

Tercera entrega Proyecto Ingeniería de datos

Link GitHub: <https://github.com/josed-mejia/Proyecto-Ing.-de-Datos---Saber-Pro>

1. Discusión y gráficos

- La relación que hay entre si el estudiante tiene beca o crédito con su puntaje de lectura crítica.

Gráfica 1. Diagrama de cajas y bigotes de la relación entre beca y el puntaje Lectura Crítica



Hay evidencia para creer que el hecho de que el estudiante posea una beca para cubrir su educación tiene una correlación con el puntaje de lectura crítica. Como se puede ver en la gráfica, los estudiantes con beca tienen una mediana mayor que los que no tienen beca. La gráfica también evidencia que no es imprescindible tener una beca para obtener un buen puntaje de lectura crítica en las pruebas Saber Pro.

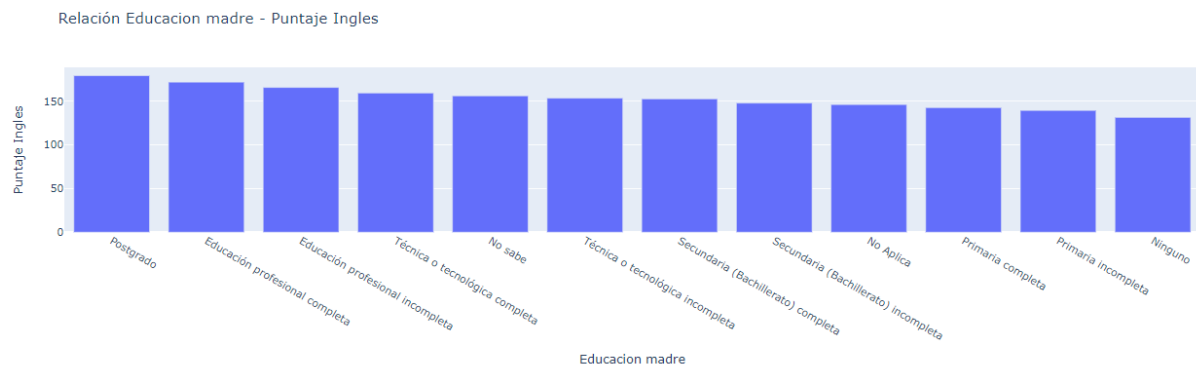
Gráfica 2. Diagrama de cajas y bigotes de la relación entre crédito y el puntaje Lectura Crítica



Hay evidencia para creer que hay una pequeña correlación entre tener un crédito para pagar la educación y el puntaje de lectura crítica. La gráfica visualiza que los estudiantes con crédito poseen una mediana menor en sus puntajes con los que no tienen crédito, ubicándolos con un promedio de puntaje inferior. Esto podría ser por la presión y el estrés por tener un crédito.

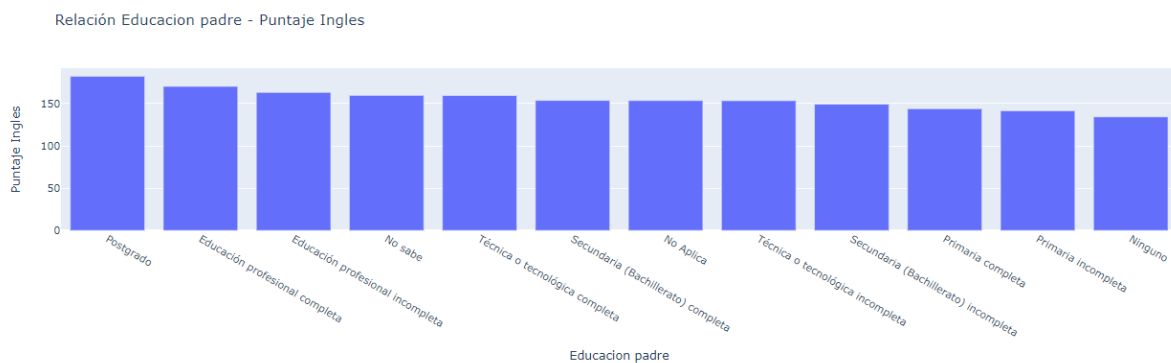
- b) La relación que hay entre el nivel de educación de padre y madre con su puntaje en inglés.

Gráfica 3. Relación entre la educación de la madre y el puntaje de inglés



Hay evidencia para creer que hay una correlación directamente proporcional entre el nivel de educación de la madre con el puntaje de inglés. La gráfica muestra una tendencia creciente, donde a medida que la madre tenga un mayor nivel de educación el promedio en el puntaje de inglés aumenta.

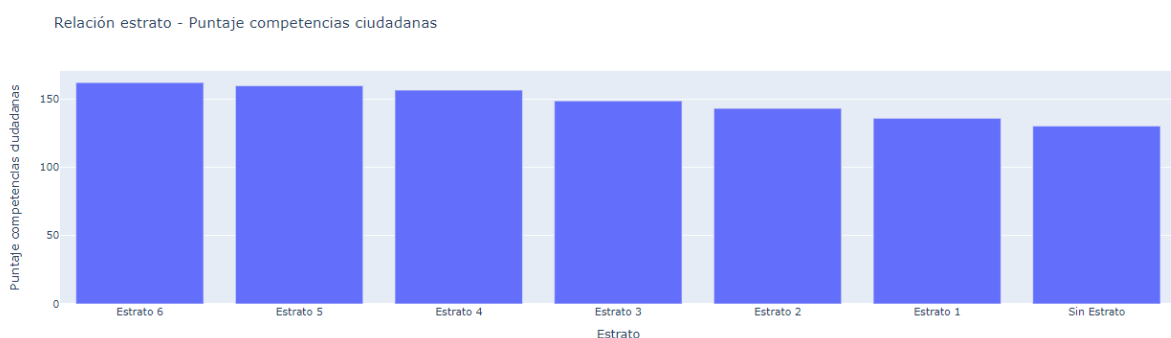
Gráfica 4. Relación entre la educación del padre y el puntaje de inglés



Hay evidencia para creer que hay una correlación directamente proporcional entre el nivel de educación del padre con el puntaje de inglés. La gráfica muestra una tendencia creciente, donde a medida que el padre tenga un mayor nivel de educación el promedio en el puntaje de inglés aumenta.

- c) La relación que hay entre el estrato con su puntaje en competencias ciudadanas.

Gráfica 5. Relación entre el estrato y el puntaje de competencias ciudadanas



Hay evidencia para creer que hay una correlación directamente proporcional entre el estrato de la vivienda del estudiante con el puntaje de competencias ciudadanas. La gráfica muestra una tendencia creciente, donde a medida que el estrato socioeconómico sea mayor el promedio en el puntaje de competencias ciudadanas aumenta.

- d) La relación que hay entre si el estudiante tiene internet con su puntaje en razonamiento cuantitativo.

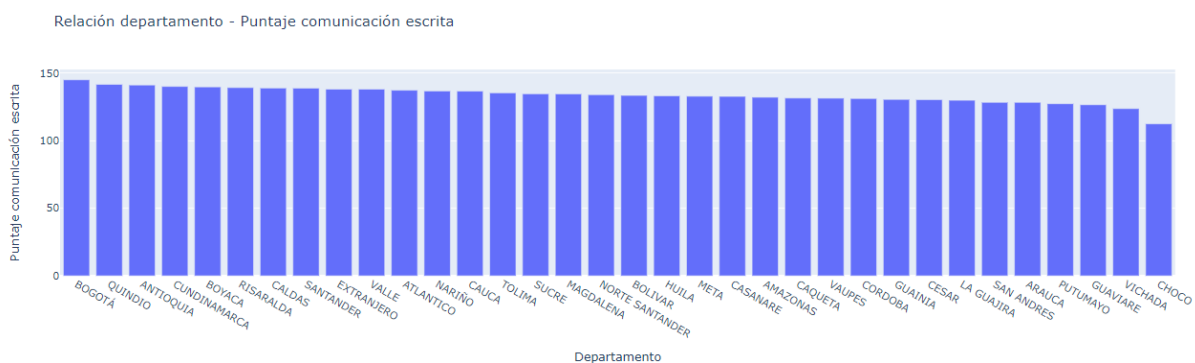
Gráfica 6. Diagrama de violín de la relación entre tener internet y el puntaje de razonamiento cuantitativo



Hay evidencia para creer que hay una pequeña correlación entre que el estudiante tenga el servicio de internet en su vivienda y el puntaje de razonamiento cuantitativo. La gráfica visualiza que los estudiantes que no tienen internet poseen una mediana y cuartiles superiores a los que tienen internet, ubicándolos con un promedio de puntaje inferior. Esto podría indicar que el internet sirve como herramienta para desarrollar el razonamiento cuantitativo.

- e) La relación que hay entre el departamento en el que reside con su puntaje en comunicación escrita.

Gráfica 7. Relación entre el departamento de la vivienda y el puntaje de comunicación escrita



La gráfica muestra que Bogotá obtuvo el mayor promedio en el puntaje de comunicación escrita y Chocó tuvo el peor promedio en esta sección. Sin embargo, el rango es menor a los 50 puntos por lo que no se ve una brecha tan grande en el desempeño de esta área del Saber Pro. Adicionalmente, se puede observar que los departamentos de la región Andina son los que tienen un mejor desempeño mientras que los departamentos en la región Amazonía y Orinoquía son los que tienen un promedio más bajo.

2. Conclusiones

En la selección de los datos, encontramos diversas fuentes en las cuales no pudimos orientar un problema. De esta manera, no fue un proceso de corto plazo. Además, tuvimos cierta dificultad en hallar una base de datos que tuviera suficientes columnas para armar entidades y relaciones. Sin embargo, la base del Saber Pro nos brindó facilidad de diseño y nos puso a explorar nuestra creatividad.

Ahora bien, al tener 1'209.043 datos, la carga masiva resultó ser todo un reto. En primer lugar, intentamos con el método de Excel, pero fracasamos al intentar dividir el archivo csv grande en las tablas organizadas para nuestra base de datos porque nuestros computadores no aguantaron la tarea. Entonces, decidimos cargar el archivo csv original a una base de datos en PostgreSQL para luego dividir la información dentro de esa misma base de datos creando las tablas que necesitábamos y llenándolas utilizando subconsultas.

Para el desarrollo de conexiones de base de datos no tuvimos problema alguno, se nos facilitó el código y la creación de las consultas. Así mismo, no presentamos dificultad usando Dash. No obstante, hubo dos escenarios de análisis que tuvimos que modificar ya que las gráficas resultantes no se podían desarrollar bien. El primer escenario fue el que relacionaba los municipios con el puntaje de comunicación escrita, en este había muchos municipios que no se mostraban en la gráfica ya que no poseíamos pantallas tan grandes. Por otra parte, el segundo escenario relacionaba el desempeño en inglés con la educación del padre y la madre, en este el desempeño no era un dato cuantitativo y al graficarlo el diagrama de barras quedaba con múltiples dimensiones, lo que hacía imposible su desarrollo.