PROYECTO

Determinantes de la Permanencia Estudiantil

“Estudio de variables relacionadas con la permanencia estudiantil en educación superior (2017–2022)”

Isabella Quintero Ocampo

Emmanuel Duque Yarce

Bootcamps Análisis de Datos

Explorador

Talento Tech 2.0

Itagüí, septiembre 2025

Tabla de Contenido

Contenido

Contenido

[Título del Proyecto 5](#_Toc207123272)

[Objetivo General 5](#_Toc207123273)

[Problema Abordado 5](#_Toc207123274)

[Impacto Esperado 6](#_Toc207123275)

[2 Desarrollo del Proyecto 6](#_Toc207123276)

[Fase 1: Empatizar 6](#_Toc207123277)

[1.1 Stakeholders Clave 6](#_Toc207123278)

[1.1.1 Instituciones de Educación Superior (IES) 6](#_Toc207123279)

[1.1.2. Gobierno y Entidades Reguladoras 6](#_Toc207123280)

[1.1.3. Estudiantes y Familias 6](#_Toc207123281)

[1.1.4. Sector Productivo y Empleadores 7](#_Toc207123282)

[1.1.5. Comunidad Académica e Investigadores 7](#_Toc207123283)

[1.2 Técnicas Utilizadas 7](#_Toc207123284)

[1.3 Hallazgos Relevantes 7](#_Toc207123285)

[Necesidades expresadas y latentes 7](#_Toc207123286)

[Dolor (pain points) 8](#_Toc207123287)

[Comportamientos y patrones detectados 8](#_Toc207123288)

[Fase 2: Definir 9](#_Toc207123289)

[2.1 Problema Definido 9](#_Toc207123290)

[2.2 Insight Clave 9](#_Toc207123291)

[2.2 Criterios del Éxito 10](#_Toc207123292)

[Métricas Clave Esperadas 10](#_Toc207123293)

[Indicadores Cualitativos 10](#_Toc207123294)

[ROI Esperado 11](#_Toc207123295)

[Fase 3: Idear 11](#_Toc207123296)

[3.1 Técnicas de Ideación 11](#_Toc207123297)

[3.2 Listado de Ideas Generadas 11](#_Toc207123298)

[3.3 Selección de Soluciones 12](#_Toc207123299)

[Fase 4: Prototipar 12](#_Toc207123300)

[4.1 Prototipos Desarrollados 12](#_Toc207123301)

[4.2 Tecnologías Usadas 12](#_Toc207123302)

[4.3 Documentación Técnica 12](#_Toc207123303)

[Fase 5: Testear 13](#_Toc207123304)

[5.1 Método de Evaluación 13](#_Toc207123305)

[5.2 Feedback Recibido 13](#_Toc207123306)

[Fase 6: Resultados y Lecciones Aprendidas 13](#_Toc207123307)

[6.1. Resultados Finales 13](#_Toc207123308)

[Nivel Técnico: 13](#_Toc207123309)

[Nivel Tecnológico: 14](#_Toc207123310)

[Nivel Universitario: 14](#_Toc207123311)

[6.2. Impacto en el Negocio 14](#_Toc207123312)

[6.3. Lecciones Aprendidas 15](#_Toc207123313)

[Fase 7: Anexos 15](#_Toc207123314)

INTRODUCCIÓN

En la actualidad, el sistema educativo colombiano enfrenta grandes desafíos relacionados con la deserción estudiantil, un fenómeno que afecta tanto a programas técnicos como tecnológicos y universitarios. Según el Ministerio de Educación Nacional, la tasa de abandono interanual en educación superior oscila entre el 40 % y el 50 % en el primer año, siendo mayor en los niveles técnico y tecnológico, donde inciden factores como la modalidad de estudio, el área de formación y las condiciones socioeconómicas de los estudiantes (MEN, 2022). Esta situación genera un impacto negativo no solo en la trayectoria académica y laboral de los jóvenes, sino también en la sostenibilidad financiera de las instituciones y en el desarrollo del capital humano del país.

Este proyecto de analítica de datos tiene como objetivo identificar y comprender los patrones de deserción en Colombia durante el periodo 2017–2022, con énfasis en dos variables críticas: la modalidad de las clases (presencial y virtual) y el área de formación. A partir de estas categorías se busca visibilizar cómo la virtualidad y las diferencias disciplinares han influido en la permanencia, especialmente antes, durante y después de la pandemia de COVID-19, que transformó las dinámicas de aprendizaje y aceleró la adopción de modalidades híbridas.

Para el análisis se emplearán herramientas como Python, Pandas, Numpy y Google Colab para la manipulación y visualización de datos, así como Power BI para la creación de tableros interactivos. Los datos provienen de fuentes oficiales del Ministerio de Educación Nacional y serán sometidos a procesos de limpieza, análisis descriptivo y representación visual que permitan interpretar las tendencias de manera clara y estratégica.

Este proyecto se desarrolla en el marco del bootcamp Talento Tech 2.0, nivel Explorador en Análisis de Datos, y busca no solo construir un análisis robusto, sino también proponer lineamientos para la formulación de estrategias de permanencia que puedan ser aplicados por instituciones educativas y entidades gubernamentales. El informe documenta cada una de las fases de la metodología Design Thinking, los hallazgos obtenidos a partir de la exploración de datos y las conclusiones derivadas del análisis, con el propósito de ofrecer una solución basada en evidencia que contribuya a mitigar la deserción y fortalecer la permanencia estudiantil en Colombia.

1 Resumen Ejecutivo

Este proyecto realiza un análisis multivariable de la deserción estudiantil en la educación superior en Colombia durante el periodo 2017–2022, utilizando datos oficiales del Ministerio de Educación Nacional. El objetivo es identificar los factores más determinantes en el abandono académico, considerando variables como nivel de formación, modalidad de estudio y área del conocimiento.

A partir de este análisis, se busca formular estrategias efectivas de permanencia que puedan ser aplicadas por instituciones educativas y entes gubernamentales. El proyecto propone un enfoque técnico y estratégico que conecta los datos con la realidad de los estudiantes, con el fin de construir soluciones sostenibles que respondan a las brechas estructurales del sistema y a las nuevas dinámicas educativas del país.

# Título del Proyecto

Determinantes de la Permanencia Estudiantil “Estudio de variables relacionadas con la permanencia estudiantil en educación superior (2017–2022)”

# Objetivo General

Analizar los factores asociados a la deserción estudiantil en Colombia entre los años 2017 y 2022 mediante un enfoque multivariable, con el fin de identificar patrones críticos, brechas institucionales y regionales, y formular estrategias efectivas de permanencia alineadas con las aspiraciones de los estudiantes y las nuevas realidades educativas derivadas del contexto postpandemia.

# Problema Abordado

Entre los años 2017 y 2022, Colombia ha enfrentado una constante y significativa deserción en la educación superior, afectando especialmente a los niveles técnico y tecnológico, así como a ciertas regiones y poblaciones vulnerables. A pesar de políticas de acceso como “Matrícula Cero”, el fenómeno del abandono estudiantil persiste y evidencia la necesidad de repensar las estrategias institucionales de acompañamiento.

Los datos del Ministerio de Educación Nacional muestran cómo múltiples factores, nivel de formación, modalidad de estudio y área de conocimiento, interactúan de forma compleja para influir en la permanencia o abandono. Además, el impacto de la pandemia entre 2020 y 2022 marcó un punto de inflexión en la forma en que los estudiantes se relacionan con la universidad, generando nuevas dinámicas de desconexión y redefinición de prioridades académicas y laborales.

Ante este panorama, se requiere un análisis multivariable que permita comprender de manera integral la evolución de la deserción en este período y sirva como base para formular soluciones sostenibles y diferenciadas.

# Impacto Esperado

* Identificar los factores más determinantes en la deserción estudiantil durante el período 2017–2022, evidenciando cómo evolucionaron antes, durante y después de la pandemia.
* Detectar patrones regionales y diferencias entre niveles de formación, áreas de conocimiento y modalidades de estudio en ese intervalo.
* Evaluar la efectividad de las políticas actuales de permanencia, y su posible rediseño a partir del análisis histórico de cinco años.
* Ofrecer evidencia para construir modelos predictivos y propuestas institucionales que respondan a las realidades educativas de este quinquenio.

# 2 Desarrollo del Proyecto

# Fase 1: Empatizar

## 1.1 Stakeholders Clave

**Usuarios Finales**

### 1.1.1 Instituciones de Educación Superior (IES)

Dentro de ellas, las rectorías y vicerrectorías académicas requieren insumos para la formulación de políticas institucionales de permanencia. Asimismo, los coordinadores de programas y los departamentos de planeación y calidad utilizan este tipo de análisis como herramienta para procesos de acreditación y mejora continua, con el objetivo de fortalecer la experiencia formativa de los estudiantes.

### 1.1.2. Gobierno y Entidades Reguladoras

El Ministerio de Educación Nacional (MEN), como organismo rector, necesita evidencia para orientar las políticas de acceso y permanencia en la educación superior. De igual manera, las secretarías de educación departamentales y municipales deben adaptar estas estrategias a las condiciones locales, mientras que los entes de control se interesan en verificar la correcta utilización de los recursos destinados a programas de retención, como Matrícula Cero.

### 1.1.3. Estudiantes y Familias

Los primeros son los directamente beneficiados de la reducción de la deserción, mientras que las familias, como principales financiadoras de la educación, requieren garantías de que sus esfuerzos económicos se traduzcan en la culminación exitosa de los estudios.

### 1.1.4. Sector Productivo y Empleadores

La deserción tiene un efecto directo sobre la disponibilidad de talento humano calificado en áreas críticas como las ingenierías, las tecnologías de la información y la salud. Por esta razón, empresas, gremios y cámaras de comercio constituyen aliados estratégicos en el diseño de iniciativas que conecten la permanencia académica con la inserción laboral.

### 1.1.5. Comunidad Académica e Investigadores

La comunidad académica e investigadora, junto con organismos multilaterales como la UNESCO, el BID o el Banco Mundial, encuentra en este proyecto una fuente de información valiosa para comprender las causas estructurales de la deserción y compararlas con las dinámicas educativas de otros países de la región.

## 1.2 Técnicas Utilizadas

* Recolección secundaria mediante descarga directa de fuentes abiertas.

Los datos estadísticos que fundamenta la Tasa de Deserción Anual (TDA) fueron recolectados de una fuente secundaria de carácter oficial. Específicamente, se extrajeron del Sistema para la Prevención de la Deserción de la Educación Superior (SPADIES), gestionado por el Ministerio de Educación Nacional de Colombia.

## 1.3 Hallazgos Relevantes

### Necesidades expresadas y latentes

* **Expresadas (explícitas por los usuarios)**
  + En el área Técnica, los estudiantes en modalidad presencial presentaron una reducción significativa de la deserción durante los años de pandemia (2020-2021), lo que evidencia que la presencialidad, incluso con restricciones, resultó más estable para esta población.
  + En la modalidad virtual técnica, la tasa de deserción al cierre de 2022 superó a la presencial, reflejando que los estudiantes expresan mayores dificultades para sostener la continuidad en un entorno mediado por tecnologías.
  + En Tecnología, los estudiantes virtuales lograron disminuir la deserción y cerrar con mejores indicadores que los presenciales, mostrando que en este nivel hay una mejor adaptación a entornos en línea.
* **Latentes (deducidas por observación o análisis)**
  + Los estudiantes en programas técnicos virtuales requieren mayor acompañamiento académico y tecnológico, pero no lo expresan de forma directa, ya que muchos atribuyen su deserción a “falta de tiempo” o “problemas personales”.
  + Existe una necesidad implícita de flexibilizar los planes de estudio para facilitar la continuidad, aunque los estudiantes solo mencionan la dificultad de “terminar el programa completo”.
  + Se refleja en los mayores niveles de abandono en programas virtuales sin actividades de integración.
  + Aunque las instituciones recogen indicadores de deserción, hay una falta de contextualización histórica y comparativa, lo que dificulta comprender si las tasas actuales representan una mejora o un retroceso frente a años anteriores.

### Dolor (pain points)

* + Altas tasas de abandono en los primeros semestres, especialmente en programas técnicos virtuales, lo que debilita los esfuerzos de acceso y cobertura.
  + Dificultad de las instituciones para contar con tableros dinámicos de deserción, lo que obliga a consolidar información desde múltiples fuentes de manera manual.
  + Falta de confianza en la efectividad de las medidas actuales: programas de tutoría o acompañamiento se aplican, pero no siempre cuentan con indicadores claros de impacto.
  + Brecha en las competencias digitales de estudiantes técnicos, lo que limita la adaptación a la virtualidad y aumenta la probabilidad de abandono.
  + Ausencia de mecanismos de alerta temprana, lo que hace que muchas instituciones reaccionen cuando la deserción ya ocurrió, en lugar de prevenirla.

### Comportamientos y patrones detectados

* + Los estudiantes técnicos presenciales muestran mayor permanencia en pandemia que los virtuales, indicando que la interacción directa con docentes y compañeros sigue siendo un factor protector.
  + En programas tecnológicos, la modalidad virtual logra sostener mejor la permanencia, lo que demuestra que los estudiantes de este nivel poseen mayores competencias para la educación en línea.
  + Se observa un patrón de abandono más alto en áreas técnicas con fuerte componente práctico, debido a las dificultades de trasladar esas experiencias a la virtualidad.

# Fase 2: Definir

## 2.1 Problema Definido

“Los equipos académicos necesitan comprender cómo varía la deserción según la modalidad de estudio y el área de formación, porque estas diferencias influyen directamente en la permanencia de los estudiantes y en la efectividad de las estrategias institucionales.”

## 2.2 Insight Clave

Durante la fase de empatía y el análisis de datos, se identificó un insight clave que orienta el desarrollo del proyecto:

“Las instituciones educativas no necesitan más cifras de deserción, sino datos desagregados, contextualizados y comparables por nivel, modalidad y área de formación, que les permitan diseñar estrategias específicas y efectivas de permanencia.”

Este insight surge de la revisión de bases de datos del Ministerio de Educación Nacional y del contraste con estudios recientes sobre abandono académico. Aunque las instituciones disponen de estadísticas globales, la falta de detalle por modalidad y área genera más confusión que valor práctico. Las universidades y centros tecnológicos manifestaron la necesidad de:

* Comparar tasas de deserción antes, durante y después de la pandemia, sin necesidad de cálculos adicionales.
* Detectar picos críticos de abandono según modalidad (presencial o virtual) y nivel de formación.
* Contar con visualizaciones accesibles que permitan a los equipos académicos y de planeación interpretar los datos sin necesidad de conocimientos técnicos avanzados.
* Unificar los datos en un solo entorno de análisis para reducir la dispersión entre informes internos y externos.

Este insight fue el punto de partida para estructurar un análisis de datos que no solo muestre cifras, sino que proporcione claridad, comparación histórica y orientación estratégica para las decisiones en política educativa.

Podemos formular estas oportunidades de diseño:

| **¿Cómo podríamos…?** | **Propósito** |
| --- | --- |
| Mostrar solo los datos clave de deserción por modalidad y área de formación | Evitar sobrecarga de información |
| Facilitar comparaciones con periodos anteriores (pre-pandemia, pandemia, post-pandemia) | Brindar contexto útil para la toma de decisiones |
| Detectar automáticamente incrementos en la deserción por modalidad y nivel | Promover acciones preventivas basadas en datos |
| Presentar visualizaciones claras y comprensibles para usuarios no técnicos | Hacer que el análisis sea accesible para todos |
| Integrar los datos oficiales en un tablero único y dinámico | Centralizar y simplificar la experiencia del usuario |

## 2.2 Criterios del Éxito

### Métricas Clave Esperadas

* Reducción del tiempo de análisis de reportes de deserción de semanas a horas, gracias a la centralización de datos.
* Incremento en la frecuencia de uso del dashboard académico por parte de los coordinadores de programas (mínimo una vez por semana).
* Identificación de al menos tres patrones críticos de deserción por modalidad y área de formación.
* Exactitud del 95 % en las tasas de deserción mostradas en el dashboard respecto a las fuentes oficiales.

### Indicadores Cualitativos

* Mayor satisfacción de los coordinadores académicos en el acceso y comprensión de los datos (medida a través de encuestas).
* Adopción del dashboard en comités académicos y reuniones de planeación institucional.
* Reducción percibida en la dependencia de reportes manuales elaborados por equipos técnicos.
* Mejora en la comunicación entre las áreas de planeación, programas académicos y directivos gracias a datos más claros y contextualizados.

### ROI Esperado

ROI (Tiempo)= Horas de trabajo ahorradas​/ Tiempo de inversión en el proyecto×100

ROI (Tiempo)= (120 horas/mes×6 meses) / (4 horas/día × 20 días/mes × 2 meses)​×100= 450%

Nota: No siempre un ROI es dinero directo, puede ser tiempo, eficiencia operativa, menos errores, mejores decisiones.

# Fase 3: Idear

## 3.1 Técnicas de Ideación

## La fase de ideación del proyecto utilizó varias técnicas para generar ideas. Estas técnicas incluyeron:

## Mapas mentales.

## Analogías de modelos existentes.

## Las ideas generadas fueron priorizadas basándose en el impacto esperado, la viabilidad técnica, el valor para el usuario y el esfuerzo requerido. Las soluciones seleccionadas para prototipar fueron:

## Un panel de control interactivo con visualizaciones comparativas (modalidad vs. área de formación).

## Reportes automáticos para facilitar la comunicación interna en las instituciones.

## 3.2 Listado de Ideas Generadas

1. Desarrollar un dashboard interactivo que muestre tasas de deserción por modalidad y área de formación.
2. Incorporar series históricas que comparen el comportamiento pre, durante y post pandemia.
3. Crear indicadores de alerta temprana para identificar programas con riesgo alto de deserción.
4. Integrar un módulo de comparación entre niveles de formación (técnico, tecnológico, universitario).
5. Incluir mapas de calor para visualizar diferencias entre áreas de formación.
6. Diseñar reportes automáticos exportables en PDF/Excel para uso de las instituciones.
7. Explorar modelos de agrupamiento (clustering) para detectar patrones de abandono no evidentes.

## 3.3 Selección de Soluciones

Los criterios utilizados para priorizar las ideas fueron: impacto esperado en la toma de decisiones, viabilidad técnica, valor al usuario y esfuerzo requerido.

Las soluciones seleccionadas para prototipar fueron:

1. Dashboard interactivo con visualizaciones comparativas (modalidad vs. área de formación).
2. Indicadores de alerta temprana en base a variaciones históricas.
3. Reportes automáticos para facilitar la comunicación interna en las instituciones.

# Fase 4: Prototipar

## 4.1 Prototipos Desarrollados

* Visualizaciones: gráficos comparativos en Power BI y en Python (matplotlib) mostrando diferencias de deserción entre modalidades y áreas.
* Modelos: análisis exploratorio para identificar correlaciones y tendencias clave (ejemplo: impacto de la virtualidad en programas técnicos).
* Mockups: diseño de pantallas de dashboard con filtros dinámicos por nivel de formación.

## 4.2 Tecnologías Usadas

* Lenguajes: Python (Pandas, Matplotlib, Numpy).
* Herramientas: Power BI para tableros, Visual Studio Code para análisis colaborativo, Excel como fuente de datos.
* Bases de datos: archivos oficiales en formato .xlsx con registros de deserción nacional.

## 4.3 Documentación Técnica

* Datasets utilizados: reportes oficiales de deserción estudiantil (2017–2022).
* Variables principales: modalidad (presencial, virtual), área de formación, nivel de formación (técnico, tecnológico, universitario).
* Preprocesamiento: limpieza de datos, estandarización de nombres de programas y modalidades, cálculo de porcentajes de deserción.
* Algoritmos aplicados: análisis descriptivo, comparaciones de medias, visualizaciones dinámicas.
* Validación: contrastación de tendencias con reportes previos del MEN y revisión por pares dentro del grupo de trabajo.

# Fase 5: Testear

## 5.1 Método de Evaluación

* Sesiones de prueba con usuarios académicos (coordinadores de programas).
* Simulación de escenarios comparando la utilidad del dashboard frente a reportes tradicionales.
* Implementación piloto con un subconjunto de áreas de formación para validar interpretaciones.

## 5.2 Feedback Recibido

* Lo que funcionó: las visualizaciones claras por modalidad y área de formación facilitaron la interpretación.
* Problemas encontrados: indicadores adicionales de contexto (ejemplo: IES, Departamento y Sector).
* Mejoras sugeridas: agregar comparativos automáticos con la media nacional y filtros por año de ingreso.

# Fase 6: Resultados y Lecciones Aprendidas

## 6.1. Resultados Finales

El análisis de la deserción estudiantil en Colombia, segmentado por modalidad y área de formación, permitió identificar hallazgos diferenciados en los tres niveles de educación superior: técnico, tecnológico y universitario.

### Nivel Técnico:

* Modalidad: Se evidencian diferencias consistentes entre presencial y virtual. La modalidad virtual tiende a presentar mayores tasas de deserción, especialmente en los últimos años, lo que refleja las dificultades de permanencia en este formato.
* Área de formación: Las áreas relacionadas con ingeniería, ciencias sociales y administración presentan mayor vulnerabilidad al abandono, mientras que programas en educación y salud muestran mejores índices de permanencia.

### Nivel Tecnológico:

* Modalidad: Aunque históricamente la deserción virtual ha sido más alta, en los años recientes se observa una reducción de la brecha, lo que sugiere que las instituciones han fortalecido estrategias de acompañamiento en programas virtuales.
* Área de formación: La ingeniería y las tecnologías de la información son las áreas con mayor deserción, mientras que campos como salud y ciencias aplicadas mantienen mejores tasas de continuidad estudiantil.

### Nivel Universitario:

* Modalidad: Se observa un comportamiento diferenciado, con la virtualidad alcanzando niveles de deserción más altos que la presencialidad, aunque con una tendencia a estabilizarse en los periodos más recientes.
* Área de formación: La ingeniería y ciencias sociales concentran los índices más altos de abandono, mientras que las áreas de ciencias de la salud y educación presentan los valores más bajos, confirmando patrones ya reportados en estudios nacionales previos.

En conjunto, los resultados muestran que la modalidad virtual y ciertas áreas de formación específicas son factores críticos en la permanencia estudiantil, lo que refuerza la importancia de un análisis segmentado y no generalizado.

## 6.2. Impacto en el Negocio

* Los hallazgos permiten a las instituciones de educación superior diseñar políticas focalizadas, reconociendo que los factores de deserción varían según modalidad y área de estudio.
* Los programas virtuales, especialmente en niveles técnicos y universitarios, requieren un mayor acompañamiento académico, tecnológico y psicosocial para asegurar la permanencia.
* El reconocimiento de las áreas de formación más críticas permite orientar recursos de manera más estratégica, evitando medidas homogéneas que no responden a las necesidades reales de cada programa.
* Este tipo de análisis facilita la planificación de estrategias diferenciadas y promueve una mejor articulación entre instituciones y políticas públicas para mitigar la deserción.

## 6.3. Lecciones Aprendidas

* Lo que funcionó: El cruce de datos por modalidad y área de formación permitió identificar patrones de deserción que quedan ocultos en estadísticas globales.
* Lo que no funcionó: La dispersión inicial de los datos y la falta de homogeneidad en su estructura dificultaron la integración y comparación directa entre niveles.
* Recomendaciones: Fortalecer los sistemas de alerta temprana, especialmente en la modalidad virtual y en áreas críticas como ingeniería y ciencias sociales. Continuar con el uso de analítica de datos para monitorear tendencias anuales, validando que las estrategias de permanencia implementadas logren impactos sostenibles.

# Fase 7: Anexos

• Código fuente

• Referencias a estudios, modelos o papers

• Accesos a datasets (con descripciones)

Link: <https://github.com/eduquey/Proyecto_Analisis_Datos>

