



Índice

1. Descripción de la temática de los datos	2
2. Hipótesis	3
2.1. Comentarios:	3
3. Alcance del proyecto	3
4. Usuario final y nivel de aplicación	4
5. Diagrama entidad-relación	4
6. Tablas contenidas en la base de datos	5
6.1. Tabla 1: Estudiantes Saber Pro Info	6
6.2. Tablas 2 a 5: SaberPro 2019 a SaberPro 2022	6
7. Listado de columnas de cada tabla	7
8. Análisis preliminar de datos	9
9. CONCLUSIONES	10

ANÁLISIS DE RESULTADOS PRUEBAS SABER-PRO 2022 EN RELACIÓN AL EFECTO DE LA PANDEMIA, EL ACCESO A INTERNET Y FACTORES SOCIO-ECONÓMICOS

Resumen

La calidad de la educación es una preocupación de carácter prioritario en las políticas de gobierno a nivel mundial y, en particular, en los países en vía de desarrollo, especialmente en la situación de profunda disrupción (disrupción positiva) que se ha dado debido pandemia de **COVID** que llevó a un proceso acelerado de la introducción y desarrollo del *e-learning* y las *TICS* innovadoras. No obstante, en países como Colombia, el acceso a estos recursos se vio profundamente limitado [1] en regiones del país correspondientes a centros poblados y regiones rurales dispersas¹. Según reportes del **DANE**², en estos territorios el porcentaje de hogares que poseían computador de escritorio o dispositivos similares es de tan solo el 9,7 %, comparado con el 46,3 % en las cabeceras, mientras que en relación a conexión a internet para el mismo periodo documentado correspondía al 60,5 % del total nacional, concentrado en el 70,0 % en cabeceras y tan solo el 28,8 % en centros poblados y rurales³. En este orden de ideas, la educación en Colombia enfrentó un escenario donde el desempeño y desarrollo de competencias de los estudiantes se pudo ver expuesto a una “región de caos” lo que llevó a un nuevo status quo donde se están incorporando de forma nativa nuevos paradigmas de educación y evaluación que abrazan la educación online y las inteligencias artificiales. En este contexto, el presente proyecto de analítica de datos pretende realizar un estudio descriptivo de la incidencia de la variables socioeconómicas y sociodemográficas en las pruebas de calidad de la educación superior conocidas como “**SaberPro**” durante el periodo que abarca la pandemia y la pospandemia.

1. Descripción de la temática de los datos

La base de datos que se presenta en esta propuesta corresponde a los datos del ministerio de educación de Colombia, conocido como **Icfes**, entidad encargada de la evaluación de la calidad de la educación en todos los niveles a lo largo del territorio nacional y asimismo de la realización de investigaciones relacionadas con el desarrollo y mejoramiento de la educación en el país [2] de los resultados en las competencias generales y específicas en la prueba **SaberPro** del primer semestre del año 2022⁴. La base de datos contiene información detallada de los estudiantes y campos asociados a información socioeconómica y sociodemográfica de los mismos, salvo sus nombres que han sido codificados por ser información sensible. A continuación se detallan algunos aspectos de la base de datos:

1. La base de datos corresponde a los resultados en competencias genéricas en la prueba **Saberpro** entre los periodos 2019 a 2022.

¹A estas regiones se les conoce como “Colombia profunda”

²Departamento Administrativo Nacional de Estadísticas

³Es importante mencionar que, de acuerdo al reporte del **DANE**, la conexión en zonas rurales está principalmente concentrada en conexión móvil, la cual no es competitiva para el óptimo acceso al *e-learning*

⁴La prueba **SaberPro** se aplica hoy en día como requisito a estudiantes de pregrado de instituciones de educación superior que tienen un número de inscritos superior a 1000 estudiantes y evalúa competencias generales (lectura crítica, razonamiento cuantitativo, comunicación escrita, inglés y competencias ciudadanas) y otro set de competencias, denominadas como “específicas”

2. Las dimensiones de las tablas de la base de datos están entre los 92 campos y 98561 registros y los 97 campos y 260757 registros.
3. Los tipos de registros contenidos en la base de datos son: cadenas de texto (strings), fechas, y valores numéricos tanto tipo *int*, como tipo *float*.
4. La fuente de los datos es la página del **Icfes** DataIcfes

2. Hipótesis

El presente proyecto busca estudiar y comprender los efectos en los resultados obtenidos por estudiantes de instituciones de educación superior en las pruebas **Saberpro** respecto a su acceso a recursos tecnológicos y de conectividad a lo largo de la pandemia de **Covid** y la pospandemia, partiendo del supuesto que dichos recursos fueron o son más limitados en territorios nacionales conocidos como centros poblados o zonas rurales. En este orden de ideas, se usarán como parámetros de comparación la información socioeconómica y sociodemográfica de los estudiantes que presentaron la prueba entre los periodos 2019 a 2022 con el fin de establecer un primer acercamiento al estudio del fenómeno.

2.1. Comentarios:

- A continuación se comparte el link a la base de datos de los resultados en competencias generales de la prueba **Saberpro** entre los periodos 2019 a 2022 a ser usada en el proyecto: Base de datos proyecto JLSR. El mismo link se comparte en la entrega de la tarea.
- Aunque el objetivo a mediano.largo plazo es poder hacer un análisis de evolución temporal, considerando datos previos a la pandemia (2019-2020), durante la pandemia (2020-2022) y lo que se puede denominar como actualidad o pospandemia, se realizará inicialmente un análisis restringido entre los periodos 2019 a 2022 para establecer criterios de análisis y comparaciones locales.

3. Alcance del proyecto

El proyecto de analítica de datos tiene como objetivo principal hacer una descripción detallada de los resultados obtenidos por los estudiantes en las pruebas SABER PRO realizadas en Colombia entre los años 2019 y 2022. Además, busca plantear explicaciones hipotéticas acerca del impacto que pudo haber tenido el haber estado expuesto a una apropiada transición a la educación en modalidad remota.

En particular, el proyecto se enfocará en estudiar las variables socioeconómicas y sociodemográficas que pueden influir en los resultados de las pruebas, con un enfoque especial en la comparación de los resultados obtenidos por estudiantes de zonas rurales de la Colombia profunda y estudiantes de cascos urbanos y ciudades principales. También se analizará el acceso a recursos tecnológicos durante la pandemia de COVID-19, para determinar si esto ha afectado los resultados de las pruebas.

Se espera que los resultados del proyecto proporcionen información valiosa sobre cómo las diferencias socioeconómicas y sociodemográficas pueden influir en el rendimiento académico de los estudiantes en Colombia, y cómo el acceso a recursos tecnológicos durante la pandemia de COVID-19 ha afectado los resultados de las pruebas. Los hallazgos podrían ser útiles para el diseño de políticas públicas y programas educativos que busquen reducir las brechas de desigualdad entre los estudiantes de diferentes regiones del país.

En resumen, el proyecto de analítica de datos tiene como objetivo describir los resultados obtenidos por los estudiantes en las pruebas SABER PRO y plantear hipótesis sobre el impacto de la transición a la educación en modalidad remota. A través de la aplicación de técnicas avanzadas de análisis de datos, se espera obtener información valiosa que pueda ser utilizada para mejorar la calidad de la educación y reducir las brechas de desigualdad en Colombia.

4. Usuario final y nivel de aplicación

El presente proyecto de análisis de datos tiene como objetivo proporcionar información relevante y útil para futuros estudios y estrategias enfocadas en mejorar el desarrollo académico de estudiantes provenientes de regiones rurales. A través del análisis de resultados y necesidades de estas poblaciones, el proyecto busca ser un insumo valioso para la formulación de políticas y programas que fomenten la inclusión y el éxito académico de estos estudiantes.

Este estudio está dirigido a instituciones de educación superior y entidades gubernamentales relacionadas con el Ministerio de Educación Nacional, con el fin de que puedan utilizar los hallazgos y recomendaciones del proyecto para informar y mejorar sus propias iniciativas y programas en este ámbito.

5. Diagrama entidad-relación

El diagrama entidad-relación para esta base de datos consta de una entidad principal llamada “Estudiantes Saber Pro Info” y cuatro entidades secundarias llamadas “SaberPro 2022”, “SaberPro 2021”, “SaberPro 2020” y “SaberPro 2019”. Cada entidad secundaria contiene información sobre los resultados de los exámenes de cada estudiante para un año específico incluyendo información sociodemográfica y socioeconómica relevante para el proyecto.

En la entidad primaria y en cada entidad secundaria, ESTU CONSECUTIVO es la clave primaria. Además, las entidades secundarias tienen otras claves foráneas como ESTU COD RESIDE-DEPTO, ESTU COD RESIDE MCPIO, ESTU AREARESIDE, ESTU PRGM ACADEMICO que permiten establecer la relación entre tablas para acceder a la información sociodemográfica de los estudiantes.

En el diagrama entidad-relación, la entidad “Estudiantes Saber Pro Info” se representa como un rectángulo con los atributos ESTU CONSECUTIVO, ESTU ESTUDIANTE, ESTU PAIS RESIDE y ESTU DEPTO RESIDE dentro. Cada entidad secundaria se representa como otro rectángulo con los atributos específicos de cada año y los atributos respectivos. Las relaciones entre las entidades se representan como líneas que conectan las claves primarias y foráneas correspondientes.

En resumen, el diagrama entidad-relación para esta base de datos muestra cómo la información sobre los resultados de los exámenes de cada estudiante se almacena en una entidad integrada con la información sociodemográfica de los estudiantes, y cómo estas entidades se relacionan entre sí mediante claves primarias y foráneas. Cada entidad secundaria representa información sobre los resultados de los exámenes de un año específico y para una población específica. El diagrama entidad-relación se dispone en la figura 1

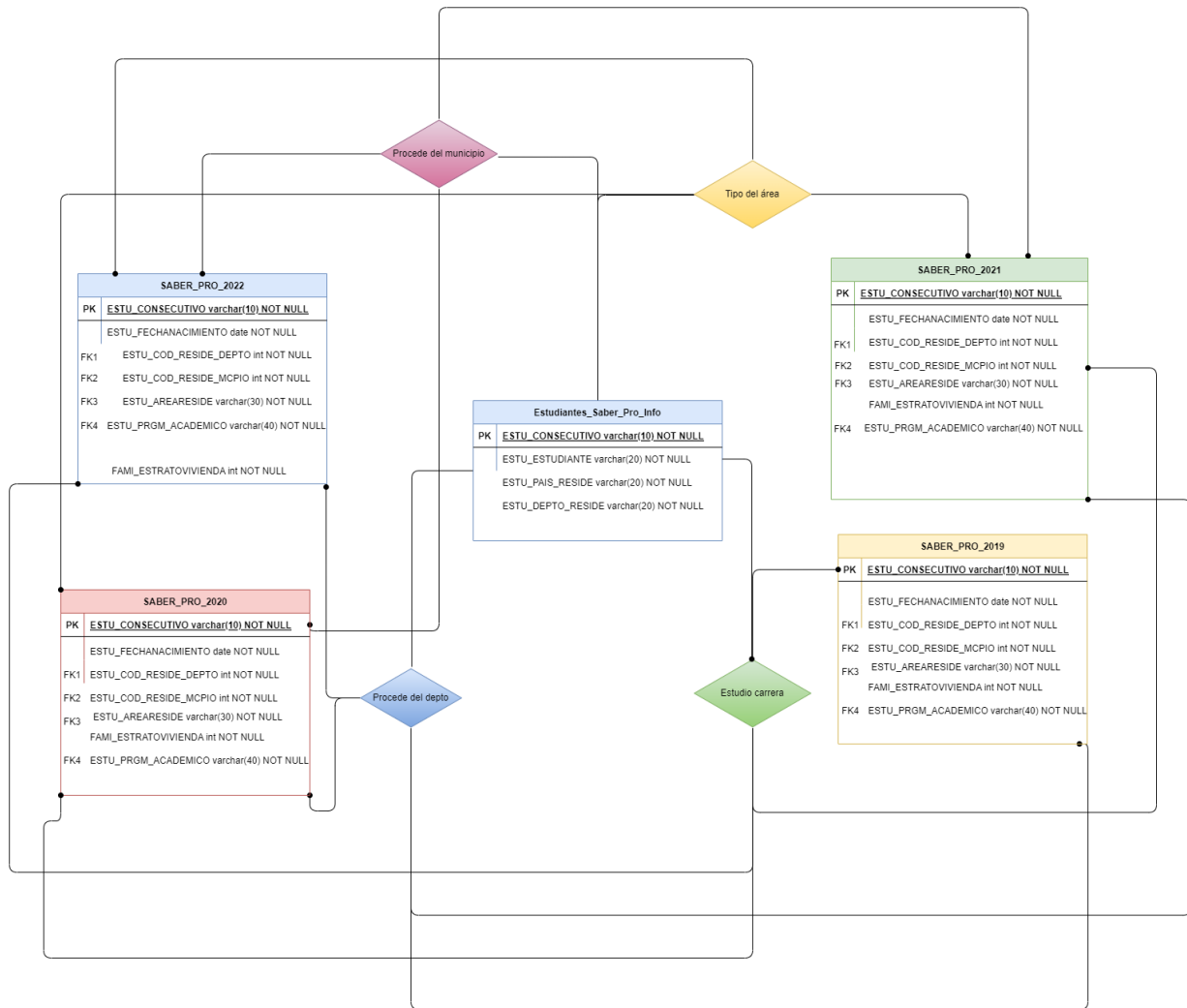


Figura 1: Diagrama entidad-objeto para la base de datos de tipo relacional que incluye las tablas de datos recopilados por el ICFES de las pruebas SABERPRO correspondientes a los periodos consecutivos entre los años 2019-2022

6. Tablas contenidas en la base de datos

La base de datos cuenta con un total de 5 tablas, de las cuales 4 son tablas oficiales emitidas por el ICFES sin modificación alguna. Estas tablas contienen información detallada sobre los resultados de la evaluación de competencias genéricas para las diferentes carreras ofrecidas por las instituciones de educación superior en el territorio nacional que cuentan con al menos 1.000 inscritos y están reconocidas por el Ministerio de Educación. Las tablas tienen entre 98 y 100

campos y entre 98,500 y 240,000 registros.

La información contenida en estas tablas es de gran importancia para analizar y comprender los resultados de las pruebas SABER PRO en Colombia. Algunas de las variables incluidas en estas tablas son: información demográfica de los estudiantes, información sobre su universidad y programa de estudio, resultados en cada una de las áreas evaluadas, puntajes totales, entre otros.

Además de estas 4 tablas, también se incluye una tabla adicional que se genera a partir de la información correspondiente al país de residencia de los estudiantes, el departamento de residencia y el estado del estudiante al momento de presentar la prueba (si es un estudiante regular o se encuentra en una condición diferente). Esta tabla adicional puede ser de gran utilidad para analizar cómo los factores geográficos y el estado del estudiante influyen en los resultados de las pruebas SABER PRO.

En resumen, las tablas que se disponen en este proyecto, contienen datos importantes sobre la información demográfica y de desempeño de los estudiantes, información sobre sus universidades y programas de estudio, y resultados en cada una de las áreas evaluadas. La tabla adicional que se genera a partir de la información geográfica y el estado del estudiante puede ayudar a analizar cómo estos factores influyen en los resultados de las pruebas.

6.1. Tabla 1: Estudiantes Saber Pro Info

Esta es una tabla adicional que fue generada con datos extraídos de las cuatro tablas oficiales de los resultados en competencias genéricas de la prueba SaberPro en los años consecutivos entre el 2019 y el 2022. La tabla fue integrada debido a la necesidad practica del análisis que se pretende hacer, pues se busca establecer una base de datos de tipo relacional para “un mismo set de datos” que se mide de manera anual con poblaciones diferentes. Dado que cada una de las 4 tablas oficiales consigna y recopila información socioeconómica y sociodemográfica sin requerir a tablas adicionales para hacer los registros, se evito hacer una división en tablas adicionales a cada una de las tablas oficiales dispuestas y se optó por crear esta tabla principal para establecer relaciones entre la misma y cada una de las correspondientes a cada uno de los años incluidos en el estudio. Los campos de la tabla en cuestión se presentan a continuación:

NOMBRE DEL CAMPO	TIPO DE REGISTRO
ESTU_CONSECUTIVO	PK varchar(10) NOT NULL
ESTU_ESTUDIANTE	varchar(10) NOT NULL
ESTU_PAIS_RESIDE	varchar(10) NOT NULL
ESTU_DEPTO_RESIDE	varchar(10) NOT NULL

Cuadro 1: Listado de campos de la tabla Estudiantes Saber Pro Info

6.2. Tablas 2 a 5: SaberPro 2019 a SaberPro 2022

Por razones de forma no se presentará el detalle de los campos de estas tablas, pues las mismas cuentan con entre 98 a 100 campos, no obstante en la figura 2 se disponen la mayoría de los mismos. La clave primaria de estas tablas coincide con la clave primaria de la tabla 1, debido a que el único ID unívoco en estos datos es el código que el ICFES asignó a los estudiantes. Las

tablas 2 a 5 continen 4 claves foráneas que permiten establecer relaciones entre las mismas y la tabla 1 (ver diagrama entidad-relación 1). A continuación se dispone un cuadro sintetizando dicha información:

CAMPOS PRINCIPALES	TIPO DE REGISTRO
ESTU	PK varchar(10) NOT NULL
ESTU_COD RESIDE_DEPTO	FK int NOT NULL
ESTU_COD.RESIDE_MCPIO	FK int NOT NULL
ESTU_AREARESIDE_	FK varchar(30) NOT NULL
ESTU_PRGM_ACADEMICO	FK varchar(40) NOT NULL

Cuadro 2: Listado de campos que constituyen claves primarias y foráneas de las tablas Saber Pro 2019 a Saber pro 2022. No se presentan todos los campos de estas tablas debido al elevado número de estos.

De es esta forma se tiene que el listado de tablas de la base de datos, es el siguiente:

NOMBRE DE LA TABLA	PK	NUMERO DE CAMPOS	NUMERO DE REGISTROS
Estudiantes Saber Pro Info	ESTU_CONSECUTIVO	4	834840
SaberPro 2022	ESTU_CONSECUTIVO	92	98566
SaberPro 2021	ESTU_CONSECUTIVO	88	229083
SaberPro 2020	ESTU_CONSECUTIVO	97	246437
SaberPro 2019	ESTU_CONSECUTIVO	71	260757

Cuadro 3: Listado de tablas incluidas en la base de datos a trabajar en el proyecto de analítica de datos.

En este listado se especifica la información relevante de las tablas en cuestión, en particular las claves primarias de las tablas y el número de campos y registros.

7. Listado de columnas de cada tabla

En esta sección se comparten algunas de las columnas presentes en la base de datos; la razón por la cual se añaden todas es simplemente de forma, debido al gran volumen de campos contenidos en las tablas. Si desea conocer todos los campos, puede acceder a la base de datos en el siguiente link [Base de datos proyecto JLSR](#) y dirigirse a la pestaña denominada “Coincidencia Campos 18 22”

SABERPRO 2022	SABERPRO 2021	SABERPRO 2020	SABERPRO 2019
ESTU_TIPODOCUMENTO	ESTU_TIPODOCUMENTO	ESTU_TIPODOCUMENTO	ESTU_TIPODOCUMENTO
ESTU_NACIONALIDAD	ESTU_NACIONALIDAD	ESTU_NACIONALIDAD	ESTU_NACIONALIDAD
ESTU_GENERO	ESTU_GENERO	ESTU_GENERO	ESTU_GENERO
ESTU_FECHANACIMIENTO	ESTU_FECHANACIMIENTO	ESTU_FECHANACIMIENTO	ESTU_FECHANACIMIENTO
ESTU_EXTERIOR	ESTU_EXTERIOR	ESTU_EXTERIOR	ESTU_EXTERIOR
PERIODO	PERIODO	PERIODO	PERIODO
ESTU_CONSECUTIVO	ESTU_CONSECUTIVO	ESTU_CONSECUTIVO	ESTU_CONSECUTIVO
ESTU_ESTUDIANTE	ESTU_ESTUDIANTE	ESTU_ESTUDIANTE	ESTU_ESTUDIANTE
ESTU_PAIS_RESIDE	ESTU_PAIS_RESIDE	ESTU_PAIS_RESIDE	ESTU_PAIS_RESIDE
ESTU_DEPTO_RESIDE	ESTU_DEPTO_RESIDE	ESTU_DEPTO_RESIDE	ESTU_DISC_FISICA
ESTU_COD_RESIDE_DEPTO	ESTU_COD_RESIDE_DEPTO	ESTU_COD_RESIDE_DEPTO	ESTU_DISC_SISTEM
ESTU_MCPIO_RESIDE	ESTU_MCPIO_RESIDE	ESTU_MCPIO_RESIDE	ESTU_DISC_AUDITCASTELL
ESTU_COD_RESIDE_MCPIO	ESTU_COD_RESIDE_MCPIO	ESTU_COD_RESIDE_MCPIO	ESTU_DISC_AUDITLSC
ESTU_AREARESIDE	ESTU_AREARESIDE	ESTU_AREARESIDE	ESTU_DISC_VISUALCEGUE
ESTU_ESTADOCIVIL	ESTU_ESTADOCIVIL	ESTU_ESTADOCIVIL	ESTU_DISC_VISUALBAJA
ESTU_COLE_TERMINO	ESTU_CODICFESCOLE_TERMINO	ESTU_CODICFESCOLE_TERMINO	ESTU_DISC_VOZHABLA
ESTU_CODDANE_COLE_TERMINO	ESTU_COLE_TERMINO	ESTU_COLE_TERMINO	ESTU_DISC_INTELEC
ESTU_COD_COLE_MCPIO_TERMINO	ESTU_CODDANE_COLE_TERMINO	ESTU_CODDANE_COLE_TERMINO	ESTU_DISC_SORDOCEGUERA
ESTU_TITULOBTENIDOBACHILLER	ESTU_COD_COLE_MCPIO_TERMINO	ESTU_COD_COLE_MCPIO_TERMINO	ESTU_DISC_PSICOSOCIAL
ESTU_VALORMATRICULAUNIVERSIDAD	ESTU_TITULOBTENIDOBACHILLER	ESTU_TITULOBTENIDOBACHILLER	ESTU_DISC_ASPPERGER
ESTU_PAGOMATRICULABECA	ESTU_VALORMATRICULAUNIVERSIDAD	ESTU_VALORMATRICULAUNIVERSIDAD	ESTU_DEPTO_RESIDE
ESTU_PAGOMATRICULACREDITO	ESTU_PAGOMATRICULABECA	ESTU_PAGOMATRICULABECA	ESTU_COD_RESIDE_DEPTO
ESTU_PAGOMATRICULAPADRES	ESTU_PAGOMATRICULACREDITO	ESTU_PAGOMATRICULACREDITO	ESTU_MCPIO_RESIDE
ESTU_PAGOMATRICULAPROPIO	ESTU_PAGOMATRICULAPADRES	ESTU_PAGOMATRICULAPADRES	ESTU_COD_RESIDE_MCPIO
ESTU_COMOCAPACITOEAMENS811	ESTU_PAGOMATRICULAPROPIO	ESTU_PAGOMATRICULAPROPIO	ESTU_AREARESIDE
ESTU_CURSODOCENTESIES	ESTU_COMOCAPACITOEAMENS811	ESTU_COMOCAPACITOEAMENS811	ESTU_COLE_TERMINO
ESTU_CURSIOESAPOYOEXTERNO	ESTU_TIPODOCUMENTOS811	ESTU_TIPODOCUMENTOS811	ESTU_OTROCOLE_TERMINO
ESTU_CURSIOESEXTERNA	ESTU_SEMESTRECURSA	ESTU_SEMESTRECURSA	ESTU_CODDANE_COLE_TERMINO
ESTU_SIMULACROTIPICFES	FAMI_EDUCACIONPADRE	FAMI_EDUCACIONPADRE	ESTU_COD_COLE_MCPIO_TERMINO
ESTU_ACTIVIDADREFUERZOAREAS	FAMI_EDUCACIONMADRE	FAMI_EDUCACIONMADRE	ESTU_VALORMATRICULAUNIVERSIDAD
ESTU_ACTIVIDADREFUERZOGENERIC	FAMI_OCUPACIONPADRE	FAMI_OCUPACIONPADRE	ESTU_PAGOMATRICULABECA
ESTU_TIPODOCUMENTOS811	FAMI_OCUPACIONMADRE	FAMI_OCUPACIONMADRE	ESTU_PAGOMATRICULACREDITO
ESTU_SEMESTRECURSA	FAMI_ESTRATOVIENDA	FAMI_ESTRATOVIENDA	ESTU_PAGOMATRICULAPADRES
FAMI_EDUCACIONPADRE	FAMI_TIENINTERNET	FAMI_TIENINTERNET	ESTU_PAGOMATRICULAPROPIO
FAMI_EDUCACIONMADRE	FAMI_TIENECOMPUTADOR	FAMI_TIENECOMPUTADOR	ESTU_COMOCAPACITOEAMENS811
FAMI_TRABAJOLABORPADRE	FAMI_TIENELAVADORA	FAMI_TIENELAVADORA	ESTU_CURSODOCENTESIES
FAMI_TRABAJOLABORMADRE	FAMI_TIENEHORNOMICROOGAS	FAMI_TIENEHORNOMICROOGAS	ESTU_CURSIOESAPOYOEXTERNO
FAMI_ESTRATOVIENDA	FAMI_TIENESERVICIOTV	FAMI_TIENESERVICIOTV	ESTU_CURSIOESEXTERNA
ESTU_OTROCOLE_TERMINO	FAMI_TIENEAUTOMOVIL	FAMI_TIENEAUTOMOVIL	ESTU_SIMULACROTIPICFES
FAMI_TIENINTERNET	FAMI_TIENEMOTOCICLETA	FAMI_TIENEMOTOCICLETA	ESTU_ACTIVIDADREFUERZOAREAS
FAMI_TIENECOMPUTADOR	FAMI_TIENCONSOAVIDEJUEGOS	FAMI_TIENCONSOAVIDEJUEGOS	ESTU_ACTIVIDADREFUERZOGENERIC
FAMI_TIENELAVADORA	FAMI_TRABAJOLABORPADRE	FAMI_TRABAJOLABORPADRE	ESTU_TIPODOCUMENTOS811
FAMI_TIENEHORNOMICROOGAS	FAMI_TRABAJOLABORMADRE	FAMI_TRABAJOLABORMADRE	ESTU_SEMESTRECURSA
FAMI_TIENESERVICIOTV	ESTU_HORASSEMANTRABAJA	ESTU_HORASSEMANTRABAJA	FAMI_EDUCACIONPADRE
FAMI_TIENEAUTOMOVIL	FAMI_CUANTOSCOMPARTEBAÑO	FAMI_CUANTOSCOMPARTEBAÑO	FAMI_EDUCACIONMADRE
FAMI_TIENEMOTOCICLETA	ESTU_PAGOMATRICULA	ESTU_PAGOMATRICULA	FAMI_ESTRATOVIENDA
FAMI_TIENCONSOAVIDEJUEGOS	ESTU_PRESENTACIONCASA	ESTU_PRESENTACIONCASA	FAMI_TIENINTERNET
FAMI_CUANTOSCOMPARTEBAÑO	INST_COD_INSTITUCION	INST_COD_INSTITUCION	FAMI_TIENECOMPUTADOR
ESTU_HORASSEMANTRABAJA	INST_NOMBRE_INSTITUCION	INST_NOMBRE_INSTITUCION	FAMI_TIENELAVADORA
INST_COD_INSTITUCION	ESTU_PRGM_ACADEMICO	ESTU_PRGM_ACADEMICO	FAMI_TIENEHORNOMICROOGAS
INST_NOMBRE_INSTITUCION	ESTU_SNIES_PRGMACADEMICO	ESTU_SNIES_PRGMACADEMICO	FAMI_TIENESERVICIOTV
ESTU_PRGM_ACADEMICO	GRUPOREFERENCIA	GRUPOREFERENCIA	FAMI_TIENEAUTOMOVIL
ESTU_SNIES_PRGMACADEMICO	ESTU_PRGM_CODMUNICIPIO	ESTU_PRGM_CODMUNICIPIO	FAMI_TIENEMOTOCICLETA
GRUPOREFERENCIA	ESTU_PRGM_MUNICIPIO	ESTU_PRGM_MUNICIPIO	FAMI_TIENCONSOAVIDEJUEGOS
ESTU_PRGM_CODMUNICIPIO	ESTU_PRGM_DEPARTAMENTO	ESTU_PRGM_DEPARTAMENTO	FAMI_CUANTOSCOMPARTEBAÑO
ESTU_PRGM_MUNICIPIO	ESTU_NIVEL_PRGM_ACADEMICO	ESTU_NIVEL_PRGM_ACADEMICO	FAMI_TRABAJOLABORPADRE
ESTU_PRGM_DEPARTAMENTO	ESTU_METODO_PRGM	ESTU_METODO_PRGM	FAMI_TRABAJOLABORMADRE
ESTU_NIVEL_PRGM_ACADEMICO	ESTU_NUCLEO_PREGRAO	ESTU_NUCLEO_PREGRAO	ESTU_HORASSEMANTRABAJA
ESTU_METODO_PRGM	ESTU_INST_CODMUNICIPIO	ESTU_INST_CODMUNICIPIO	INST_COD_INSTITUCION
ESTU_NUCLEO_PREGRAO	ESTU_INST_MUNICIPIO	ESTU_INST_MUNICIPIO	INST_NOMBRE_INSTITUCION
ESTU_INST_CODMUNICIPIO	ESTU_INST_DEPARTAMENTO	ESTU_INST_DEPARTAMENTO	ESTU_PRGM_ACADEMICO
ESTU_INST_MUNICIPIO	INST_CARACTER_ACADEMICO	INST_CARACTER_ACADEMICO	ESTU_SNIES_PRGMACADEMICO
ESTU_INST_DEPARTAMENTO	INST_ORIGEN	INST_ORIGEN	ESTU_PRGM_CODMUNICIPIO
INST_CARACTER_ACADEMICO	ESTU_PRIVADO_LIBERTAD	ESTU_PRIVADO_LIBERTAD	ESTU_PRGM_MUNICIPIO
INST_ORIGEN	ESTU_COD_MCPIO_PRESENTACION	ESTU_COD_MCPIO_PRESENTACION	ESTU_PRGM_DEPARTAMENTO
ESTU_PRIVADO_LIBERTAD	ESTU_MCPIO_PRESENTACION	ESTU_MCPIO_PRESENTACION	ESTU_NIVEL_PRGM_ACADEMICO
ESTU_COD_MCPIO_PRESENTACION	ESTU_DEPTO_PRESENTACION	ESTU_DEPTO_PRESENTACION	ESTU_METODO_PRGM
ESTU_MCPIO_PRESENTACION	ESTU_COD_DEPTO_PRESENTACION	ESTU_COD_DEPTO_PRESENTACION	ESTU_NUCLEO_PREGRAO
ESTU_DEPTO_PRESENTACION	MOD_RAZONA_CUANTITAT_PUNT	MOD_RAZONA_CUANTITAT_PUNT	ESTU_INST_CODMUNICIPIO
ESTU_COD_DEPTO_PRESENTACION	MOD_RAZONA_CUANTITAT_DESEM	MOD_RAZONA_CUANTITAT_DESEM	ESTU_INST_MUNICIPIO
MOD_RAZONA_CUANTITAT_PUNT	MOD_RAZONA_CUANTITATIVO_PNAL	MOD_RAZONA_CUANTITATIVO_PNAL	ESTU_INST_DEPARTAMENTO
MOD_RAZONA_CUANTITAT_DESEM	MOD_RAZONA_CUANTITATIVO_PNBC	MOD_RAZONA_CUANTITATIVO_PNBC	INST_CARACTER_ACADEMICO

Figura 2: Listado de tablas

8. Análisis preliminar de datos

El tratamiento de datos realizado consistió en la extracción y transformación de los datos de interés, limitando la base de datos únicamente al período correspondiente al año 2022 para los resultados de la prueba Saber Pro. Se construyeron dos tablas relacionales a partir de los datos, utilizando como clave primaria y foránea el campo “ESTU CONSECUTIVO” asignado por el ICFES.

El proceso de construcción de las tablas se llevó a cabo haciendo uso de la librería Pandas en Python, y posteriormente se exportó la base de datos acotada a Power BI para su visualización y análisis.

La primera tabla construida, llamada “student info”, corresponde a la tabla que contiene información relevante sobre los estudiantes que realizaron la prueba Saber Pro en el período 2022. Tiene los siguientes campos:

- **ID:** Identificador único para cada registro de la tabla, generado automáticamente y utilizado como clave primaria.
- **ESTU_CONSECUTIVO:** Campo que almacena el identificador asignado por el ICFES a cada estudiante que realizó la prueba Saber Pro en el período 2022. Se utiliza como clave foránea en la segunda tabla.
- **ESTU_GENERO:** Campo que indica el género del estudiante (M: Masculino, F: Femenino).
- **ESTU_DEPTO_RESIDE:** Campo que almacena el nombre del departamento donde reside el estudiante.
- **ESTU_COD_RESIDE_DEPTO:** Campo que almacena el código asignado al departamento donde reside el estudiante.
- **ESTU_MCPIO_RESIDE:** Campo que almacena el nombre del municipio donde reside el estudiante.
- **ESTU_COD_RESIDE_MCPIO:** Campo que almacena el código asignado al municipio donde reside el estudiante.
- **FAMI_TIENEINTERNET:** Campo que indica si el estudiante cuenta con acceso a internet en su hogar (S: Sí, N: No).
- **FAMI_TIENECOMPUTADOR:** Campo que indica si el estudiante cuenta con un computador en su hogar (S: Sí, N: No).

La segunda tabla construida, llamada “estudiantes”, está relacionada con la tabla “student” a través del campo “ESTU_CONSECUTIVO”. Contiene información detallada sobre los resultados obtenidos por cada estudiante en la prueba Saber Pro del período 2022, y cuenta, entre otros con los siguientes campos:

- **ESTU_CONSECUTIVO:** Campo que actúa como clave foránea y que identifica a cada estudiante en la tabla “student_info”.
- **ESTU_PRGM_ACADEMICO:** Campo que almacena el nombre del programa académico en el que está matriculado el estudiante.
- **ESTU_SNIES_PRGMACADEMICO:** Campo que almacena el código asignado al programa académico por el SNIES (Sistema Nacional de Información de la Educación Superior).

- **MOD_RAZONA_CUANTITAT_PUNT**: Campo que almacena el puntaje obtenido por el estudiante en el módulo de razonamiento cuantitativo.
- **MOD_RAZONA_CUANTITAT_DESEM**: Campo que almacena el nivel de desempeño alcanzado por el estudiante en el módulo de razonamiento cuantitativo.

Además de la construcción de las tablas relacionales a partir de los datos del período 2022 para los resultados de la prueba Saber Pro y su posterior análisis en Power BI, se realizaron medidas comparativas entre el promedio obtenido en el resultado global por estudiantes que viven en cabeceras municipales y zonas rurales. También se tuvo en cuenta el estrato al cual pertenecen y si contaban con acceso a internet y si poseían o no computador.

Estas medidas comparativas permitieron una mejor comprensión de las diferencias en el rendimiento académico de los estudiantes en relación con su lugar de residencia y su acceso a recursos tecnológicos. Los resultados obtenidos a partir de estas comparativas pueden ser útiles para identificar las áreas en las que se requiere una mayor atención y para desarrollar políticas que promuevan la igualdad de oportunidades educativas para todos los estudiantes.

9. CONCLUSIONES

1. Tras de hacer un primer acercamiento de un análisis para los resultados del saber pro-2022, se presume un impacto negativo en los promedios alcanzados en regiones por fuera del centro del país vinculado con el acceso nulo o limitado a recursos TIC y conectividad.
2. Se observan variaciones porcentuales crecientes entre los resultados obtenidos por estudiantes con acceso a recursos TIC y conectividad a partir del estrato 3: a partir del análisis presentado en este documento no es posible discernir si dicho comportamiento está asociado a la varianza debido al tamaño diferencial de las poblaciones (el volumen de estudiantes con acceso a educación superior en estratos medios y altos es mayor que en estratos bajos) o a que el poseer acceso a recursos TIC y conectividad es más diferenciador en estratos más altos.
3. Se observa una tendencia “creciente” en los promedios globales de estratos bajos a estratos altos (creciente con el estrato); no obstante no se establece un modelo o ajuste que describa la tendencia ni se establece una hipótesis que describa relación de causalidad alguna en este comportamiento.
4. Se observa una variación porcentual en el promedio global constante en los estratos 1 y 2 (como se mencionó previamente).

Referencias

- [1] Departamento Administrativo Nacional de Estadística (DANE). Boletín técnico - encuesta nacional de tecnología y ciencia (entic) - resultados principales de hogares - 2021. Technical report, DANE, 2021.
- [2] Instituto Colombiano para la Evaluación de la Educación (ICFES). Misión y visión. <https://www.icfes.gov.co/web/guest/mision-vision>. Accessed: March 4, 2023.

Marzo 06 de 2023.

Chía, Cundinamarca

Documento escrito en L^AT_EXpor:

Jorge Luis Sánchez Ruiz
Físico, Universidad Nacional de Colombia
Magíster en Ciencias-Física, Universidad de los Andes