

TRABAJO DE GRADO  
Opción Seminario-Diplomado.

Análisis y visualización de datos  
Seminario ruta Crehana ingenierías

Corporación Universitaria Remington  
Facultad de ingenierías  
Ingeniería de sistemas  
Seminario de grado

Luis Albeiro Múnera Cárdenas  
John Fredy Mira Mejía  
Seminario Diplomado  
2023

## Contenido

<b>Resumen</b>	3
<b>Introducción</b>	3
<b>Palabras claves</b>	3
<b>Marco conceptual y contextual</b>	3
Big data	3
Análítica de datos	4
Importancia de la Analítica de Datos para el Crecimiento Empresarial	4
Historia y Evolución	4
Herramientas Reconocidas	4
<b>Objetivo general</b>	5
<b>Objetivos específicos</b>	5
<b>Desarrollo e implementación del aprendizaje</b>	6
<b>Figuras y tablas</b>	16
<b>Conclusiones</b>	17
<b>Referencias</b>	18
Capítulo 1 Bibliografía	18

# Aplicación analítica para el estudio de los resultados de las pruebas saber pro (2018 - 2022)

## 1. **Resumen**

El big data ha estado tomando mucha fuerza en las organizaciones que desde hace mucho tiempo se han encargado de almacenar grandes volúmenes de información sobre sus clientes, movimientos y demás datos que puedan ser de gran utilidad en la administración de las empresas, con esta ha surgido la necesidad de crear aplicaciones o plataformas que puedan realizar un análisis de estos grandes volúmenes de datos para obtener estrategias que puedan ayudar a las empresas en el crecimiento. La analítica de datos crece con la necesidad de procesar estas grandes cantidades de información. En la actualidad existen muchas plataformas que se encargan de facilitar el análisis y visión de los datos almacenados.

Este proyecto se centra en analizar exhaustivamente los resultados de las pruebas Saber Pro (2018-2022) utilizando la herramienta POWER BI. Se aplican conocimientos adquiridos en cursos de analítica de datos, Excel avanzado y visualización de datos para proporcionar insights detallados en el ámbito educativo. La investigación se contextualiza en la creciente importancia de la analítica de datos para la toma de decisiones informadas en la educación.

1.

## 2. **Introducción**

El análisis de datos educativos se ha vuelto fundamental para comprender el rendimiento estudiantil y mejorar la calidad de la educación. Este proyecto utiliza la herramienta POWER BI para explorar y visualizar los resultados de las pruebas Saber Pro, contribuyendo a la creciente literatura sobre la aplicación de la analítica de datos en el ámbito educativo.

## 3. **Palabras claves**

Big data, Analítica de Datos, POWER BI, Pruebas Saber Pro, Visualización, Filtros.

## 4. **Marco conceptual y contextual**

### 5. **Big data**

El término "big data" se refiere a los datos que son tan grandes, rápidos o complejos que es difícil o imposible procesarlos con los métodos tradicionales. El acto de acceder y almacenar grandes cantidades de información para la analítica ha existido desde hace mucho tiempo. Pero el concepto de big data cobró impulso a principios de la década de 2000 cuando el analista de la industria, Doug Laney, articuló la definición actual de big data como las tres V: Volumen, velocidad, variedad. (Oracle, Colombia, 2023)

## 6. Analítica de datos

La analítica de datos tiene su origen en la necesidad de comprender al consumidor y sus gustos, forma de pensar, preferencias, etcétera. La transformación de estas variables cualitativas y cuantitativas en datos capaces de ser procesados y analizados ha sido el eje principal de la revolución digital que ha experimentado el mundo de los negocios en los últimos años. (Treviño-Reyes, Rivera-Rodríguez, & Garza-Alonso, 2020)P.1065

Respecto a la alta cantidad de datos que se procesan hoy en día en distintas aplicaciones, dispositivos y sistemas, vale la pena adicionar el concepto relacionado con la tecnología que hace posible el tratamiento y procesamiento de información con estas características Big Data. Hace referencia a las soluciones de hardware y software que permiten capturar, almacenar y organizar el acceso a conjuntos de datos tan grandes donde los medios tradicionales de procesamiento de datos son insuficientes. El problema ha pasado de recopilar grandes cantidades de datos a la capacidad tecnológica para almacenarlos y comprenderlos, convirtiéndolos en conocimiento. (Rozo-García, 2020, p.12).

7. **Importancia de la Analítica de Datos para el Crecimiento Empresarial**  
En el ámbito empresarial, la analítica de datos ha cobrado gran relevancia en los últimos años. Se ha convertido en una herramienta estratégica para la toma de decisiones y el impulso del crecimiento. La capacidad de recopilar, analizar y visualizar datos permite a las empresas identificar patrones, tendencias y oportunidades de mejora.
  
8. **Historia y Evolución**  
La analítica de datos ha experimentado un crecimiento significativo en las últimas décadas. A medida que las empresas han acumulado grandes cantidades de datos, la necesidad de convertir estos datos en información útil ha impulsado el desarrollo de herramientas analíticas avanzadas.
  
9. **Herramientas Reconocidas**  
Algunas de las herramientas más reconocidas en el campo de la analítica de datos son:  
  
Tableau: Una plataforma líder en visualización de datos que permite crear dashboards interactivos y análisis en tiempo real.  
  
Microsoft Power BI: Proporciona capacidades de análisis de datos y visualización, facilitando la creación de informes interactivos y paneles de control.  
  
Google Analytics: Especializada en análisis web, proporciona información detallada sobre el comportamiento de los usuarios en sitios web.

**10.**

**11. Objetivo general**

Desarrollar un análisis detallado de los resultados de las pruebas Saber Pro utilizando la herramienta POWER BI, integrando conocimientos de cursos previos realizados desde la plataforma CREHANA y aplicándolos al contexto educativo.

**12. Objetivos específicos**

Utilizar las funciones avanzadas de POWER BI para analizar y visualizar datos educativos.

Aplicar filtros para segmentar la información según diferentes variables (departamento, ciudad, institución educativa, programa, género, estrato, etc.).

Generar insights específicos para mejorar el desempeño estudiantil y la calidad educativa.

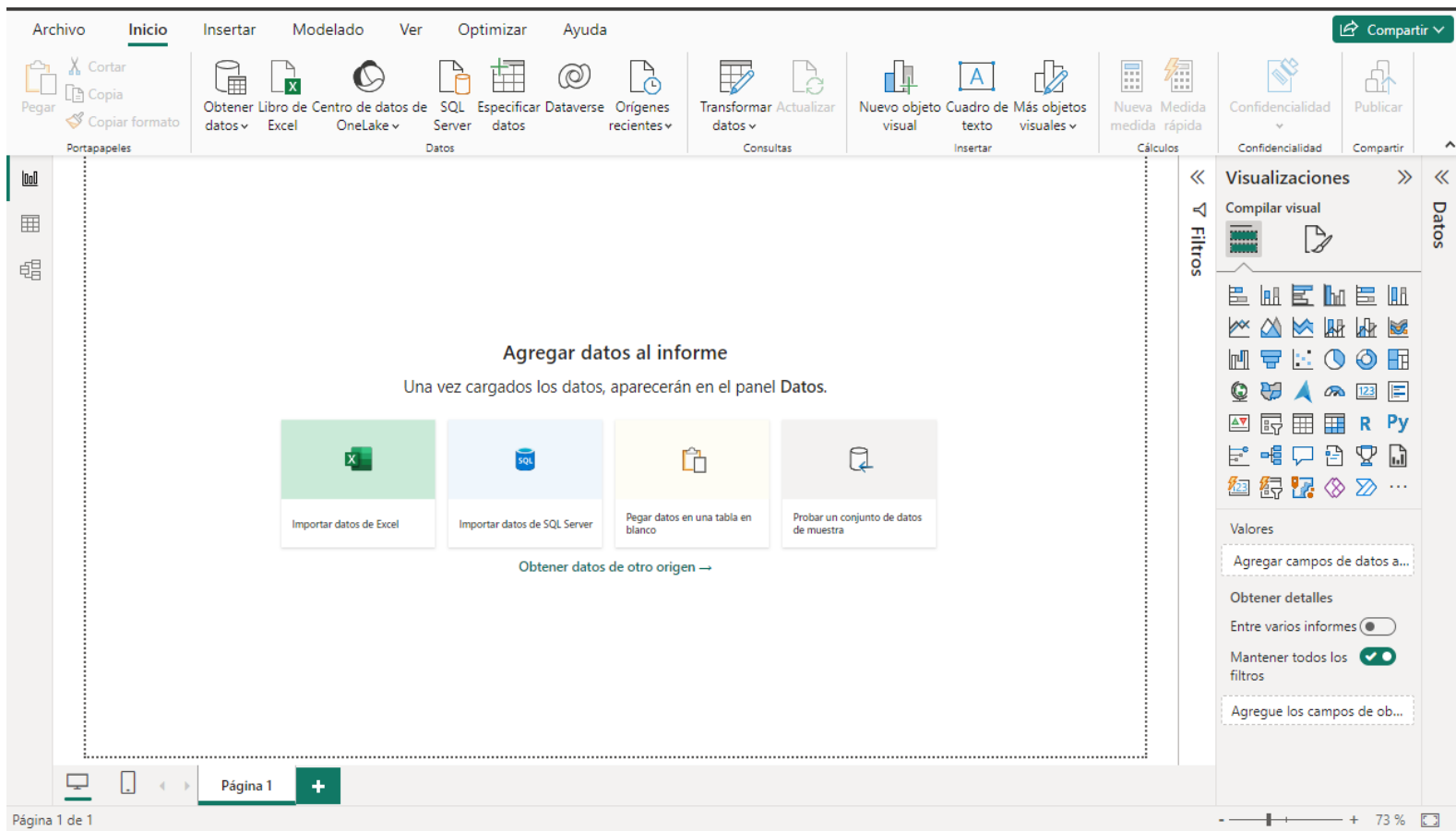
### 13. Desarrollo e implementación del aprendizaje

Para realizar este análisis se realiza la descarga del archivo en formato csv obtenido de la página [www.datos.gov.co](http://www.datos.gov.co) la cual consta de 1'217.482 filas y 86 columnas. Por su tamaño, resulta muy difícil obtener estadísticas sin la ayuda de una herramienta como power BI u otra de las muchas que existen en la actualidad.

Lo primero que debemos hacer es descargar la aplicación de Power BI, luego de abrir la aplicación se podrá visualizar la vista principal como lo muestra la gráfica de la tabla #1 donde se puede obtener los datos desde diferentes fuentes como archivos xlsx, csv, txt, SQL entre otras, dando clic en la opción obtener datos.

Luego de cargar los datos aparece en la parte superior izquierda una pestaña llamada datos en donde se cargan todos los encabezados de las columnas y otra pestaña visualizaciones donde se podrán generar diferentes tipos de gráficas, tablas, filtros y muchas otras funcionalidades que permiten generar un análisis de datos muy fácilmente.

Tabla 1. Plataforma Power BI



Después de cargar los datos a la plataforma, se puede ver los encabezados de las columnas en la parte izquierda los cuales tienen una casilla de check para seleccionar una o varios tipos de cruces que se desean realizar. Para conocer la cantidad de estudiantes que realizaron las pruebas durante los años (2018-2022), se selecciona el tipo de gráfica con la que se quiere representar el segmento y se selecciona las columnas involucradas, en este caso se toma la columna cod\_estudiante y la de género para conocer la suma total de todos los estudiantes del género masculino y femenino

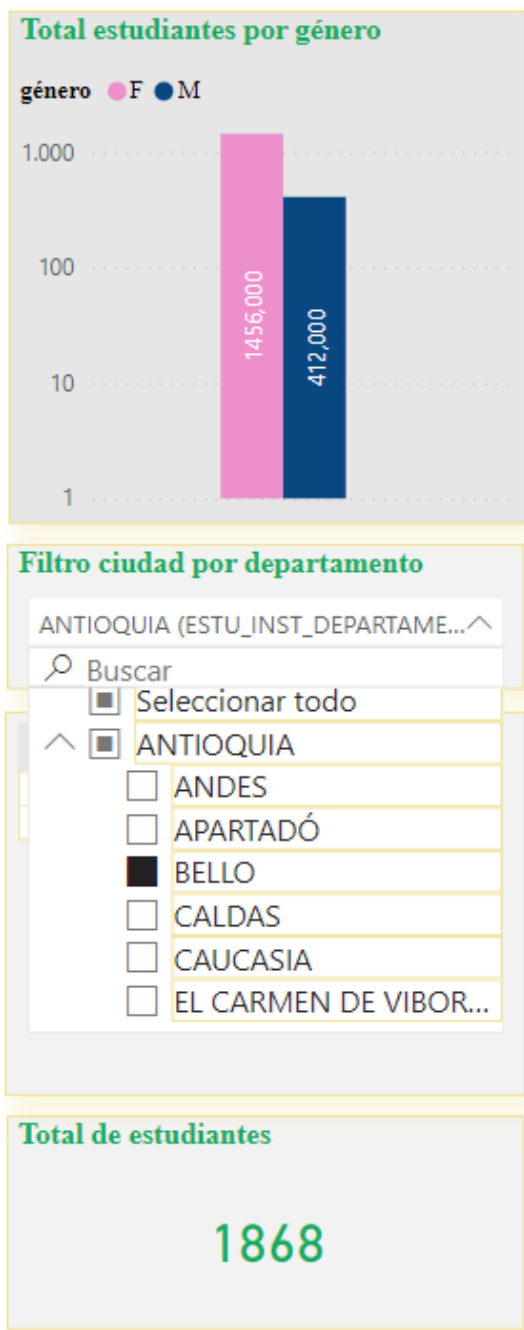
Tabla 2. Total, estudiantes por género



Se puede ver que hay más de la mitad de los estudiantes son mujeres con un total de 716.347 vs 495.990 de hombre

Se puede realizar un filtro de ciudades por departamento para obtener los datos específicos por cada ciudad

Tabla 3. Estudiantes por ciudad y género

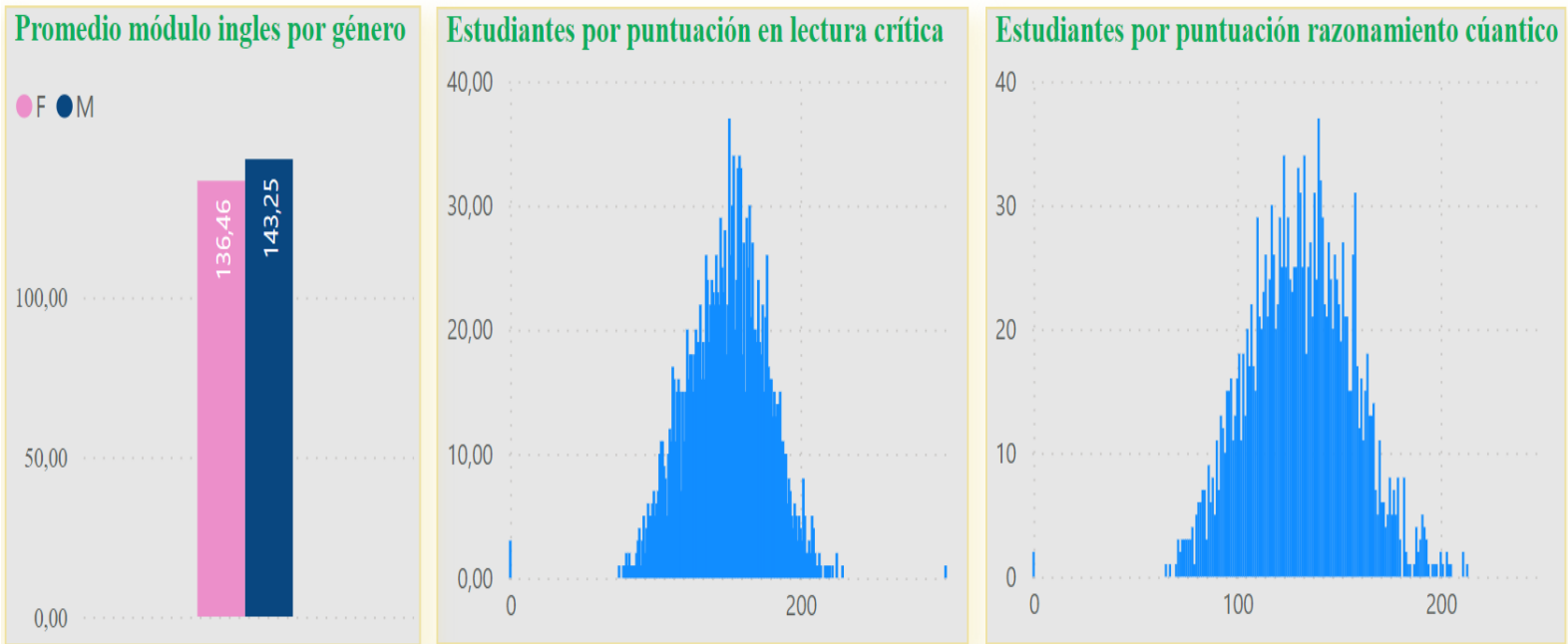


En este caso se puede contar solo los estudiantes que realizaron las pruebas de una ciudad en específico. En el ejemplo de la tabla 3, se observa que la cantidad de estudiantes totales es de 1868 personas, 1456 de género femenino y 412 de género masculino por lo que se evidencia que más de la tercera parte de las personas fueron mujeres



Luego se pueden agregar más filtros y gráficos que ayuden a conocer más detalles de los resultados, se crean 5 gráficos con las puntuaciones y promedios obtenidos por cada módulo y también se incluyó un mapa para visualizar con mayor facilidad las ciudades con un mayor número de estudiantes que realizaron los exámenes

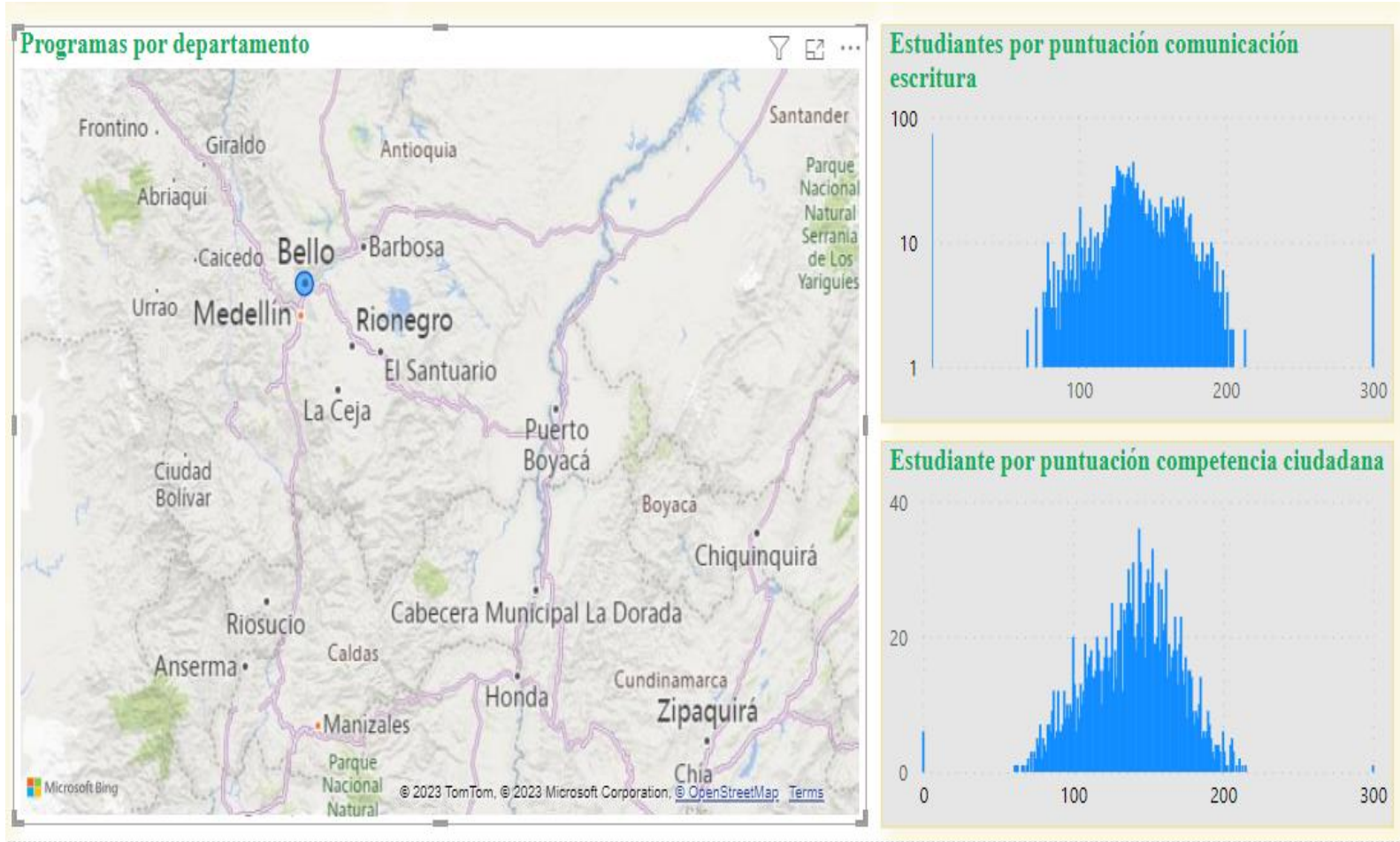
Tabla . Promedio de notas modulo inglés y puntuación razonamiento cuántico y lectura crítica



Con estos gráficos de barras agrupadas se puede ver que el promedio de nota en el módulo de inglés es mayor al promedio de las mujeres, también se evidencia que la gran mayoría de estudiantes obtuvo una puntuación entre 113 y 186 puntos o que solo 1 persona obtuvo el puntaje perfecto y también que 3 estudiantes sacaron un puntaje de 0 en lectura crítica.

En el caso del razonamiento cuántico, ninguna persona obtuvo el resultado completo y 2 personas obtuvieron puntaje cero, como también se nota que la mayoría de los estudiantes obtuvieron un puntaje mayor a 100

Tabla . Mapa , puntuación comunicación escrita y competencias ciudadanas



Se evidencia en el mapa que efectivamente solo se marca el filtro por la ciudad de bello, también se puede visualizar que el gráfico puntuación comunicación escrita, tuvo la mayor cantidad de puntajes perfectos con 8 estudiantes

Se pueden personalizar los resultados agregando más filtros donde se pueda visualizar los resultados por código de la persona, programa, universidad entre otros que estaremos validando con la próxima gráfica

Tabla . Resultado promedio de notas por módulo de la UDA



En esta gráfica circular se puede obtener los promedios de las pruebas por módulo según la institución seleccionada. Se puede apreciar un promedio general en todos los módulos entre el 150% y el 172% para los estudiantes de la universidad de Antioquia, Podemos comparar con otras instituciones de la misma ciudad o de otra de diferentes departamentos o ciudades.

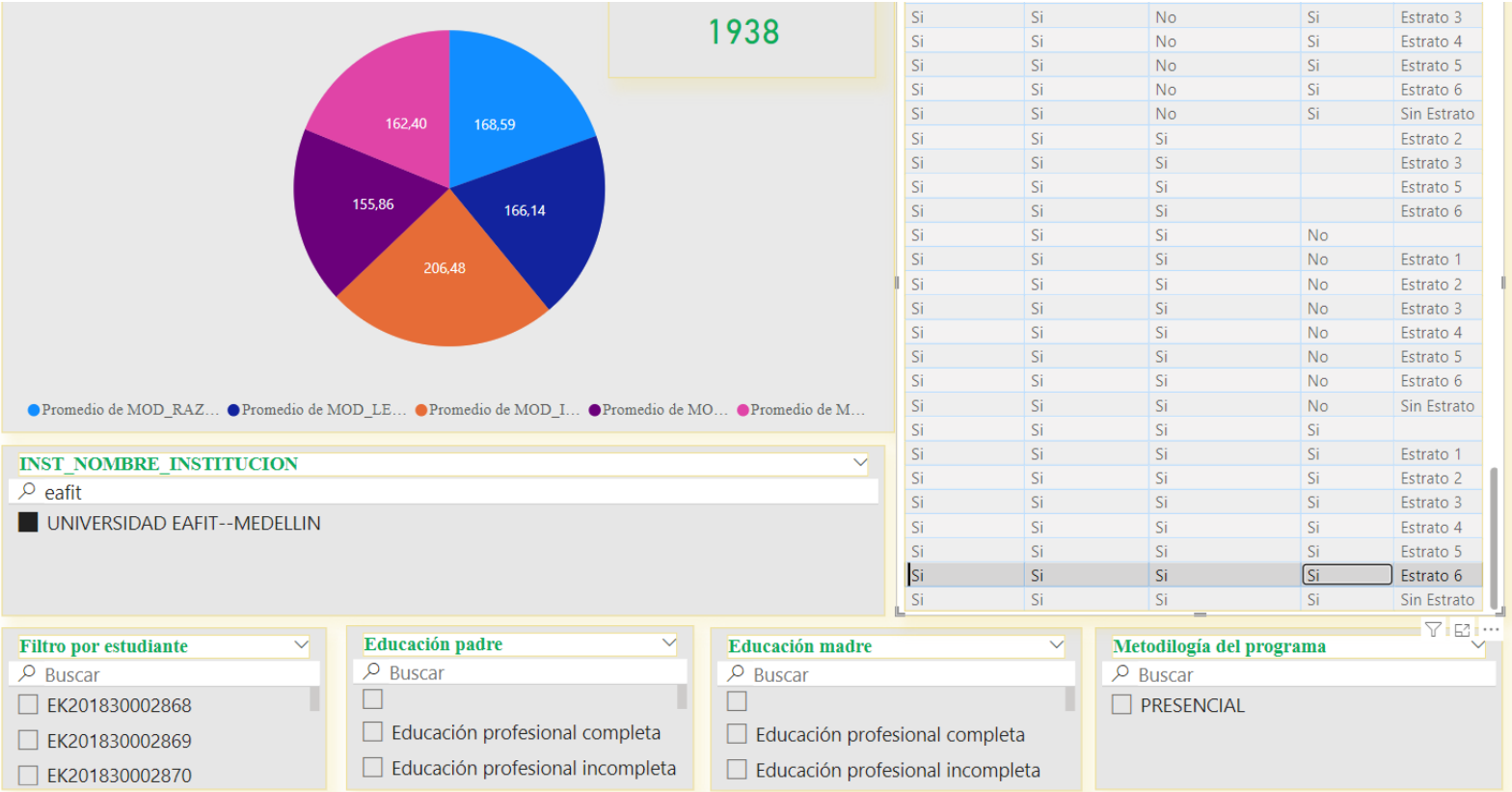
En la próxima gráfica se consultó por la Corporación Universitaria de la costa, CUC-Barranquilla y se evidencia un menor promedio en general en comparación con la UDA

Tabla . Promedio de notas por modula de la universidad de la costa, CUC-Barranquilla



En la siguiente gráfica, se realizaron diferentes tipos de filtros para realizar una búsqueda más personalizada en donde se pueden obtener los resultados por estudiante y validar el estrato económico, si tiene auto, internet, computador, lavadora, internet

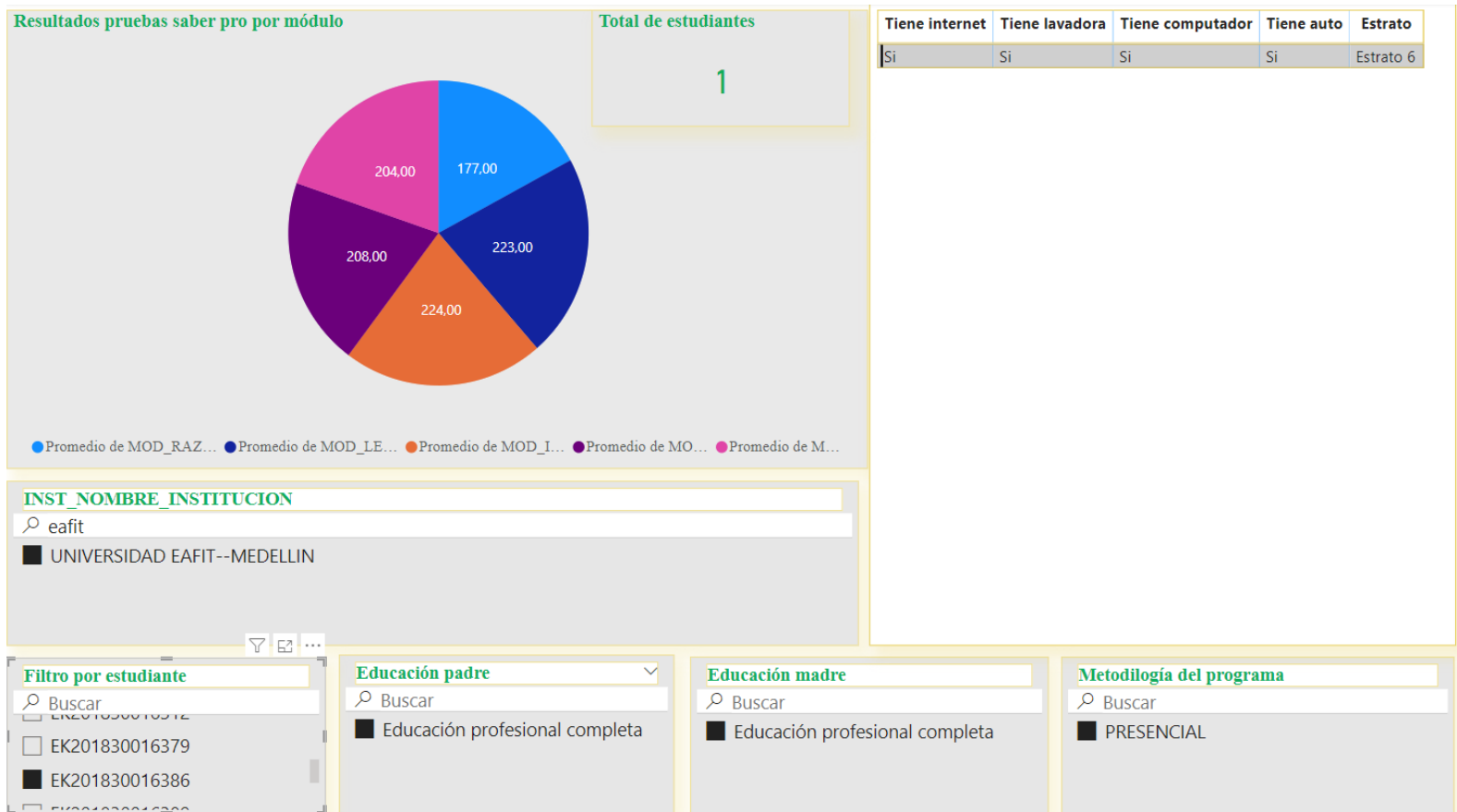
Tabla . Resultados de las pruebas con filtros personalizados



En esta gráfica se evidencia que, de los 8061 estudiantes de la Universidad EAFIT de Medellín, hay 1938 estudiantes que tienen todas las opciones y pertenecen a un extracto 6

Finalmente, en la siguiente gráfica se realizará un filtro completo para obtener todos los resultados de un estudiante en específico

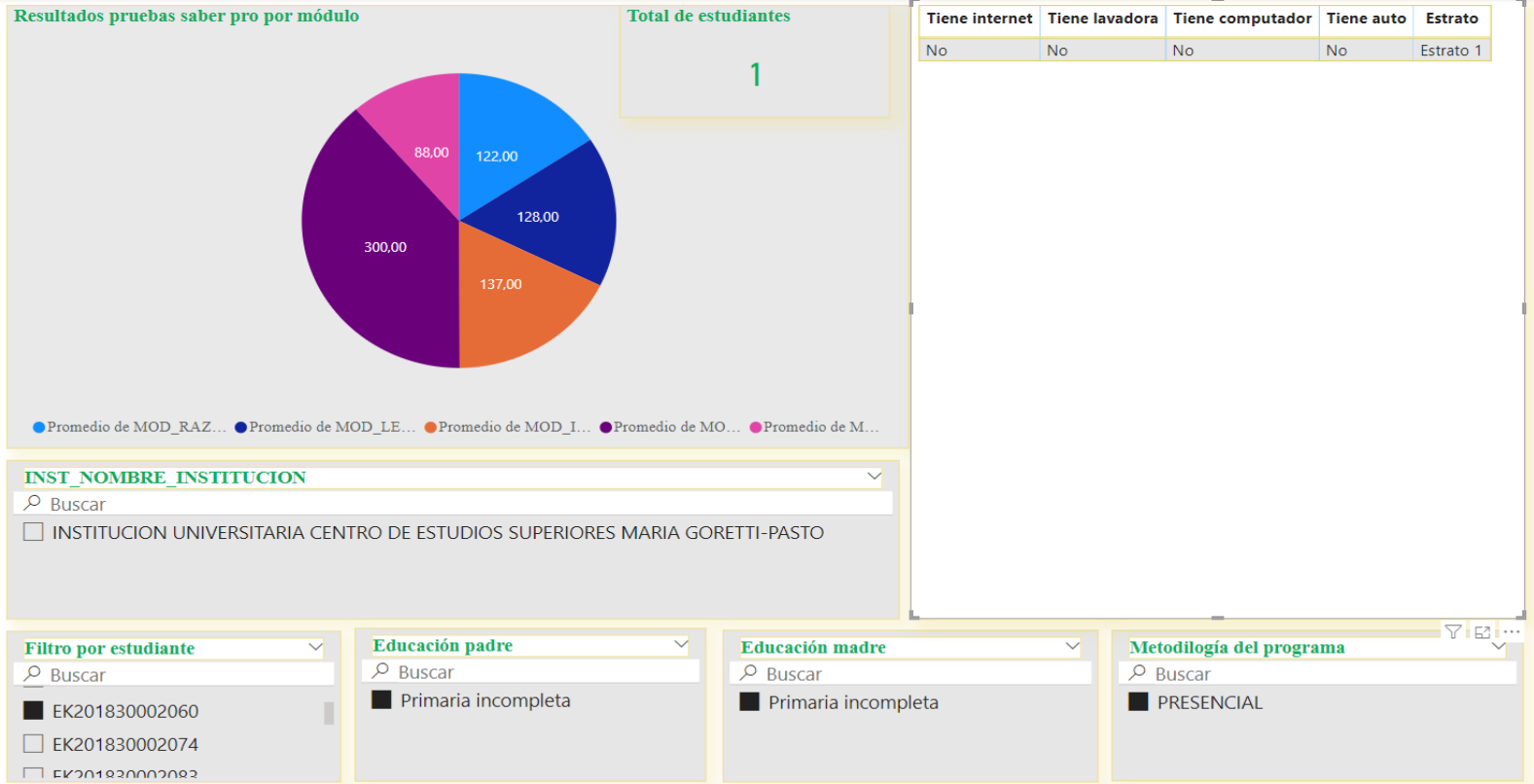
Tabla 9. Grafica de estudiante de extracto 6 de la Eafit en Medellín



En esta gráfica podemos ver un estudiante de la EAFIT Medellín de extracto 6 y con padres profesionales y además cuenta con internet, auto, computador, lavadora. Se evidencia un rendimiento mayor en sus estudios con un promedio muy alto en todos los módulos

En la siguiente gráfica se comparó con otros estudiantes de estrato más bajo y con padres sin estudio y sin las oportunidades que disfruta el estudiante de extracto 6 y se evidencia una diferencia importante en su rendimiento y al realizar la consulta en general a los extractos 1, padres sin estudios y sin las comodidades de los anteriores y se nota un promedio en general muy bajo

Tabla . Promedio de estudiante de extracto bajo y sin las comodidades



## **14. Figuras y tablas**

### **15.**

Tabla 1. Plataforma Power BI	6
Tabla 2. Total, estudiantes por género	7
Tabla 3. Estudiantes por ciudad y género	8
Tabla 4. Promedio de notas módulo inglés y puntuación razonamiento cuántico y lectura crítica	9
Tabla 5. Mapa, puntuación comunicación escrita y competencias ciudadanas	10
Tabla 6. Resultado promedio de notas por módulo de la UDA	11
Tabla 7. Promedio de notas por modula de la universidad de la costa, CUC-Barranquilla	12
Tabla 8. Resultados de las pruebas con filtros personalizados	13
Tabla 9. Gráfica de estudiante de extracto 6 de la Eafit en Medellín	14
Tabla 10. Promedio de estudiante de estrato bajo y sin las comodidades	15

### **16.**



## **18. Conclusiones**

La aplicación analítica en POWER BI ha demostrado ser una herramienta valiosa para comprender los resultados de las pruebas Saber Pro. Los insights generados pueden influir directamente en la implementación de estrategias educativas personalizadas, siguiendo las mejores prácticas identificadas por investigaciones previas (Rodríguez & Pérez, 2022).

Este proyecto contribuye a la creciente literatura sobre la aplicación de analítica de datos en el ámbito educativo, brindando evidencia adicional de su efectividad en la toma de decisiones informadas para la mejora continua.

## **19.**

## 21. **Referencias**

Oracle, Colombia. (2023). *Big data*. Obtenido de ¿Que es big data?:  
<https://www.oracle.com/co/big-data/what-is-big-data/>

Treviño-Reyes, R., Rivera-Rodríguez, F. S., & Garza-Alonso, J. A. (2020). *La analítica de datos como ventaja competitiva en las organizaciones*. Nuevo León, México.