

DESCRIPTIVE ANALYSIS OF THE GENDER GAPS IN THE SABER 11, SABER PRO AND SABER TYT TESTS

ANÁLISIS DESCRIPTIVO DE LAS BRECHAS DE GÉNERO EN LAS PRUEBAS SABER 11, SABER PRO Y SABER TYT

**Alisson Catalina Celeita Gutierrez, Angie Paola Manrique Ravelo,
Cesar Yesid Barahona Rodriguez,**

Universidad de Cundinamarca
Grupo de investigación GISTFA
Facatativá, Cundinamarca, Colombia
Calle 14 con Avenida 15 | (+57 1) 892 0706

E-mail: apaolamanrique@ucundinamarca.edu.co, acceleita@ucundinamarca.edu.co,
cbarahona@ucundinamarca.edu.co

Abstract: For many years there have been gender gaps in different fields of society and nowadays there has been an increase in interest in learning about their behavior, especially in the educational field; For this reason, it has been decided to carry out an investigation that analyzes the data provided by the ICFES from the Saber tests, finding indicators of gender gaps and, in turn, being able to relate the common thematic nuclei, from the tests presented in secondary education to education. superior, in order to predict these behaviors in the future. Thanks to this, the user will be presented, by means of software, a detailed representation of the indicators found in order to be used in future investigations of entities that are interested in the study and reduction of this problem.

Keywords: Data mining, Gender gaps, Education, Descriptive analysis.

Resumen: Desde hace muchos años se han presentado brechas de género en diferentes campos de la sociedad y en la actualidad ha aumentado el interés de conocer el comportamiento de estas, en especial en el ámbito educativo; por esa razón se ha decidido realizar una investigación donde se analicen los datos brindados por el ICFES de las pruebas Saber, encontrando indicadores de brechas de género y a su vez poder relacionar los núcleos temáticos en común, desde las pruebas presentadas en la educación media hasta la educación superior, para lograr predecir estos comportamientos a futuro. Gracias a esto, se le presentará al usuario, por medio de un software, una representación detallada de los indicadores hallados con el fin de que sean usados en futuras investigaciones de entes que se interesen en el estudio y la reducción de esta problemática.

Palabras clave: Minería de datos, Brechas de género, Educación, Análisis descriptivo.

1. INTRODUCCIÓN

Las brechas de género son una problemática que en la actualidad se encuentra en auge. Cada vez son más los grupos y organizaciones que se unen a la causa de disminuir las diferencias existentes entre hombres y mujeres. A través del tiempo, se han presentado injusticias para el género femenino, tanto en los ámbitos educativos, laborales, familiares como en los socioculturales observando ventajas competitivas para los hombres, las cuales se convierten en múltiples indicadores que contribuyen directamente con esta disparidad.

A nivel mundial se presentan diversas diferencias, por ejemplo, en los países de América Latina y el Caribe (ALC) es preocupante ver como los hombres siguen ganando más que las mujeres, la violencia doméstica es elevada, las mujeres indígenas y rurales son quienes menos educación reciben (López-Bassols et al., 2018). Este proyecto centra su interés en la identificación de indicadores de brechas de género en Colombia, más específicamente en el ámbito estudiantil, para poder contrarrestar su crecimiento.

Para mirar el comportamiento de las mismas, se está realizando un análisis descriptivo por medio de técnicas de minería de datos aplicadas a las pruebas Saber realizadas en la educación media y superior, haciendo uso de los resultados entregados por el ICFES de cada una de las pruebas en sus respectivos periodos de presentación. Específicamente hablando, del año 2016 a la actualidad.

2. DESARROLLO

Antes de dar inicio con el análisis de los datos existe una fase de preprocesamiento que tiene como fin seleccionar aquellos datos relevantes para su estudio, más conocida como ETL. Vale la pena aclarar, que no hay un procedimiento determinado a seguir, sino que es realizado en función del objetivo que se va a llevar a cabo. Cuenta con 3 etapas de funcionamiento, mostradas a continuación:

• Etapa 1: Extracción

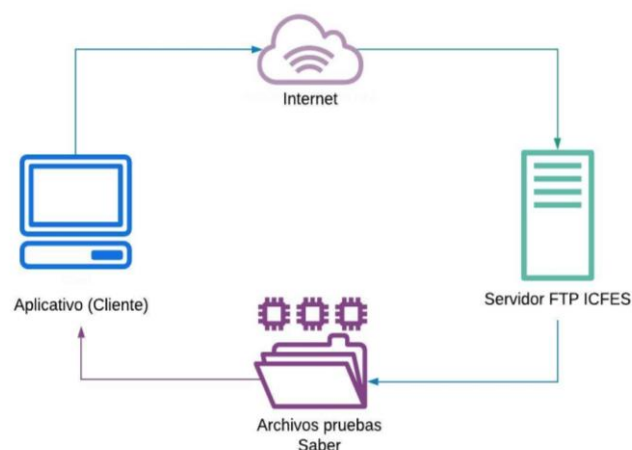


Fig. 1. Extracción de información

Consiste en la obtención de la información desde su origen, en este caso el servidor FTP proporcionado por el ICFES, hasta el repositorio local, donde se desea hacer uso de la misma. Para llevar a acabo esto, se realizó una conexión por medio del software libre WinSCP, de forma que facilitara la tarea de descarga de los archivos por parte del administrador del sistema. Con este proceso de almacenamiento se evita que en caso de caída del servidor haya pérdida de los datos y se vea interrumpido el proceso de análisis. WinSCP fue escogido por la practicidad que tiene en la sincronización de elementos, puesto que se encarga de hacer una comparación entre los existentes y faltantes, permitiendo así que no se genere redundancia y/o duplicación en la información.

• Etapa 2: Transformación



Fig. 2. Transformación de información

Esta etapa es de vital importancia debido a que se encarga de verificar la integridad y consistencia de los datos. Posteriormente, se realiza una selección de los atributos que cumplen con los estándares requeridos. Haciendo énfasis en nuestros requerimientos, en primer lugar, fue necesario realizar la eliminación de todos aquellos registros que no contenían los campos mínimos diligenciados para su evaluación. En segundo lugar, se hizo la transformación de todo aquel carácter que pudiese presentar conflicto en su

lectura y su carga. Por último, para evitar una sobrecarga en la base de datos se excluyeron las columnas o atributos que presentaban datos constantes e identificadores que no suman relevancia al estudio.

- Etapa 3: Carga

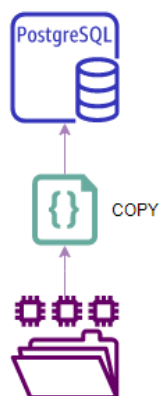


Fig. 3. Carga de archivos a base de datos

Para el almacenamiento se decidió hacer uso del gestor de base de datos relacional PostgreSQL (SGBD), por su flexibilidad en cuanto a lenguajes de programación, su robustez, eficiencia y estabilidad. Además, que es uno de los más recomendados para trabajar con grandes volúmenes de información y ampliamente usado a nivel empresarial. El proceso de carga se realiza por medio del comando “COPY” propio del SGBD, que permite la subida de la información directamente desde el archivo .txt, el cual proviene de descomprimir el archivo rar obtenido del repositorio FTP del ICFES, donde se encuentran alojados los documentos de cada una de las pruebas.

3. RESULTADOS

Ya con las pruebas cargadas a la base de datos, y su respectivo proceso de transformación se procede al análisis descriptivo de las mismas. En primer lugar, las pruebas Saber 11 respecto al género se encontraron diferencias en los puntajes de Lectura crítica, Matemáticas, Sociales y Ciencias Naturales; se puede evidenciar que los puntajes más altos están en mayor proporción para los hombres que para las mujeres. Mientras que, en inglés y Competencias ciudadanas los puntajes se muestran muy parejos para ambos géneros.

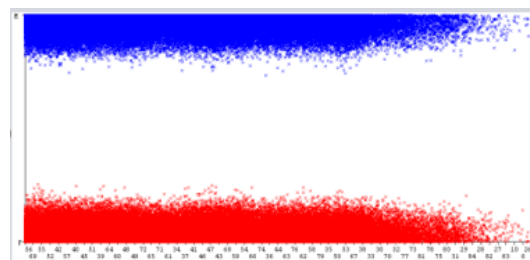


Fig. 4. Puntajes de lectura crítica según género

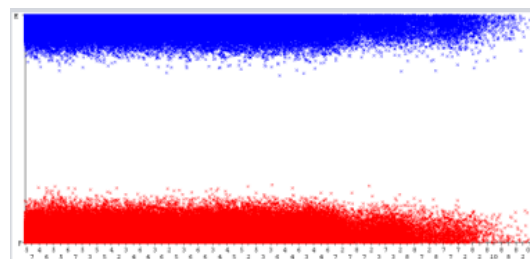


Fig. 5. Puntajes de matemáticas según género

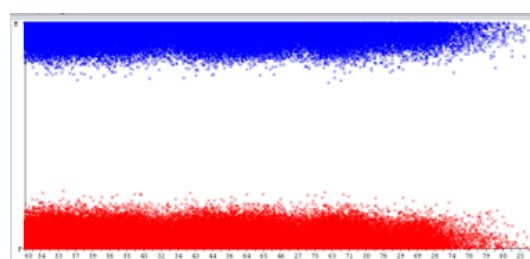


Fig. 6. Puntajes ciencias naturales según género

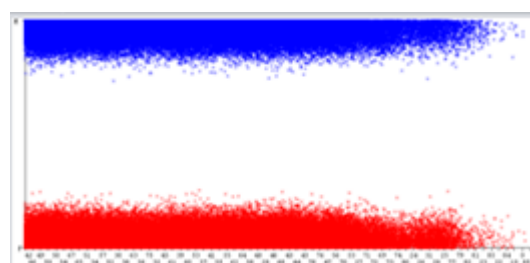


Fig. 7. Puntajes competencias ciudadanas según género

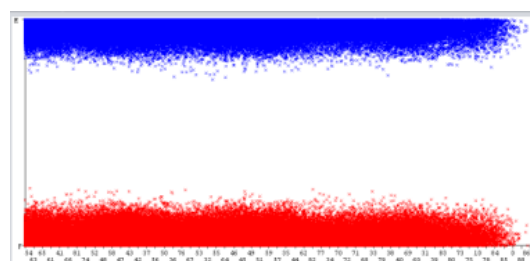


Fig. 8. Puntajes ingles según género

Aparte de los puntajes obtenidos en cada núcleo, se encontraron unos indicadores que afectan un poco el resultado final. Como lo son: el estrato, género del colegio, nivel de Sisbén, ingreso familiar y jornada.

Con **el estrato**, se nota diferencia en los 2 más altos niveles, es decir que en los estratos 5 y 6 se encuentra que hay más cantidad de mujeres que de hombres. Además, la mayoría de la población en ambos géneros está concentrada en el estrato 1 y 2, lo que nos indica que sus ingresos económicos a nivel familiar son bajos o reducidos.

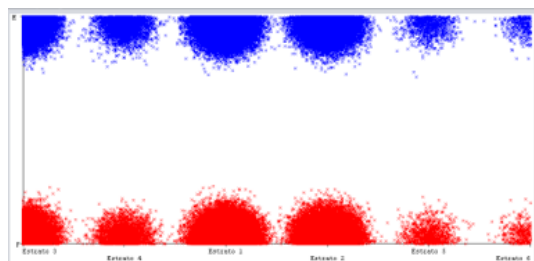


Fig. 9. Estrato según género

Al evaluar los resultados obtenidos en cada uno de los núcleos respecto a los estratos, se puede observar como los estratos mas bajos logran obtener puntajes más sobresalientes en lectura crítica, matemáticas y competencias ciudadanas. Cosa contraria pasa con el núcleo de inglés, en el cual los puntajes mas altos fueron obtenidos por los estratos más altos.

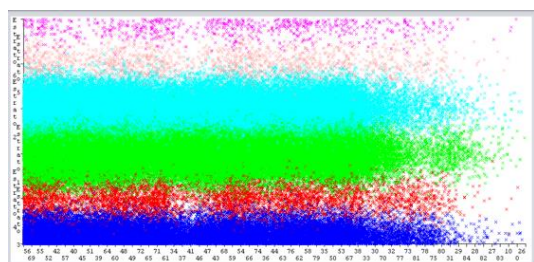


Fig. 10. Puntaje lectura crítica según estrato

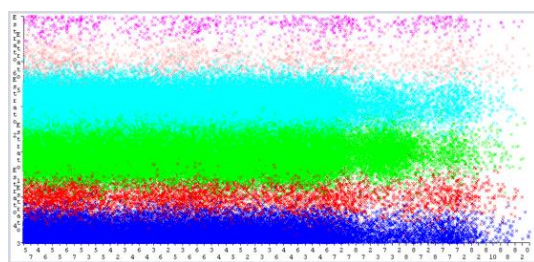


Fig. 11. Puntaje matemáticas según estrato

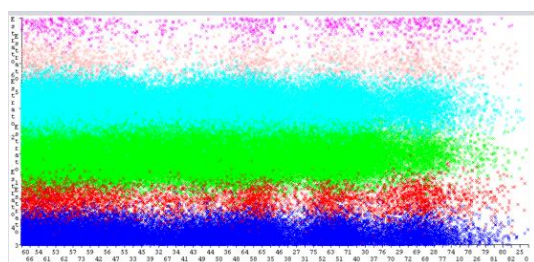


Fig. 12. Puntaje ciencias naturales según estrato

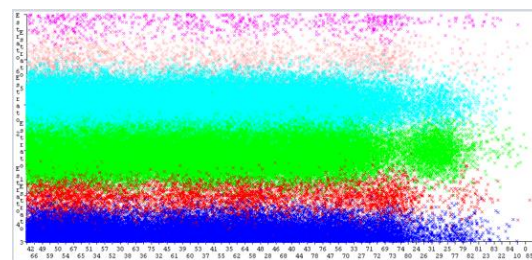


Fig. 13. Puntaje competencias ciudadanas según estrato

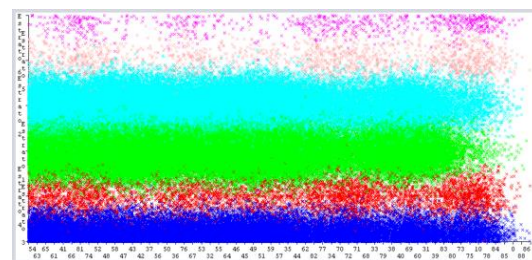


Fig. 14. Puntaje inglés según estrato

La situación que se ve reflejada con el **género del colegio** es la siguiente: Son pocos los colegios de carácter masculino, sin embargo, se destacan mujeres en esta clasificación, mientras que, en los colegios de tipo femenino no se ve una presencia significativa de hombres, y la gran mayoría son de tipo mixto. La cantidad de colegios femeninos a nivel nacional es 5 veces mayor con respecto a la de colegios masculinos.

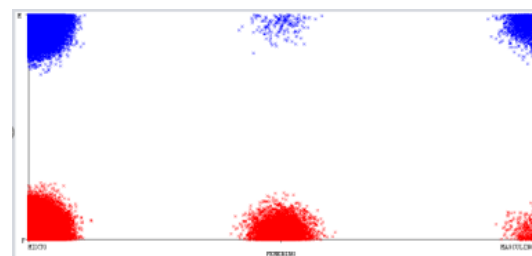


Fig. 15. Género del colegio

Además, dentro de este mismo indicador se encontraron departamentos donde no hay presencia de colegios ni femeninos ni masculinos, solo hay mixtos. Los departamentos en mención son: Guaviare, Vaupés, Vichada, Casanare, Arauca, San Andrés y Guainía. Y donde la presencia de colegios femeninos es mínima son Quindío, Nariño, La Guajira, Sucre, Cesar y Amazonas. Como es mostrado en la **Fig. 16**, a continuación.

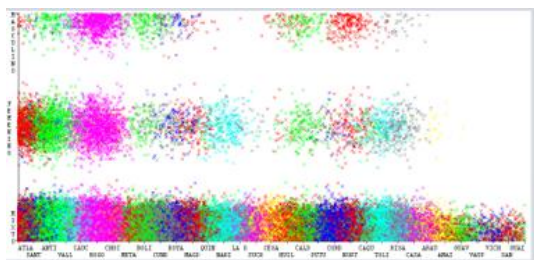


Fig. 16. Género del colegio según departamento

En el análisis de este indicador respecto a los puntajes podemos visualizar como los colegios clasificados por género tienen una gran desventaja en los puntajes obtenidos. En el caso puntual de inglés se ve un poco equilibrada la situación de las calificaciones, pero en los otros núcleos la desventaja aumenta, exceptuando matemáticas, en el cual los colegios masculinos logran emparejar a mixtos dejando atrás a las instituciones femeninas.

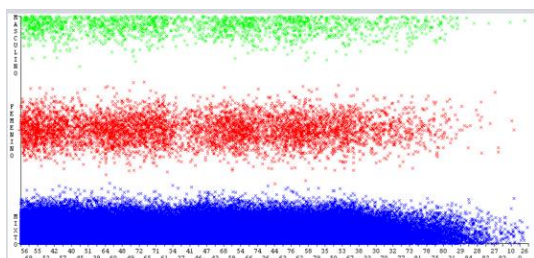


Fig. 17. Puntajes lectura crítica según género del colegio

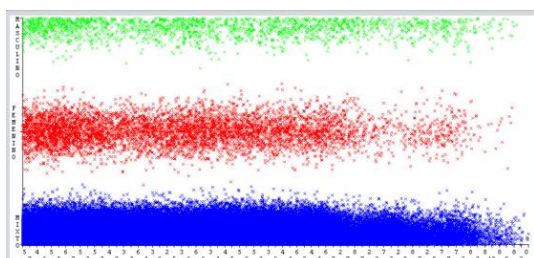


Fig. 18. Puntajes de matemáticas según género del colegio

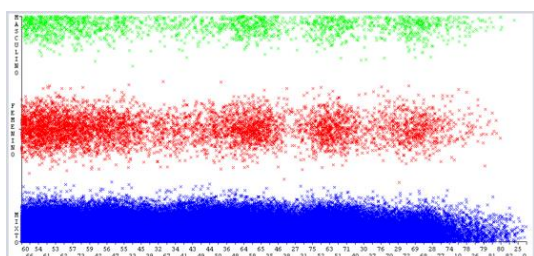


Fig. 19. Puntajes ciencias naturales según género del colegio

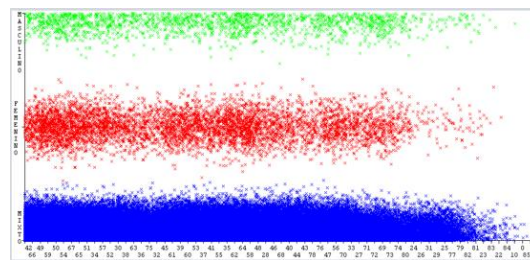


Fig. 20. Puntajes competencias ciudadanas según género del colegio

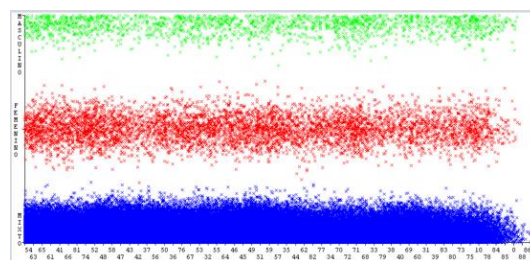


Fig. 21. Puntajes ingles según género del colegio

Otro indicador es **el nivel de Sisbén**, la cantidad de mujeres es mayor en los niveles superiores a 3, o como lo muestra la **Fig. 22** está clasificado en otro nivel, este comportamiento es bastante similar al presentado con el estrato. También en el nivel 2 se puede ver más presencia de mujeres que de hombres.

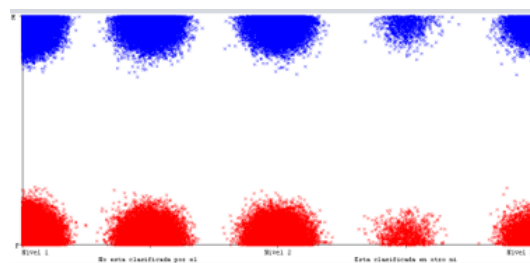


Fig. 22. Nivel de Sisbén según género

En el caso del **ingreso familiar**, quienes reciben menos de 1 SMMV son en mayor proporción los del género femenino, entre 2 y menos de 3 el comportamiento es parejo, entre 3 y menos de 5 hay más hombres que mujeres, entre 5 y menos de 7 también son menos las mujeres, entre 7 y menos de 10 pasa lo mismo y para 10 o más salarios el comportamiento es similar en ambos géneros.

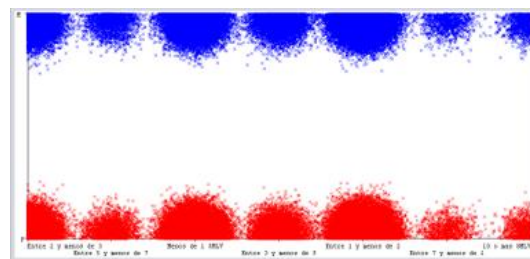


Fig. 23. Ingreso familiar según género

Su comportamiento frente a las valoraciones de las materias demuestra que este indicador solamente afecta los resultados en inglés, donde las personas con mayor ingreso familiar obtienen puntajes más altos.

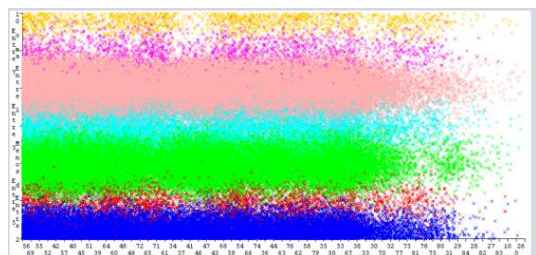


Fig. 24. Puntajes lectura crítica según nivel de Sisbén

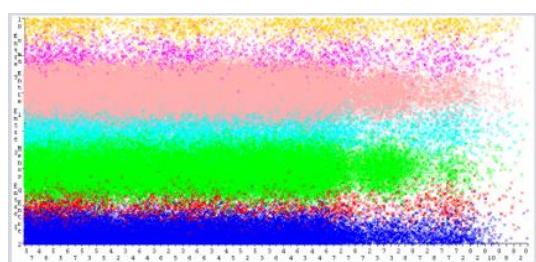


Fig. 25. Puntaje matemáticas según nivel de Sisbén

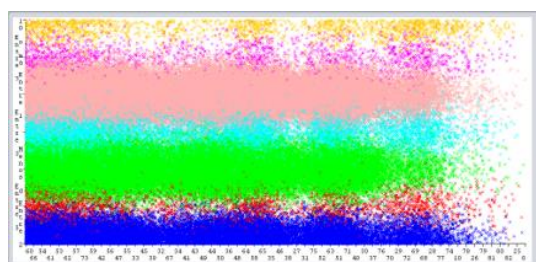


Fig. 26. Puntaje ciencias naturales según nivel de Sisbén

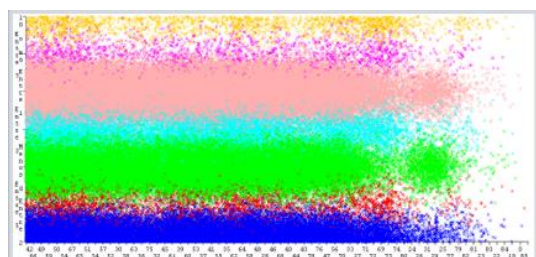


Fig. 27. Puntaje competencia ciudadana según nivel del Sisbén

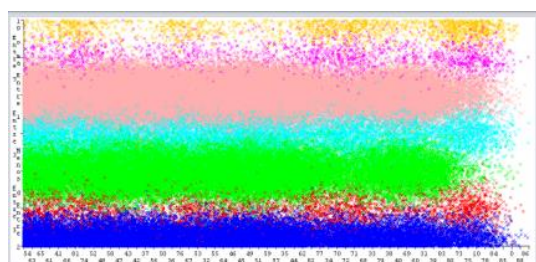


Fig. 28. Puntaje ingles según nivel del Sisbén

El último indicador encontrado en esta prueba es la **jornada**, para las 3 primeras categorías que son de tipo completa, tarde y mañana el comportamiento es parejo en ambos géneros. Las diferencias se encuentran en las jornadas sabatina, noche y única donde hay mayor presencia femenina que masculina.

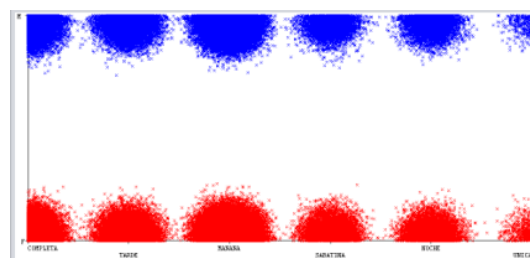


Fig. 29. Jornada académica según género

Se demuestra que este indicador afecta en gran proporción los resultados en cada núcleo. En los casos de matemáticas, lectura crítica, ciencias naturales e inglés las jornadas mañana y completa presentan mejores resultados, dejando en últimos lugares a jornada sabatina y tarde. Competencia ciudadana por su parte tiene puntajes mas equilibrados frente a este indicador.

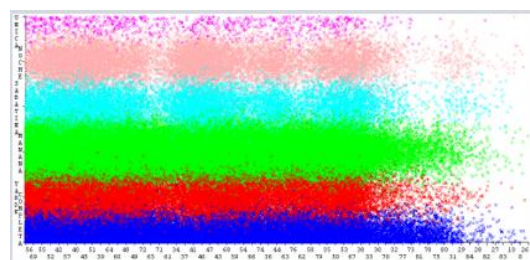


Fig. 30. Puntaje lectura crítica según jornada

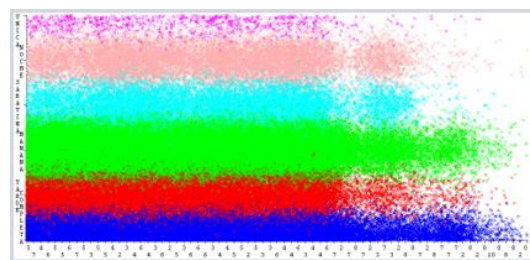


Fig. 31. Puntaje matemáticas según jornada

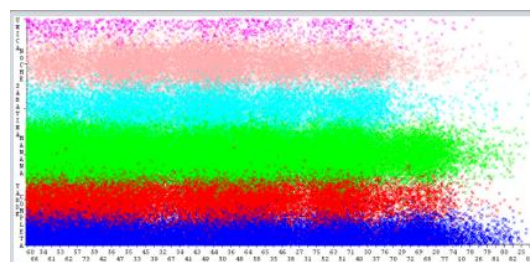


Fig. 32. Puntaje ciencias naturales según jornada

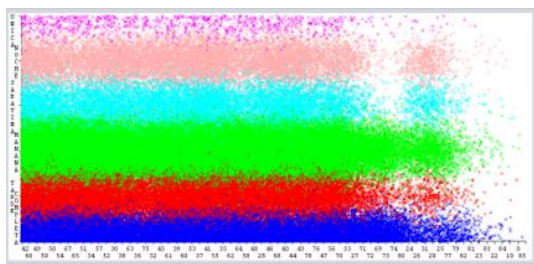


Fig. 33. Puntaje competencia ciudadana según jornada

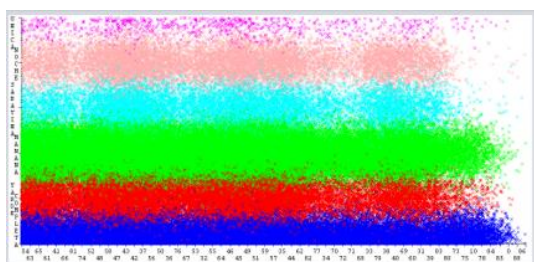


Fig. 34. Puntaje ingles según jornada

Ahora, al analizar la prueba Saber TyT, se encontraron como nuevos indicadores: el estado civil, el valor de matrícula, el tipo de pago matrícula y grupo de referencia de la carrera. Para el **estado civil**, llama la atención que las mujeres con respecto a los hombres son en mayor cantidad casadas, separadas y/o viudas, esto nos indica que ya con un hogar formado bastantes mujeres deciden estudiar carreras técnicas y tecnológicas.

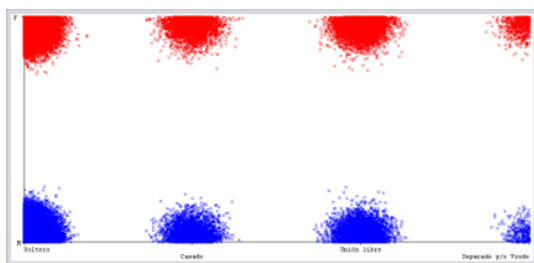


Fig. 35. Estado civil según género

Con el **valor de la matrícula**, podemos observar que el 58% del total de personas que presentaron la prueba no pagaron nada por la matrícula. De los que, si pagaron, más hombres que mujeres pagaron un valor menor a 500mil, entre 1 y 4 millones el balance está equilibrado para ambos géneros, entre 4 millones y 5.5 pagaron más hombres que mujeres, entre 5.5 y 7 millones es parejo y más de 7 millones el género femenino tuvo más pagos.

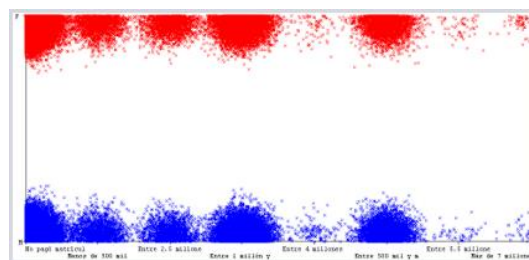


Fig. 36. Valor de la matrícula según género

Para el **tipo de pago de matrícula**, en el caso específico de crédito son más las mujeres que afirman pagar de este modo que los hombres que lo hacen.

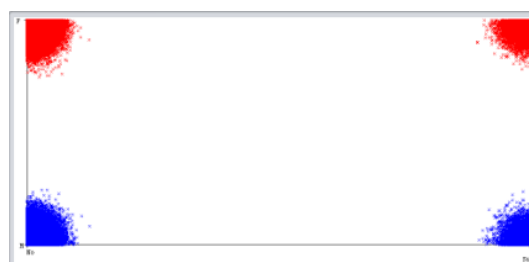


Fig. 37. Pago por crédito según género

Y como último indicador para esta prueba se tiene el **grupo de referencia**, que es el atributo que menciona las carreras de los estudiantes siendo agrupadas entre sus semejantes. Aquí se encontró que las carreras afines a la agronomía, electricidad, mecánica y minas son estudiadas en mayor cantidad por los hombres. Mientras que las mujeres prefieren estudiar en áreas afines a la salud y la administración.

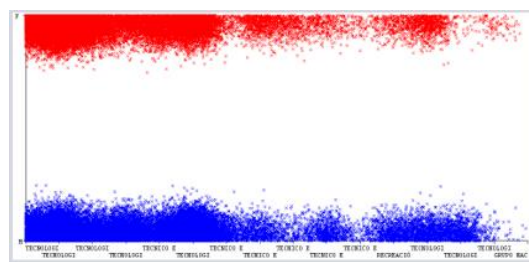


Fig. 38. Grupo de referencia según género

Para finalizar el análisis, fue aplicado el mismo algoritmo para las pruebas Saber Pro, que se presentan en las instituciones de educación superior. Allí los nuevos indicadores hallados son: Cabeza de familia, personas a cargo y lectura diaria. En el primero de ellos, **cabeza de familia**, se puede ver que hay mayor cantidad de mujeres que lo son con respecto a los hombres.

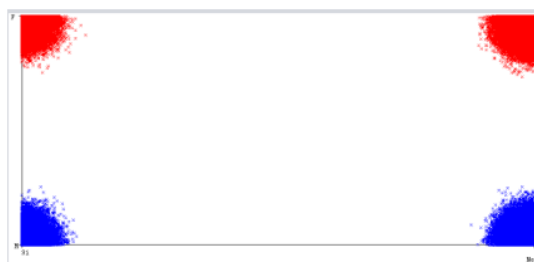


Fig. 39. Cabeza de familia según género

Por el **número de personas a cargo**, se puede apreciar que gran cantidad de los estudiantes no tienen personas a cargo, en especial las mujeres. Entre los que sí tienen encontramos que entre 1 y 8 personas el comportamiento entre ambos géneros no presenta diferencias significativas, para 9 personas en los hombres casi no hay registros mientras que en las mujeres sí, para 11 hay mayor cantidad en los hombres que en las mujeres y doce o más el género femenino vuelve a tener mayoría.

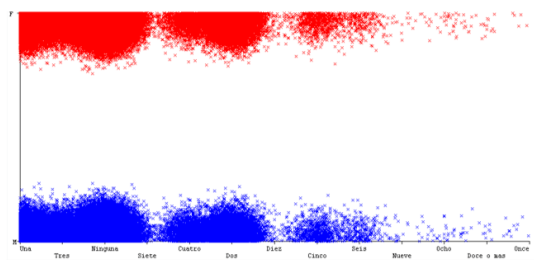


Fig. 40. Número de personas a cargo según género

Por último, con el indicador de **lectura diaria** se logró observar como las mujeres tienen mejores hábitos de lectura diaria, ya que en la mayoría de las opciones se ve un porcentaje significativo del género femenino.

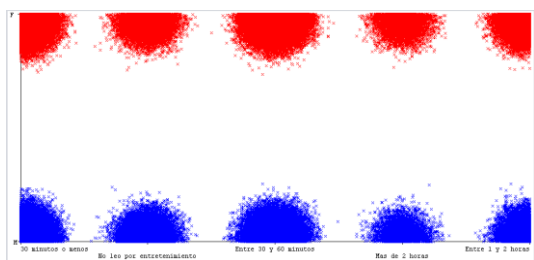


Fig. 41. Lectura diaria según género

4. CONCLUSIONES

Del análisis realizado con las 3 pruebas se puede concluir que los factores que influyen en las brechas de género están relacionados con los campos socioeconómicos y familiares.

Hay áreas donde se destacan las mujeres, hablando desde el punto de vista de participación en carreras afines a la administración, ciencias de

la salud, entre otras; mientras que hay grupos como las ingenierías y mecánica o afines donde el hombre se destaca. Esto está directamente relacionado con los puntajes obtenidos en estos mismos núcleos temáticos

El estrato es un indicador común para las 3 pruebas, es el principal factor de medición económica para los estudiantes, y está bastante relacionado con otros como el nivel de Sisbén, valor de matrícula, ingreso familiar, entre otros.

Los colegios y su departamento de ubicación son 2 de los indicadores que más movimiento presentaron, el caso de los colegios masculinos llama la atención de manera significativa para el estudio, puesto que presenta un comportamiento mucho más sesgado que el de los colegios femeninos. Además, que no tiende a ser muy común hallar este tipo de instituciones en varias partes del país.

Con la aplicación del algoritmo K-means se descartaron atributos que no presentaran cambios o diferencias siendo evaluados con respecto al género. Solo quedaron los que de alguna manera influyeron en las brechas.

En la actualidad, los jóvenes universitarios dependen económicamente de sus familias (Rojas, 2011) por lo cual se da a entender que las mujeres, al pedir mayores créditos es porque tienen más apoyo familiar.

REFERENCIAS

- López-Bassols, V., Grazzi, M., Guillard, C., & Salazar, M. (2018). Las brechas de género en ciencia, tecnología e innovación en América Latina y el Caribe. *Banco Interamericano de Desarrollo*, 52. <https://doi.org/10.18235/0001082>
- Rojas, Mauricio (2011) "Autonomía postergada: jóvenes, familia y educación superior". Recuperado el 20 de Junio de 2016. <http://www.redalyc.org/pdf/1942/194218961010.pdf>.