a  $90^\circ$ , entonces la clase es A. Para un vector  $x_2$ , el ángulo será mayor a  $90^\circ$ , por lo que el producto escalar será negativo y la clase será B. Una muestra ubicada en el plano es ortogonal a  $w$ , por lo que el producto escalar será cero, lo que significa que no podemos determinar su clase, es en ese punto que se introduce un hiperplano para clasificar los elementos en el plano cartesiano. El procedimiento SVM consiste en colocar el hiperplano lo más lejos posible de la muestra más cercana como se muestra en la Figura 3.

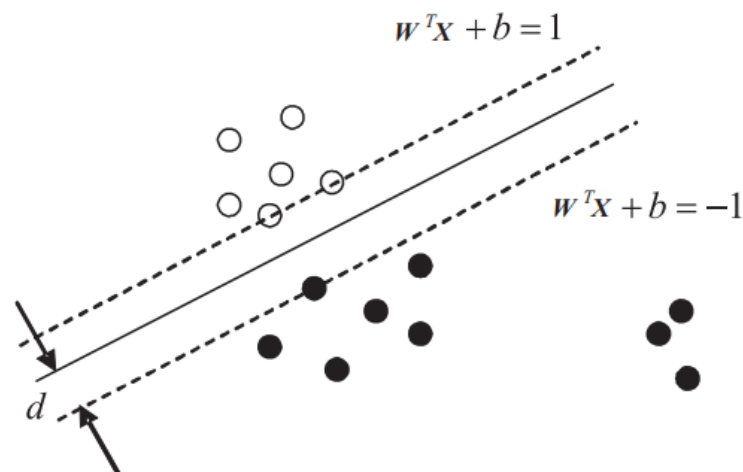


Figura 3. HiperPlano SVM [9]

#### 1.4.4. Linear Regression (LR)

Quizás uno de los algoritmos estadísticos y de aprendizaje automático más comunes y completos es la regresión lineal, el análisis de regresión se puede utilizar en algunos casos para determinar las relaciones causales entre las variables independientes y dependientes. El análisis de regresión estima el valor de la variable dependiente 'y' debido al rango de valores de la variable independiente 'x' [10], la estimación de regresión lineal sobre  $Y' = b_1 + b_2 X$  [11] se representa en la Figura 4.

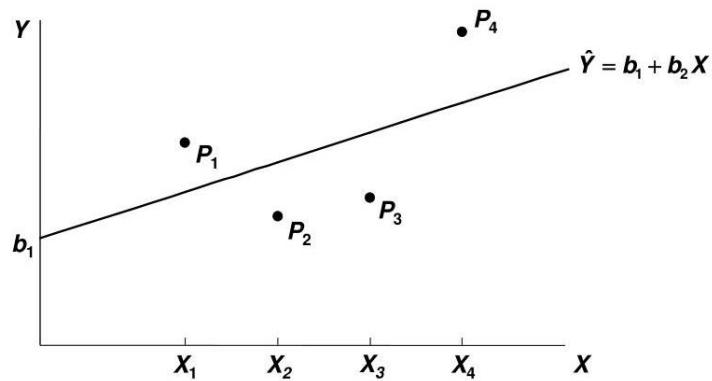


Figura 4. Estimación LR

#### 1.4.5. Artificial Neural Networks (ANN)

Las redes neuronales artificiales son una técnica de aprendizaje automático inspirada en la red neuronal biológica del cerebro humano. Su principio de funcionamiento consiste en enviar valores para cada neurona artificial, los procesa y envía a la siguiente capa después de procesar con entradas de neuronas en la capa anterior. Este modelo es bastante bueno en patrones de aprendizaje, puede adaptarse fácilmente a nuevos valores en los datos, pero el sistema puede mostrar una convergencia lenta y tiene el riesgo de un óptimo local o sobre-entrenamiento también conocido como “overfitting” [12]. La Figura 5 muestra un ejemplo de topologías ANN; izq. 7.a Red Monocapa, centro 7.b. Red Multicapa, der. 7.c Red recurrente [13].

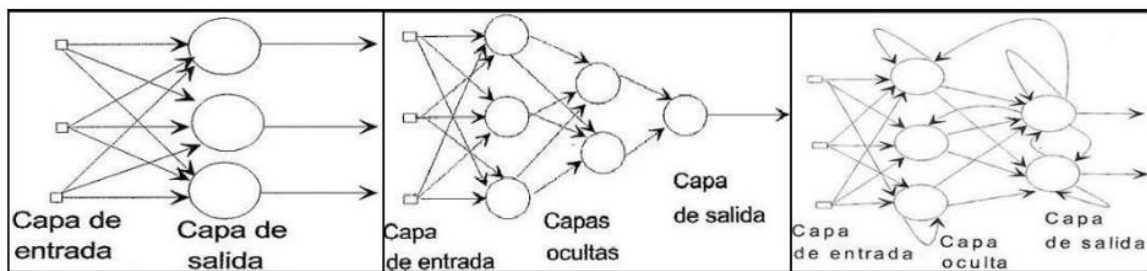


Figura 5. Posibles topologías ANN

#### 1.4.6. Decision Trees (DT)

Decision Trees es un método de aprendizaje automático intuitivo y ampliamente utilizado para crear modelos de predicción basados en declaraciones lógicas

básicas. A diferencia de SVM o la regresión polinomial, la relación entre las funciones y la salida no está predeterminada, lo que significa que no utiliza una función para ajustar los datos. DT crea un conjunto de preguntas como 'es igual' o 'es mayor' utilizando el conjunto de funciones, con las respuestas 'sí' o 'no'. Este proceso se repite hasta que no quedan más preguntas por responder, así se obtiene el resultado. DT crece dividiendo repetidamente los datos en secciones binarias [13], lo cual puede ser visualizado en la Figura 6, la cual muestra un esquema Decision Tree de 3 variables [ $x_1$ ,  $x_2$ ,  $x_3$ ] [15].

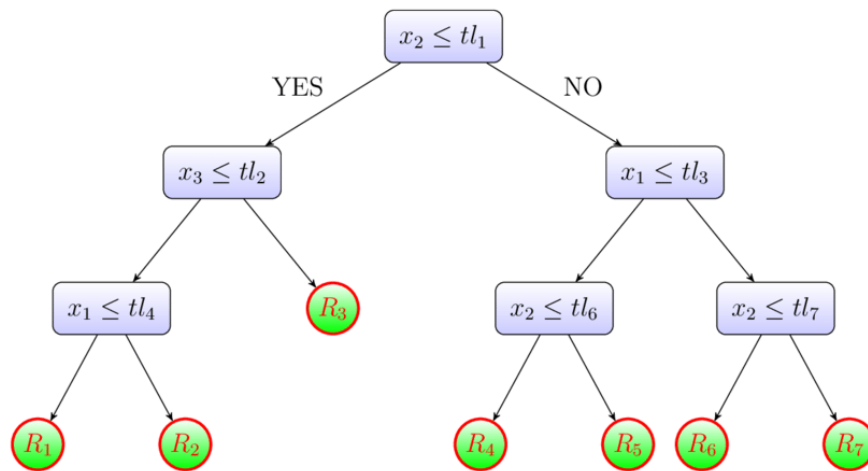


Figura 6. Esquema Decision Tree

#### 1.4.7. Trabajos Relacionados

Desde hace varios años, se han publicado numerosos trabajos en los cuales se emplea técnicas de aprendizaje de máquina con enfoque en factores socioeconómicos. En Argentina, se crearon perfiles de rendimiento académico con el objetivo principal de utilizar aquellos tendientes al fracaso o deserción como base a la determinación de futuras políticas de gestión académica que podrían implementarse para reducir dicho fenómeno [14]. Los altos índices de deserción académica también son una problemática en Colombia donde mediante un modelo de Árbol de Decisión implementando el algoritmo J48 con la herramienta WEKA se identificó varias causas de deserción [15]. Ecuador no es ajeno a esa problemática; por lo que se realizó un estudio analizando factores como raza, edad y etnia de los estudiantes y su influencia en diferentes contextos socioeconómicos utilizando algoritmos de clasificación para determinar cuáles son los más influyentes en el

rendimiento académico. Cabe destacar que los factores socioeconómicos son diversos y numerosos, sin embargo, en base a investigaciones previas se pueden clasificar en tres áreas: factores demográficos, capital humano y estabilidad residencial [16], y en base a esas investigaciones se concluye que el presente proyecto presenta variables y factores que no han sido analizados en otras investigaciones.

## **2. METODOLOGÍA**

En este capítulo se describe y desarrolla la metodología seleccionada, Investigación-Acción (IA) , la cual tiene como objetivo resolver problemas prácticos actuales al tiempo que amplía el conocimiento científico [17]. Esta metodología está fuertemente orientada hacia la colaboración y el cambio involucrando tanto a investigadores y a los sujetos de investigación (estudiantes objeto de evaluación socioeconómica) en un proceso iterativo que capitaliza el aprendizaje tanto de los investigadores como de los sujetos dentro del contexto del sistema social de los sujetos [18].

Debido a que el resultado esperado de nuestro proceso de IA es la generación de un modelo predictivo en base a técnicas Machine Learning, en las fases de Planificación de acciones y Toma de acciones se seguirá lo establecido en el marco de investigación CRISP-DM.

### **2.1. Investigación Acción (IA)**

IA constituye una opción metodológica práctica y social ya que por un lado permite la expansión del conocimiento, y, por la otra, va dando respuestas concretas a problemáticas que van planteando los participantes de la investigación [19].

La metodología IA representa un proceso por medio del cual los sujetos investigados se convierten en coinvestigadores, y participan activamente en el planteamiento del problema a ser investigado inclusive en otras fases permitiendo la prueba y evaluación de los resultados [20].

Existen diversas variaciones de la metodología entre las cuales se consolidó el resumen que se representa en la Figura 7.

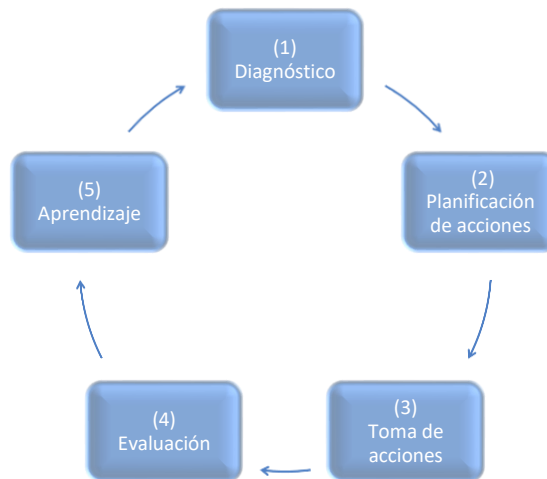


Figura 7. Ciclo de Investigación Acción aplicado al presente caso de estudio

## 2.2. Cross-Industry Standard Process for Data Mining CRISP-DM

CRISP-DM, es un método diseñado por IBM para dirigir proyectos de minería de datos. Como metodología incluye descripciones de las fases normales de un proyecto, las tareas necesarias en cada fase y una explicación de las relaciones entre las tareas, las mismas que se detallan en la Figura 8.



Figura 8. Metodología CRISP – DM [21]

Como modelo de proceso, CRISP-DM ofrece un resumen del ciclo vital de minería de datos y se representa en la Figura 9.

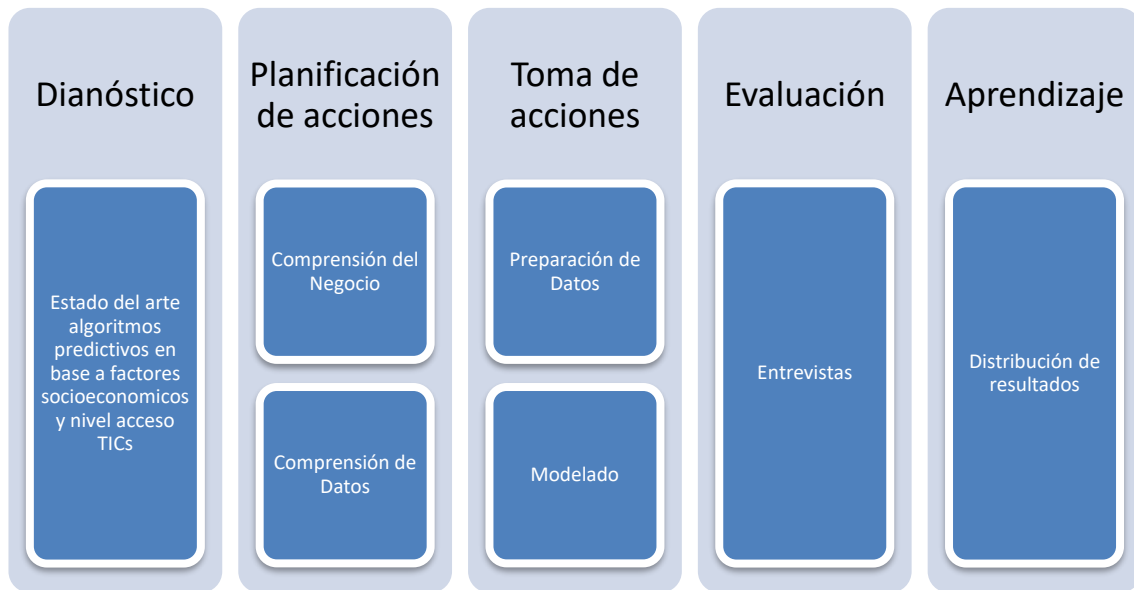


Figura 9. Modelo de proceso CRISP-DM

A continuación, se realiza una descripción breve de la aplicación de las fases de la metodología en este proyecto.

## 2.2.1 Comprensión del Negocio

Esta fase inicial se enfoca en comprender los objetivos y requisitos del proyecto tomando en consideración enunciados y conceptos propios de proceso de evaluación de la EPN, posteriormente, este conocimiento se convirtió en la definición de problema de minería de datos y un plan preliminar diseñado para cumplir los objetivos planteados.

### 2.2.1.1 Determinar los Objetivos del Negocio

El objetivo de la presente investigación es desarrollar un modelo de datos que permita predecir el rendimiento académico de estudiantes de pregrado de la EPN en base a su nivel de acceso a TICs y factores socioeconómicos, usando algoritmos de aprendizaje de máquina supervisados, aportando insumos para la toma estratégica de decisiones por las autoridades de la EPN.



### **2.2.1.2 Criterios de éxito del negocio**

Desde el punto de vista del negocio, se establece como criterio de éxito la posibilidad de realizar predicciones sobre el rendimiento académico de alumnos con un elevado porcentaje de fiabilidad, tomando en consideración sus factores socio académicos y nivel de acceso a TICs. Lo cual permitirá a las autoridades de la EPN la toma de decisiones estratégicas que beneficien a los intereses institucionales.

### **2.2.1.3 Evaluación de la Situación**

Para el desarrollo de esta investigación, se hizo uso de un documento Excel en el que constan los datos generales de los alumnos de la EPN desde el periodo 1970 hasta el 2020-A, cabe destacar que utilizaron los registros que cuentan de manera simultánea datos socioeconómicos, académicos y nivel de acceso a TICs, lo cual aseguró una cantidad de datos suficiente para ejecutar el procesamiento de información.

### **2.2.1.4 Inventario de recursos**

En cuanto a recursos de software, debido a su versatilidad, en todas las fases del ciclo de vida CRISP-DM se hizo uso de lenguaje Open Source Python como eje principal para la investigación dentro de la plataforma Google Colab.

Con relación a recursos de hardware, se utilizó un ordenador portátil con las siguientes características técnicas:

- Marca: ASUS©
- Modelo: TP50LA
- Procesador: Intel © Core i5 2.40 GHz
- Memoria RAM: 12,00 GB
- Capacidad de almacenamiento: 1TB SD
- Sistema operativo: Microsoft Windows 10 Professional ©

### **2.2.1.5 Requisitos, supuestos y restricciones**

Para esta investigación se hizo uso de datos reales de alumnos de la EPN, los mismos que fueron proporcionados en un archivo de Excel como se muestra en la

Figura 10. Los datos se encuentran identificados por el Código Único De Estudiante (CódigoEnc), el cual identifica los registros en tres pestañas: Datos Socioeconómicos, Académicos y Encuesta TICs

CódigoEnc	FecNacim	EstadoCiv	Genero	Etnia	País	Provincia	Canton	Discapaci	NDiscapaci	Parroquia	Telefono	Celular	Correo	Correosms	UnidadEdu	TituloCol	TipoUnida	PA
3AA4A4848ATEENC	11/10/1998	S	HOMBRE	MESTIZO/A	BELGICA	SIN REGISTR	SIN REGISTR	NULL	NULL	NULL	S	S	S	S	ABELARDO Y BACHILLER E FISCAL	EC		
3AA4A4868BLEOU	04/12/2000	S	HOMBRE	MESTIZO/A	VENEZUELA	SIN REGISTR	SIN REGISTR	NULL	NULL	NULL	S	S	S	S	COLEGIOS - BACHILLER E EXTRANJERO N			
3AA4A4887CAUM	02/01/2001	S	HOMBRE	MESTIZO/A	VENEZUELA	SIN REGISTR	SIN REGISTR	NULL	NULL	NULL	S	S	S	S	COLEGIOS - BACHILLER E EXTRANJERO N			
3AA4A4897BCOLA	21/06/1999	S	HOMBRE	MESTIZO/A	VENEZUELA	SIN REGISTR	SIN REGISTR	NULL	NULL	NULL	S	S	S	S	COLEGIOS - CIENCIAS - EXTRANJERO N			
3AA4A4897BCOLA	29/10/2001	S	HOMBRE	MESTIZO/A	ESPAÑA	SIN REGISTR	SIN REGISTR	NULL	NULL	NULL	S	S	S	S	CENTRAL TBI BACHILLER T FISCAL	EC		
3AA4A4897BCOLA	18/10/2001	S	HOMBRE	MESTIZO/A	ESPAÑA	SIN REGISTR	SIN REGISTR	NULL	NULL	NULL	S	S	S	S	TECNICO AS BACHILLER FISCAL	EC		
3AA4A4897BCOLA	06/08/2000	S	HOMBRE	MESTIZO/A	ESPAÑA	SIN REGISTR	SIN REGISTR	NULL	NULL	NULL	S	S	S	S	CAMILO PON BACHILLER E FISCAL	EC		
3AA4A4897BCOLA	02/03/1999	S	HOMBRE	MESTIZO/A	VENEZUELA	SIN REGISTR	SIN REGISTR	NULL	NULL	NULL	S	S	S	S	NUMA POMI BACHILLER E FISCAL	EC		
47ABC8ARETY	09/05/1989	S	MUJER	NULL	PERU	SIN REGISTR	SIN REGISTR	NULL	NULL	NULL	S	S	S	N	GALLEGO GAL QUIMICO BI PARTICULAR EC			
8A4A47ABAREN	28/08/1997	S	MUJER	MESTIZO/A	COLOMBIA	SIN REGISTR	SIN REGISTR	NULL	NULL	ITCHIMBIA	S	S	S	S	COLEGIOS - BACHILLER E EXTRANJERO N			
8A4A47ABAREN	24/05/1999	S	MUJER	MESTIZO/A	COLOMBIA	SIN REGISTR	SIN REGISTR	NULL	NULL	NULL	S	S	S	S	MANTIA ANG BACHILLER E FISCAL	EC		
8B3AAA83AMAMA	24/10/1994	S	HOMBRE	MESTIZO/A	COLOMBIA	SIN REGISTR	SIN REGISTR	NULL	NULL	CENTRO HIS	S	S	S	S	SANTIAGO ADMINISTRA PARTICULAR EC			
8B747886BIALO	04/10/1994	S	HOMBRE	MESTIZO/A	COLOMBIA	SIN REGISTR	SIN REGISTR	NULL	NULL	NULL	S	S	S	S	COLEGIOS - ADMINISTRA EXTRANJERO N			
8B74A4C8BIALO	28/05/1989	S	HOMBRE	NULL	COLOMBIA	SIN REGISTR	SIN REGISTR	NULL	NULL	NULL	S	S	S	N	AMAZONAS ADMINISTRA FISCAL	EC		
8B74A4C8BIALO	18/10/1988	S	MUJER	MESTIZO/A	COLOMBIA	SIN REGISTR	SIN REGISTR	NULL	NULL	CACELÉN	S	S	S	S	COLEGIOS - BACHILLER E EXTRANJERO N			
8B74A4C8BIALO	10/03/1996	S	HOMBRE	MESTIZO/A	COLOMBIA	SIN REGISTR	SIN REGISTR	NULL	NULL	COTOCOLLAS	S	S	S	S	MINOR UNI CIENCIAS FISCAL	EC		
8B74A4C8BIALO	11/11/1993	S	HOMBRE	NULL	COLOMBIA	SIN REGISTR	SIN REGISTR	NULL	NULL	NULL	S	S	S	S	ROCAFUERTI BACHILLER T FISCOMISOR EC			
A3A4A8888AINTO	27/05/2000	S	HOMBRE	MESTIZO/A	ECUADOR	CAÑAR	AZOGUES	NULL	NULL	NULL	S	S	S	S	INDEF BACHILLER E PARTICULAR EC			
A3A4A8888AINTO	19/09/2001	S	HOMBRE	MESTIZO/A	ECUADOR	EL ORO	HUACULLAS	NULL	NULL	NULL	S	S	S	S	VICTORIA BI BACHILLER PARTICULAR EC			
A3A4A8888AINTO	11/01/1997	S	HOMBRE	MESTIZO/A	ECUADOR	AZUAY	CUENCA	NULL	NULL	NULL	S	S	S	S	UNIDAD EDJ BACHILLER E PARTICULAR EC			
A3A4A8888AINTO	05/03/1999	S	HOMBRE	NULL	ECUADOR	CAÑAR	SUSCAL	NULL	NULL	NULL	S	S	S	N	SIN CODIGO N/E NO ESPECIF EC			
A3A4A8888AINTO	19/01/1998	S	HOMBRE	MESTIZO/A	ECUADOR	CAÑAR	CAÑAR	NULL	NULL	NULL	S	S	S	S	INST.TEC.SU ADMINISTRA FISCAL	EC		
A3A4A8888AINTO	18/09/1999	S	HOMBRE	INDIGENA	ECUADOR	CAÑAR	CAÑAR	NULL	NULL	NULL	S	S	S	S	INSTITUTO S BACHILLER FISCAL	EC		
A3A4A8888AINTO	25/09/2000	S	MUJER	MESTIZO/A	ECUADOR	CAÑAR	AZOGUES	NULL	NULL	NULL	S	S	S	S	JUAN BALTRI BACHILLER E FISCAL	EC		
A3A4A8888AINTO	05/04/1991	S	HOMBRE	NULL	ECUADOR	CAÑAR	AZOGUES	NULL	NULL	NULL	S	S	S	N	HERMAMO N BACHILLER E PARTICULAR EC			
A3A4A8888AINTO	09/10/1997	S	HOMBRE	MESTIZO/A	ECUADOR	CAÑAR	AZOGUES	NULL	NULL	IÑAQUITO	S	S	S	S	INSTITUTO L BACHILLER E FISCAL	EC		
A3A4A8888AINTO	07/04/1991	S	HOMBRE	NULL	ECUADOR	PICHINCHA	QUITO	NULL	NULL	SANTA PRISC	S	S	S	S	SAN GABRIEL BACHILLER E PARTICULAR EC			

Figura 10. Archivo inicial de datos EPN

### 2.2.1.6 Terminología

Dentro del desarrollo de esta investigación se hace uso de terminología específica, la misma que se puede validar en el Anexo I.

### 2.2.1.7 Costes y beneficios

Los datos de este proyecto no suponen ningún coste adicional a la universidad ya que estos datos pertenecen a la propia universidad desde el momento en el que el alumno se matricula en ella.

En cuanto a beneficios, no se puede decir que este proyecto genere algún beneficio económico para la universidad directamente, pero sí que puede suponerlo indirectamente ya que el objetivo de este proyecto es mejorar la calidad de los servicios ofrecidos a los alumnos por parte de la universidad, y por tanto la satisfacción de los clientes (los alumnos). Esto se traduce en prestigio para la universidad, lo cual hará que más alumnos consideren cursar sus estudios en esta universidad a la hora de elegir una.

### **2.2.1.8 Criterios de éxito de minería de datos**

Desde el punto de vista de la minería de datos se establece como criterio de éxito la posibilidad de realizar predicciones sobre rendimiento académico de estudiantes con porcentaje de fiabilidad no menor al 80%. El grado de fiabilidad lo determinará el algoritmo específico que se emplee a la hora de conseguir el modelo de la minería de datos, por lo que este tema se volverá a abordar más adelante en la fase de evaluación.

## **2.2.2 Comprensión de los datos**

Esta fase comprende la recolección inicial de los datos para poder establecer un primer contacto con el problema, familiarizarse con los datos y averiguar su calidad, así como identificar las relaciones más evidentes para formular las primeras hipótesis.

### **2.2.2.1 Recolección de datos iniciales**

Los datos utilizados en este proyecto corresponden a alumnos de la EPN que incluyen información personal como: fechas de nacimiento, notas, factores socio económicos y nivel de acceso a TICs.

Los datos se encuentran organizados en un archivo Excel por Código Enc y seccionados en tres pestañas: Datos Socioeconómicos, Académicos, Encuesta TICs.

### **2.2.2.2 Descripción de los datos**

El conjunto de datos inicial consta de 125 atributos con la siguiente cantidad de registros:

- Datos socioeconómicos- 26340 Registros
- Académicos -1035833 Registros
- Encuesta TICs -6483 Registros.

El detalle de los 125 atributos mostrando su significado, tipo de dato, cantidad de registros y sus posibles valores se ha consolidado en Anexo II.

### 2.2.2.3 Exploración del conjunto de datos

Una vez que se han descrito los datos, se procede a explorarlos. Esto implica revisar la distribución de datos, realizar un primer acercamiento sobre la calidad de datos y entendimiento de los mismos para una selección posterior con el objetivo de conseguir consistencia y completitud de los datos.

La Figura 11 muestra la cabecera del primer dataframe datos socioeconómicos de donde se valida que existen 123 atributos que lo componen y deben ser reducidos acorde a los objetivos de la investigación.

```
df1.head()
```

	CodigoEnc	FecNacimiento	EstadoCivil	Genero	Etnia	Pais	Provincia	Canton	Discapacidad	Discapacidad.1	...	lpf
0	3A4AAB64BATENC	1998-10-11	S	HOMBRE	MESTIZO/A	BELGICA	SIN REGISTRO	SIN REGISTRO	NaN	NaN	...	30.86
1	3AAA4B6B8BLEOU	2000-12-04	S	HOMBRE	MESTIZO/A	VENEZUELA	SIN REGISTRO	SIN REGISTRO	NaN	NaN	...	0.00
2	3AAA4CBBA3GUE	2001-01-02	S	HOMBRE	MESTIZO/A	VENEZUELA	SIN REGISTRO	SIN REGISTRO	NaN	NaN	...	1000.00
3	3AAA6AABB7CAJM	1999-06-21	S	HOMBRE	MESTIZO/A	VENEZUELA	SIN REGISTRO	SIN REGISTRO	NaN	NaN	...	1000.00
4	3AAA37BBBCOLA	2001-10-29	S	HOMBRE	MESTIZO/A	ESPAÑA	SIN REGISTRO	SIN REGISTRO	NaN	NaN	...	0.00

5 rows x 123 columns

Figura 11. Cabecera df1

Se esta revisión se observó que gran cantidad de datos están categorizados como Object por lo cual se consideró técnicas de limpieza de datos que permitan presentar los mismos en función de la variable objetivo y el tipo de algoritmo predictivo que se pretenda usar. Como resumen inicial se obtuvieron las siguientes cantidades y tipos de datos (ver Tabla 1):

**Tabla 1.** Tipos de datos DataFrame factores socioeconómicos

No	Tipo de dato	Cantidad
1	datetime64[ns]	2
2	float64	37
3	int64	2
4	object	82

Por otro lado, del análisis de los datos se identificó que existen registros con gran cantidad de datos nulos que deberán ser considerados para definir una estrategia de manejo de datos y se detallan en la Tabla 2.

**Tabla 2.** Cantidad de datos no nulos df1 factores socioeconómicos

<b>Atributo</b>	<b>No nulos</b>	<b>Tipo de dato</b>
CodigoEnc	26340	object
FecNacimiento	26340	datetime64[ns]
EstadoCivil	26340	object
Genero	26340	object
Etnia	19118	object
Pais	26340	object
Provincia	26340	object
Canton	26340	object
Discapacidad	42	object
Discapacidad.1	42	float64
Parroquia	18322	object
Teléfono	26340	object
Celular	26340	object
Correo	26340	object
CorreoIns	26340	object
UnidadEducativa	26340	object
TituloCol	26339	object
TipoUnidadEducativa	26340	object
PaisUnidadEducativa	26340	object
ProvinciaUnidadEducativa	26340	object
CantonUnidadEducativa	26313	object
UrbanResidence	26340	object
PeriodoInicioPrepo	25546	object
PeriodoFinPrepo	25546	object
NroPeriodosPro	26340	int64
Nivelacion	26170	object
PeriodoInicioCarr	26340	object
Begin_YR	26340	int64
Begin_Career_Date	26340	datetime64[ns]
AgeWhenStarting	26340	float64
PeriodoUltMatriculaCarr	26340	object
TipoUltMatriculaCarr	26340	object
FecUltMatriculaCarr	26317	object
FormaPagUltMatriculaCarr	26340	object
PagUltMatriculaCarr	12361	float64
Carrera	26340	object
Facultad	26340	object
CredAcum	19343	float64
CredRef	26337	float64
CredAcumRepro	13494	float64
PorcenAcum	26304	float64
Ingreso	18009	float64

NroMiem	18009	float64
TrabajaEstudiante	9330	object
OcupacionEstudiante	2743	object
IngresoEstudiante	26340	float64
RelLaboralEstudiante	3649	object
OcupacionConyugue	610	object
IngresoConyugue	26340	float64
AfiliadoEstudiante	9330	object
TrabajaConyugue	9330	object
RelLaboralConyugue	720	object
OcupacionPadre	20549	object
IngresoPadre	26340	float64
AfiliadoConyugue	9330	object
RelLaboralPadre	17590	object
OcupacionMadre	22736	object
IngresoMadre	26340	float64
AfiliadoPadre	9330	object
RelLaboralMadre	21058	object
AfiliadoMadre	9330	object
TrabajaHermano	9330	object
RelLaboralHermano	554	object
IngresoHermano	9330	float64
AfiliadoHermano	9330	object
Casadounion	9330	object
TieneHijos	9330	object
Hijos	9330	float64
TieneHijosMenores	9330	object
HijosMenores	9330	float64
Lactancia	9330	object
MadreSoltera	9330	object
PadreSoltero	9330	object
NroMiembrosDependen	9330	float64
Vive	9330	object
TieneDiscapacidad	9330	object
DiscapacidadFamiliar	9330	object
NroDiscapacidad	9330	float64
DiscapacidadTotal	9330	object
Enfermedad	9330	object
Embarazo	9330	object
Adultomayor	9330	object
Adolescente	9330	object
Violencia	9330	object
FinanciamientoEstudios	26340	object
OtrosIngresos	9330	object
Propiedad	9330	object

AvalúoPropiedad	9330	float64
NegocioPropio	9330	object
Impuesto	9330	object
Vehiculo	9330	object
VehículoPublico	9330	object
AvalúoVehiculo	9330	float64
Rebaja	9330	float64
IngresoArr	17022	float64
IngresoOtros	17022	float64
TotalRebaja	24325	float64
ValorRebajaTransporte	7198	float64
ValorRebajaServiciosBasicos	24310	float64
ValorRebajaInternet	8074	float64
ValorRebajaArriendo	5497	float64
ValorRebajaEnfermedad	1011	float64
TipoPropiedad	11865	object
ValorPropiedad	11865	float64
TipoPropiedadTerr	3695	object
ValorPropiedadTerr	3695	float64
TipoAuto	6528	object
AñoAuto	6524	float64
ValorAuto	6528	float64
TipoCamioneta	2485	object
AñoCamioneta	2479	float64
ValorCamioneta	2485	float64
QuintilPeriodoActual	18749	float64
lpf	26340	float64
EstadoEstudiante	6225	object
BecaExcelenciaAcad	26340	object
UltPerExcelenciaAcad	26340	object
BecaSituacionEcon	1857	object
UltPerSituacionEcon	1857	object
BecaMerito	284	object
UltPerMerito	284	object
BecaColegiatura	3	object
UltPerColegiatura	3	object

La Figura 12 muestra la cabecera del segundo dataframe (df2) de datos académicos de donde se aprecia que existen 9 atributos que lo componen.

```
df2.head()
```

	CodigoEnc	Periodo	ltComat	Materia	Nrom	Calif	Créditos	Aprobacion	Carrera
0	3A4AAB64BATENC	2019-A	CNIC010	FUNDAMENTOS DE MATEMATICA	1	9.40	0.00	F	INGENIERIA, CIENCIAS Y CIENCIAS ADMINISTRATIVAS
1	3A4AAB64BATENC	2019-A	CNIC020	GEOMETRIA Y TRIGONOMETRIA	1	10.50	0.00	F	INGENIERIA, CIENCIAS Y CIENCIAS ADMINISTRATIVAS
2	3A4AAB64BATENC	2019-A	CNIC030	FISICA	1	8.70	0.00	F	INGENIERIA, CIENCIAS Y CIENCIAS ADMINISTRATIVAS
3	3A4AAB64BATENC	2019-A	CNIC040	FUNDAMENTOS DE QUIMICA	1	10.30	0.00	F	INGENIERIA, CIENCIAS Y CIENCIAS ADMINISTRATIVAS
4	3A4AAB64BATENC	2019-A	CNIC050	LENGUAJE Y COMUNICACION	1	6.10	0.00	F	INGENIERIA, CIENCIAS Y CIENCIAS ADMINISTRATIVAS

Figura 12. Cabecera df2

Dentro del df2 se puede observar que los 9 atributos están categorizados con el tipo de dato Object por lo cual, se deberá considerar técnicas de limpieza de datos que permitan presentar los mismos en función de la variable objetivo que en este caso es el atributo Calif.

Cabe destacar que en este dataframe se cuenta con 1'035.832 registros. Esto se debe a que cada registro está asociado a una materia de cada estudiante. Por ello, un estudiante puede tener calificaciones en varias materias, esto hace que la cantidad de datos sea directamente proporcional a la cantidad de materias que cada estudiante "CodigoEnc" está cursando.

Por otro lado, del análisis de los datos en la Tabla 3 se puede verificar que no existe gran cantidad de datos nulos, lo cual se validará a posterior en el proceso de limpieza de datos.

**Tabla 3.** Cantidad de datos no nulos df2 factores académicos

Atributo	No nulos	Tipo de dato
CodigoEnc	1035832	object
Periodo	1035121	object
Comat	1035832	object
Materia	1035832	object
Nrom	1035832	object
Calif	970515	object
Créditos	1035832	object
Aprobacion	1035121	object
Carrera	1035832	object

La Figura 13 muestra la cabecera del tercer dataframe (df3) de Encuesta TICs de donde se señala que existen 16 atributos y 6482 registros que lo componen.



```
df2.head()
```

	CodigoEnc	Periodo	ltComat	Materia	Nrom	Calif	Créditos	Aprobacion	Carrera
0	3A4AAB64BATENC	2019-A	CNIC010	FUNDAMENTOS DE MATEMATICA	1	9.40	0.00	F	INGENIERIA, CIENCIAS Y CIENCIAS ADMINISTRATIVAS
1	3A4AAB64BATENC	2019-A	CNIC020	GEOMETRIA Y TRIGONOMETRIA	1	10.50	0.00	F	INGENIERIA, CIENCIAS Y CIENCIAS ADMINISTRATIVAS
2	3A4AAB64BATENC	2019-A	CNIC030	FISICA	1	8.70	0.00	F	INGENIERIA, CIENCIAS Y CIENCIAS ADMINISTRATIVAS
3	3A4AAB64BATENC	2019-A	CNIC040	FUNDAMENTOS DE QUIMICA	1	10.30	0.00	F	INGENIERIA, CIENCIAS Y CIENCIAS ADMINISTRATIVAS
4	3A4AAB64BATENC	2019-A	CNIC050	LENGUAJE Y COMUNICACION	1	6.10	0.00	F	INGENIERIA, CIENCIAS Y CIENCIAS ADMINISTRATIVAS

Figura 13. Cabecera df3

Dentro del df3 se puede observar que los 16 atributos que lo componen están categorizados con el tipo de dato Object por lo cual se deberá considerar dentro del proceso de limpieza de datos.

Por otro lado, del análisis de los datos, se valida que existe datos completos sin nulos hasta la pregunta 7 y a partir de la pregunta 8 a la 14 existe la misma cantidad de registros nulos. Esto se debe a que existe dependencia entre las respuestas a partir de la pregunta 7.

**Tabla 4.** Cantidad de datos no nulos df3 Encuesta factores socioeconómicos

Atributo	No nulos	Tipo de dato
CodigoEnc	6482	object
PREG1	6482	object
PREG2	6482	object
PREG3	6482	object
PREG4	6482	object
PREG5	6482	object
PREG6	6482	object
PREG7	6482	object
PREG8	708	object
PREG9	708	object
PREG10	708	object
PREG11	708	object
PREG12	708	object
PREG13	708	object
PREG14	708	object
PREG15	6482	object

Como se puede verificar, existe gran cantidad de atributos, por lo que en esta fase de la investigación se plantea reducir el número de columnas para poder trabajar con los datos que son relevantes para la investigación.

#### **2.2.2.4 Verificar la calidad de los datos**

Una vez realizados los procesos antes descritos y la exploración inicial de los datos, se puede afirmar que estos datos están completos. Se verifica que los mismos cubren los casos requeridos para la obtención de los resultados necesarios para cumplir los objetivos del proyecto.

En cuanto a los valores nulos, una vez que se haya seleccionado todos los atributos que conformarán el algoritmo predictivo, se tomará en consideración el manejo de nulos en cada caso particular para completar los datos o descartar el atributo por no aportar considerablemente para la investigación.

### **2.2.3 Preparación de los datos**

En esta fase de la metodología se prepara los datos para adecuarlos a las técnicas de minería de datos que se van a emplear sobre ellos. Esto implica seleccionar el subconjunto de datos que se va a utilizar, limpiarlos para mejorar su calidad, añadir nuevos datos a partir de los existentes y darles el formato requerido por la herramienta de modelado.

#### **2.2.3.1 Selección de datos**

Para la selección de datos se ha consolidado los registros por CódigoEnc de los tres dataframes. En fases previas se ha determinado que existen atributos no relevantes para los objetivos de la investigación, por ello se puede prescindir de algunos de ellos. En esta etapa se identificó al atributo objetivo el cual se denomina Calif.

En términos de registros, se hace uso de 6482 registros de todo el dataframe ya que son los estudiantes que han respondido la encuesta de factores socioeconómicos, académicos y encuesta TICs de manera simultánea.

Es así como, se ha escogido a los siguientes atributos para su posterior tratamiento:

FecNacimiento	IngresoEstudiante	EstadoEstudiante
EstadoCivil	RelLaboralEstudiante	UltPerExcelenciaAcad
Genero	IngresoConyugue	UltPerSituacionEcon
Etnia	RelLaboralPadre	UltPerMerito
Pais	AfiliadoPadre	Periodo

Provincia	RelLaboralMadre	Comat
Discapacidad	AfiliadoMadre	CodigoEnc
Discapacidad	Casadounion	PREG1
UnidadEducativa	Hijos	PREG2
TituloCol	Lactancia	PREG3
TipoUnidadEducativa	MadreSoltera	PREG4
ProvinciaUnidadEducativa	PadreSoltero	PREG5
UrbanResidence	NroMiembrosDependen	PREG6
PeriodoInicioPrepo	Vive	PREG7
NroPeriodosPro	TieneDiscapacidad	PREG8
Nivelacion	Enfermedad	PREG9
PeriodoInicioCarr	Embarazo	PREG10
AgeWhenStarting	Adultomayor	PREG11
Carrera	Adolescente	PREG12
Facultad	Violencia	PREG13
CredAcumRepro	FinanciamientoEstudios	PREG14
PorcenAcum	OtrosIngresos	PREG15
Ingreso	Vehiculo	lpf
NroMiem	QuintilPeriodoActual	Calif
TrabajaEstudiante		

El motivo para la inclusión o exclusión de algunos campos se debe a la importancia o pertinencia de dichos atributos en relación con los objetivos de la investigación, que se definieron en la fase 1 (comprensión del negocio) de la metodología.

### **2.2.3.2 Limpieza de datos**

El dataset consolidado contiene suficiente información para cumplir el objetivo de la investigación, no obstante, los mismos deben ser objeto de un proceso de calidad de datos, tomando en consideración que en algunos casos existe gran cantidad de nulos o NAN. Por otro lado; se ha detectado que en algunos atributos las respuestas almacenadas no son claras, son ambiguas o repetitivas por lo que se debe estandarizar resultados.

Dentro de la limpieza de datos se verifica que algunos datos tienen gran cantidad de registros nulos, por lo que se pondrá a consideración caso por caso la pertinencia de aplicar técnicas estadísticas para completar datos con la media, mediana o moda. El proceso de limpieza y sus comandos en Python se encuentran detallados en Anexo III.

### 2.2.3.3 Construir Datos

Dentro del proceso de limpieza de datos, se identificó atributos que tuvieron que ser procesados e integrados al dataframe y son los siguientes:

- Edad – Edad calculada del estudiante según la fecha de nacimiento
- NivelDiscapacidad. - Categoriza en 5 niveles a estudiantes según su nivel de discapacidad
- Nivel Ingreso. - Categoriza en 9 niveles a estudiantes según su nivel de ingresos
- Calif- Se realizó el promedio de notas por estudiante, agrupado por CodigoEnc.

### 2.2.3.4 Integrar datos

Conforme se estableció en la etapa de comprensión de datos, el conjunto de datos está dividido en 3 Dataframes (df). Cada df cuenta con información de estudiantes con una llave primaria "CodigoEnc", esto permite que se integren los datos en un df general para que puedan ser procesados. Este procedimiento se ha registrado en Anexo IV.

Una vez que se ha consolidado los datos en un df se procede a validar nuevamente su integridad en base al código establecido en el Anexo V, del cual se desprende la Tabla 5. Dicha tabla muestra la cantidad de valores nulos, cero y combinados, esto brinda una perspectiva amplia de cómo se manejará los datos a posterior.

**Tabla 5.** Resumen de valores nulos

Nombre Atributo	Valores Cero	Valores Nan	% de Nan	Total Cero y Nan	% Total Cero y Nan	Data Type
OcupacionPadre	0	1	0.0	1	0.0	object
OcupacionMadre	0	1	0.0	1	0.0	object
Nivelacion	0	1	0.0	1	0.0	object
NroMiem	0	1	0.0	1	0.0	float64
Enfermedad	0	5	0.1	5	0.1	object
AfiliadoPadre	0	5	0.1	5	0.1	object
DiscapacidadTotal	0	5	0.1	5	0.1	object
AfiliadoMadre	0	5	0.1	5	0.1	object
TrabajaHermano	0	5	0.1	5	0.1	object
Embarazo	0	5	0.1	5	0.1	object
TieneDiscapacidad	0	5	0.1	5	0.1	object

AfiliadoHermano	0	5	0.1	5	0.1	object
Casadounion	0	5	0.1	5	0.1	object
TieneHijos	0	5	0.1	5	0.1	object
Vive	0	5	0.1	5	0.1	object
TieneHijosMenores	0	5	0.1	5	0.1	object
PadreSoltero	0	5	0.1	5	0.1	object
AfiliadoConyugue	0	5	0.1	5	0.1	object
DiscapacidadFamiliar	0	5	0.1	5	0.1	object
Adultomayor	0	5	0.1	5	0.1	object
TrabajaConyugue	0	5	0.1	5	0.1	object
Rebaja	0	5	0.1	5	0.1	float64
VehiculoPublico	0	5	0.1	5	0.1	object
Vehiculo	0	5	0.1	5	0.1	object
Lactancia	0	5	0.1	5	0.1	object
NegocioPropio	0	5	0.1	5	0.1	object
Impuesto	0	5	0.1	5	0.1	object
TrabajaEstudiante	0	5	0.1	5	0.1	object
OtrosIngresos	0	5	0.1	5	0.1	object
Violencia	0	5	0.1	5	0.1	object
Adolescente	0	5	0.1	5	0.1	object
AfiliadoEstudiante	0	5	0.1	5	0.1	object
Propiedad	0	5	0.1	5	0.1	object
MadreSoltera	0	5	0.1	5	0.1	object
CantonUnidadEducativa	0	17	0.3	17	0.3	object
PeriodoFinPrepo	0	229	3.5	229	3.5	object
PeriodoInicioPrepo	0	229	3.5	229	3.5	object
Calif	7	958	14.8	965	14.9	float64
RelLaboralMadre	0	1069	16.5	1069	16.5	object
-TotalRebaja	0	1678	25.9	1678	25.9	float64
ValorRebajaServiciosBasicos	0	1680	25.9	1680	25.9	float64
RelLaboralPadre	0	2006	30.9	2006	30.9	object
CredAcum	797	1259	19.4	2056	31.7	float64
Ingreso	2459	1	0.0	2460	38.0	float64
Parroquia	0	2912	44.9	2912	44.9	object
CredAcumRepro	0	2913	44.9	2913	44.9	float64
AvalúoPropiedad	3047	5	0.1	3052	47.1	float64
ValorPropiedad	0	3848	59.4	3848	59.4	float64
TipoPropiedad	0	3848	59.4	3848	59.4	object
AvalúoVehiculo	3884	5	0.1	3889	60.0	float64
ValorRebajaInternet	0	4321	66.7	4321	66.7	float64
PagUltMatriculaCarr	0	4375	67.5	4375	67.5	float64
ValorRebajaTransporte	0	4712	72.7	4712	72.7	float64
TipoAuto	0	4990	77.0	4990	77.0	object
ValorAuto	0	4990	77.0	4990	77.0	float64
AñoAuto	0	4992	77.0	4992	77.0	float64
ValorRebajaArriendo	0	5249	81.0	5249	81.0	float64
ValorPropiedadTerr	0	5600	86.4	5600	86.4	float64
TipoPropiedadTerr	0	5600	86.4	5600	86.4	object
PREG8	0	5774	89.1	5774	89.1	object
PREG13	0	5774	89.1	5774	89.1	object
PREG10	0	5774	89.1	5774	89.1	object
PREG11	0	5774	89.1	5774	89.1	object
PREG14	0	5774	89.1	5774	89.1	object
PREG9	0	5774	89.1	5774	89.1	object
PREG12	0	5774	89.1	5774	89.1	object
UltPerSituacionEcon	0	5854	90.3	5854	90.3	object
BecaSituacionEcon	0	5854	90.3	5854	90.3	object
AñoCamioneta	0	5893	90.9	5893	90.9	float64
ValorCamioneta	0	5893	90.9	5893	90.9	float64
TipoCamioneta	0	5893	90.9	5893	90.9	object
RelLaboralEstudiante	0	5992	92.4	5992	92.4	object
RelLaboralHermano	0	6081	93.8	6081	93.8	object

IngresoHermano	6080	5	0.1	6085	93.9	float64
NroMiembrosDependen	6083	5	0.1	6088	93.9	float64
NroDiscapacidad	6148	5	0.1	6153	94.9	float64
ValorRebajaEnfermedad	1	6248	96.4	6249	96.4	float64
Hijos	6264	5	0.1	6269	96.7	float64
HijosMenores	6311	5	0.1	6316	97.4	float64
BecaMerito	0	6398	98.7	6398	98.7	object
UltPerMerito	0	6398	98.7	6398	98.7	object
RelLaboralConyugue	0	6422	99.1	6422	99.1	object
Discapacidad.1	0	6469	99.8	6469	99.8	float64
Discapacidad	0	6469	99.8	6469	99.8	object
EstadoEstudiante	0	6476	99.9	6476	99.9	object
+IngresoOtros	7	6473	99.9	6480	100.0	float64
+IngresoArr	9	6473	99.9	6482	100.0	float64
BecaColegiatura	0	6482	100.0	6482	100.0	float64
UltPerColegiatura	0	6482	100.0	6482	100.0	float64
OcupacionEstudiante	0	6482	100.0	6482	100.0	float64
OcupacionConyugue	0	6482	100.0	6482	100.0	float64

Con el objetivo de visualizar de mejor manera la información, se presenta la Figura 14 donde los espacios o columnas marcadas en negro representan nulos o Nan. Del análisis de la información se llega a la conclusión que existen atributos que contienen un alto porcentaje de valores nulos y Nan. Por ello, se establece como criterio y plan de acción para limpieza de datos trabajar únicamente con atributos con el 80% de datos utilizables, el resto son descartados. Con los datos con un porcentaje cercano al 10% de ceros y Nan se establecerán reglas y procedimientos para completar datos.

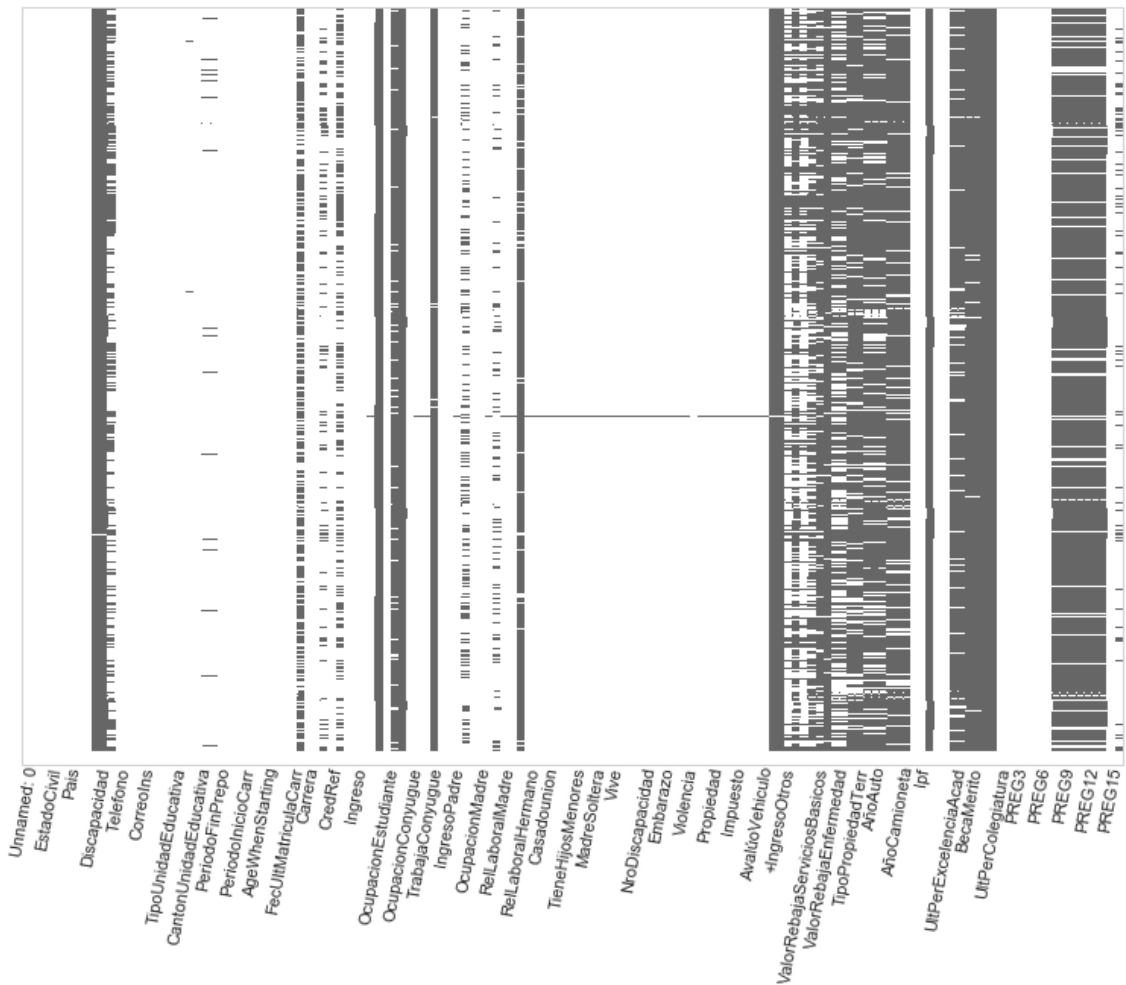


Figura 14. Representación gráfica de nulos y ceros

Como punto final en el proceso de limpieza de datos se identificó 5615 registros válidos los mismos que serán utilizados como base para la implementación de los modelos seleccionados.

## 2.2.4 Modelado

En esta sección se describe las tareas que permitieron crear el modelo.

### 2.2.4.1 Selección de técnica de modelado

Para el desarrollo de esta investigación se hizo uso de los siguientes algoritmos de inteligencia artificial:

- Random Forest Regressor
- Lineal SVM
- Lineal Regressor

- ANN
- Decision Tree Regressor

En cada uno de ellos se realizó pruebas con datos normalizados y los métodos Get Dummies y One Hot Encoder para el manejo de datos categóricos.

Get Dummies convierte una variable categórica en variables Dummy. Según la cantidad valores diferentes que posea la variable original getdummies presentará la misma cantidad de columnas dummy con un prefijo para identificarlas [22].

One Hot Encoder codifica las características categóricas como una matriz numérica única. Las funciones se codifican mediante un esquema de codificación one-hot (también conocido como "uno de K" o "ficticio"). Esto crea una columna binaria para cada categoría y devuelve una matriz dispersa o una matriz densa (según el parámetro *sparse*) [23].

#### 2.2.4.2 Construir un modelo

Una vez preparados todos los conjuntos de datos, se ejecutó los modelos bajo la herramienta Google Colab y se generó 10 modelos funcionales que fueron analizados, de los cuales se desprenden los siguientes hallazgos.

Basados en el set de datos resultado de la fase de limpieza de datos, se realizó pruebas con cada algoritmo en donde se pudo identificar que el tiempo de procesamiento era muy alto y en casos como SVM no se contaba con suficientes recursos de hardware que soporte procesamiento. Debido a ello, se hizo uso de la librería *sklearn.feature\_selection.SelectKBest* la cual muestra una valoración de los valores k más altos [24], permitiendo seleccionar los atributos que tienen correlación directa con la variable objetivo "Calif", obteniendo los resultados que se muestran en la Tabla 6. Con esta valoración se realizó una nueva selección con los 15 atributos más representativos del dataset, el nuevo df que se presenta en la Tabla 7.

**Tabla 6.** Atributos correlacionados variable objetivo

No	Atributo	score
1	PeriodoInicioCarr_2019-B	1.366.973.327
2	Facultad_NIVELACION	1.286.013.507
3	Carrera_INGENIERIA, CIENCIAS Y CIENCIAS ADMINI...	937.082.337
4	PorcenAcum	507.077.001



5	PeriodoInicioCarr_2017-A	183.298.135
6	Carrera_NIVEL TECNOLOGICO SUPERIOR	161.413.573
7	CredRef	117.963.840
8	lpf	98.066.448
9	Begin_YR	95.106.956
10	Facultad_INGENIERIA QUIMICA Y AGROINDUSTRIA	90.561.739
11	ComputadorEsPersonal_Si	88.436.482
12	PeriodoInicioCarr_2018-A	78.787.638
13	Vehiculo_SI	76.997.337
14	Facultad_INGENIERIA CIVIL	75.938.725
15	Facultad_INGENIERIA ELECTRICA Y ELECTRONICA	68.140.935
16	Carrera_INGENIERIA QUIMICA	60.134.909
17	Propiedad_SI	60.082.210
18	PeriodoInicioCarr_2016-A	59.778.917
19	IngresoMadre	57.027.071
20	AfiliadoMadre_SI	56.560.084
21	Carrera_INGENIERIA CIVIL	55.351.203
22	IngresoPadre	39.795.067
23	Carrera_INGENIERIA AMBIENTAL	33.717.624
24	PeriodoInicioCarr_2015-A	32.425.452
25	Carrera_INGENIERIA EN SISTEMAS INFORMATICOS Y ...	30.228.860
26	PeriodoInicioCarr_2019-A	30.029.629
27	Carrera_(RRA) INGENIERIA QUIMICA	29.864.614
28	Facultad_INGENIERIA DE SISTEMAS	27.799.911
29	Carrera_INGENIERIA EMPRESARIAL	26.191.683
30	NroMiem	22.318.704
31	AfiliadoPadre_SI	20.106.004
32	Carrera_INGENIERIA ELECTRONICA Y CONTROL	19.957.570
33	PeriodoInicioCarr_2017-B	19.166.257
34	Carrera_(RRA) ELECTRONICA Y AUTOMATIZACION	18.989.825
35	PeriodoInicioCarr_2015-B	16.873.522
36	PeriodoInicioCarr_2016-B	15.787.396
37	Carrera_INGENIERIA AGROINDUSTRIAL	14.998.115
38	Facultad_CIENCIAS ADMINISTRATIVAS	14.470.885
39	Carrera_INGENIERIA EN ELECTRONICA Y TELECOMUNI...	14.281.898
40	Carrera_(RRA) SOFTWARE	13.775.322
41	ComputadorEsPersonal_No tiene computador	13.403.563
42	Carrera_INGENIERIA EN CIENCIAS ECONOMICAS Y FI...	12.565.413
43	Facultad_GEOLOGIA Y PETROLEOS	11.661.006
44	Carrera_(RRA) FISICA	10.579.796
45	PeriodoInicioCarr_2018-B	9.745.669
46	PeriodoInicioCarr_2014-B	9.158.775
47	Facultad_INGENIERIA MECANICA	8.771.951
48	Carrera_INGENIERIA MECANICA	8.681.339
49	Carrera_(RRA) ELECTRICIDAD	7.989.162

50	Facultad_TECNOLOGICA	7.890.669
51	Carrera_(RRA) TECNOLOGIA SUPERIOR EN AGUA Y SA...	7.033.689
52	Carrera_INGENIERIA ELECTRICA	6.468.999
53	Carrera_INGENIERIA EN PETROLEOS	6.094.509
54	AgeWhenStarting	5.488.040
55	PeriodoInicioCarr_2014-A	5.384.497
56	Carrera_(RRA) INGENIERIA CIVIL	4.531.924
57	Carrera_(RRA) TECNOLOGIA SUPERIOR EN REDES Y T...	4.494.283
58	Carrera_INGENIERIA GEOLOGICA	4.350.171
59	PeriodoInicioCarr_2013-B	4.179.795
60	Carrera_(RRA) TELECOMUNICACIONES	4.114.117

**Tabla 7.** Atributos más representativos, base de pruebas con todos los algoritmos

Column	Non-Null Count	Data Type
PeriodoInicioCarr	5615 non-null	object
Begin_YR	5615 non-null	float64
AgeWhenStarting	5615 non-null	float64
Carrera	5615 non-null	object
Facultad	5615 non-null	object
CredRef	5615 non-null	float64
PorcenAcum	5615 non-null	float64
NroMiem	5615 non-null	float64
IngresoPadre	5615 non-null	float64
IngresoMadre	5615 non-null	float64
AfiliadoPadre	5615 non-null	object
AfiliadoMadre	5615 non-null	object
Propiedad	5615 non-null	object
Vehiculo	5615 non-null	object
lpf	5615 non-null	float64
ComputadorEsPersonal	5615 non-null	object
Calif	5615 non-null	float64

### 3. RESULTADOS Y DISCUSIÓN

En esta sección se muestra los resultados obtenidos de los diferentes modelos puestos a prueba en base al conjunto de datos preparado. Este dataframe fue sometido a los procesos de calidad descritos en la fase de preparación de datos. Luego de ello se procedió a realizar un resumen de sus métricas de precisión para su discusión.

En este experimento se probó el rendimiento de 5 algoritmos de aprendizaje de máquina: Random Forest Regressor, SVM, Linear Regressor, Neuronal Network, DecisionTree Regressor. Para cada uno de ellos se convirtió los datos categóricos del dataframe bajo las metodologías Dummies y Encoder, dando como resultado 5 conjuntos de métricas de precisión y rendimiento diferentes.

Cabe recalcar que el objetivo de este estudio es determinar el algoritmo con mejor calidad de predicción del rendimiento académico de estudiantes de la EPN en base a sus factores socioeconómicos y nivel de acceso a las TICs.

A continuación, en la Tabla 8 se presentan los resultados alcanzados posterior a las etapas de entrenamiento, validación y pruebas con los algoritmos antes descritos:

**Tabla 8.** Resultados de Entrenamiento, Validación y Prueba

Algoritmo	Método	Métricas			
		Precisión	MAE	R2	RMSE
Random Forest Regressor	Dummies	83,70%	1,43	4,48	2,11
	Encoder	65,80%	1,44	4,37	2,09
Lineal SVM	Dummies	16,00%	1204	-4,35	1455,55
	Encoder	14,79%	7,83	70,52	8,9
Lineal Regressor	Dummies	36,00%	3,22	0,37	4,15
	Encoder	26,00%	7,7	2,28	9,48
ANN	Dummies	4,85%	2,73	0,54	3,54
	Encoder	12,56%	2,91	0,41	3,99
Decision Tree Regressor	Dummies	34,70%	3,03	0,34	4,23
	Encoder	69,17%	1,95	8,46	2,91

Una vez establecido el algoritmo con mayor precisión (Random Forest Regressor) se generó un archivo .pkl para la implementación del modelo, con el cual se realizó la validación con resultados en un esquema de pruebas seleccionando aleatoriamente 10 registros de estudiantes con su promedio de notas conocidas.

Se creó un dataframe "x\_val" con datos de prueba separando la variable objetivo Calif en un dataframe independiente "y\_val". Con el set de datos cargado se procedió a llamar al modelo de datos ejecutando la predicción sobre un set de datos

vacio “y\_predVal” obteniendo como resultado la predicción de notas en base al modelo entrenado. Datos obtenidos que se presentan en la Tabla 9.

**Tabla 9.** Resumen de resultados de ejecución

Begin_YR	AgeWhenStarting	...	ComputadorEsPersonal	ValReal	ValPredict
2017	18,88	...	1	34,19	33,92
2015	18,80	...	1	29,82	31,13
2017	19,01	...	1	32,76	32,12
2018	32,40	...	0	30,07	28,74
2011	18,51	...	1	28,87	23,98
2019	21,56	...	0	29,58	29,40
2018	19,73	...	1	33,94	30,07
2015	18,87	...	1	32,78	31,81
2020	18,82	...	0	24,99	24,99

Se puede validar que los valores ValReal vs ValPredict se encuentran dentro del rango de tolerancia establecido en métrica MAE 1.43, lo cual demuestra el funcionamiento del modelo desarrollado dentro del margen de precisión 83,70% señalado anteriormente.

Con el objetivo de reproducir los resultados se ha colocado el código del algoritmo con mayor precisión Random Forest Regressor bajo el proceso Get Dummies en un repositorio GitHub de acceso libre con el siguiente URL: <https://github.com/andresreinosoq/ModeloRandomForestRegressor.git>

### 3.1. Discusión

Los resultados obtenidos de la Tabla 8 muestran diversas métricas, no obstante, se resalta que el algoritmo con el mejor rendimiento es: Random Forest Regressor haciendo uso de la metodología *Dummies* con 83,7%, de igual manera se puede verificar que la métrica MAE muestra un valor de 1.43 lo cual representa un valor aceptable considerando que el rango base de la variable objetivo *Calif* va de 0 a 40 puntos.

De la información presentada se puede validar que el algoritmo Decision Tree Regressor con la metodología *enconder* también presenta un porcentaje de precisión elevado con 69,17% y MAE de 1,95. No obstante, se puede considerar como un rendimiento no satisfactorio.

Para el desarrollo de esta investigación se utilizó una gran cantidad de datos, la cual fue sometida a un proceso de calidad de datos, de donde se identificó los 15 atributos con directa relación con la variable objetivo Calif. Para asegurar la reproducibilidad de los resultados para cada algoritmo se realizó el balanceo de datos y se particionó del 75% para entrenamiento, 25% para validación y test.

## **4. CONCLUSIONES Y RECOMENDACIONES**

### **4.1. CONCLUSIONES**

Sin lugar a duda, la etapa que conlleva un mayor tiempo en desarrollo es la fase de análisis y comprensión de los datos del negocio, con su posterior etapa de calidad de datos. En esta fase se ocupó el gran parte del tiempo de ejecución de este proyecto, dejando como resultado un dataframe confiable y estructurado para continuar con las siguientes fases de investigación. De igual manera, se concluye que el éxito o fracaso de este tipo de proyectos depende de obtener resultados de calidad para poder trabajar.

Para el desarrollo de este proyecto se hizo uso de la metodología CRISP-DM, la cual brindó un marco de acción en cada una de las fases de investigación y un modelo de documentación e información entregable.

De esta manera se pudo comprender estructura de los datos, su funcionamiento y aplicación en la investigación, por otro lado, facilitó la consolidación de resultados en la parte final del proyecto.

Dentro de la implementación de los algoritmos de prueba, se encontró varias incidencias que debieron ser superadas en su debido momento. Por ejemplo, se observó que la ejecución de los diferentes algoritmos es un reto, ya que, a pesar de contar con el mismo set de datos para todas las pruebas, cada algoritmo funciona de manera diferente, tienen diferentes metodologías para procesar datos, consumen diferentes cantidades de recursos en hardware y se debió realizar desarrollos independientes en cada uno de ellos.

Debido a la diversidad de algoritmos de prueba, se encontró una variación amplia en las métricas de precisión de cada uno de ellos. Como se determinó en el marco teórico el principio de funcionamiento, base matemática y variables de entrada de cada algoritmo es diferente por ello es de esperarse variación en sus resultados.

De los resultados obtenidos se identifica que el algoritmo. Random Forest Regressor transformando sus datos categóricos en *Dummies* es el más eficiente y presenta el mejor rendimiento con 83,7%, con un MAE de 1.43. Esta métrica representa un valor aceptable considerando que el rango de datos variable objetivo Calif comprende de 0 a 40 puntos. En ese sentido, y posterior a la implementación con datos reales se concluye que el algoritmo es eficiente y presenta resultados confiables.

## **4.2. RECOMENDACIONES**

En este proyecto, la fase de comprensión de negocio y su posterior análisis de calidad de datos fue la base sobre la cual se obtuvo los resultados. En este sentido, se recomienda estandarizar los cuestionarios de recolección de información en la EPN para proveer información amplia que permita obtener mayor cantidad de atributos y registros para fortalecer la aplicación de algoritmos de aprendizaje de máquina.

Para la ejecución de los algoritmos, en primera instancia se utilizó un recurso local con Jupyter, no obstante, se tuvo inconvenientes en el procesamiento de gran cantidad de datos, de tal manera que se hizo uso de la herramienta Google Colab. Esta resultó ser una herramienta útil al prestar recursos de procesamiento sin costo adicional, por lo que, se recomienda su uso ya que esta ofrece flexibilidad y buen rendimiento.

Toda vez que se ha identificado los factores socioeconómicos y nivel de acceso a las TICs que influyen en las calificaciones de estudiantes o rendimiento académico. La recomendación principal es realizar un análisis de los resultados presentados en este estudio, con ello se podrá revisar procesos y tomar decisiones que permitan a los estudiantes superar dificultades académicas.

## REFERENCIAS BIBLIOGRÁFICAS

- [1] C. E. para A. L. y el C. CEPAL, “Aspectos conceptuales de los censos de población y vivienda Gracias por su interés en esta publicación de la CEPAL,” 2019.
- [2] F. R. Aguilar-Gamboa, “Desafíos para el manejo y detección de pacientes con COVID-19 en Latinoamérica,” *Revista Experiencia en Medicina del Hospital Regional Lambayeque*, vol. 6, no. 1, 2020, doi: 10.37065/rem.v6i1.436.
- [3] S. A. Portillo Peñuelas, L. I. Castellanos Pierra, O. U. Reynoso Gonzáles, and O. I. Gavotto Nogales, “Enseñanza remota de emergencia ante la pandemia Covid-19 en Educación Media Superior y Educación Superior,” *Propósitos y Representaciones*, vol. 8, no. 3, pp. 1–17, 2020.
- [4] C. Russo, H. Ramón, N. Alonso, B. Cicerchia, L. Esnaola, and J. P. Tessore, “Tratamiento Masivo de Datos Utilizando Técnicas de Machine Learning,” p. 131, 2015.
- [5] R. Siddartha and R. Pulgarín, “tarjeta de crédito y ser cancelada a los 90 días . Model for forecasting cardholder ’ s attrition when acquiring a credit card and being canceled after 90 days .,” 2020.
- [6] J. W. Coulston, C. E. Blinn, V. A. Thomas, and R. H. Wynne, “Approximating prediction uncertainty for random forest regression models,” *Photogramm Eng Remote Sensing*, vol. 82, no. 3, pp. 189–197, Mar. 2016, doi: 10.14358/PERS.82.3.189.
- [7] “Random Forests Definition | DeepAI.” <https://deepai.org/machine-learning-glossary-and-terms/random-forest> (accessed Oct. 09, 2022).
- [8] L. Auria, R. A. M. Berlin, and R. A. Moro, “Support Vector Machines (SVM) as a Technique for Solvency Analysis,” 2008.
- [9] M. Martínez-Ramón and C. Christodoulou, “Support vector machines for antenna array processing and electromagnetics,” *Synthesis Lectures on Computational Electromagnetics*, vol. 5, pp. 1–120, Jan. 2006, doi: 10.2200/S00020ED1V01Y200604CEM005.
- [10] D. Maulud and A. M. Abdulazeez, “A Review on Linear Regression Comprehensive in Machine Learning,” *Journal of Applied Science and Technology Trends*, vol. 1, no. 4, pp. 140–147, Dec. 2020, doi: 10.38094/jastt1457.
- [11] M. Carmen Carollo Limeres Página, “Regresión lineal simple REGRESIÓN LINEAL SIMPLE,” 2011.



- [12] M. M. Saritas and A. Yasar, "International Journal of Intelligent Systems and Applications in Engineering Performance Analysis of ANN and Naive Bayes Classification Algorithm for Data Classification," *Original Research Paper International Journal of Intelligent Systems and Applications in Engineering IJISAE*, vol. 7, no. 2, pp. 88–91, 2019, doi: 10.1039/b000000x.
- [13] F. Elmaz, Ö. Yücel, and A. Y. Mutlu, "Predictive modeling of biomass gasification with machine learning-based regression methods," *Energy*, vol. 191, Jan. 2020, doi: 10.1016/j.energy.2019.116541.
- [14] D. la Red Martínez, M. Karanik, M. Giovannini, M. Báez, and J. Torre, "Descubrimiento de perfiles de rendimiento estudiantil: un modelo de integración de datos académicos y socioeconómicos," *Campus Virtuales*, vol. 5, no. 2, pp. 70–83, 2016.
- [15] J. E. Sotomonte, C. C. Rodríguez, P. A. Gaona, and J. G. Castellanos, "Hacia la construcción de un modelo predictivo de deserción académica basado en técnicas de minería de datos," *Revista científica*, vol. 3, no. 26, pp. 35–52, 2016, Accessed: Feb. 14, 2021. [Online]. Available: <https://revistas.udistrital.edu.co/index.php/revcie/article/view/11089/11931>
- [16] M. Segura-Morales and E. Loza-Aguirre, "Using Decision Trees for Predicting Academic Performance Based on Socio-Economic Factors," *Proceedings - 2017 International Conference on Computational Science and Computational Intelligence, CSCI 2017*, pp. 1132–1136, 2017, doi: 10.1109/CSCI.2017.197.
- [17] R. L. Baskerville and M. D. Myers, "Making IS research relevant to practice," *MIS Quarterly*, vol. 28, no. 3, pp. 329–335, 2004, [Online]. Available: <http://dl.acm.org/citation.cfm?id=2017229>
- [18] R. L. Baskerville and A. T. Wood-Harper, "A critical perspective on action research as a method for information systems research," *Journal of Information Technology*, vol. 11, no. 3, pp. 235–246, 1996, doi: 10.1080/026839696345289.
- [19] A. M. C. E and M. L. P. M, "LA INVESTIGACIÓN ACCIÓN. Una herramienta metodológica heurística para la comprensión y transformación de realidades y prácticas socio-educativas," *Laurus*, vol. 14, no. 27, pp. 96–114, 2008.
- [20] M. Martínez Miguélez, "La investigación-acción en el aula," 2000.
- [21] I. B. Machines. IBM, "Manual CRISP-DM de IBM SPSS Modeler," *IBM Corporation*, p. 56, 2012, [Online]. Available:

- <http://www.ibm.com/spss.%0Aftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/15.0/es/CRISP-DM.pdf>
- [22] “pandas.get\_dummies — pandas 1.5.1 documentation.”  
[https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.get\\_dummies.html](https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.get_dummies.html) (accessed Oct. 22, 2022).
- [23] “sklearn.preprocessing.OneHotEncoder — scikit-learn 1.1.2 documentation.”  
<https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.OneHotEncoder.html> (accessed Oct. 22, 2022).
- [24] “sklearn.feature\_selection.SelectKBest — scikit-learn 1.1.2 documentation.”  
[https://scikit-learn.org/stable/modules/generated/sklearn.feature\\_selection.SelectKBest.html](https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.SelectKBest.html) (accessed Oct. 22, 2022).

## ANEXOS

## Anexo I – Glosario de términos

- **Dimensión:** En una base de datos relacional o plana, cada campo en un registro representa una dimensión. En una base de datos multidimensional, una dimensión es un conjunto de entidades similares; por ejemplo, una base de datos multidimensional de ventas podría incluir las dimensiones Producto, Tiempo y Ciudad.
- **Modelo analítico:** Una estructura y proceso para analizar un conjunto de datos. Por ejemplo, un árbol de decisión es un modelo para la clasificación de un conjunto de datos.
- **Modelo lineal:** Un modelo analítico que asume relaciones lineales entre una variable seleccionada (dependiente) y sus predictores (variables independientes).
- **Modelo no lineal:** Un modelo analítico que no asume una relación lineal en los coeficientes de las variables que son estudiadas.
- **Modelo predictivo:** Estructura y proceso para predecir valores de variables especificadas en un conjunto de datos.
- **Navegación de datos:** Proceso de visualizar diferentes dimensiones, "fetas" y niveles de una base de datos multidimensional.
- **OLAP Procesamiento analítico on-line (On Line Analytic Processing):** Se refiere a aplicaciones de bases de datos orientadas a arrays que permite a los usuarios ver, navegar, manipular y analizar bases de datos multidimensionales.
- **Outlier:** Un ítem de datos cuyo valor cae fuera de los límites que encierran a la mayoría del resto de los valores correspondientes de la muestra. Puede indicar datos anormales. Deberían ser examinados detenidamente, pueden dar importante información.
- **Procesamiento paralelo:** Uso coordinado de múltiples procesadores para realizar tareas computacionales. El procesamiento paralelo puede ocurrir en una computadora con múltiples procesadores o en una red de estaciones de trabajo o PCs.
- **Regresión lineal:** Técnica estadística utilizada para encontrar la mejor relación lineal que encaja entre una variable seleccionada (dependiente) y sus predicados (variables independientes).
- **Regresión logística:** Una regresión lineal que predice las proporciones de una variable seleccionada categórica, tal como Tipo de Consumidor, en una población.
- **SMP Multiprocesador simétrico (Symmetric Multiprocessor):** Tipo de computadora con multiprocesadores en la cual la memoria es compartida entre los procesadores.

**Anexo II – Descripción de datos**

**Datos Socioeconómicos**

<b>Atributo</b>	<b>Descripcion</b>	<b>Tipo de dato</b>	<b>Nulos</b>	<b>Valores</b>
CodigoEnc	Codigo único de estudiante	object	0	Codigo único de estudiante
FecNacimiento	Fecha nacimiento de estudiante	datetime 64[ns]	0	
EstadoCivil	EstadoCivil estudiante	object	0	S', 'C', 'D', 'V', 'U', ''
Genero	Genero Estudiante	object	0	HOMBRE', 'MUJER'
Etnia	Indica la etnia estudiante	object	7222	MESTIZO/A', nan, 'INDÍGENA', 'OTRO', 'AFROECUATORIANO/AFRODESCENDIENTE', 'BLANCO/A', 'MULATO/A', 'MONTUBIO/A', 'NEGRO/A', 'NO REGISTRA'
Pais	Indica el país de procedencia del estudiante	object	0	BELGICA', 'VENEZUELA', 'ESPAÑA', 'PERU', 'COLOMBIA', 'ECUADOR', 'REINO UNIDO', 'ESTADOS UNIDOS', 'EL SALVADOR', 'ARGENTINA', 'SIN REGISTRO', 'BRASIL', 'CUBA', 'GUATEMALA', 'MEXICO', 'CHILE', 'BOLIVIA', 'KAZAJISTAN', 'COSTA RICA', 'FRANCIA', 'HAITI', 'ITALIA', 'JAPON', 'CHINA', 'HONDURAS', 'CANADA', 'UCRANIA', 'SUIZA', 'ISRAEL', 'RUSIA', 'URUGUAY', 'ALEMANIA'
Provincia	Indica la provincia de procedencia del estudiante	object	0	SIN REGISTRO', 'CAÑAR', 'EL ORO', 'AZUAY', 'PICHINCHA', 'TUNGURAHUA', 'MORONA SANTIAGO', 'CARCHI', 'IMBABURA', 'CHIMBORAZO', 'GUAYAS', 'SUCUMBIOS', 'COTOPAXI', 'PASTAZA', 'SANTO DOMINGO DE LOS TSACHILAS', 'ORELLANA', 'LOJA', 'BOLIVAR', 'GALAPAGOS', 'ZAMORA CHINCHIPE', 'ESMERALDAS', 'MANABI', 'LOS RIOS', 'SANTA ELENA', 'NAPO', 'SIN REGISTRO', 'AZOGUES', 'HUAQUILLAS', 'CUENCA', 'SUSCAL', 'CAÑAR', 'QUITO', 'EL TAMBO', 'GUALACEO', 'AMBATO', 'GUALAQUIZA', 'SANTIAGO', 'BOLIVAR', 'TULCAN', 'MONTUFAR', 'ESPEJO', 'OTAVALO', 'IBARRA', 'BAÑOS DE AGUA SANTA', 'ANTONIO ANTE', 'RIOBAMBA', 'SAN PEDRO DE HUACA', 'GUAYAQUIL', 'SUCUMBIOS', 'RUMIÑAHUI', 'COTACACHI', 'MIRA', 'SANTIAGO DE PILLARO', 'LATACUNGA', 'MERA', 'SANTO DOMINGO', 'PASTAZA', 'PUJILI', 'CAMILO PONCE ENRIQUEZ', 'GUANO', 'ALASI', 'COLTA', 'GUAMOTE', 'PASAJE', 'ORELLANA', 'LOJA', 'PENIPE', 'CELICA', 'SAN MIGUEL', 'CHUNCHI',

'PALLATANGA','SANTA ROSA', 'MACHALA',  
'PIÑAS', 'ZARUMA', 'EL EMPALME','SANTA  
CRUZ', 'CHINCHIPE', 'ATAHUALPA',  
'PORTOVELO', 'ARENILLAS','MARCABELI',  
'PUYANGO', 'LAS LAJAS', 'BALSAS', 'ELOY  
ALFARO','ESMERALDAS', 'QUININDE', 'EL  
CARMEN', 'SAN LORENZO', 'ATACAMES','LA  
CONCORDIA', 'SALCEDO', 'RIOVERDE',  
'MUISNE', 'CHONE','GUARANDA', 'SIGCHOS',  
'PANGUA', 'MEJIA', 'LA MANA',  
'VALENCIA','SAQUISILI', 'VENTANAS',  
'MILAGRO', 'QUERO', 'LAGO  
AGRIO','SALINAS', 'SUCUA', 'GONZANAMA',  
'SIGSIG', 'SEVILLA DE ORO','PAUTE',  
'CHAGUARPAMBA', 'BABAHOYO', 'PUCARA',  
'CATAMAYO','CHILLANES', 'ECHEANDIA',  
'CHIMBO', 'URDANETA', 'QUEVEDO',  
'DAULE','SUCRE', 'NARANJITO', 'DURAN',  
'SANTA ELENA', 'MANTA','LA LIBERTAD',  
'BUENA FE', 'CORONEL MARCELINO  
MARIDUEÑA','PEDRO CARBO', 'EL TRIUNFO',  
'MONTECRISTI', 'PORTOVIEJO','SANTA  
ISABEL', 'TOSAGUA', 'SAN MIGUEL DE LOS  
BANCOS','PUERTO QUITO', 'CAYAMBE',  
'FLAVIO ALFARO', 'JIPIJAPA','SANTA ANA',  
'ROCAFUERTE', 'PEDERNALES', 'PAJAN',  
'JARAMIJO','SAN VICENTE', '24 DE MAYO',  
'PICHINCHA', 'MORONA', 'EL  
PAN','HUAMBOYA', 'SAN PEDRO DE  
PELILEO', 'PUTUMAYO',  
'CEVALLOS','CUMANDA', 'PALORA',  
'ARAJUNO', 'QUIJOS', 'PEDRO  
MONCAYO','SHUSHUFINDI', 'PALANDA',  
'TISALEO', 'PIMAMPIRO', 'CHILLA','SANTO  
DOMINGO DE LOS COLORADOS',  
'MACARA','PEDRO VICENTE MALDONADO',  
'ESPINDOLA', 'PALTAS', 'NABON',  
'JAMA','SAN MIGUEL DE URCUQUI', 'EL  
CHACO', 'BALZAR', 'SARAGURO','LA JOYA DE  
LOS SACHAS', 'CASCALES', 'GONZALO  
PIZARRO','ALFREDO BAQUERIZO MORENO  
(JUJAN)', 'CALUMA', 'PINDAL',  
'QUILANGA','YANTZAZA (YANZATZA)',  
'ARCHIDONA', 'CALVAS','SAN JACINTO DE  
SAN JACINTO DE YAGUACHI',  
'MONTALVO','SAN CRISTOBAL', 'OLMEDO',  
'TENA', 'CHAMBO', 'SAN  
FERNANDO','ISABELA', 'NARANJAL',  
'ZAMORA', 'PUEBLOVIEJO',  
'QUINSALOMA','JUNIN', 'PATATE', 'SAN  
PEDRO DE SAN PEDRO DE PELILEO',  
'OÑA','SANTIAGO DE SANTIAGO DE  
PILLARO', 'VINCES', 'MOCHA',

			'SOZORANGA','NANGARITZA', 'EL PANGUI', 'MOCACHE', 'ZAPOTILLO', 'BALAO','CENTINELA DEL CONDOR'
Canton	Canton estudiante	de object	0 SIN REGISTRO', 'CAÑAR', 'EL ORO', 'AZUAY', 'PICHINCHA','TUNGURAHUA', 'MORONA SANTIAGO', 'CARCHI', 'IMBABURA','CHIMBORAZO', 'GUAYAS', 'SUCUMBIOS', 'COTOPAXI', 'PASTAZA','SANTO DOMINGO DE LOS TSACHILAS', 'ORELLANA', 'LOJA', 'BOLIVAR','GALAPAGOS', 'ZAMORA CHINCHIPE', 'ESMERALDAS', 'MANABI','LOS RIOS', 'SANTA ELENA', 'NAPO','SIN REGISTRO', 'AZOGUES', 'HUAQUILLAS', 'CUENCA', 'SUSCAL','CAÑAR', 'QUITO', 'EL TAMBO', 'GUALACEO', 'AMBATO', 'GUALAQUIZA','SANTIAGO', 'BOLIVAR', 'TULCAN', 'MONTUFAR', 'ESPEJO', 'OTAVALO','IBARRA', 'BAÑOS DE AGUA SANTA', 'ANTONIO ANTE', 'RIOBAMBA','SAN PEDRO DE HUACA', 'GUAYAQUIL', 'SUCUMBIOS', 'RUMIÑAHUI','COTACACHI', 'MIRA', 'SANTIAGO DE PILLARO', 'LATACUNGA', 'MERA','SANTO DOMINGO', 'PASTAZA', 'PUJILI', 'CAMILO PONCE ENRIQUEZ','GUANO', 'ALAUSSI', 'COLTA', 'GUAMOTE', 'PASAJE', 'ORELLANA','LOJA', 'PENIPE', 'CELICA', 'SAN MIGUEL', 'CHUNCHI', 'PALLATANGA','SANTA ROSA', 'MACHALA', 'PIÑAS', 'ZARUMA', 'EL EMPALME','SANTA CRUZ', 'CHINCHIPE', 'ATAHUALPA', 'PORTOVELO', 'ARENILLAS','MARCABELI', 'PUYANGO', 'LAS LAJAS', 'BALSAS', 'ELOY ALFARO','ESMERALDAS', 'QUININDE', 'EL CARMEN', 'SAN LORENZO', 'ATACAMES','LA CONCORDIA', 'SALCEDO', 'RIOVERDE', 'MUISNE', 'CHONE','GUARANDA', 'SIGCHOS', 'PANGUA', 'MEJIA', 'LA MANA', 'VALENCIA','SAQUISILI', 'VENTANAS', 'MILAGRO', 'QUERO', 'LAGO AGRIO','SALINAS', 'SUCUA', 'GONZANAMA', 'SIGSIG', 'SEVILLA DE ORO','PAUTE', 'CHAGUARPAMBA', 'BABAHOYO', 'PUCARA', 'CATAMAYO','CHILLANES', 'ECHEANDIA', 'CHIMBO', 'URDANETA', 'QUEVEDO', 'DAULE','SUCRE', 'NARANJITO', 'DURAN', 'SANTA ELENA', 'MANTA','LA LIBERTAD', 'BUENA FE', 'CORONEL MARCELINO MARIDUEÑA','PEDRO CARBO', 'EL TRIUNFO', 'MONTECRISTI', 'PORTOVIEJO','SANTA

				ISABEL', 'TOSAGUA', 'SAN MIGUEL DE LOS BANCOS', 'PUERTO QUITO', 'CAYAMBE', 'FLAVIO ALFARO', 'JIPIJAPA', 'SANTA ANA', 'ROCAFUERTE', 'PEDERNALES', 'PAJAN', 'JARAMIJO', 'SAN VICENTE', '24 DE MAYO', 'PICHINCHA', 'MORONA', 'EL PAN', 'HUAMBOYA', 'SAN PEDRO DE PELILEO', 'PUTUMAYO', 'CEVALLOS', 'CUMANDA', 'PALORA', 'ARAJUNO', 'QUIJOS', 'PEDRO MONCAYO', 'SHUSHUFINDI', 'PALANDA', 'TISALEO', 'PIMAMPIRO', 'CHILLA', 'SANTO DOMINGO DE LOS COLORADOS', 'MACARA', 'PEDRO VICENTE MALDONADO', 'ESPINDOLA', 'PALTAS', 'NABON', 'JAMA', 'SAN MIGUEL DE URCUQUI', 'EL CHACO', 'BALZAR', 'SARAGURO', 'LA JOYA DE LOS SACHAS', 'CASCALES', 'GONZALO PIZARRO', 'ALFREDO BAQUERIZO MORENO (JUJAN)', 'CALUMA', 'PINDAL', 'QUILANGA', 'YANTZAZA (YANZATZA)', 'ARCHIDONA', 'CALVAS', 'SAN JACINTO DE SAN JACINTO DE YAGUACHI', 'MONTALVO', 'SAN CRISTOBAL', 'OLMEDO', 'TENA', 'CHAMBO', 'SAN FERNANDO', 'ISABELA', 'NARANJAL', 'ZAMORA', 'PUEBLOVIEJO', 'QUINSALOMA', 'JUNIN', 'PATATE', 'SAN PEDRO DE SAN PEDRO DE PELILEO', 'OÑA', 'SANTIAGO DE SANTIAGO DE PILLARO', 'VINCES', 'MOCHA', 'SOZORANGA', 'NANGARITZA', 'EL PANGUI', 'MOCACHE', 'ZAPOTILLO', 'BALAO', 'CENTINELA DEL CONDOR'
Discapacidad	indica el tipo de discapacidad que tiene el estudiante.	object	26298	nan, 'AUDITIVA', 'VISUAL', 'FISICA', 'MENTAL PSICOSOCIAL', 'INTELECTUAL', 'LENGUAJE'
Discapacidad	Indica el porcentaje de discapacidad que posee el estudiante	float64	26298	nan, 32., 58., 39., 45., 72., 30., 80., 35., 41., 42., 51., 36., 46., 37., 70., 88., 64., 40., 43., 60., 61., 34., 62., 47., 77.
Parroquia	Señala la parroquia donde vive el estudiante	object	8018	SIN REGISTRO', 'CAÑAR', 'EL ORO', 'AZUAY', 'PICHINCHA', 'TUNGURAHUA', 'MORONA SANTIAGO', 'CARCHI', 'IMBABURA', 'CHIMBORAZO', 'GUAYAS', 'SUCUMBOS', 'COTOPAXI', 'PASTAZA', 'SANTO DOMINGO DE LOS TSACHILAS', 'ORELLANA', 'LOJA', 'BOLIVAR', 'GALAPAGOS', 'ZAMORA CHINCHIPE', 'ESMERALDAS', 'MANABI', 'LOS RIOS', 'SANTA ELENA', 'NAPO', 'SIN REGISTRO', 'AZOGUES', 'HUAQUILLAS', 'CUENCA', 'SUSCAL', 'CAÑAR', 'QUITO', 'EL



TAMBO', 'GUALACEO', 'AMBATO',  
 'GUALAQUIZA', 'SANTIAGO', 'BOLIVAR',  
 'TULCAN', 'MONTUFAR', 'ESPEJO',  
 'OTAVALO', 'IBARRA', 'BAÑOS DE AGUA  
 SANTA', 'ANTONIO ANTE', 'RIOBAMBA', 'SAN  
 PEDRO DE HUACA', 'GUAYAQUIL',  
 'SUCUMBIOS', 'RUMIÑAHUI', 'COTACACHI',  
 'MIRA', 'SANTIAGO DE PILLARO',  
 'LATACUNGA', 'MERA', 'SANTO DOMINGO',  
 'PASTAZA', 'PUJILI', 'CAMILO PONCE  
 ENRIQUEZ', 'GUANO', 'ALAUSI', 'COLTA',  
 'GUAMOTE', 'PASAJE', 'ORELLANA', 'LOJA',  
 'PENIPE', 'CELICA', 'SAN MIGUEL', 'CHUNCHI',  
 'PALLATANGA', 'SANTA ROSA', 'MACHALA',  
 'PIÑAS', 'ZARUMA', 'EL EMPALME', 'SANTA  
 CRUZ', 'CHINCHIPE', 'ATAHUALPA',  
 'PORTOVELO', 'ARENILLAS', 'MARCABELI',  
 'PUYANGO', 'LAS LAJAS', 'BALSAS', 'ELOY  
 ALFARO', 'ESMERALDAS', 'QUININDE', 'EL  
 CARMEN', 'SAN LORENZO', 'ATACAMES', 'LA  
 CONCORDIA', 'SALCEDO', 'RIOVERDE',  
 'MUISNE', 'CHONE', 'GUARANDA', 'SIGCHOS',  
 'PANGUA', 'MEJIA', 'LA MANA',  
 'VALENCIA', 'SAQUISILI', 'VENTANAS',  
 'MILAGRO', 'QUERO', 'LAGO  
 AGRIO', 'SALINAS', 'SUCUA', 'GONZANAMA',  
 'SIGSIG', 'SEVILLA DE ORO', 'PAUTE',  
 'CHAGUARPAMBA', 'BABAHOYO', 'PUCARA',  
 'CATAMAYO', 'CHILLANES', 'ECHEANDIA',  
 'CHIMBO', 'URDANETA', 'QUEVEDO',  
 'DAULE', 'SUCRE', 'NARANJITO', 'DURAN',  
 'SANTA ELENA', 'MANTA', 'LA LIBERTAD',  
 'BUENA FE', 'CORONEL MARCELINO  
 MARIDUEÑA', 'PEDRO CARBO', 'EL TRIUNFO',  
 'MONTECRISTI', 'PORTOVIEJO', 'SANTA  
 ISABEL', 'TOSAGUA', 'SAN MIGUEL DE LOS  
 BANCOS', 'PUERTO QUITO', 'CAYAMBE',  
 'FLAVIO ALFARO', 'JIPIJAPA', 'SANTA ANA',  
 'ROCAFUERTE', 'PEDERNALES', 'PAJAN',  
 'JARAMIJO', 'SAN VICENTE', '24 DE MAYO',  
 'PICHINCHA', 'MORONA', 'EL  
 PAN', 'HUAMBOYA', 'SAN PEDRO DE  
 PELILEO', 'PUTUMAYO',  
 'CEVALLOS', 'CUMANDA', 'PALORA',  
 'ARAJUNO', 'QUIJOS', 'PEDRO  
 MONCAYO', 'SHUSHUFINDI', 'PALANDA',  
 'TISALEO', 'PIMAMPIRO', 'CHILLA', 'SANTO  
 DOMINGO DE LOS COLORADOS',  
 'MACARA', 'PEDRO VICENTE MALDONADO',  
 'ESPINDOLA', 'PALTAS', 'NABON',  
 'JAMA', 'SAN MIGUEL DE URCUQUI', 'EL  
 CHACO', 'BALZAR', 'SARAGURO', 'LA JOYA DE  
 LOS SACHAS', 'CASCALES', 'GONZALO

PIZARRO', 'ALFREDO BAQUERIZO MORENO (JUJAN)', 'CALUMA', 'PINDAL', 'QUILANGA', 'YANTZAZA (YANZATZA)', 'ARCHIDONA', 'CALVAS', 'SAN JACINTO DE SAN JACINTO DE YAGUACHI', 'MONTALVO', 'SAN CRISTOBAL', 'OLMEDO', 'TENA', 'CHAMBO', 'SAN FERNANDO', 'ISABELA', 'NARANJAL', 'ZAMORA', 'PUEBLO', 'INAN', 'ITCHIMBIA', 'CENTRO HISTÓRICO', 'CARCELÉN', 'COTOCOLLAO', 'IÑAQUITO', 'SANTA PRISCA', 'ATAHUALPA', 'MARISCAL SUCRE', 'EL CONDADO', 'POMASQUI', 'PUENGASÍ', 'CUTUGLAHUA', 'CHIMBACALLE', 'QUITUMBE', 'LA MAGDALENA', 'LA ARGELIA', 'HUAYNACAPAC', 'SAN BLAS', 'CHAUPICRUZ', 'CALDERÓN', 'BENALCAZAR', 'SAN JUAN', 'RUMIPAMBA', 'LA VICENTINA', 'CONOCOTO', 'MACHACHI', 'KENNEDY', 'CHILLOGALLO', 'BELISARIO QUEVEDO', 'COCHAPAMBA', 'TUMBACO', 'PONCEANO', 'PUEMBO', 'LA CONCEPCIÓN', 'PIFO', 'SAN ISIDRO DEL INCA', 'LA FERROVIARIA', 'SAN ROQUE', 'COMITÉ DEL PUEBLO', 'SAN BARTOLO', 'CALDERON', 'CENTRO HISTORICO', 'GUAMANÍ', 'SOLANDA', 'CAYAMBE', 'ZÁMBIZA', 'CUMBAYÁ', 'YARUQUÍ', 'LA MERCED', 'JIPIJAPA', 'LA ECUATORIANA', 'PÍNTAG', 'GONZALEZ SUAREZ', 'SAN JOSÉ DE MINAS', 'CARCELEN', 'LA MENA', 'LA LIBERTAD', 'SANGOLQUI', 'SAN RAFAEL', 'SANTA BARBARA', 'SAN ANTONIO', 'LA FLORESTA', 'ALANGASÍ', 'LA ESPERANZA', 'SANGOLQUÍ', 'GUAYLLABAMBA', 'AMAGUAÑA', 'GUAMANI', 'CHILIBULO', 'NANEGALITO', 'TURUBAMBA', 'PEDRO VICENTE MALDONADO', 'SAN PEDRO DE TABOADA', 'ALANGASI', 'PUENGASI', 'CONCEPCION', 'LLANO CHICO', 'CHECA', 'COTOGCHOA', 'VILLA FLORA', 'ALOASI', 'TAMBILLO', 'COMITE DEL PUEBLO', 'GUANGOPOLO', 'ALOAG', 'RICAURTE', 'EL VECINO', 'CAÑARIBAMBA', 'NAYON', 'NAYÓN', 'TABABELA', 'QUINCHE', 'NANEGAL', 'PUERTO QUITO', 'EL QUINCHE', 'SAN JOSE DE AYORA', 'CANGAHUA', 'YARUQUI', 'ASCÁZUBI', 'CUMBAYA', 'EL SALVADOR', 'GUAPULO', 'ALÓAG', 'SAN SEBASTIAN', 'OTÓN', 'CALACALÍ', 'MALCHINGUÍ', 'CUTUGLAGUA', 'CHAVEZPAMBA', 'OLMEDO', 'MINDO', 'SAN MARCOS',

				'TABACUNDO', 'EL BATAN', 'ZAMBIZA', 'PINTAG', 'PUELLARO', 'SAN MIGUEL DE LOS BANCOS', 'UYUMBICHO', 'PERUCHO', 'LLOA', 'CHAUPI', 'SAN JOSE DE MINAS', 'ASCAZUBI', 'TOCACHI', 'GUALEA', 'SAN JOSÉ DE AYORA', 'TUPIGACHI', 'SANTA ROSA DE CUZUBAMBA', 'PUÉLLARO', 'PACTO', 'JUAN MONTALVO', 'CALACALI', 'EL CHAUPI', 'STA ROSA DE CUSUBAMBA', 'NONO', 'MANUEL CORNEJO ASTORGA'
Teléfono	Teléfono de su residencia	object	0	Teléfono de su residencia
Celular	teléfono celular personal	object	0	teléfono celular personal
Correo	correo personal	object	0	correo personal
Correolns	correo institucional	object	0	correo institucional
UnidadEducativa	Indica el nombre de la Unidad Educativa de origen	object	0	Unidad Educativa de origen
TituloCol	señala el título de colegio del estudiante	object	1	SIN REGISTRO', 'CAÑAR', 'EL ORO', 'AZUAY', 'PICHINCHA', 'TUNGURAHUA', 'MORONA SANTIAGO', 'CARCHI', 'IMBABURA', 'CHIMBORAZO', 'GUAYAS', 'SUCUMBIOS', 'COTOPAXI', 'PASTAZA', 'SANTO DOMINGO DE LOS TSACHILAS', 'ORELLANA', 'LOJA', 'BOLIVAR', 'GALAPAGOS', 'ZAMORA CHINCHIPE', 'ESMERALDAS', 'MANABI', 'LOS RIOS', 'SANTA ELENA', 'NAPO', 'SIN REGISTRO', 'AZOGUES', 'HUAQUILLAS', 'CUENCA', 'SUSCAL', 'CAÑAR', 'QUITO', 'EL TAMBO', 'GUALACEO', 'AMBATO', 'GUALAQUIZA', 'SANTIAGO', 'BOLIVAR', 'TULCAN', 'MONTUFAR', 'ESPEJO', 'OTAVALO', 'IBARRA', 'BAÑOS DE AGUA SANTA', 'ANTONIO ANTE', 'RIOBAMBA', 'SAN PEDRO DE HUACA', 'GUAYAQUIL', 'SUCUMBIOS', 'RUMIÑAHUI', 'COTACACHI', 'MIRA', 'SANTIAGO DE PILLARO', 'LATACUNGA', 'MERA', 'SANTO DOMINGO', 'PASTAZA', 'PUJILI', 'CAMILO PONCE ENRIQUEZ', 'GUANO', 'ALASI', 'COLTA', 'GUAMOTE', 'PASAJE', 'ORELLANA', 'LOJA', 'PENIPE', 'CELICA', 'SAN MIGUEL', 'CHUNCHI', 'PALLATANGA', 'SANTA ROSA', 'MACHALA', 'PIÑAS', 'ZARUMA', 'EL EMPALME', 'SANTA CRUZ', 'CHINCHIPE', 'ATAHUALPA', 'PORTOVELO', 'ARENILLAS', 'MARCABELI', 'PUYANGO', 'LAS LAJAS', 'BALSAS', 'ELOY ALFARO', 'ESMERALDAS', 'QUININDE', 'EL CARMEN', 'SAN LORENZO', 'ATACAMES', 'LA CONCORDIA', 'SALCEDO', 'RIOVERDE',

'MUISNE', 'CHONE', 'GUARANDA', 'SIGCHOS',  
 'PANGUA', 'MEJIA', 'LA MANA',  
 'VALENCIA', 'SAQUISILI', 'VENTANAS',  
 'MILAGRO', 'QUERO', 'LAGO  
 AGRIO', 'SALINAS', 'SUCUA', 'GONZANAMA',  
 'SIGSIG', 'SEVILLA DE ORO', 'PAUTE',  
 'CHAGUARPAMBA', 'BABAHOYO', 'PUCARA',  
 'CATAMAYO', 'CHILLANES', 'ECHEANDIA',  
 'CHIMBO', 'URDANETA', 'QUEVEDO',  
 'DAULE', 'SUCRE', 'NARANJITO', 'DURAN',  
 'SANTA ELENA', 'MANTA', 'LA LIBERTAD',  
 'BUENA FE', 'CORONEL MARCELINO  
 MARIDUEÑA', 'PEDRO CARBO', 'EL TRIUNFO',  
 'MONTECRISTI', 'PORTOVIEJO', 'SANTA  
 ISABEL', 'TOSAGUA', 'SAN MIGUEL DE LOS  
 BANCOS', 'PUERTO QUITO', 'CAYAMBE',  
 'FLAVIO ALFARO', 'JIPIJAPA', 'SANTA ANA',  
 'ROCAFUERTE', 'PEDERNALES', 'PAJAN',  
 'JARAMIJO', 'SAN VICENTE', '24 DE MAYO',  
 'PICHINCHA', 'MORONA', 'EL  
 PAN', 'HUAMBOYA', 'SAN PEDRO DE  
 PELILEO', 'PUTUMAYO',  
 'CEVALLOS', 'CUMANDA', 'PALORA',  
 'ARAJUNO', 'QUIJOS', 'PEDRO  
 MONCAYO', 'SHUSHUFINDI', 'PALANDA',  
 'TISALEO', 'PIMAMPIRO', 'CHILLA', 'SANTO  
 DOMINGO DE LOS COLORADOS',  
 'MACARA', 'PEDRO VICENTE MALDONADO',  
 'ESPINDOLA', 'PALTAS', 'NABON',  
 'JAMA', 'SAN MIGUEL DE URQUQUI', 'EL  
 CHACO', 'BALZAR', 'SARAGURO', 'LA JOYA DE  
 LOS SACHAS', 'CASCALES', 'GONZALO  
 PIZARRO', 'ALFREDO BAQUERIZO MORENO  
 (JUJAN)', 'CALUMA', 'PINDAL',  
 'QUILANGA', 'YANTZAZA (YANZATZA)',  
 'ARCHIDONA', 'CALVAS', 'SAN JACINTO DE  
 SAN JACINTO DE YAGUACHI',  
 'MONTALVO', 'SAN CRISTOBAL', 'OLMEDO',  
 'TENA', 'CHAMBO', 'SAN  
 FERNANDO', 'ISABELA', 'NARANJAL',  
 'ZAMORA', 'PUEBLO VInan, 'ITCHIMBIA',  
 'CENTRO HISTÓRICO', 'CARCELÉN',  
 'COTOCOLLAO', 'IÑAQUITO', 'SANTA PRISCA',  
 'ATAHUALPA', 'MARISCAL SUCRE', 'EL  
 CONDADO', 'POMASQUI', 'PUENGASÍ',  
 'CUTUGLAHUA',  
 'CHIMBACALLE', 'QUITUMBE', 'LA  
 MAGDALENA', 'LA ARGELIA',  
 'HUAYNACAPAC', 'SAN BLAS', 'CHAUPICRUZ',  
 'CALDERÓN', 'BENALCAZAR', 'SAN  
 JUAN', 'RUMIPAMBA', 'LA VICENTINA',  
 'CONOCOTO', 'MACHACHI',  
 'KENNEDY', 'CHILLOGALLO', 'BELISARIO

QUEVEDO', 'COCHAPAMBA',  
 'TUMBACO','PONCEANO', 'PUEMBO', 'LA  
 CONCEPCIÓN', 'PIFO','SAN ISIDRO DEL  
 INCA', 'LA FERROVIARIA', 'SAN  
 ROQUE','COMITÉ DEL PUEBLO', 'SAN  
 BAROLO', 'CALDERON', 'CENTRO  
 HISTORICO','GUAMANÍ', 'SOLANDA',  
 'CAYAMBE', 'ZÁMBIZA', 'CUMBAYÁ',  
 'YARUQUÍ','LA MERCED', 'JIPIJAPA', 'LA  
 ECUATORIANA', 'PÍNTAG','GONZALEZ  
 SUAREZ', 'SAN JOSÉ DE MINAS',  
 'CARCELEN', 'LA MENA','LA LIBERTAD',  
 'SANGOLQUI', 'SAN RAFAEL', 'SANTA  
 BARBARA','SAN ANTONIO', 'LA FLORESTA',  
 'ALANGASÍ', 'LA ESPERANZA','SANGOLQUÍ',  
 'GUAYLLABAMBA', 'AMAGUAÑA', 'GUAMANI',  
 'CHILIBULO','NANEGALITO', 'TURUBAMBA',  
 'PEDRO VICENTE MALDONADO','SAN  
 PEDRO DE TABOADA', 'ALANGASI',  
 'PUENGASI', 'CONCEPCION','LLANO CHICO',  
 'CHECA', 'COTOGCHOA', 'VILLA FLORA',  
 'ALOASI','TAMBILLO', 'COMITE DEL PUEBLO',  
 'GUANGOPOLO', 'ALOAG', 'RICAURTE','EL  
 VECINO', 'CAÑARIBAMBA', 'NAYON',  
 'NAYÓN', 'TABABELA','QUINCHE', 'NANEGAL',  
 'PUERTO QUITO', 'EL QUINCHE','SAN JOSE  
 DE AYORA', 'CANGAHUA', 'YARUQUI',  
 'ASCÁZUBI', 'CUMBAYA','EL SALVADOR',  
 'GUAPULO', 'ALÓAG', 'SAN SEBASTIAN',  
 'OTÓN','CALACALÍ', 'MALCHINGUÍ',  
 'CUTUGLAGUA', 'CHAVEZPAMBA',  
 'OLMEDO','MINDO', 'SAN MARCOS',  
 'TABACUNDO', 'EL BATAN',  
 'ZAMBIZA','PINTAG', 'PUELLARO', 'SAN  
 MIGUEL DE LOS BANCOS',  
 'UYUMBICHO','PERUCHO', 'LLOA', 'CHAUPI',  
 'SAN JOSE DE MINAS', 'ASCAZUBACHILLER  
 EN CIENCIAS', 'BACHILLER EN CIENCIAS  
 (GENERAL)','CIENCIAS', 'BACHILLER  
 TECNICO INDUSTRIAL INSTALACIONES,  
 EQUIPOS','BACHILLER TECNICO', 'QUIMICO  
 BIOLOGO','ADMINISTRACION Y  
 CONTABILIDAD', 'N/E','BACHILLER EN  
 CIENCIAS ESPECIALIZACION FISICO  
 MATEMATICAS','FISICO  
 MATEMATICO','BACHILLER TECNICO  
 INDUSTRIAL ESPECIALIDAD  
 ELECTRONICA','BACHILLER TECNICO  
 INDUSTRIAL ESPECIALIZACION  
 ELECTRONICA','BACHILLER EN COMERCIO  
 Y ADMINISTRACION ESPECIALIZACION  
 INFOR','BACHILLER EN HUMANIDADES  
 MODERNAS ESPECIALIZACION FISICO-

			<p>MAT', 'BACHILLER TECNICO INDUSTRIAL ESPECIALIZACION ELECTRONICA DE', 'BACHILLER TECNICO EN AGROPECUARIA ESPECIALIZACION AGROPECUAR', 'BACHILLER EN CIENCIAS ESPECIALIZACION INFORMATICA', 'BACHILLER EN CIENCIAS ESPECIALIZACION QUIMICO BIOLOGICAS', 'HUMANIDADES MODERNAS', 'BACHILLER EN CIENCIAS ESPECIALIDAD FISICO MATEMATICAS', 'BACHILLER EN CIENCIAS ESPECIALIDAD QUIMICO-BIOLOGICAS', 'INGENIERO EN SISTEMAS INFORMATICOS Y COMPUTACION', 'BACHILLER EN CIENCIAS ESPECIALIZACION FISICO MATEMATICO', 'INGENIERO MECÁNICO', 'INGENIERO EN ELECTRONICA Y CONTROL', 'BACHILLER EN CIENCIAS FISICO MATEMATICAS', 'BACHILLER TECNICO ESPECIALIZACION MECANICA AUTOMOTRIZ', 'BACHILLERATO GENERAL UNIFICADO', 'BACHILLER EN CIENCIAS EN LA ESPECIALIZACION DE FISICO MATEMA', 'BACHILLER EN CIENCIAS GENERALES', 'BACHILLER TÉCNICO INDUSTRIAL ESPECIALIZACIÓN ELECTROMECÁNICA', 'FISICO MATEMATICAS', 'BACHILLER TECNICO INDUSTRIAL ESPECIALIZACION MECANICA AUTOMO', 'BACHILLER TECNICO INDUSTRIAL ESPECIALIDAD MECANICA INDUSTRIAL', 'BACHILLER TECNICO ESPECIALIZACION INSTALACIONES EQUIPOS Y MA', 'INFORMATICA', 'BACHILLER TECNICO INDUSTRIAL ESPECIALIZACION ELECTRICIDAD', 'BACHILLER EN CIENCIAS DE COMERCIO Y ADMINISTRACION', 'INGENIERA QUÍMICA', 'BACHILLER TECNICO INDUSTRIAL ESPECIALIDAD INSTALACIONES, EQU', 'CONTADOR BACHILLER EN COMERCIO Y ADMINISTRACION ESPECIALIDAD', 'BACHILLER DE SERVICIOS APLICACIONES INFORMATICAS', 'INGENIERA QUIMICA', 'BACHILLER TECNICO ESPECIALIZACION ELECTRONICA', 'BACHILLER TECNICO INDUSTRIAL ESPECIALIDAD ELECTRONICA DE CON', 'BACHILLER TECNICO EN</p>
--	--	--	---

			<p> COMERCIO Y ADMINISTRACION  ESPECIALIDAD', BACHILLER TECNICO  INDUSTRIAL ESPECIALIDAD  ELECTRICIDAD-ELECTRONICA', INGENIERO  INDUSTRIAL', BACHILLER TECNICO  ESPECIALIZACION  ELECTRICIDAD', BACHILLER TECNICO EN  INFORMATICA ESP APLICACIONES  INFORMATIC', BACHILLER TECNICO  ESPECIALIDAD MECANICA  INDUSTRIAL', INGENIERIA EN  ELECTRONICA Y COMTROL', CIENCIAS  SOCIALES', BACHILLER TECNICO  INDUSTRIAL EN LA ESPECIALIDAD  ELECTRICIDAD', BACHILLER TECNICO EN  COMERCIO Y  ADMINISTRACION', BACHILLER TÉCNICO EN  COMERCIO Y ADMINISTRACIÓN  INFORMÁTICA', INGENIERA EN  PETROLEOS', BACHILLER BILINGÜE EN  CIENCIAS', BACHILLER EN CIENCIAS  GENERAL', BACHILLER TECNICO  INDUSTRIAL', BACHILLER TECNICO EN  COMERCIO Y ADMINISTRACION  ESPECIALIZACION  INFORMATICA', BACHILLER EN CIENCIAS  ESPECIALIZACION SOCIALES', BACHILLER  TECNICO INDUSTRIAL ESPECIALIDAD  ELECTRICIDAD', BACHILLER TECNICO  POLIVALENTE EN  TELEINFORMATICA', TECNICO  INDUSTRIALES ELECTRONICA DE  CONSUMO', BACHILLER TECNICO  POLIVALENTE EN  ADMINISTRACION', BACHILLER EN  HUMANIDADES ESPECIALIZACION FISICO  MATEMATICAS', BACHILLER TECNICO  ESPECIALIZACION  ELECTROMECHANICA', BACHILLER TECNICO  INDUSTRIAL EN LA ESPECIALIZACION DE  ELECTR', BACHILLER TECNICO INDUSTRIAL  ESPECIALIZACION INSTALACIONES  E', BACHILLER TECNICO POLIVALENTE EN  INFORMATICA ESPECIALIDAD  AP', BACHILLER EN CIENCIAS  ESPECIALIDAD EN CIENCIAS ECONOMICAS  Y SOCIALES', BACHILLER TECNICO  INDUSTRIAL ESPECIALIZACION MECANICA  INDUST', BACHILLER TECNICO EN  COMERCIO Y ADMINISTRACION ESP. APL.  INF', BACHILLER EN COMERCIO Y  ADMINISTRACION', BACHILLER TECNICO  INDUSTRIAL ESPEC. </p>
--	--	--	--

			<p>MECANIZADO', 'APLICACIONES INFORMATICAS', 'BACHILLER TECNICO EN GESTION ADMINISTRATIVA Y CONTABLE', 'BACHILLER TECNICO EN COMERCIO Y ADMINISTRACION DE SISTEMAS', 'BACHILLER TECNICO EN CIENCIAS ESPECIALIDAD INFORMATICA', 'DE SERVICIOS INFORMACION Y COMERCIALIZACION TURISTICA', 'CONTADORA, BACHILLER EN CIENCIAS DE COMERCIO Y ADMINISTRACION', 'FISICO MATEMATICO', 'BACHILLER TECNICO EN COMERCIO Y ADMINISTRACION ESPECIALIZACION', 'FÍSICO', 'INGENIERA DE EMPRESAS', 'BACHILLER TECNICO INDUSTRIAL ELECTRONICA', 'BACHILLER EN INDUSTRIAS ESPECIALIZACION ELECTRONICA', 'BACHILLER TÉCNICO EN MECÁNICA AUTOMOTRIZ', 'INGENIERO AMBIENTAL', 'INGENIERA EN ELECTRONICA Y CONTROL', 'INGENIERO QUIMICO', 'PRACTICO EN AUXILIAR EN MANEJO DE EQUIPOS DE COMPUTO', 'INGENIERO CIVIL', 'AUXILIAR EN MANEJO DE EQUIPOS DE COMPUTO', 'BACHILLER TECNICO INDUSTRIAL ESPECIALIZACION MECANIZADO', 'POLIVANTE', 'BACHILLER EN CIENCIAS ESPECIALIDAD CIENCIAS', 'BACHILLER TECNICO EN ELECTRICIDAD Y ELECTRONICA ESPECIALIZACION', 'CONTADOR BACHILLER EN CIENCIAS DE COMERCIO Y ADMINISTRACION', 'FISICO', 'BACHILLER EN CIENCIAS DE COMERCIO Y ADMINISTRACION ESPECIALIZACION', 'BACHILLER TECNICO EN ARTES INDUSTRIALES ESPECIALIZACION ELEC', 'BACHILLER EN CIENCIAS EN LA ESPECIALIZACION DE FISICO MATEMAT', 'BACHILLER EN COMERCIO Y ADMINISTRACION ESPECIALIDAD CONTABIL', 'MATEMÁTICO', 'SECRETARIADO', 'BACHILLER BILINGÜE EN CIENCIAS ESPECIALIDAD FÍSICO MATEMÁTICO', 'CARRERAS CORTAS (PRÁCTICO) ESPECIALIZACION ELECTRICIDAD', 'BACHILLER TECNICO DE SEVICIOS ADMINISTRACION DE SISTEMAS', 'BACHILLER TÉCNICO EN INFORMÁTICA', 'BACHILLER TECNICO INDUSTRIAL ESPECIALIDAD ELECTRONICA DE CONSUMO', 'NO SE OTORGA</p>
--	--	--	--



			TITULACION', 'BACHILLER EN HUMANIDADES MODERNAS', 'PROGRAMACIÓN DE SISTEMAS', 'BACHILLER TECNICO INDUSTRIAL ESPECIALIZACION INDUSTRIA DEL V', 'BACHILLER TECNICO INDUSTRIAL ESP. MECANIZADO Y', 'BACHILLER EN CIENCIAS BASICAS', 'BACHILLER EN CIENCIAS ESPECIALIZACION FISICO MATEMATICAS-QUI', 'BACHILLER TÉCNICO EN ARTES INDUSTRIALES', 'BACHILLER EN COMERCIO Y ADMINISTRACION ESPECIALIDAD COMPUTAC', 'TECNICA EN EXPLOTACION DE SISTEMAS INFORMATICOS', 'CONTADOR BACHILLER EN COMERCIO Y ADMINISTRACIÓN', 'BACHILLER EN COMERCIO Y ADMINISTRACION ESPECIALIZACION COMER', 'CONTADOR PUBLICO BACHILLER EN CIENCIAS DE COMERCIO Y ADMINIS', 'INGENIERO ELÉCTRICO', 'INGENIERA MATEMATICA', 'HUMANIDADES MODERNAS ', 'BACHILLER EN GENERAL EN CIENCIAS', 'BACHILLER TECNICO EN COMERCIO Y ADMINISTRACION ESPECIALIDAD INFORMATICA', 'INGENIERA AMBIENTAL', 'INGENIERA EN SISTEMAS INFORMATICOS Y DE COMPUTACION', 'N/E ', 'BACHILLER EN CIENCIAS ESPECIALIDAD FISICO MATEMATICAS-QUIMIC', 'BACHILLER EN CIENCIAS ESPECIALIZACION FISICO MATEMATICAS QUI', 'BACHILLER TECNICO INDUSTRIAL ESPECIALIZACION MECANICA AUTOMOTRIZ', 'INGENIERA MECANICA', 'BACHILLER EN HUMANIDADES MODERNAS ESP QUIM-BIOL', 'BACHILLER TECNICO ', 'BACHILLER TECNICO EN ARTES INDUSTRIALES ESPECIALIDAD ELECTRI', 'CONTADORA, BACHILLER EN CIENCIAS DE COMERCIO Y ADMINISTRACIÓN', 'LICENCIADO EN RESTAURACION Y MUSEOLOGIA', 'INGENIERO EN ELECTRONICA Y TELECOMUNICACIONES', 'BACHILLER INDUSTRIALES MECANICA DE AVIACION', 'CARRERA TECNICA EN ELECTRONICA BASICA', nan, 'INGENIERO EN CIENCIAS ECONOMICAS Y FINANCIERAS', 'BACHILLER TECNICO EN MECANICA INDUSTRIAL', 'BACHILLER EN COMERCIO Y ADMINISTRACION
--	--	--	---

			<p>'BACHILLER POLIVALENTE EN CIENCIAS ESPECIALIZACION EXACTAS Y','PRÁCTICO EN MANTENIMIENTO Y ENSAMBLAJE DE COMPUTADORAS','BACHILLER EN CIENCIAS ESPECIALIZACION EXACTAS Y NATURALES','INGENIERO EN SISTEMAS INFORMATICOS Y COMPUTACION','INGENIERA EMPRESARIAL','BACHILLER EN COMERCIO Y ADMINISTRACION ESPECIALIZACION COMERCIALIZACION','BACHILLER EN COMERCIO Y ADMINISTRACION ESPECIALIDAD SECRETAR','INGENIERA EN ELECTRÓNICA Y REDES DE INFORMACIÓN','BACHILLER BILINGÜE EN CIENCIAS ESPECIALIDAD FÍSICO MATEMÁTICAS','INGENIERA EN SISTEMAS DE COMPUTACION E INFORMATICA','PSICOLOGO ADMINISTRATIVO Y ORGANIZACIONAL','INGENIERO MECANICO','BACHILLER EN CIENCIAS ESPECIALIDAD EXACTAS Y BIOLOGICAS','INGENIERO COMERCIAL','BACHILLER TECNICO INDUSTRIAL ESPECIALIZACION MOTORES DE AVIA','INGENIERA CIVIL','PRACTICO CARRERA CORTA EN LABORATORIO DE LA PEQUEÑA INDUSTRI','BACHILLER TECNICO EN AGROPECUARIA','BACHILLER EN CIENCIAS ESPECIALIZACION CIENCIAS SOCIALES','ABOGADO','CONTADOR BACHILLER TECNICO INDUSTRIAL EN MECANICA INDUSTRIAL','BACHILLER TECNICO INDUSTRIAL EN LA ESPECIALIZACION DE ELECTRONICA','COMUNICADOR CORPORATIVO','INGENIERA EN ELECTRONICA Y TELECOMUNICACIONES','PRÁCTICO (CARRERA CORTA) EN ELECTRÓNICA BÁSICA','BACHILLER EN CIENCIAS ESPECIALIZACION FISICO MATEMATICAS QUIMICO BIOLOGICAS','INGENIERO AGROINDUSTRIAL','BACHILLER TECNICO INDUSTRIAL ESPECIALIZACION MECANICA INDUSTRIAL','PRACTICO CARRERA CORTA EN ELECTRONICA BASICA','BACHILLER EN CIENCIAS ESPECIALIZACION','INGENIERO DE SISTEMAS','INGENIERO EN SISTEMAS INFORMATICOS Y DE COMPUTACION','ADMINISTRADOR DE EMPRESAS','BACHILLER EN CIENCIAS ESPECIALIZACIÓN CIENCIAS EXACTAS Y</p>
--	--	--	---

				<p>NATURALES', 'BACHILLER EN CIENCIAS ESPECIALIZACION FISICO MATEMATICAS-QUIMICO BIOLÓGICAS', 'BACHILLER TECNICO INDUSTRIAL ESPECIALIZACION INDUSTRIA DEL VESTIDO', 'CARRERA TECMOCA ESPECIALIZACION PROGRAMACION DE SISTEMAS', 'BACHILLER POLIVALENTE EN CIENCIAS ESPECIALIZACION EXACTAS Y NATURALES', 'BACHILLER EN CIENCIAS ESPECIALIDAD ELECTRONICA', 'INGENIERO EN ELECTRONICA Y REDES DE INFORMACION', 'PRÁCTICO CARRERA TÉCNICA MANEJO EQUIPOS COMPUTACIÓN', 'INGENIERA AGROINDUSTRIAL', 'INGENIERO QUÍMICO', 'INGENIERO ELECTROMECHANICO', 'BACHILLER EN COMERCIO Y ADMINISTRACION ESPECIALIZACION INFORMATICA', 'BIOLOGO', 'CONTADOR BACHILLER EN COMERCIO Y ADMINISTRACION ESPECIALIDAD INFORMATICA', 'LICENCIADO EN CIENCIAS DE LA EDUCACION', 'BACHILLER EN CIENCIAS EN LA ESPECIALIZACION DE FISICO MATEMATICAS', 'BACHILLER EN CIENCIAS EN LA SPECIALIZACION DE FISICO MATEMATICAS', 'BACHILLER EN CIENCIAS ESPECIALIZACION CIENCIAS BÁSICAS'</p>
TipoUnidadEducativa	señala el tipo de unidad educativa de procedencia del estudiante.	object	0	<p>FISCAL', 'EXTRANJERO', 'PARTICULAR LAICO', 'PARTICULAR RELIGIOSO', 'FISCOMISIONAL', 'PARTICULAR', 'NO ESPECIFICA', 'MUNICIPAL'</p>
PaisUnidadEducativa	Señala el país de origen de la unidad educativa de donde proviene el estudiante	object	0	<p>'ECUADOR', 'N/E', 'COLOMBIA'</p>
ProvinciaUnidadEducativa	Señala la provincia de donde proviene la unidad educativa del estudiante	object	0	<p>PICHINCHA', 'N/E', 'ESMERALDAS', 'IMBABURA', 'CAÑAR', 'AZUAY', 'CHIMBORAZO', 'COTOPAXI', 'MORONA SANTIAGO', 'CARCHI', 'LOS RIOS', 'PASTAZA', 'NAPO', 'SUCUMBOS', 'SANTO DOMINGO DE LOS TSACHILAS', 'GUAYAS', 'TUNGURAHUA', 'ORELLANA', 'BOLIVAR', 'SANTA ELENA', 'EL ORO', 'GALAPAGOS', 'LOJA', 'MANABI', 'ZAMORA CHINCHIPE'</p>

CantonUnidadEducativa	Señala el cantón de donde proviene la unidad educativa del estudiante	object	<p>QUITO', 'N/E', 'CAYAMBE', 'RUMIÑAHUI', 'PUERTO QUITO', 'RIOVERDE', 'IBARRA', 'AZOGUES', 'CAÑAR', 'CUENCA', 'COTACACHI', 'RIOBAMBA', 'LATACUNGA', 'SUCUA', 'MONTUFAR', 'TULCAN', 'ESPEJO', 'OTAVALO', 'BABAHOYO', 'ESMERALDAS', 'SAN PEDRO DE HUACA', 'MIRA', 'MERA', 'BOLIVAR', 'TENA', 'SHUSHUFINDI', 'SANTO DOMINGO DE LOS TSACHILAS', 'GUAYAQUIL', 'CHUNCHI', 'ALAUSI', 'AMBATO', 'ORELLANA', 'SAN MIGUEL', 'COLTA', 'GUARANDA', 'DAULE', 'PASTAZA', 'SANTA ELENA', 'GENERAL ANTONIO ELIZALDE (BUCAY)', 'MEJIA', 'GUANO', 'GUAMOTE', 'PALLATANGA', 'PASAJE', 'MACHALA', 'ZARUMA', 'SANTA ROSA', 'PIÑAS', 'HUAQUILLAS', 'ANTONIO ANTE', 'SAN MIGUEL DE LOS BANCOS', 'SANTA CRUZ', 'MARCABELI', 'EL GUABO', 'NARANJAL', 'BALSAS', 'ARENILLAS', 'PORTOVELO', 'LOJA', 'LAGO AGRIO', 'LAS LAJAS', 'PUYANGO', 'CAMILO PONCE ENRIQUEZ', 'ELOY ALFARO', 'SAN LORENZO', 'QUININDE', 'MANTA', 'ATACAMES', 'MUISNE', 'QUEVEDO', 'SALCEDO', 'SIGCHOS', 'SAQUISILI', 'PANGUA', 'PUJILI', 'LA MANA', 'PLAYAS', 'MONTALVO', 'VENTANAS', 'MORONA', 'ECHEANDIA', 'SIGSIG', 'GUALACEO', 'SANTA ISABEL', 'CHAGUARPAMBA', 'GUALAQUIZA', 'LIMON INDANZA', 'PUCARA', 'CALUMA', 'CHILLANES', 'MILAGRO', 'EL EMPALME', 'SANTA ANA', 'PORTOVIEJO', 'NARANJITO', 'SALINAS', 'SAN JACINTO DE YAGUACHI', 'LA LIBERTAD', 'LA TRONCAL', 'NOBOL', 'SAMBORONDON', 'CHONE', 'EL CARMEN', 'PEDERNALES', 'TOSAGUA', 'MONTECRISTI', 'SAN VICENTE', 'FLAVIO ALFARO', 'JIPIJAPA', 'PICHINCHA', 'JAMA', 'SUCRE', 'ROCAFUERTE', 'PAJAN', 'PUERTO LOPEZ', 'JARAMIJO', 'PALORA', 'BAÑOS DE AGUA SANTA', 'ARAJUNO', 'DURAN', 'QUIJOS', 'PEDRO MONCAYO', 'PEDRO VICENTE MALDONADO', 'PALTAS', 'ESPINDOLA', 'GONZANAMA', 'CALVAS', 'PALESTINA', '24 DE MAYO', 'ZAMORA', 'SANTIAGO DE PILLARO', 'CELICA', 'VINCES', 'CHINCHIPE', 'LOGROÑO', 'SAN MIGUEL DE URCUQUI', 'ATAHUALPA', 'SANTA CLARA', 'BUENA FE', 'EL CHACO', 'LORETO', 'CATAMAYO', 'SARAGURO', 'PATATE', 'SAN PEDRO DE</p>
-----------------------	---	--------	--

				PELILEO', 'MOCHA', 'CEVALLOS', 'QUERO', 'PIMAMPIRO', 'ARCHIDONA', 'SAN CRISTOBAL', 'MACARA', 'OLMEDO', 'URDANETA', 'ZAPOTILLO', 'TISALEO', 'BABA', 'PUEBLOVIEJO', 'PINDAL', 'CASCALES', 'LA JOYA DE LOS SACHAS', 'GONZALO PIZARRO', 'CUYABENO', 'SUCUMBIO', 'PALANDA', 'QUILANGA', 'VALENCIA', 'SOZORANGA', 'YANTZAZA', 'CENTINELA DEL CONDOR'
UrbanResidence	Señala si estudiante vive en una zona urbana	object	0	'S', 'N'
PeriodoInicioPrepo	Señala el periodo en el cual inicio el Pre Politécnico	object	794	'2019-A', '2018-B', '2018-A', '2019-B', '2017-B', '2011-1', '2016-A', '2015-A', '2010-2', '2015-B', '2014-A', '2017-A', '2020-A', '2009-1', '2010-1', '2012-1', '2016-B', '2014-B', '2004-1', '2011-2', '1987-1', '2012-B', nan, '2008-1', '1981-1', '1990-1', '2009-2', '1982-1', '2013-B', '1991-1', '2003-1', '2006-1', '2004-2', '2005-1', '2007-1', '2008-2', '2000-1', '2007-2', '2006-2', '1998-1', '2001-2', '2000-2', '2012-A', '2002-1', '2003-2', '2005-2', '1997-1', '1996-1', '1993-1', '2002-2', '1999-1', '1995-1', '1985-1', '1979-1', '1983-1', '1997-2', '1999-2', '2001-1', '1998-2', '1978-1', '1978-2', '1995-2', '1999-3', '1986-1', '1971-1', '1971-2', '1970-1', '1973-2', '1976-1', '1977-1', '1980-1', '1992-1', '1980-2', '1984-2', '1992-2', '1983-2', '1982-2', '1986-2', '1985-2', '1984-1', '1989-1', '1993-2', '1994-2', '1991-2', '1994-1', '2013-A', '1996-2', '1989-2', '1988-2'
PeriodoFinPrepo	Señala el periodo en el cual finalizó el Pre Politécnico	object	794	'2019-B', '2019-A', '2018-B', '2018-A', '2020-A', '2017-B', '2011-1', '2016-B', '2015-B', '2015-A', '2010-2', '2016-A', '2014-A', '2017-A', '2009-1', '2010-1', '2012-1', '2005-2', '2014-B', '1989-2', '2012-B', nan, '2011-2', '2008-1', '1983-2', '1994-2', '2009-2', '1984-2', '2013-B', '1993-1', '2003-2', '2006-2', '2004-2', '2013-A', '2000-2', '2007-2', '2008-2', '1999-2', '2002-1', '2001-2', '2012-A', '2007-1', '2005-1', '2006-1', '1999-3', '2000-3', '1993-2', '2003-1', '2004-1', '1999-1', '1998-1', '1997-1', '2000-1', '1995-1', '1988-2', '1982-1', '1998-2', '1989-1', '1980-1', '1981-2', '2002-2', '1985-2', '2001-1', '1996-1', '1992-1', '1972-2', '1974-1', '1985-1', '1971-1', '1978-1', '1979-2', '1977-1', '1991-1', '1983-1', '1987-2', '1996-2', '1980-2', '1987-1', '1986-1', '1986-2', '1988-1', '1984-1', '1992-2', '1991-2', '1979-1', '1990-2', '1982-2', '1994-1', '1995-2', '1997-2'
NroPeriodosPro	Indica el número de periodos que permaneció en el prepo	int64	0	2, 1, 3, 6, 0, 10, 5, 4, 7, 11, 8, 9, 12

Nivelacion	Indica si el estudiante ha cursado periodos de nivelación	object	170	SNNA', 'PREPO', 'EXONERA', 'CAMBIO CARRERA', 'EXONERA_CAMBIOUNIVERSIDAD', 'GAR', 'GARC', 'CAMBIO UNIVERSIDAD', 'INTERCAMBIO EXTRANJERO'
PeriodoInicioCarr	Señala el periodo en el cual el estudiante inicio la carrera	object	0	'2019-A', '2019-B', '2018-A', '2017-B', '2011-1', '2016-A', '2020-A', '2015-A', '2010-2', '2016-B', '2014-A', '2017-A', '2009-2', '2012-A', '2015-B', '1990-1', '2018-B', '2012-1', '2008-2', '1984-1', '1998-1', '1995-1', '2010-1', '2001-2', '1993-2', '2004-1', '2007-1', '2005-1', '2005-2', '2013-B', '2011-2', '2014-B', '2001-1', '2008-1', '2000-1', '2012-B', '2002-2', '2002-1', '2013-A', '2007-2', '2009-1', '2006-1', '2006-2', '1994-1', '2004-2', '2003-2', '1999-2', '1998-2', '1997-2', '2002-4', '1995-2', '1989-1', '1982-2', '1985-1', '1999-1', '1989-2', '1981-1', '1980-1', '2003-1', '1986-1', '2000-2', '1996-2', '1972-1', '1992-2', '1973-1', '1974-1', '1975-1', '1988-1', '1982-1', '1979-1', '1978-1', '1991-2', '1983-2', '1987-1', '1984-2', '1987-2', '1986-2', '1990-2', '1988-2', '1997-1', '1992-1', '1985-2', '1991-1', '1994-2', '1996-1', '2006-4', '1993-1', '2000-3'
Begin_YR	Señala el año de inicio de la carrera	int64	0	2019, 2018, 2017, 2011, 2016, 2020, 2015, 2010, 2014, 2009, 2012, 1990, 2008, 1984, 1998, 1995, 2001, 1993, 2004, 2007, 2005, 2013, 2000, 2002, 2006, 1994, 2003, 1999, 1997, 1989, 1982, 1985, 1981, 1980, 1986, 1996, 1972, 1992, 1973, 1974, 1975, 1988, 1979, 1978, 1991, 1983, 1987, 2019-04-01T00:00:00.000000000', '2019-09-23T00:00:00.000000000', '2018-04-09T00:00:00.000000000', '2017-10-16T00:00:00.000000000', '2010-09-06T00:00:00.000000000', '2016-04-18T00:00:00.000000000', '2020-06-01T00:00:00.000000000', '2015-04-06T00:00:00.000000000', '2010-03-01T00:00:00.000000000', '2016-10-10T00:00:00.000000000', '2014-02-03T00:00:00.000000000', '2017-04-17T00:00:00.000000000', '2009-03-02T00:00:00.000000000', '2012-01-23T00:00:00.000000000', '2015-09-28T00:00:00.000000000', '1990-01-01T00:00:00.000000000', '2018-10-01T00:00:00.000000000', '2011-08-22T00:00:00.000000000', '2008-03-17T00:00:00.000000000', '1984-01-01T00:00:00.000000000', '1998-01-01T00:00:00.000000000', '1995-01-01T00:00:00.000000000', '2009-09-07T00:00:00.000000000', '2001-04-

			16T00:00:00.000000000', '1993-01-01T00:00:00.000000000', '2003-10-27T00:00:00.000000000', '2006-10-02T00:00:00.000000000', '2004-10-04T00:00:00.000000000', '2005-03-28T00:00:00.000000000', '2013-07-29T00:00:00.000000000', '2011-02-28T00:00:00.000000000', '2014-09-29T00:00:00.000000000', '2000-10-10T00:00:00.000000000', '2007-09-24T00:00:00.000000000', '2000-01-01T00:00:00.000000000', '2012-07-30T00:00:00.000000000', '2002-04-01T00:00:00.000000000', '2001-10-01T00:00:00.000000000', '2013-01-21T00:00:00.000000000', '2007-03-19T00:00:00.000000000', '2008-09-15T00:00:00.000000000', '2005-10-03T00:00:00.000000000', '2006-03-27T00:00:00.000000000', '1994-01-01T00:00:00.000000000', '2004-04-12T00:00:00.000000000', '2003-05-19T00:00:00.000000000', '1999-01-01T00:00:00.000000000', '1997-01-01T00:00:00.000000000', '2002-01-01T00:00:00.000000000', '1989-01-01T00:00:00.000000000', '1982-01-01T00:00:00.000000000', '1985-01-01T00:00:00.000000000', '1981-01-01T00:00:00.000000000', '1980-01-01T00:00:00.000000000', '2002-12-02T00:00:00.000000000', '1986-01-01T00:00:00.000000000', '1996-01-01T00:00:00.000000000', '1972-01-01T00:00:00.000000000', '1992-01-01T00:00:00.000000000', '1973-01-01T00:00:00.000000000', '1974-01-01T00:00:00.000000000', '1975-01-01T00:00:00.000000000', '1988-01-01T00:00:00.000000000', '1979-01-01T00:00:00.000000000', '1978-01-01T00:00:00.000000000', '1991-01-01T00:00:00.000000000', '1983-01-01T00:00:00.000000000', '1987-01-01T00:00:00.000000000', '2006-01-01T00:00:00.000000000'
--	--	--	---

Begin_Career_Date	indica la fecha en timestamp de inicio de carrera	datetime64[ns]	<p>2019, 2018, 2017, 2011, 2016, 2020, 2015, 2010, 2014, 2009, 2012, 1990, 2008, 1984, 1998, 1995, 2001, 1993, 2004, 2007, 2005, 2013, 2000, 2002, 2006, 1994, 2003, 1999, 1997, 1989, 1982, 1985, 1981, 1980, 1986, 1996, 1972, 1992, 1973, 1974, 1975, 1988, 1979, 1978, 1991, 1983, 1987</p> <p>2019-04-01T00:00:00.000000000', '2019-09-23T00:00:00.000000000', '2018-04-09T00:00:00.000000000', '2017-10-16T00:00:00.000000000', '2010-09-06T00:00:00.000000000', '2016-04-18T00:00:00.000000000', '2020-06-01T00:00:00.000000000', '2015-04-06T00:00:00.000000000', '2010-03-01T00:00:00.000000000', '2016-10-10T00:00:00.000000000', '2014-02-03T00:00:00.000000000', '2017-04-17T00:00:00.000000000', '2009-03-02T00:00:00.000000000', '2012-01-23T00:00:00.000000000', '2015-09-28T00:00:00.000000000', '1990-01-01T00:00:00.000000000', '2018-10-01T00:00:00.000000000', '2011-08-22T00:00:00.000000000', '2008-03-17T00:00:00.000000000', '1984-01-01T00:00:00.000000000', '1998-01-01T00:00:00.000000000', '1995-01-01T00:00:00.000000000', '2009-09-07T00:00:00.000000000', '2001-04-16T00:00:00.000000000', '1993-01-01T00:00:00.000000000', '2003-10-27T00:00:00.000000000', '2006-10-02T00:00:00.000000000', '2004-10-04T00:00:00.000000000', '2005-03-28T00:00:00.000000000', '2013-07-29T00:00:00.000000000', '2011-02-28T00:00:00.000000000', '2014-09-29T00:00:00.000000000', '2000-10-10T00:00:00.000000000', '2007-09-24T00:00:00.000000000', '2000-01-01T00:00:00.000000000', '2012-07-30T00:00:00.000000000', '2002-04-01T00:00:00.000000000', '2001-10-01T00:00:00.000000000', '2013-01-21T00:00:00.000000000', '2007-03-19T00:00:00.000000000', '2008-09-15T00:00:00.000000000', '2005-10-03T00:00:00.000000000', '2006-03-27T00:00:00.000000000', '1994-01-01T00:00:00.000000000', '2004-04-12T00:00:00.000000000', '2003-05-19T00:00:00.000000000', '1999-01-</p>
-------------------	---	----------------	---



				01T00:00:00.000000000', '1997-01-01T00:00:00.000000000', '2002-01-01T00:00:00.000000000', '1989-01-01T00:00:00.000000000', '1982-01-01T00:00:00.000000000', '1985-01-01T00:00:00.000000000', '1981-01-01T00:00:00.000000000', '1980-01-01T00:00:00.000000000', '2002-12-02T00:00:00.000000000', '1986-01-01T00:00:00.000000000', '1996-01-01T00:00:00.000000000', '1972-01-01T00:00:00.000000000', '1992-01-01T00:00:00.000000000', '1973-01-01T00:00:00.000000000', '1974-01-01T00:00:00.000000000', '1975-01-01T00:00:00.000000000', '1988-01-01T00:00:00.000000000', '1979-01-01T00:00:00.000000000', '1978-01-01T00:00:00.000000000', '1991-01-01T00:00:00.000000000', '1983-01-01T00:00:00.000000000', '1987-01-01T00:00:00.000000000', '2006-01-01T00:00:00.000000000'
AgeWhenStarting	Señala la edad en la cual inició la carrera	float64	0	20.48493151, 18.81369863, 17.27671233, ..., 16.65753425, 31.13150685, 25.56164384
PeriodoUltMatriculaCarr	Señala el periodo en el cual el estudiante realizó su última matrícula	object	0	'2019-B', '2020-A', '2018-B', '2018-A', '2017-B', '2011-1', '2016-B', '2015-A', '2010-2', '2014-A', '2019-A', '2017-A', '2014-B', '2012-1', '2016-A', '2013-B', '2015-B', '2013-A', '2010-1', '2011-2', '2012-B', '2012-A', '2009-2', '2008-1', '2008-2', '2009-1', '2007-1', '2007-2', '2000-1', '1998-1', '2006-1', '2003-1'
TipoUltMatriculaCarr	Indica el tipo de matrícula que tuvo el estudiante en su último periodo	object	0	Ordinaria', 'Extraordinaria/Especial
FecUltMatriculaCarr	Indica la fecha de matrícula	object	23	Fecha de matrícula
FormaPagUltMatriculaCarr	Señala la forma del último pago de la matrícula	object	0	TARJETA CREDITO', 'N/A ', 'TRANSFERENCIA', 'BANCO ', 'EFECTIVO', 'ROL'
PagUltMatriculaCarr	Fecha de ultimo pago de matrícula	float64	13979	Fecha de ultimo pago de matrícula
Carrera	Señala la carrera que está cursando el estudiante.	object	0	INGENIERIA, CIENCIAS Y CIENCIAS ADMINISTRATIVAS', '(RRA) PETROLEOS', '(RRA) INGENIERIA CIVIL', 'PROPEDEUTICO DE INGENIERIA Y CIENCIAS', 'NIVELACION DE INGENIERIA Y CIENCIAS', '(RRA) INGENIERIA QUIMICA', 'INGENIERIA EN SISTEMAS INFORMATICOS Y DE COMPUTACION', 'PROPEDEUTICO DE EMPRESARIAL Y ECONOMICAS', 'NIVELACION DE

			TECNOLOGOS', (RRA) COMPUTACION', 'INGENIERIA ELECTRONICA Y CONTROL', (RRA) ELECTRICIDAD', 'INGENIERIA EN ELECTRONICA Y REDES DE INFORMACION', 'FISICA', 'INGENIERIA EN CIENCIAS ECONOMICAS Y FINANCIERAS', 'INGENIERIA CIVIL', 'INGENIERIA EN PETROLEOS', (RRA) ELECTRONICA Y AUTOMATIZACION', 'TECNOLOGIA EN ELECTROMECHANICA', 'INGENIERIA MECANICA', 'INGENIERIA EN ELECTRONICA Y TELECOMUNICACIONES', 'INGENIERIA ELECTRICA', (RRA) MECANICA', (RRA) SOFTWARE', 'PROPEDEUTICO TECNOLOGIA EN AGUA Y SANEAMIENTO AMBIENTAL', 'TECNOLOGIA EN ELECTRONICA Y TELECOMUNICACIONES', 'TECNOLOGIA EN AGUA Y SANEAMIENTO AMBIENTAL', (RRA) INGENIERIA DE LA PRODUCCION', 'INGENIERIA INFORMATICA', (RRA) TELECOMUNICACIONES', 'NIVEL TECNOLOGICO SUPERIOR', 'MATEMATICA', (RRA) TECNOLOGIAS DE LA INFORMACION', 'TECNOLOGIA EN PROCESOS DE PRODUCCION MECANICA', 'INGENIERIA AGROINDUSTRIAL', 'INGENIERIA AMBIENTAL', 'INGENIERIA EMPRESARIAL', 'INGENIERIA GEOLOGICA', 'INGENIERIA QUIMICA', 'TECNOLOGIA EN ANALISIS DE SISTEMAS INFORMATICOS', 'INGENIERIA MATEMATICA', 'NIVELACION DE EMPRESARIAL Y ECONOMICAS', 'PROPEDEUTICO DE TECNOLOGOS', (RRA) TECNOLOGIA SUPERIOR EN REDES Y TELECOMUNICACIONES', (RRA) FISICA', (RRA) GEOLOGIA', (RRA) AGROINDUSTRIA', (RRA) TECNOLOGIA SUPERIOR EN ELECTROMECHANICA', (RRA) TECNOLOGIA SUPERIOR EN DESARROLLO DE SOFTWARE', 'TECNOLOGIA EN MANTENIMIENTO INDUSTRIAL', (RRA) TECNOLOGIA SUPERIOR EN AGUA Y SANEAMIENTO AMBIENTAL', 'TECNOLOGIA EN ADMINISTRACION DE PROYECTOS DE LA CONSTRUCCION', (RRA) MATEMATICA', (RRA) ECONOMIA', (RRA) MATEMATICA APLICADA', (RRA) INGENIERIA AMBIENTAL', 'INGENIERIA EN
--	--	--	--

				GEOTECNIA', 'ESTUDIANTES EXTRANJEROS'
Facultad	Señala la facultad en la que estudia la carrera.	object	0	NIVELACION', 'GEOLOGIA Y PETROLEOS', 'INGENIERIA CIVIL', 'INGENIERIA QUIMICA Y AGROINDUSTRIA', 'INGENIERIA DE SISTEMAS', 'INGENIERIA ELECTRICA Y ELECTRONICA', 'CIENCIAS', 'TECNOLOGICA', 'INGENIERIA MECANICA', 'CIENCIAS ADMINISTRATIVAS', 'ADMISIONES'
CredAcum	Créditos acumulados	float64	6997	Créditos acumulados
CredRef	Indica los créditos referenciales del estudiante	float64	3	Lista de créditos con los que cuenta el estudiante
CredAcumRepro	Créditos acumulados reprobados	float64	12846	Créditos acumulados reprobados
PorcenAcum	Porcentaje de créditos acumulados	float64	36	porcentaje de acumulados
Ingreso	Ingreso económico del estudiante	float64	8331	Ingreso económico
NroMiem	Número de miembros del núcleo familiar	float64	8331	Número de miembros del núcleo familiar
TrabajaEstudiante	Indica si el estudiante trabaja	object	17010	NO', nan, 'SI'
OcupacionEstudiante	Señala la ocupación del estudiante	object	23597	nan, 'QUEHACERES DOMESTICOS', 'EMPLEADO DE SERVICIO', 'INFORMAL', 'VENDEDOR', 'TECNICO', 'EJECUTIVO ADMINISTRATIVO', 'EMPLEADO ADMINISTRATIVO', 'PROFESIONAL', 'TRABAJADOR DEL DEPORTE', 'TRABAJADOR DEL TRANSPORTE', 'AGRICULTOR/PESCADOR/CAZADOR/TRABAJADOR FORESTAL', 'ARTESANO', 'POLICIA/MILITAR', 'JUBILADO'
IngresoEstudiante	Indica el ingreso económico del estudiante	float64	0	Ingreso económico del estudiante
RelLaboralEstudiante	Señala la relación laboral del estudiante	object	22691	nan, 'INDEPENDIENTE', 'DEPENDENCIA', 'DEPENDIENTE'

OcupacionConyugue	Señala la ocupación de la cónyuge estudiante.	object	25730	nan, 'QUEHACERES DOMESTICOS', 'EJECUTIVO ADMINISTRATIVO', 'PROFESIONAL', 'TECNICO', 'EMPLEADO ADMINISTRATIVO', 'EMPLEADO DE SERVICIO', 'VENDEDOR', 'TRABAJADOR DEL DEPORTE', 'ARTESANO', 'POLICIA/MILITAR', 'INFORMAL', 'JUBILADO', 'TRABAJADOR DEL TRANSPORTE', 'AGRICULTOR/PESCADOR/CAZADOR/TRABAJADOR FORESTAL'
IngresoConyugue	Señala el ingreso económico del estudiante	float64	0	ingreso económico del cónyuge
AfiliadoEstudiante	Señala si es estudiante afiliado o no	object	17010	NO', 'SI', nan
TrabajaConyugue	Señala si la cónyuge del estudiante trabaja	object	17010	NO', 'SI', nan
RelLaboralConyugue	Señala la relación laboral de la cónyuge estudiante	object	25620	nan, 'DEPENDIENTE', 'INDEPENDIENTE', 'DEPENDENCIA'
OcupacionPadre	Señala la ocupación del padre estudiante.	object	5791	INFORMAL', 'QUEHACERES DOMESTICOS', 'PROFESIONAL', nan, 'AGRICULTOR/PESCADOR/CAZADOR/TRABAJADOR FORESTAL', 'NO APLICA', 'ARTESANO', 'DIFUNTO', 'NO DEPENDO ECONOMICAMENTE DE PADRE/MADRE', 'EMPLEADO DE SERVICIO', 'JUBILADO', 'TRABAJADOR DEL TRANSPORTE', 'VENDEDOR', 'POLICIA/MILITAR', 'TECNICO', 'EMPLEADO ADMINISTRATIVO', 'EJECUTIVO ADMINISTRATIVO', 'TRABAJADOR DEL DEPORTE', 'MINERO/CANTERO'
IngresoPadre	Señala el ingreso económico del padre del estudiante	float64	0	ingresos económicos del padre
AfiliadoConyugue	Indica si la cónyuge estudiante es afiliado	object	17010	'NO', nan, 'SI'
RelLaboralPadre	Señala la relación laboral del padre del estudiante	object	8750	'INDEPENDIENTE', 'DEPENDENCIA', nan, 'DEPENDIENTE'

OcupacionMadre	Señala la ocupación de la madre del estudiante	object	3604	'PROFESIONAL', 'INFORMAL', nan, 'ARTESANO', 'QUEHACERES DOMESTICOS', 'NO APLICA', 'NO DEPENDO ECONOMICAMENTE DE PADRE/MADRE', 'VENDEDOR', 'EMPLEADO DE SERVICIO', 'EMPLEADO ADMINISTRATIVO', 'POLICIA/MILITAR', 'AGRICULTOR/PESCADOR/CAZADOR/TRABAJADOR FORESTAL', 'EJECUTIVO ADMINISTRATIVO', 'JUBILADO', 'TECNICO', 'DIFUNTO', 'TRABAJADOR DEL TRANSPORTE', 'MINERO/CANTERO', 'TRABAJADOR DEL DEPORTE'
IngresoMadre	Señala el ingreso económico de la madre del estudiante	float64	0	ingreso económico de la madre
AfiliadoPadre	Señala si el padre del estudiante es afiliado o no	object	17010	'NO', nan, 'SI'
RelLaboralMadre	Señala la relación laboral de la madre del estudiante	object	5282	'INDEPENDIENTE', 'DEPENDENCIA', nan, 'DEPENDIENTE'
AfiliadoMadre	Señala si la madre del estudiante es afiliada.	object	17010	'NO', nan, 'SI'
TrabajaHermano	Señala si el hermano del estudiante trabaja o no	object	17010	'NO', nan, 'SI'
RelLaboralHermano	Señala la relación laboral del hermano del estudiante	object	25786	nan, 'INDEPENDIENTE', 'DEPENDIENTE'
IngresoHermano	Señala el ingreso económico del hermano	float64	17010	Ingreso económico hermano
AfiliadoHermano	Señala si el hermano del estudiante está afiliado	object	17010	'NO', nan, 'SI'
Casadounion	indica si el estudiante está casado o en unión libre	object	17010	'NO', nan, 'SI'
TieneHijos	Indica si el estudiante tiene hijos	object	17010	'NO', nan, 'SI'
Hijos	cantidad de hijos del estudiante	float64	17010	cantidad de hijos
TieneHijosMenores	Indica si el estudiante tiene hijos menores	object	17010	'NO', nan, 'SI'

HijosMenores	Indica la cantidad de hijos menores que tiene el estudiante.	float64	17010	0., nan, 1., 10., 4., 2., 3.
Lactancia	Indica si el estudiante se encuentra en periodo de lactancia	object	17010	'NO', nan, 'SI'
MadreSoltera	Indica si el estudiante es madre soltera	object	17010	'NO', nan, 'SI'
PadreSoltero	Indica si el estudiante es padre soltero	object	17010	'NO', nan, 'SI'
NroMiembrosDependen	Indica el número de miembros dependientes en el núcleo familiar	float64	17010	Número de miembros dependientes
Vive	Indica con quien vive el estudiante	object	17010	FAMILIARES', 'AMBOS', nan, 'SOLO', 'HERMANOS', 'MADRE', 'OTROS', 'CONYUGEHIJOS', 'PADRE', 'CONYUGE', 'HIJOS', '0'
TieneDiscapacidad	Indica si el estudiante tiene discapacidad	object	17010	'NO', nan, 'SI'
DiscapacidadFamiliar	Señala si vive con un familiar que tiene discapacidad	object	17010	'NO', nan, 'SI'
NroDiscapacidad	Indica el porcentaje de discapacidad que posee el familiar del estudiante	float64	17010	porcentaje de discapacidad
DiscapacidadTotal	Indica si el familiar tiene discapacidad total o no	object	17010	'NO', nan, 'SI'
Enfermedad	Señala si el estudiante tiene alguna enfermedad	object	17010	'NO', nan, 'SI'
Embarazo	Señala si la estudiante está en periodo de embarazo	object	17010	'NO', nan, 'SI'
Adultomayor	indica si el estudiante es adulto mayor	object	17010	'NO', nan, 'SI'
Adolescente	indica si el estudiante es adolescente	object	17010	'NO', nan, 'SI'
Violencia	Señala si el estudiante ha sufrido violencia intrafamiliar	object	17010	'NO', nan, 'SI'

Financiamiento Estudios	Señala como el estudiante financia sus estudios.	object	0	PADRES TUTORES', 'NO REGISTRA', 'HERMANOS', 'RECURSOS PROPIOS', 'OTROS FAMILIARES', 'CONYUGE', 'CREDITO EDUCATIVO', 'BECA ESTUDIO', 'OTROS MIEMBROS HOGAR', 'PAREJA ENTIMENTAL'
OtrosIngresos	Indica si el estudiante tiene ingresos adicionales	object	17010	'NO', nan, 'SI'
Propiedad	indica su el estudiante tiene propiedades	object	17010	'NO', nan, 'SI'
AvalúoPropiedad	Señala el avaluado de la propiedad del estudiante	float64	17010	avalúo de propiedad de vivienda
NegocioPropio	indica si el estudiante tiene negocio propio	object	17010	'NO', nan, 'SI'
Impuesto	Indica si el estudiante para impuestos	object	17010	'NO', nan, 'SI'
Vehiculo	indica si el estudiante tiene vehículo	object	17010	'NO', nan, 'SI'
VehículoPublico	Indica si el estudiante utiliza vehículo publico	object	17010	'NO', nan, 'SI'
AvalúoVehiculo	Señala el avalúo del vehículo del estudiante	float64	17010	avalúo de vehículo
Rebaja	Indica el porcentaje de rebaja en beca	float64	17010	503.15, nan
+IngresoArr	Indica el nivel de ingresos del estudiante	float64	9318	Lista de valores de Nivel de ingreso
+IngresoOtros	Indica el nivel de ingreso del estudiante por otros conceptos	float64	9318	Lista de valores de Nivel de ingreso por otros
-TotalRebaja	Señala el total de rebaja	float64	2015	40., 380., 120., ..., 3330., 1158.96, 1097.1
ValorRebajaTransporte	Indica el valor de rebaja por transporte	float64	19142	ValorRebajaTransporte
ValorRebajaServiciosBasicos	Indica el valor de rebaja por servicios básicos	float64	2030	ValorRebajaServiciosBasicos
ValorRebajaInternet	Indica el Valor de rebaja por internet	float64	18266	ValorRebajaInternet
ValorRebajaArriendo	Indica el valor de rebaja por arriendo	float64	20843	ValorRebajaArriendo

ValorRebaja Enfermedad	Indica el valor de rebaja por enfermedad	float64	25329	ValorRebajaEnfermedad
TipoPropiedad	Indica que tipo de propiedad es el propietario estudiante	object	14475	[nan, 'CASA', 'DEPARTAMENTO']
ValorPropiedad	Indica el valor de la propiedad del estudiante	float64	14475	ValorPropiedad
TipoPropiedadTerr	indica el tipo de propiedad del estudiante, si es terreno	object	22645	[nan, 'TERRENO']
ValorPropiedadTerr	Señala el valor del terreno	float64	22645	ValorPropiedadTerr
TipoAuto	Señala el tipo de auto del estudiante	object	19812	nan, 'AUTOMOVIL'
AñoAuto	Señala el año del vehículo de propiedad del estudiante	float64	19816	AñoAuto
ValorAuto	indica el valor del auto del estudiante	float64	19812	ValorAuto
TipoCamioneta	indica el tipo de camioneta que tiene el estudiante	object	23855	nan, 'CAMIONETA'
AñoCamioneta	indica el año de fabricación de la camioneta del estudiante	float64	23861	AñoCamioneta
ValorCamioneta	indica el valor de la camioneta del estudiante	float64	23855	ValorCamioneta
QuintilPeriodoActual	indica el quintil del periodo actual	float64	7591	QuintilPeriodoActual
lpf	indica el código lpf del estudiante	float64	0	lpf
EstadoEstudiante	Indica el estado de la titulación del estudiante	object	20115	nan, 'Graduado', 'Tramite Graduación-Anillados', 'ComplexivoAnt'
BecaExcelenciaAcad	indica si el estudiante ha tenido beca por excelencia académica	object	0	'N', 'S'
UltPerExcelenciaAcad	Señala el último periodo en el cual el estudiante tuvo beca por excelencia académica	object	0	'N', '2012-A', '2019-B', '2015-A', '2018-B', '2011-2', '2013-B', '2015-B', '2012-B', '2013-A', '2016-B', '2012-1', '2017-A', '2014-A', '2016-A', '2017-B', '2019-A', '2010-1', '2018-A', '2010-2', '2009-2', '2014-B', '2011-1', '2003-1'



BecaSituacionEcon	indica si el estudiante ha tenido beca por situación económica	object	24483	nan, 'S'
UltPerSituacionEcon	Señala el último periodo en el cual el estudiante tuvo beca por situación económica	object	24483	nan, '2017-A', '2015-A', '2014-A', '2018-B', '2012-B', '2007-1', '2016-A', '2006-2', '2013-A', '2015-B', '2013-B', '2009-2', '2016-B', '2019-B', '2014-B', '2020-A', '2018-A', '2011-2', '2019-A', '2012-A', '2010-2', '2010-1', '2011-1', '2012-1', '2017-B', '2003-2', '2008-2', '2009-1', '2008-1', '2003-1', '2005-2', '2005-1', '2004-2', nan, 'S'
BecaMerito	indica si el estudiante ha tenido beca por merito	object	26056	nan, '2017-A', '2015-A', '2014-A', '2018-B', '2012-B', '2007-1', '2016-A', '2006-2', '2013-A', '2015-B', '2013-B', '2009-2', '2016-B', '2019-B', '2014-B', '2020-A', '2018-A', '2011-2', '2019-A', '2012-A', '2010-2', '2010-1', '2011-1', '2012-1', '2017-B', '2003-2', '2008-2', '2009-1', '2008-1', '2003-1', '2005-2', '2005-1', '2004-2', nan, 'S'
UltPerMerito	Señala el último periodo en el cual el estudiante tuvo beca por merito	object	26056	nan, '2009-1', '2018-B', '2018-A', '2010-2', '2008-2', '2017-A', '2007-1', '2015-B', '2013-B', '2019-A', '2011-1', '2020-A', '2019-B', '2014-B', '2016-B', '2009-2', '2016-A', '2014-A', '2017-B'
BecaColegiatura	indica si el estudiante ha tenido beca por colegiatura	object	26337	nan, 'S'
UltPerColegiatura	Señala el último periodo en el cual el estudiante tuvo beca por colegiatura	object	26337	nan, 'S'

### Académicos

Atributo	Descripcion	Tipo de dato	Nulos	Valores
CodigoEnc	Codigo único de estudiante	object	0	Codigo único de estudiante
Periodo	Periodo académico del estudiante	object	711	'2019-A', '2019-B', '2018-B', '2018-A', '2017-B', '2011-1', '2016-A', '2016-B', '2020-A', '2015-A', '2015-B', '2017-A', '2010-2', '2014-A', '2009-1', '2009-2', '2010-1', '2011-2', '2012-1', '2012-A', '2012-B', '2013-A', '2013-B', '2014-B', '2013-R', '1987-1', '1987-2', '1988-1', '1988-2', '1989-1', '1989-2', '1990-1', '1990-2', '1991-1', '1991-2', '1992-1', '1992-2', '1993-1', '1993-2', '1994-1', '2004-1', '2008-1', '2008-2', '1981-1', '1981-2', '1982-1', '1982-2', '1983-1', '1983-2', '1985-1', '1985-2', '1986-1', '1986-2', '1998-1', '1998-2', '1999-1', '1999-2', '2000-1', '2000-2', '2001-1', '2001-2', '2002-1', '2002-2', '2003-1', '2003-2', '1994-2', '1995-1', '1995-2', '1996-1', '1996-2',

				'1997-1', '1997-2','2005-1', '2006-1', '2006-2', '2007-1', '2002-4', '2004-2','2005-2', '2007-2', nan, '2012-R', '2011-R', '2019-R', '2015-R','2014-R', '2016-R', '1999-3', '2000-3', '2017-R', '1979-1','1979-2', '1984-1', '1984-2', '1978-1', '1978-2', '1980-1','1971-1', '1971-2', '1973-2', '1974-1', '1974-2', '1975-1','1975-2', '1976-1', '1976-2', '1977-1', '1977-2', '1980-2','2018-R', 'Periodo'
Comat	Codigo de materia	object	0	CNIC010', 'CNIC020', 'CNIC030', ..., 'EPD21', 'EPD41', 'Codmat'
Materia	Nombre de la materia	object	0	FUNDAMENTOS DE MATEMATICA', 'GEOMETRIA Y TRIGONOMETRIA', 'FISICA', 'CENTRALES ELECTRICAS', 'DISTRIBUCION III', 'Materia'
Nrom	numero de materia	object	0	1', '2', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12', '13', '14', '15', '16', '17', '18', '19', '20', '0', '21', '22'
Calif	calificación del estudiante	object	65317	Lista de calificaciones del estudiante
Créditos	Créditos que cuenta el estudiante	object	0	créditos del estudiante
Aprobacion	Estado de aprobación de la materia	object	711	F', 'A', 'E', 'N', 'FS', nan, 'FA', 'Aprobación'
Carrera	Carrea que está cursando el estudiante	object	0	INGENIERIA, CIENCIAS Y CIENCIAS ADMINISTRATIVAS','(RRA) PETROLEOS', '(RRA) INGENIERIA CIVIL','PROPEDEUTICO DE INGENIERIA Y CIENCIAS','NIVELACION DE INGENIERIA Y CIENCIAS','(RRA) INGENIERIA DE LA PRODUCCION','INGENIERIA EN SISTEMAS INFORMATICOS Y DE COMPUTACION','PROPEDEUTICO DE EMPRESARIAL Y ECONOMICAS','NIVELACION DE TECNOLOGOS', '(RRA) COMPUTACION','INGENIERIA ELECTRONICA Y CONTROL','INGENIERIA EN ELECTRONICA Y TELECOMUNICACIONES','(RRA) TELECOMUNICACIONES', '(RRA) ELECTRICIDAD','INGENIERIA EN ELECTRONICA Y REDES DE INFORMACION', 'FISICA','NIVELACION DE EMPRESARIAL Y ECONOMICAS','INGENIERIA EN CIENCIAS ECONOMICAS Y FINANCIERAS','INGENIERIA CIVIL', 'INGENIERIA EN PETROLEOS','(RRA) ELECTRONICA Y AUTOMATIZACION','TECNOLOGIA EN ELECTRONICA Y TELECOMUNICACIONES','TECNOLOGIA EN ELECTROMECHANICA', 'INGENIERIA MECANICA','INGENIERIA ELECTRICA', 'PROPEDEUTICO DE TECNOLOGOS','(RRA)

			SOFTWARE','PROPEDEUTICO TECNOLOGIA EN AGUA Y SANEAMIENTO AMBIENTAL','TECNOLOGIA EN AGUA Y SANEAMIENTO AMBIENTAL','INGENIERIA INFORMATICA', '(RRA) MECANICA', 'MATEMATICA','(RRA) TECNOLOGIAS DE LA INFORMACION', 'INGENIERIA QUIMICA','TECNOLOGIA EN PROCESOS DE PRODUCCION MECANICA','PROPEDEUTICO DE AGROINDUSTRIA', 'INGENIERIA AGROINDUSTRIAL','INGENIERIA AMBIENTAL', 'INGENIERIA EMPRESARIAL','INGENIERIA GEOLOGICA','TECNOLOGIA EN ANALISIS DE SISTEMAS INFORMATICOS','(RRA) INGENIERIA QUIMICA', 'INGENIERIA MATEMATICA','NIVEL TECNOLOGICO SUPERIOR', '(RRA) GEOLOGIA','(RRA) AGROINDUSTRIA', '(RRA) FISICA', '(RRA) MATEMATICA APLICADA','TECNOLOGIA EN MANTENIMIENTO INDUSTRIAL','(RRA) TECNOLOGIA SUPERIOR EN ELECTROMECHANICA','TECNOLOGIA EN ADMINISTRACION DE PROYECTOS DE LA CONSTRUCCION','(RRA) MATEMATICA', '(RRA) ECONOMIA','(RRA) TECNOLOGIA SUPERIOR EN REDES Y TELECOMUNICACIONES','(RRA) INGENIERIA AMBIENTAL','(RRA) TECNOLOGIA SUPERIOR EN AGUA Y SANEAMIENTO AMBIENTAL','MAESTRIA EN INGENIERIA AMBIENTAL','(RRA) TECNOLOGIA SUPERIOR EN DESARROLLO DE SOFTWARE','INGENIERIA EN GEOTECNIA', 'ESTUDIANTES EXTRANJEROS', 'Carrera'
--	--	--	---

### Encuesta Tics

Atributo	Descripcion	Tipo dato	Nulos	Valores
CodigoEnc	Codigo único de estudiante	object	0	código único de estudiante
PREG1	G1P1[SQ001]. 1. ¿Cuál es la dirección en la cual estima se encontrará al iniciar las clases, el día 11 de mayo de 2020?: Siga las instrucciones definidas en el siguiente enlace para obtener su ubicación: CLIC AQUÍ. Una vez obtenida la ubicación, copia las coordenadas en el campo f) Ubicación (un ejemplo de coordenadas se presenta a continuación: -0.197030, -78.495096) [a) Provincia] (grupos)	object	0	COLOMBIA', 'PICHINCHA', 'AZUAY', 'LOJA', 'EL ORO', 'BOLIVAR','CHIMBORAZO', 'CAÑAR', 'IMBABURA', 'CARCHI', 'TUNGURAHUA','COTOPAXI', 'PASTAZA', 'SANTO DOMINGO DE LOS TSACHILAS','ESMERALDAS', 'GUAYAS', 'MANABI', 'SANTA ELENA', 'LOS RIOS','GALÁPAGOS', 'PERU', 'ESTADOS UNIDOS', 'MORONA SANTIAGO', 'NAPO','LATACUNGA',

				'ESPAÑA', 'SUCUMBIOS', 'ORELLANA', 'ZAMORA CHINCHIPE'
PREG2	G1P1[SQ002]. 1. ¿Cuál es la dirección en la cual estima se encontrará al iniciar las clases, el día 11 de mayo de 2020?: Siga las instrucciones definidas en el siguiente enlace para obtener su ubicación: CLIC AQUÍ. Una vez obtenida la ubicación, copia las coordenadas en el campo f) Ubicación (un ejemplo de coordenadas se presenta a continuación: -0.197030, -78.495096) [b) Cantón] (grupos)	object	0	Lista de coordenadas seleccionadas.
PREG3	G1P2. 2. En la ubicación que indicó en el punto anterior, ¿Tiene servicio de Internet?	object	0	array (['Si', 'No']
PREG4	G1P3. 3. Si la respuesta a la pregunta 2 fue SI, ¿de qué tipo es su servicio de Internet?	object	0	Plan de Datos Fijos a través de operadores como Netlife, CNT, TVCable, entre otros.', '-', 'Plan de Datos Móviles a través de operadores de telefonía móvil como Claro, Movistar, CNT.'
PREG5	G1P6. 6. Si la respuesta a la pregunta 2 fue SI, ¿cuál es la velocidad de su proveedor de Internet? Use el siguiente enlace para obtener la velocidad: CLIC AQUÍ (grupos)	object	0	'<=5', '<=30', '<=20', '<=10', '<=2', '<=0.5', '<=3', '<=1', '<=3.5', '<=4', '<=50', '(En blanco)', '<=40', '<=2.5', '<=1.5', '>50'
PREG6	G1P7. 7. En la ubicación indicada en la pregunta 1, ¿dispone de computador?	object	0	Si', 'No'
PREG7	G1P9. 9. Si la respuesta a la pregunta 7 fue SI, ¿Dicho computador es de uso exclusivo (es decir, no lo comparte con padres, hermanos, etc.)?	object	0	Si', 'No'
PREG8	G1P11[SQ001]. 11. Si la respuesta a la pregunta 7 fue NO, ¿Tiene algún otro dispositivo para poder acceder a clases de forma virtual? [a) Tengo celular de uso exclusivo]	object	5774	[nan, 'No', 'Si']
PREG9	G1P11[SQ003]. 11. Si la respuesta a la pregunta 7 fue NO, ¿Tiene algún otro dispositivo para poder acceder a clases de forma virtual? [c) Tengo celular, pero lo comparto con mi familia]	object	5774	[nan, 'No', 'Si']

PREG10	G1P11[SQ004]. 11. Si la respuesta a la pregunta 7 fue NO, ¿Tiene algún otro dispositivo para poder acceder a clases de forma virtual? [d] Tengo tablet, pero la comparto con mi familia]	object	5774	[nan, 'No', 'Si']
PREG11	G1P11[SQ001]. 11. Si la respuesta a la pregunta 7 fue NO, ¿Tiene algún otro dispositivo para poder acceder a clases de forma virtual? [a] Tengo celular de uso exclusivo]	object	5774	[nan, 'No', 'Si']
PREG12	G1P11[SQ002]. 11. Si la respuesta a la pregunta 7 fue NO, ¿Tiene algún otro dispositivo para poder acceder a clases de forma virtual? [b] Tengo tablet de uso exclusivo]	object	5774	[nan, 'No', 'Si']
PREG13	G1P11[SQ003]. 11. Si la respuesta a la pregunta 7 fue NO, ¿Tiene algún otro dispositivo para poder acceder a clases de forma virtual? [c] Tengo celular, pero lo comparto con mi familia]	object	5774	[nan, 'No', 'Si']
PREG14	G1P11[SQ004]. 11. Si la respuesta a la pregunta 7 fue NO, ¿Tiene algún otro dispositivo para poder acceder a clases de forma virtual? [d] Tengo tablet, pero la comparto con mi familia]	object	5774	[nan, 'No', 'Si']
PREG15	G1P12. 12. Sin tener que salir del sitio indicado en el punto 1, y en caso de que NO tenga Internet, ¿puede acceder a Internet de alguna otra forma?	object	0	[nan, 'No', 'Si']

### Anexo III – Proceso de limpieza de datos

El paso inicial para empezar con la limpieza es calcular la edad del estudiante a partir de su fecha de nacimiento para lo cual se utilizaron los siguientes comandos.

#### **#Obtenemos la fecha actual para el cálculo de la edad**

```
now = pd.Timestamp('now')
```

#### **#se genera una copia del data frame para trabajar con la fecha**

```
df=df1
```

```
df['FecNacimiento'] = pd.to_datetime(df['FecNacimiento'], format='%m%d%y') # 1
```

#### **# comando que resta la fecha actual de la fecha de nacimiento**

```
df['FecNacimiento'] = df['FecNacimiento'].where(df['FecNacimiento'] < now,  
df['FecNacimiento'] - np.timedelta64(100, 'Y')) # 2
```

#### **#cálculo de la edad e impresión en el df\n",**

```
df['Edad'] = (now - df['FecNacimiento']).astype('<m8[Y]') # 3
```

```
df['Edad'].unique()
```

```
array([23., 21., 22., 20., 32., 24., 27., 33., 25., 28., 30., 36., 26., 29., 52., 35., 68., 59., 58., 53., 60., 1  
9., 49., 37., 39., 40., 34., 31., 41., 38., 42., 44., 47., 43., 45., 51., 46., 55., 61., 57., 18., 54., 64., 48.,  
62., 75., 69., 67., 71., 66., 65., 56., 50., 63., 86., 73.]
```

#### **REEMPLAZAR VALORES VACIOS CON NAN**

Dentro del proceso de limpieza se realiza el cambio de datos nulos o repetidos con NAN para su tratamiento posterior

#### **#Se han encontrado valores de estado civil con valores en blanco, por ello se ha establecido que esos valores sean reemplazados por nan**

```
df1['EstadoCivil']=df1['EstadoCivil'].replace(' ', np.nan)
```

```
df1['EstadoCivil'].unique()
```

```
array(['S', 'C', 'D', 'V', 'U', nan], dtype=object)
```

#### **#cambiamos los valores de no registra por nan**

```
df1['Etnia']=df1['Etnia'].replace('NO REGISTRA', np.nan)
```

```
df1['Etnia'].unique()
```

```
array(['MESTIZO/A', nan, 'INDÍGENA', 'OTRO',  
'AFROECUATORIANO/AFRODESCENDIENTE', 'BLANCO/A', 'MULATO/A',  
'MONTUBIO/A', 'NEGRO/A'], dtype=object)
```

#### **#Se categoriza a los países diferentes de Ecuador como extranjeros**

```
df1.loc[df1.Pais != 'ECUADOR', 'Pais'] = 'EXTRANJERO'
```

```
df1['Pais'].unique()
array(['EXTRANJERO', 'ECUADOR'], dtype=object)
```

**#Categorizamos a las provincias que no tienen registros como provincia extranjera. se toma en consideración que los datos que no tienen registros se deben a que no son del país**

```
df1['Provincia']=df1['Provincia'].replace('SIN REGISTRO','PROVINCIA_EXTRANJERO')
```

```
df1['Provincia'].unique()
array(['PROVINCIA_EXTRANJERO', 'CAÑAR', 'EL ORO', 'AZUAY', 'PICHINCHA', 'TUNGURAHUA',
'MORONA SANTIAGO', 'CARCHI', 'IMBABURA', 'CHIMBORAZO', 'GUAYAS', 'SUCUMBIOS', 'CO
TOPAXI', 'PASTAZA', 'SANTO DOMINGO DE LOS TSACHILAS', 'ORELLANA', 'LOJA', 'BOLIVAR',
GALAPAGOS', 'ZAMORA CHINCHIPE', 'ESMERALDAS', 'MANABI', 'LOS RIOS', 'SANTA ELENA',
'NAPO'], dtype=object)
```

**#Se categoriza a los estudiantes que no han registrado valores de discapacidad como SIN\_DISCAPACIDAD**

```
df1['Discapacidad']=df1['Discapacidad'].replace(np.nan,'SIN_DISCAPACIDAD')
```

```
df1['Discapacidad'].unique()
array(['SIN_DISCAPACIDAD', 'AUDITIVA', 'VISUAL', 'FISICA', 'MENTAL PSICOSOCIAL',
'INTELECTUAL', 'LENGUAJE'], dtype=object)
```

**# Los estudiantes que no cuentan con ningún tipo de discapacidad cuentan con valores nan por lo que se los categoriza con el valor de 0**

```
df1['Discapacidad.1']=df1['Discapacidad.1'].replace(np.nan,0)
```

```
df1['Discapacidad.1'].unique()
array([ 0., 32., 58., 39., 45., 72., 30., 80., 35., 41., 42., 51.,
36.,46., 37., 70., 88., 64., 40., 43., 60., 61., 34., 62., 47., 77
.])
```

**# Debido a que hay gran variedad de datos en el porcentaje de discapacidad se ha seccionado los registros en 5 categorías.**

**#funcion para categorizar a estudiantes según su nivel de discapacidad en 5 categorías diferentes**

```
def conditions(x):
    if x >= 0 and x < 20:
        return "0-20"
    elif x >= 21 and x < 40:
        return "21-40"
    elif x >= 41 and x < 60:
        return "41-60"
    elif x >= 61 and x < 80:
        return "61-80"
    else:
        return "81-100"
```

**#Se ejecuta la función en base a las condiciones establecidas anteriormente y se asigna los valores en la columna correspondiente**

```
func = np.vectorize(conditions)
NivelDiscapacidad = func(df1["Discapacidad.1"])
df1["NivelDiscapacidad"] = NivelDiscapacidad
```

```
df1['NivelDiscapacidad'].unique()
array(['0-20', '21-40', '41-60', '61-80', '81-100'], dtype=object)
```

**#Se realiza la eliminación de espacios vacíos en campos de titulo de colegio y posteriormente se reemplaza valores N/E y NO SE OTORGA TITULACION por NAN**

```
df1['TituloCol']=df1['TituloCol'].replace('N/E',np.nan)
df1['TituloCol']=df1['TituloCol'].replace('N/E',np.nan)
df1['TituloCol']=df1['TituloCol'].replace('NO SE OTORGA TITULACION',np.nan)
```

**#Se cambia los valores categorizados como NO ESPECIFICA por nan**

```
df1['TipoUnidadEducativa']=df1['TipoUnidadEducativa'].replace('NO ESPECIFICA',np.nan)
```

```
df1['TipoUnidadEducativa'].unique()
array(['FISCAL', 'EXTRANJERO', 'PARTICULAR LAICO', 'PARTICULAR RELIGIOSO', 'FISCOMISIONAL', 'PARTICULAR', nan, 'MUNICIPAL'], dtype=object)
```

**# Dentro de ProvinciaUnidadEducativa se cambia los valores categorizados como N/E por nan**

```
df1['ProvinciaUnidadEducativa']=df1['ProvinciaUnidadEducativa'].replace('N/E',np.nan)
df1['ProvinciaUnidadEducativa'].unique()
array(['PICHINCHA', nan, 'ESMERALDAS', 'IMBABURA', 'CAÑAR', 'AZUAY', 'CHIMBORAZO', 'COTOPAXI', 'MORONA SANTIAGO', 'CARCHI', 'LOS RIOS', 'PASTAZA', 'NAPO', 'SUCUMBIOS', 'SANTO DOMINGO DE LOS TSACHILAS', 'GUAYAS', 'TUNGURAHUA', 'ORELLANA', 'BOLIVAR', 'SANTA ELENA', 'EL ORO', 'GALAPAGOS', 'LOJA', 'MANABI', 'ZAMORA CHINCHIPE'], dtype=object)
```

**#categorizacion de ingreso económico de estudiantes \n",**

**#funcion para categorizar ingreso de estudiantes en categorías con rangos de un salario mínimo vital\n"**

```
def conditions1(x):
    if x >= 0 and x < 400:
        return "0-400"
    elif x >= 401 and x < 800:
        return "401-800"
    elif x >= 801 and x < 1200:
        return "801-1200"
    elif x >= 1201 and x < 1600:
        return "1201-1600"
    elif x >= 1601 and x < 2000:
```



```

    return "1601-2000"
elif x >= 2001 and x < 2400:
    return "2001-2400"
elif x >= 2401 and x < 2800:
    return "2401-2800"
elif x >= 2801 and x < 3200:
    return "2801-3200"
else:
    return "Superior a 3201"

```

**#Se ejecuta la función en base a las condiciones establecidas anteriormente y se asigna los valores en la columna correspondiente**

```

func = np.vectorize(conditions1)
NivelIngreso1 = func(df1["Ingreso"])
array(['0-400', '0-400', 'Superior a 3201', ..., 'Superior a 3201', 'Superior a 3201',
'Superior a 3201'], dtype='<U15')
df1["NivelIngreso"] = NivelIngreso1

```

```

df1["NivelIngreso"].unique()
array(['0-400', 'Superior a 3201', '401-800', '2801-3200', '2001-2400', '801-1200',
'1201-1600', '1601-2000', '2401-2800'], dtype=object)

```

**#campo TrabajaEstudiante validar cuantos son SI para definir si se puede imputar nos Nan = NO**

```

freqTrabajaEstudiante = df1.groupby(['TrabajaEstudiante']).count()
print(freqTrabajaEstudiante)

```

```

df1["TrabajaEstudiante"].unique()
array(['NO', nan, 'SI'], dtype=object)
df1["TrabajaEstudiante"].isna().sum()
out: 17010

```

**#se puede verificar que los valores de NaN representan casi las 90% de los datos por lo que el atributo deberá ser descartado por falta de datos nivelados.**

```

df1["TrabajaEstudiante"].value_counts(dropna=False)

```

```

NaN      17010
NO        8416
SI         914
Name: TrabajaEstudiante, dtype: int64

```

**#campo IngresoEstudiante, se debe confirmar la similitud con el campo ingreso, se hace uso de la misma función con los mismos rangos de sueldos**

**#Se ejecuta la función en base a las condiciones establecidas anteriormente y se asigna los valores en la columna correspondiente**

```

func = np.vectorize(conditions1)

```

```
NivelIngreso2 = func(df1["IngresoEstudiante"])
df1["NivelIngresoEstudiante"] = NivelIngreso2
```

```
df1["NivelIngresoEstudiante"]
```

```
0    0-400
```

```
1    0-400
```

```
2    0-400
```

```
3    0-400
```

```
4    0-400
```

```
...
```

```
26335 0-400
```

```
26336 0-400
```

```
26337 0-400
```

```
26338 0-400
```

```
26339 0-400
```

```
Name: NivelIngresoEstudiante, Length: 26340, dtype: object
```

```
# campo RelLaboralEstudiante
```

```
#Se cambia los valores categorizados como DEPENDENCIA por DEPENDIENTE
```

```
df1['RelLaboralEstudiante']=df1['RelLaboralEstudiante'].replace('DEPENDENCIA','DEPENDIENTE')
```

```
df1['RelLaboralEstudiante'].unique()
```

```
array([nan, 'INDEPENDIENTE', 'DEPENDIENTE'], dtype=object)
```

```
df1['RelLaboralEstudiante'].value_counts(dropna=False)
```

```
NaN      22691
```

```
DEPENDIENTE  2317
```

```
INDEPENDIENTE  1332
```

```
Name: RelLaboralEstudiante, dtype: int64
```

```
# Campo IngresoConyugue
```

```
#Se ejecuta la función en base a las condiciones establecidas anteriormente y se asigna los valores en la columna correspondiente
```

```
func = np.vectorize(conditions1)
```

```
NivelIngreso3 = func(df1["IngresoConyugue"])
```

```
df1["NivelIngresoConyugue"] = NivelIngreso3
```

```
# campo RelLaboralPadre
```

```
#Se cambia los valores categorizados como DEPENDENCIA por DEPENDIENTE
```

```
df1['RelLaboralPadre']=df1['RelLaboralPadre'].replace('DEPENDENCIA','DEPENDIENTE')
```

```
df1['RelLaboralPadre'].unique()
```

```
array(['INDEPENDIENTE', 'DEPENDIENTE', nan], dtype=object)
```

```
# campo RelLaboralMadre
```

```
#Se cambia los valores categorizados como DEPENDENCIA por DEPENDIENTE
```

```
df1['RelLaboralMadre']=df1['RelLaboralMadre'].replace('DEPENDENCIA','DEPENDIENTE')
```

```
df1['RelLaboralMadre'].unique()
```

```
array(['INDEPENDIENTE', nan, 'DEPENDIENTE'], dtype=object)
```

In []:

### #Campo Casadounion

```
df1['Casadounion'].value_counts(dropna=False)
```

out:

```
NaN      17010
NO        9140
SI         190
```

```
Name: Casadounion, dtype: int64
```

### #campo EstadoEstudiante

**#Se cambia los valores categorizados como DEPENDENCIA por DEPENDIENTE**

```
df1['EstadoEstudiante']=df1['EstadoEstudiante'].replace(np.nan,'EstudianteRegular')
```

```
df1['EstadoEstudiante'].unique()
```

```
array(['EstudianteRegular', 'Graduado', 'Tramite Graduación-Anillados', 'ComplexivoAnt'], dtype=object)
```

### # Borrado de columnas que no se han considerado en la limpieza"

**#respaldo de los cambios realizados en df1**

```
dfbk1=df1
```

```
dfbk1.drop(['Canton', 'Parroquia', 'Telefono','Celular','Correo','CorreoIns','PaisUnidadEducativa','CantonUnidadEducativa','Begin_YR','Begin_Career_Date','PeriodoUltMatriculaCarr','TipoUltMatriculaCarr','FecUltMatriculaCarr','FormaPagUltMatriculaCarr','PagUltMatriculaCarr','CredAcum','CredRef','OcupacionEstudiante','OcupacionConyugue','AfiliadoEstudiante','TrabajaConyugue','RelLaboralConyugue','OcupacionPadre','IngresoPadre','AfiliadoConyugue','OcupacionMadre','IngresoMadre','TrabajaHermano','RelLaboralHermano','IngresoHermano','AfiliadoHermano','TieneHijos','TieneHijosMenores','HijosMenores','DiscapacidadFamiliar','NroDiscapacidad','DiscapacidadTotal','Propiedad','AvalúoPropiedad','NegocioPropio','Impuesto','VehículoPublico','AvalúoVehiculo','Rebaja'], axis=1)
```

```
df1=dfbk1
```

### Segundo dataframe: Notas

```
df2.isnull().sum()
```

```
CodigoEnc\t      0
Periodo          711
\tComat           0
Materia           0
Nrom              0
Calif            65317
Créditos          0
Aprobacion       711
Carrera           0
dtype: int64
```

```
df2.info(verbose=True)
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 1035832 entries, 0 to 1035831
```

```
Data columns (total 9 columns):
```

```

CodigoEnc      1035832 non-null object
Periodo      1035121 non-null object
             Comat      1035832 non-null object
Materia      1035832 non-null object
Nrom         1035832 non-null object
Calif        970515 non-null object
Créditos     1035832 non-null object
Aprobacion   1035121 non-null object
Carrera      1035832 non-null object
dtypes: object(9)
memory usage: 71.1+ MB

```

### #cantidad de atributos columnas

```

len(df2.axes[1])
9

```

### #cantidad de registros almacenados

```

len(df2.axes[0])
1035832

```

### #renombramos las columnas con caracteres especiales en nombre

```

df2.rename(columns={'CodigoEnc\t':'CodigoEnc'},inplace=True)

```

```

df2.rename(columns={'\tComat':'Comat'},inplace=True)

```

### #Cambiamos el tipo de dato para poder realizar operaciones matemáticas

```

df2['Calif'] = pd.to_numeric(df2['Calif'], errors='coerce')

```

```

df2['Calif'].dtypes
dtype('float64')

```

**#se usa esta función sacar la el promedio de las notas por cada estudiante, recordando que CodigoEnc es la llave primaria del estudiante, se realiza un promedio de todas las materias y en todos los niveles que ha cursado el estudiante**

```

df2.groupby("CodigoEnc").agg("mean").sort_values("CodigoEnc",ascending =
False).reset_index()

```

### #se eliminan atributos que no son relevantes para la investigación

```

df2.drop (['Periodo', 'Comat', 'Materia','Nrom', 'Aprobacion', 'Aprobacion', 'Carrera'],
axis=1)

```

### #Usamos una función para validar los valores máximos y mínimos calculados en el dataframe

```

df2.describe().transpose()

```

	count	mean	std	min	25%	50%	75%	max
Calif	970514.0	24.644584	10.654585	0.0	24.0	28.6	31.8	141.2

### #asignamos los valores de la consulta con la media

```
df2=df2.groupby("CodigoEnc").agg("mean").sort_values("CodigoEnc",ascending =
False).reset_index()
```

### **Tercer dataframe: Encuesta TICs**

```
len(df3.axes[1]) #cantidad de atributos
```

```
16
```

```
df3.info(verbose=True)
```

```
out: <class 'pandas.core.frame.DataFrame'> RangeIndex: 6482 entries, 0 to 6481 Data
columns (total 16 columns): CodigoEnc 6482 non-null object PREG1 6482 non-null
object PREG2 6482 non-null object PREG3 6482 non-null object PREG4 6482 non-null
object PREG5 6482 non-null object PREG6 6482 non-null object PREG7 6482 non-null
object PREG8 708 non-null object PREG9 708 non-null object PREG10 708 non-null
object PREG11 708 non-null object PREG12 708 non-null object PREG13 708 non-null
object PREG14 708 non-null object PREG15 6482 non-null object dtypes: object(16)
memory usage: 810.4+ KB
```

**#En las preguntas de la encuesta de nivel de acceso a las TICs se cuenta con preguntas concatenadas que condicionan las respuestas a posteriores preguntas, como es el caso de la pregunta 3**

Pregunta 3.- G1P2. 2. En la ubicación que indicó en el punto anterior, ¿Tiene servicio de Internet?

**#Una respuesta afirmativa a esta pregunta condiciona las posteriores preguntas, es el caso que si el estudiante responde 'NO' no se habilita la pregunta 4 por lo que se realiza el reemplazo del valor '-' por N/A no aplica**

Pregunnta 4.- G1P3. 3. Si la respuesta a la pregunta 2 fue SI, ¿de qué tipo es su servicio de Internet?

**En el caso de que no tenga valores afirmativos de la pregunta anterior actualmente se encuentra registrado el valor '-' por lo que se procede a cambiarlo por 'N/A'**

Pregunta 5.- G1P6. 6. Si la respuesta a la pregunta 2 fue SI, ¿cuál es la velocidad de su proveedor de Internet?. Use el siguiente enlace para obtener la velocidad: [CLIC AQUÍ](#) (grupos)

**En esta pregunta si en la pregunta 2 se tiene el valor 'NO' solo se almacenan los valores <=0.5 y (En blanco) por lo que se cambia el valor (En blanco) por 'N/A'**

Pregunta 7.- G1P9. 9. Si la respuesta a la pregunta 7 fue SI, ¿Dicho computador es de uso exclusivo (es decir, no lo comparte con padres, hermanos, etc.)?

**Esta pregunta esta correlacionada con la pregunta 6, si la respuesta a la pregunta 6 es 'NO' no se habilita la pregunta 7 para responder, por lo que en la base de datos se ha registrado '-' para propósitos de análisis de datos se realiza el cambio de '-' por 'N/A'**

Pregunta 15.-G1P12. 12. Sin tener que salir del sitio indicado en el punto 1, y en caso de que NO tenga Internet, ¿puede acceder a Internet de alguna otra forma?

**La pregunta 15 guarda una relación directa con la pregunta 3 se habilita únicamente al tener una respuesta 'NO', en el caso de tener una respuesta positiva se ha almacenado '-', el cual se procede con el cambio por 'N/A'**

**#En la encuesta de nivel de acceso a los tics se ha identificado la PREGUNTA 5, la cual depende directamente de la PREGUNTA 3, en el sentido de**

```
df['PREG4']=df['PREG4'].replace('-', 'N/A')
```

```
array(['Plan de Datos Fijos a través de operadores como Netlife, CNT, TVCable, entre otros.', 'N/A', 'Plan de Datos Móviles a través de operadores de telefonía móvil como Claro, Movistar, CNT.'], dtype=object)
```

**#Pregunta 7.- G1P9. 9. Si la respuesta a la pregunta 7 fue SI, ¿Dicho computador es de uso exclusivo (es decir, no lo comparte con padres, hermanos, etc.)?**

**#Esta pregunta esta correlacionada con la pregunta 6, si la respuesta a la pregunta 6 es 'NO' no se habilita la pregunta 7 para responder, por lo que en la base de datos se ha registrado '-'**

**#para propósitos de análisis de datos se realiza el cambio de '-' por 'N/A'**

```
df['PREG7']=df['PREG7'].replace('-', 'N/A')
```

```
df['PREG7'].unique()
```

```
array(['No', 'Si', 'N/A'], type=object)
```

**#En la encuesta de nivel de acceso a las tics se ha identificado la PREGUNTA 5, la cual depende directamente de la PREGUNTA 3, en el sentido de**

```
df['PREG5']=df['PREG5'].replace('(En blanco)', 'N/A')
```

```
df['PREG5'].unique()
```

```
array(['<=5', '<=10', '<=2.5', '<=0.5', '<=20', '<=3', '<=40', '<=3.5', '<=1.5', '<=4', '>50', 'N/A', '<=1', '<=2', '<=30', '<=50'], dtype=object)
```

```
df['PREG15']=df['PREG15'].replace('-', 'N/A')
```

```
array(['N/A', 'Si', 'No'], dtype=object)
```

## Anexo IV– Integración de datos

**#Se unen los dataframes por su llave primaria**

```
m1=pd.merge(df1,df3, on='CodigoEnc', how='outer')
```

**#Se unen los dataframes por su llave primaria**

```
m2=pd.merge(m1,df2, on='CodigoEnc', how='outer')
```

**# cantidad de atributos**

```
len(m2.axes[1])
```

```
out: 144
```

**# cantidad de registros del df"**

```
len(m2.axes[0])
```

```
out: 26341
```

```
df=m2
```