

Dashboard para el análisis de las pruebas Saber pro en la Universidad Mariana

Luis Carlos Benavides Betancourt
Oscar Alejandro Pantoja Pantoja
Estudiantes de Ingeniería de Sistemas
Universidad Mariana

Álvaro Alexander Martínez Navarro
Giovanni Albeiro Hernández Pantoja
Profesores de Ingeniería de Sistemas
Universidad Mariana

Introducción

En la actualidad, no hay una herramienta interactiva que brinde un análisis sobre las pruebas Saber pro en la Universidad Mariana de Pasto, por lo cual se desea implementar un tablero (*dashboard*) interactivo que permita analizar los resultados obtenidos y, que apoye la toma de decisiones por parte del usuario, para indagar más a fondo sobre los posibles fallos que pueda haber en las diferentes competencias genéricas que se evalúa en la prueba nacional.

La herramienta usada fue *Tableau*, cuya función principal es crear vistas para comparar una gran cantidad de información, además de ofrecer una interacción entre el tablero de control y el usuario, permitiendo así, poder filtrar diferentes tipos de datos que sean más relevantes y, agilizar la toma de decisiones.

En este trabajo se describe cómo se construyó el tablero, iniciando con la fase de selección de análisis y continuando con las fases de estructuración, tipos de visualizaciones, modelo de visualización de datos y, culminando con la fase de ajustes.

Fase de selección de análisis

Se hizo una revisión de documentos donde se encontró los análisis más relevantes a usar para la interpretación de las pruebas Saber pro: el análisis descriptivo y las tablas de contingencia. A continuación, se explicará cada uno, con su respectivo ejemplo.

Análisis descriptivo: es altamente usado para la interpretación de los datos de las pruebas Saber pro, dado que se encarga de analizar los datos, simplificando, recolectando y ordenando la información por medio de recursos visuales. Cuando se trabaja con grandes cantidades de datos, el análisis descriptivo es de gran ayuda porque resume la información, haciéndola más precisa, simplificada y clara (Ayala, 2017). El Instituto Colombiano para la Evaluación de la Educación Superior (ICFES) usa este tipo de análisis, ya que permite identificar los posibles patrones en las competencias genéricas y específicas.

Un ejemplo de análisis descriptivo se puede observar en la Tabla 1, donde se obtiene la cantidad de personas por facultad de una cierta universidad, que presentaron el examen Saber pro en un año específico (Ayala, 2017). Como se puede evidenciar, de manera inmediata se puede establecer varias conclusiones, por ejemplo, de la cantidad de estudiantes en cada facultad, en pro de tomar decisiones; o, se puede inferir que la Facultad de Ciencias de la Educación y Estudios a Distancia son las que mayor número de estudiantes presentan para las pruebas, con el fin de estudiar las dependencias que se puede encontrar en las variables observadas.

Tabla 1*Facultades que presentaron las pruebas Saber pro*

Facultad	2012	2013	2014	2015
Ciencias Agropecuarias	10,33	10,14	10,26	10,13
Ciencias Básicas	10,48	10,21	9,81	10,06
Ciencias de la Educación	10,35	10,22	10,25	10,31
Ciencias de la Salud	10,77	10,72	10,40	10,56
Ciencias Económicas y Administrativas	10,52	10,29	10,43	10,43
Derecho	11,04	10,77	10,78	10,78
Ingeniería	10,27	10,05	10,19	10,15
Estudios a Distancia	10,37	10,16	9,68	9,93
Duitama	10,14	10,02	10,10	10,24
Sogamoso	10,30	10,15	10,23	10,15
Chiquinquirá	10,14	10,12	10,06	10,02

Tablas de Contingencia: este análisis es usado en las pruebas Saber pro para resumir datos, por ejemplo, categóricos, que son creados al menos en dos filas y columnas, permitiendo analizar la relación entre dos o más variables que estén divididas de forma bidimensional (Acero Báez, 2017).

En la Tabla 2 se aprecia el uso de tablas de contingencia, donde se desea conocer la cantidad de personas, según el género, que se encuentran en cierto nivel de desempeño de inglés (Acero Báez, 2017).

Tabla 2*Género según el nivel de inglés*

Género	A-	A1	A2	B+	B1
F	8	29	24	3	24
M	20	28	33	5	23

Como se observa en la Tabla 2, los datos están ordenados por filas y columnas; además, se puede apreciar que cada una de las variables cumple un criterio; por ejemplo: el género femenino se encuentra en el nivel de inglés A2 con un puntaje de 24 puntos y así, con cada una de las variables, permitiendo analizar la relación entre dos o más variables.

Fase de estructuración del dashboard

Para llevar a cabo la segunda fase se hace uso de los análisis mencionados en la primera, con el objeto de determinar las visualizaciones del tablero; por ello, este tablero de control contará con dos tipos de visualizaciones, en donde el análisis descriptivo y las tablas de contingencia harán parte de las fichas técnicas, resultados globales, competencias genéricas, afirmaciones y niveles de desempeño, que permiten reducir la cantidad de datos recolectados bajo ciertos criterios, para luego ser presentados al usuario.

Para comenzar con la construcción del tablero, se ordenó los datos recolectados en una hoja de Excel y luego fueron importados en la herramienta *Tableau*; se hizo el debido relacionamiento de cada una de las tablas para poder acceder a la información de las cuatro hojas de trabajo, que fueron: ficha técnica, resultados globales, competencias genéricas, niveles de desempeño y afirmaciones.

La primera visualización es la ficha técnica; brinda información sobre los estudiantes que fueron inscritos en las pruebas Saber pro, que asistieron al examen y, los resultados publicados de la evaluación; la segunda visualización son los resultados globales; aquí se visualiza los promedios de los niveles de agregación como Colombia, institución

y el programa de Ingeniería de Sistemas; en la tercera visualización se aprecia las competencias genéricas del programa de Ingeniería de Sistemas; en la cuarta, están los niveles de desempeño de las competencias genéricas en los niveles de agregación, como: programa, institución y grupo de referencia, donde 1 es bajo, 2 medio, 3 alto y 4 muy alto. De igual manera, se tiene los niveles de desempeño de inglés, adoptando el modelo europeo donde -A1 es bajo, A2 medio, B1 alto y B2 muy alto. La quinta y última visualización son afirmaciones que miden las aseveraciones de cada una de las competencias, dando a conocer el porcentaje de estudiantes que erraron en cada criterio evaluado que, por cierto, están ordenados de menor a mayor complejidad. Para la culminación de esta fase, se organizó cada una de las visualizaciones.

Fase de tipos de visualizaciones

En esta fase se abordó los tipos de visualizaciones que se usó para el tablero; específicamente, se tuvo en cuenta el Visual Vocabulary (Digital Campbell et al., 2019), de donde se escogió tres categorías: Correlación, Cambio sobre el tiempo y Clasificación, de las cuales se seleccionó el diagrama por columna de la categoría Cambio vs. Tiempo y Clasificación, y el diagrama de burbujas que forma parte de la categoría Correlación (Kriebel, 2018). A continuación, se resume la función que cumple cada una de las visualizaciones seleccionadas.

Gráficos por columnas. Son una herramienta visual que describe la distribución de frecuencias de una variable de cada categoría. Una gran ventaja de este tipo de gráficos es que permite resumir un conjunto de datos e incluso, estos pueden ser comparados con otras **Referencias** a lo largo de un periodo, permitiendo dar un análisis más claro en la información (Kriebel, 2018).

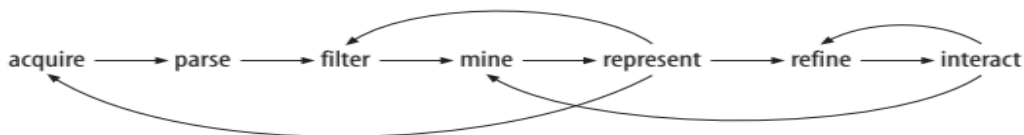
Gráfico de burbuja. Es un gráfico de dispersión que representa a varias variables en forma de burbuja y, permite mostrar tres dimensiones de datos (eje x, eje y, y el tamaño de la burbuja), lo que hace posible comparar y mostrar relaciones entre círculos categorizados, facilitando una mejor interpretación para el usuario (Kriebel, 2018).

Fase del modelo de visualización de los datos

Respecto al diseño del tablero, se optó por usar el modelo propuesto por Fry (2007), compuesto por siete etapas, como se aprecia en la Figura 1. A continuación, se explica cada una de las etapas del modelo mencionado.

Figura 1

Modelo de visualización de Ben Fry



Acquire (adquirir)

Esta fase de adquisición en el presente proyecto se hizo de forma indirecta; es decir, que se usó bases de datos públicas, como la del ICFES, para la descarga de reportes de las pruebas Saber pro en un rango de los últimos cinco años. A partir de los informes descargados se construyó matrices en Excel, de acuerdo con las necesidades del tablero propuesto.

Parse (estructurar y analizar)

En la segunda etapa se creó un repositorio de diez reportes, dividido en dos grupos: cinco por institución y cinco por programa. Una vez analizada la información, se dio paso a la organización de los datos de manera estructurada en libros de Excel, conformados por cuatro hojas de trabajo; en la primera se agrupó los datos de la ficha técnica y los resultados globales, donde se calificó las variables, de la siguiente manera: año, sede, nombre programa, inscritos, presentes, publicados, institución, Colombia y programa. En la segunda hoja se trabajó únicamente las competencias del programa de Ingeniería de Sistemas y se estableció las variables de: año, sede, nombre programa, nombre competencias y puntaje. En la tercera hoja se trabajó los niveles de desempeño, cuyas variables clasificadas



fueron: año, sede, nombre programa, institución, nombre de competencias, nivel de desempeño y valor. Para finalizar, en la cuarta hoja se trabajó afirmaciones que incluyeron las variables de: año, sede, nombre competencia, nombre programa, afirmación, institución y promedio

Filter (filtrado)

En la tercera etapa se aplicó filtros para las variables: años, que permite escoger un rango de tres años; sedes, seleccionando la sede de la institución; programas que se desea analizar y, competencia genérica en el tablero de afirmaciones. Con lo mencionado, se puede dar un mejor manejo sobre las enormes cantidades de datos que se desea analizar, permitiendo al usuario refinar los resultados de las variables que quiere estudiar.

Mine (extracción y transformación)

En la cuarta etapa se usó la estadística para la transformación de los datos, con la herramienta *Tableau*, que permite crear campos calculados sobre los datos almacenados en Excel, sin necesidad de modificarlos. Los campos calculados en las visualizaciones propuestas fueron: el dato valor de la tabla Niveles de desempeño y la variable promedio de la tabla Afirmaciones, con el fin de cambiar el tipo de dato porcentaje, a entero; por otro lado, se creó campos calculados para los *tooltip*, que fueron utilizados para brindar información más explícita al usuario, para las visualizaciones Nivel de desempeño y Afirmaciones.

Represent (representar)

En la quinta etapa, los tipos de gráficos a usar fueron: diagrama por columna y diagrama de burbujas, para una mejor interpretación de los datos por parte del usuario. El primero fue esencial para visualizar los datos de la presente investigación, ya que permitió trabajar con gran cantidad de datos para luego ser comparados con otras referencias a lo largo de un periodo de tiempo, logrando un análisis más claro. Por ello, este tipo de visualización fue usado para representar los datos de resultados globales, competencias genéricas del programa de Ingeniería de Sistemas y niveles de desempeño.

El diagrama de burbujas es un gráfico de dispersión que representa tres variables dentro de los ejes (eje x, eje y, y el tamaño de la burbuja), hecho que permitió comparar y mostrar relaciones en círculos, categorizados para una mejor interpretación del usuario; por ello, esta visualización fue usada para representar los datos Afirmaciones.

Refine (perfeccionamiento)

En la sexta etapa se habla sobre el tipo de diseño, especialmente color, que se va a abordar en las visualizaciones; se debe aclarar que, en algunos casos los colores usados fueron los mismos del ICFES, para las visualizaciones Nivel de desempeño y Afirmaciones. De esta forma, el usuario estará familiarizado con el tablero; estos colores son definidos así: marrón (muy bajo), rojo (bajo), tomate (medio), amarillo (alto) y verde (muy alto). Por otro lado, se usó colores cálidos para resaltar información y, colores naturaleza para los datos menos relevantes. (Sobre este tema se hablará a fondo en la fase de ajuste asignación del color).

Interact (interactividad)

En esta etapa, la interactividad aplicada al *dashboard* fue la exploratoria, que permite al usuario tener el control total de los datos para así poder interactuar con los rangos de años, sede, programa y competencia. (Sobre esto se hablará más a fondo en la fase 5).

Para finalizar, se puede decir que, el modelo de Fry (2007) fue de gran ayuda en este proyecto, ya que no solo aborda la estructuración de los datos, sino que también se enfoca en cómo se va a abordar las visualizaciones y cuáles serán sus interacciones con el usuario.

Fase de Ajuste del dashboard

En esta última fase se configura las propiedades de interactividad, color, formatos, campos calculados y despliegue del tablero. A continuación, se explicará cada uno de estos aspectos.

Interactividad

Un *dashboard* debe ser totalmente interactivo a la hora de su manipulación; por ello, se usó filtros en las diferentes visualizaciones, para que el usuario manejara a su antojo la información que desee analizar (Santos, 2016). Un ejemplo de interactividad fueron los rangos de años, que permitieron clasificar y presentar información en un periodo de tres años y así, brindarle al usuario la posibilidad de un mayor análisis.

Asignación de colores

Para la selección de colores se tuvo en cuenta los colores cálidos (amarillo, rojo, verde, entre otros) porque permiten mostrar información relevante para el usuario y, de esta manera, llamar su atención (Few, 2013).

Existen otros tipos de colores menos saturados, como los de la naturaleza (colores de la tierra y el cielo); su función es diferente a los colores cálidos, ya que se busca usar colores suaves y llamativos, pero sin la necesidad de opacar la información relevante (Few, 2013). De otro lado, se tiene el modelo HSL (por sus iniciales en inglés: *hue*, tono; *saturation*, saturación; y, *lightness*, luminosidad), encargado de hacer un degradado en los tonos de color de menor a mayor intensidad; su color varía con el puntaje obtenido (Few, 2013). A continuación, se explicará los colores asignados en cada una de las páginas del tablero y se dará ejemplos de su uso en las visualizaciones propuestas.

Primera página. Los tipos de colores usados fueron cálidos, como el tomate, para la gráfica de resultados globales y para las competencias genéricas, niveles de desempeño y afirmaciones. Por otra parte, se usó el modelo HSL en resultados globales y competencias del programa, para degradar el color de menor a mayor intensidad.

Segunda página. Los tipos de colores usados fueron cálidos, los mismos que usa el ICFES en su reporte como: café (muy bajo), rojo (bajo), naranja (medio), amarillo (alto) y verde (muy alto) permitiendo al usuario referenciar el color por cada nivel de desempeño.

Tercera página. Se usó lo propuesto por el ICFES: a los rangos de datos se les asigna un color cálido, dependiendo del porcentaje de errores; estos colores ya están definidos en los reportes y es por ello que se decidió usar los mismos patrones de colores: verde, si menos del 20 % de las respuestas de los estudiantes están erradas; amarillo, si están entre el 20 % y menos del 40 %; naranja, si están entre el 40 % y 70 %; y rojo, si más del 70 % de las respuestas de los estudiantes están erradas.

Formatear campos

Esta opción permitió dar formato a todos los elementos que se usa en las hojas de trabajo, como por ejemplo: alineación, utilizada para centrar los textos como los títulos, fuente y tamaño, utilizados para configurar el tipo de letra de los títulos, subtítulos de los ejes X y Y y, demás elementos de las visualizaciones; además se agregó bordes, para dar un mejor acabado en los tres tableros de control (Santos, 2016).

Campos calculados

Estos son nuevos campos que pueden ser utilizados dentro de la fuente de datos (Santos, 2016); estos datos fueron usados en los tableros de control Niveles de

desempeño, para cambiar el tipo de dato de porcentaje a entero, sin modificar el Excel; Afirmaciones, cuyo objetivo fue facilitar la interpretación dada al usuario por medio de ventanas emergentes, mostrando información adicional que ayuda a la interpretación de los datos.

Despliegue del dashboard

En esta última fase se explica cómo fue el despliegue del tablero; para ello se debe contar con una licencia para *Tableau desktop*; en este proyecto se usó la estudiantil; posterior a ello, se necesita crear una cuenta en *Tableau public* para subir el tablero de control al servidor mediante *Tableau desktop* (Santos, 2016). En el *dashboard* construido se puede interactuar con los datos de las pruebas Saber pro.

Referencias

- Acero Báez, Y. D. (2017). *Factores demográficos, socioeconómicos y académicos que influyen en los resultados del componente genérico de la prueba Saber pro. Caso Ingeniería Electromecánica de la Universidad Pedagógica y Tecnológica de Colombia, Facultad seccional Duitama* [Tesis de Pregrado, Universidad Pedagógica y Tecnológica de Colombia]. https://repositorio.uptc.edu.co/bitstream/001/2653/1/TGT_1267.pdf
- Ayala, Y. (2017). Análisis descriptivo de los resultados de las pruebas Saber pro en la UPTC del periodo 2012-2015. *Pensamiento y Acción*, (52), 59-77.
- Digital Campbell, Smithy, A., & Stabe, M. (2019). *Financial-Times / Chart-doctor*. <https://github.com/Financial-Times/chart-doctor/blob/main/visual-vocabulary/Visual-vocabulary.pdf>
- Few, S. (2013). *Information Dashboard Design: Displaying Data For At-a-Glance Monitoring* (2nd ed.). Analytics Press.
- Fry, B. (2007). *Visualizing data*. O'Reilly Media, Inc.
- Kriebel, A. (2018). *Visual Vocabulary*. <https://www.tableau.com/solutions/gallery/visual-vocabulary>
- Santos, D. (2016). *Tableau 10 Business Intelligence Cookbook*. Packt Publishing.

