

¿Cómo impactan los factores socioeconómicos en los resultados del ICFES Saber 11?

INTEGRANTES:

Jornay David Zuluaga Gomez
Juan Sebastián Delgado Caicedo
Cristian Jiovani Herrera Almanza
Ricardo Andrés Ibarra Bolívar

El ICFES se constituye como una institución gubernamental que ha tendido la misión de diseñar y administrar los exámenes de estado para evaluación del sistema de educación colombiana desde 1968.

Saber 11 es una prueba de tipo estandarizada que se aplica sobre la población que está culminando estudios de bachillerato, en todos los colegios del territorio nacional. Se evalúan diferentes módulos como lo son matemáticas, lectura crítica, inglés y sociales.

El propósito de este sitio es poder observar la influencia de ciertos factores socioeconómicos y demográficos sobre los resultados obtenidos por los estudiantes. Cabe resaltar que los datos mostrados en este estudio comprenden a los periodos académicos 2019 y 2020.

1) Variables base de datos.

La base de datos usada en este ejercicio muestra la información de 1.021.755 estudiantes de Colombia que presentaron la prueba Saber 11 entre los periodos de 2019 a 2010 y fueron encuestados sobre sus distintos factores demográficos, socioeconómicos y perceptivos. A continuación, se muestra un detalle de las variables usadas en este ejercicio:

- ESTU_GENERO: Género con el cual se identifica el estudiante.
- EDAD: Edad en años que tiene el estudiante al momento de presentar el examen.
- COLE_NATURALEZA: Naturaleza del colegio (Oficial, No Oficial).
- COLE_CHARACTER: Carácter del colegio (Técnico, Académico, No Aplica, Técnico/Académico).
- ESTU_DEPTO_RESIDE: Departamento en el cual asiste el estudiante al colegio.
- COLE_AREA: Área en el cual se encuentra ubicado el colegio (Urbano, Rural).
- FAMI ESTRATOVIVIENDA: Estrato asociado a la vivienda del estudiante.
- FAMI_TIENEINTERNET: La vivienda del estudiante tiene o no tiene internet.
- FAMI_TIENECOMPUTADOR: La vivienda del estudiante tiene o no tiene computador.
- ESTU_DEDICACIONDIARIALECTURA: Tiempo diario que dedica el estudiante a la lectura.
- FAMI_NUMLIBROS: Número de libros que tienen en el hogar del estudiante
- ESTU_HORASSEMANATRABAJA: Horas a la semana que el estudiante dedica a trabajos ya sean remunerados o no.

- FAMI_EDUCACIONPADRE: Máximo nivel educativo alcanzado por el padre del estudiante.
- FAMI_EDUCACIONMADRE: Máximo nivel educativo alcanzado por la madre del estudiante.
- PUNT_LECTURA_CRITICA: Puntaje entre 0 a 500 que obtuvo el estudiante en la prueba de lectura critica.
- FAMI_SITUACIONECONOMICA: Percepción del estudiante en la cual determina si se encuentran mejor, igual o peor económicamente respecto al año anterior.
- ESTU_DEDICACIONINTERNET: Tiempo diario que dedica el estudiante al uso de internet.
- COLE_JORNADA: Jornada académica en la cual asiste el estudiante al colegio (Mañana, Completa, Noche, Sabatina, Tarde, Única).
- PUNT_MATEMATICAS: Puntaje entre 0 a 500 que obtuvo el estudiante en la prueba de matemáticas.
- PUNT_C_NATURALES: Puntaje entre 0 a 500 que obtuvo el estudiante en la prueba de ciencias naturales.
- PUNT_SOCIALES_CIUDADANAS: Puntaje entre 0 a 500 que obtuvo el estudiante en la prueba de ciencias sociales.
- PUNT_INGLES: Puntaje entre 0 a 500 que obtuvo el estudiante en la prueba de lectura inglés.
- PUNT_GLOBAL: Puntaje entre 0 a 500 que obtuvo el estudiante en el total de la prueba.
- PERIODO: Periodo académico en el cual fue presentada la prueba.

2) Modelamiento datos.

En cuanto al modelamiento de datos, se utilizó el lenguaje de programación Python, específicamente las librerías de Pandas, para cargar la información de varios archivos de texto (un archivo por cada periodo de resultados) y consolidarlos en una sola base.



Para una mejor comprensión de la información se modelaron las siguientes columnas:

- FAMI_EDUCACIONPADRE: se recategorizaron las respuestas para que solo quedara la información del máximo nivel de escolaridad y así poder mejorar la legibilidad de estos datos
- FAMI_EDUCACIONMADRE: se recategorizaron las respuestas para que solo quedara la información del máximo nivel de escolaridad y así poder mejorar la legibilidad de estos datos

- PERIODO: Con el fin de hacer más fácil la navegación en los filtros, se estableció una columna de tipo periodo con información del año y semestre de la fecha de aplicación de la prueba

3) Visualización de datos.

La visualización se desarrolló en PowerBI, una vez preprocesada las bases de datos, la información se cargó de forma estática por medio de un archivo .csv. El reporte consta de tres páginas principales:

- Contexto demográfico
- Hábitos y contexto familiar
- Características del colegio



El desarrollo se hizo en la aplicación de escritorio y fue publicada en el servidor de PowerBI el cual nos proporciona el enlace para ingresar desde cualquier navegador y utilizarlo como insumo para la última fase de publicación de resultados, la visualización tuvo como enfoque la calificación promedio global en la prueba Icfes, esta variable se combinó con las otras para darnos una historia en como las demás influían en ella.

4) Publicación Resultados.

Para la publicación de resultados se utilizó la plataforma de GitHub Pages el cual permite un hosting para repositorios con los requerimientos de espacio en la nube necesarios para el proyecto.



La tecnología a utilizar es el framework Bootstrap que está basado en HTML, css y Javascript. Para los componentes visuales principales diseñados en PowerBI se utilizaron secciones embebidas (*iframe*) para insertar convenientemente el reporte según el contenido de la sección.



El repositorio de versiones de la página se encuentra alojado también en GitHub y puede encontrar se en el siguiente enlace:

<https://github.com/jdelgadoc201920046/Visualizacion-y-storytelling>

El resultado final se encuentra publicado en el siguiente link:

<https://jdelgadoc201920046.github.io/Visualizacion-y-storytelling/>