

Asignatura	Datos del alumno	Fecha
Diplomado en Analítica de Datos en la Gestión Empresarial	Apellidos: López Fernández	09/10/2025
	Nombre: Douglas	

Fundación Universitaria Internacional de La Rioja

Pregrado en Ingeniería Informática

Diplomado - Analítica de Datos en la Gestión Empresarial



Actividad 1: Planteamiento de un Storytelling con Datos

Douglas López Fernández

Laura Marcela Barona Marmolejo

Deisy Johanna Villarraga Cuaycan

Oscar David Bocanegra Capera

GLENN ELMER HERNANDEZ CAMELO

2025

Asignatura	Datos del alumno	Fecha
Diplomado en Analítica de Datos en la Gestión Empresarial	Apellidos: López Fernández	09/10/2025
	Nombre: Douglas	

Propuesta Detallada de Storytelling con **Datos: Brechas de Equidad en la Educación Colombiana**

Este documento presenta una ampliación del planteamiento inicial para el desarrollo de una narrativa de datos, incorporando mayor detalle en la metodología, el tratamiento de los datos y la justificación del análisis, acorde a las exigencias académicas de un diplomado.

1. Tema Elegido

El tema seleccionado es el "***Desempeño de la educación en los jóvenes en Colombia***". Este proyecto se centrará específicamente en la evaluación de la brecha de rendimiento académico entre estudiantes de colegios de naturaleza pública (Oficial) y privada (No Oficial). La elección de este tema se fundamenta en su alta relevancia social y su impacto directo en las políticas públicas orientadas a garantizar la equidad y la calidad educativa, uno de los pilares para el desarrollo sostenible del país.

2. Objetivo de la Historia

El propósito de esta narrativa no es solo mostrar datos, sino construir un relato coherente y persuasivo. Para ello, se desglosa el objetivo general en varios específicos.

- **Objetivo Principal:** Demostrar y cuantificar la existencia de una brecha de rendimiento significativa en los resultados de las pruebas Saber 11 entre estudiantes de colegios públicos y privados en las principales capitales de Colombia. El mensaje central es que la naturaleza del colegio es un factor fuertemente asociado al desempeño académico, lo que sugiere inequidades estructurales en el sistema educativo.

Asignatura	Datos del alumno	Fecha
Diplomado en Analítica de Datos en la Gestión Empresarial	Apellidos: López Fernández	09/10/2025
	Nombre: Douglas	

- **Objetivos Específicos / Preguntas de Investigación:**

- 1) **Cuantificar la Diferencia:** ¿Cuál es la diferencia promedio en el puntaje global entre ambos tipos de instituciones?
- 2) **Analizar la Distribución:** ¿Cómo se distribuyen los puntajes dentro de cada grupo? ¿Existe una mayor dispersión o consistencia en alguno de los dos?
- 3) **Explorar Factores Asociados:** ¿Cómo se relaciona el estrato socioeconómico de los estudiantes con el puntaje global dentro de cada tipo de colegio? ¿Se acentúa o se mitiga la brecha en ciertos estratos?
- 4) **Validar Estadísticamente:** ¿Es la diferencia observada entre los promedios estadísticamente significativa o podría deberse al azar?

3. Audiencia y Enfoque Comunicativo

Identificar la audiencia es crucial para adaptar el lenguaje y la profundidad del mensaje. La historia se dirigirá a los siguientes grupos con enfoques específicos:

- **Gestores de Políticas Públicas (Ministerio de Educación, Secretarías de Educación):** Para este grupo, el enfoque será cuantitativo y conclusivo. Se les presentarán los resultados clave de forma clara, con visualizaciones que comparen promedios y distribuciones. El llamado a la acción se centrará en la necesidad de focalizar recursos y diseñar estrategias para fortalecer la educación pública.
- **Comunidad Académica (Rectores, Docentes e Investigadores):** A esta audiencia se le ofrecerá un análisis más detallado, incluyendo la metodología, el tratamiento de los datos y las pruebas estadísticas.

Asignatura	Datos del alumno	Fecha
Diplomado en Analítica de Datos en la Gestión Empresarial	Apellidos: López Fernández	09/10/2025
	Nombre: Douglas	

El objetivo es que puedan usar esta evidencia para contextualizar sus propios resultados y promover mejoras pedagógicas.

- ***Público General y Medios de Comunicación:*** El mensaje se simplificará, utilizando visualizaciones de alto impacto (como gráficos de barras comparativos y mapas) y un lenguaje accesible. El objetivo es generar conciencia social sobre la inequidad educativa y fomentar un debate informado.

4. Recolección y Preparación de Datos

La calidad del análisis depende directamente de la rigurosidad en la preparación de los datos.

- **Fuentes de Datos:**
 - **Fuente Principal:** El conjunto de datos "Resultados Saber 11" del portal de Datos Abiertos del Estado Colombiano, administrado por el ICFES. Se utilizará la publicación más reciente para garantizar la actualidad del análisis.
 - **Fuente Secundaria (Opcional):** Datos demográficos del DANE para enriquecer el contexto socioeconómico de los municipios analizados.

Asignatura	Datos del alumno	Fecha
Diplomado en Analítica de Datos en la Gestión Empresarial	Apellidos: López Fernández	09/10/2025
	Nombre: Douglas	

- **Variables de Interés y su Rol:**

Variable	Descripción	Rol en el Análisis
PUNT_GLOBAL	Puntaje total obtenido en la prueba.	Variable Dependiente (el resultado a medir).
COLE_NATURALEZA	Tipo de institución (Oficial / No Oficial).	Variable Independiente Principal (el factor a comparar).
FAMI ESTRATOVIVIENDA	Estrato socioeconómico de la familia.	Variable Independiente Secundaria (para análisis de segmentación).
COLE_MCPIO_UBICACION	Municipio donde se ubica el colegio.	Variable de Control / Filtro (para acotar el estudio).
COLE_BILINGUE	Indicador de si el colegio es bilingüe.	Variable de Contexto (puede ayudar a explicar diferencias).

- **Proceso de Tratamiento de Datos (ETL - Extract, Transform, Load):**

- **Extracción:** Descarga del archivo CSV desde el portal de datos abiertos.

- **Limpieza y Transformación:**

- 1) **Manejo de Valores Nulos:** Se realizará un análisis de valores ausentes. Los registros sin PUNT_GLOBAL serán eliminados por ser inservibles para el objetivo. Para la variable FAMI ESTRATOVIVIENDA, se podría considerar la imputación por la moda del municipio si el porcentaje de nulos es bajo, o ser categorizado como "Sin Información".
- 2) **Filtrado de Datos:** Se aplicará un filtro para seleccionar únicamente los registros de las cinco principales ciudades de Colombia (Bogotá, Medellín, Cali, Barranquilla, Cartagena) para crear un estudio comparativo y manejable.
- 3) **Normalización de Texto:** Se unificarán las categorías de texto (ej. "BOGOTA D.C." y "BOGOTÁ") para evitar duplicidades en las agrupaciones.

Asignatura	Datos del alumno	Fecha
Diplomado en Analítica de Datos en la Gestión Empresarial	Apellidos: López Fernández	09/10/2025
	Nombre: Douglas	

- 4) **Creación de Variables Derivadas (Feature Engineering):** Se podría crear una variable binaria ES_PUBLICO (1 si es Oficial, 0 si es No Oficial) para facilitar su uso en modelos estadísticos.

5. Método de Análisis y Justificación

Se empleará un análisis cuantitativo mixto (descriptivo, visual y confirmatorio), ya que esta combinación es la que mejor permite no solo identificar, sino también validar y comunicar los hallazgos de forma robusta y convincente.

• Fase 1: Análisis Descriptivo

- **Técnica:** Se calcularán estadísticas descriptivas fundamentales (media, mediana, desviación estándar, cuartiles, valores mínimos y máximos) para el PUNT_GLOBAL, agrupando por COLE_NATURALEZA.
- **Justificación:** Esta fase es esencial para obtener una primera radiografía numérica de la situación. Permite responder a la pregunta: A grandes rasgos, ¿cuáles son los números que definen el rendimiento de cada grupo?

• Fase 2: Análisis Visual Comparativo

- **Técnica:** Se utilizarán múltiples gráficos para contar la historia visualmente:
- **Box Plots (Diagramas de Cajas):** Es la herramienta visual clave. Permitirá comparar de forma simultánea la mediana (centro), los rangos intercuartílicos (dispersión del 50% central) y los valores atípicos de los puntajes entre colegios públicos y privados.

Asignatura	Datos del alumno	Fecha
Diplomado en Analítica de Datos en la Gestión Empresarial	Apellidos: López Fernández	09/10/2025
	Nombre: Douglas	

- **Histogramas y Gráficos de Densidad:** Se generará un gráfico por cada tipo de colegio para observar la forma de la distribución de sus puntajes. Esto ayuda a ver si, por ejemplo, los puntajes en colegios privados se concentran en la parte alta, mientras que en los públicos la distribución es más amplia o sesgada.
- **Justificación:** Las visualizaciones son la columna vertebral del storytelling. Traducen números complejos en una narrativa intuitiva y fácil de comprender para todas las audiencias.
- **Fase 3: Análisis Inferencial (Confirmatorio)**
 - **Técnica:** Se realizará una Prueba T de Student para muestras independientes. Esta prueba estadística permite determinar si la diferencia observada en los puntajes promedio entre los dos grupos es estadísticamente significativa.
 - ✓ **Hipótesis Nula ():** La media de los puntajes globales de los colegios oficiales es igual a la media de los colegios no oficiales ().
 - ✓ **Hipótesis Alternativa ():** Las medias de los puntajes globales son diferentes ().
 - **Justificación:** Este paso eleva el rigor del estudio más allá de una simple descripción. Permite afirmar con un nivel de confianza estadística (usualmente 95%) que la brecha observada no es producto de la casualidad, lo cual es fundamental para que los hallazgos sean tomados en serio por audiencias especializadas y tomadores de decisiones.