

Integración de la Analítica de Datos en la enseñanza de Estadística I mediante el diseño de laboratorios computacionales orientados al desarrollo de competencias

Paula Juliana Hernández Herrera

Trabajo de Grado para Optar el Título de Ingeniería Industrial

Director

Iván David Ortiz Pineda

Candidato a Doctor en Ciencias de la Computación

Universidad Industrial de Santander

Facultad de Ingenierías Fisicomecánicas

Escuela de Estudios Industriales y Empresariales

Bucaramanga

2026

Dedicatoria

Dedico este trabajo, en primer lugar, a las dos mujeres que son el pilar de mi vida: mi madre, Sandra Liliana, y mi abuela, Eddy Cecilia. Gracias por darme alas a través del estudio, por sus sacrificios y por entregarme todo para que este sueño fuera posible. Este triunfo es tan mío como suyo.

A Dios, por acompañarme en cada paso de este camino, por ser mi guía y por sostenerme en los momentos más difíciles de mi vida universitaria.

De manera muy especial, a Greta, Iván y Elieth. Cuando tuve la oportunidad de realizar mi intercambio internacional, ustedes fueron mi refugio, mi familia y mi mayor soporte. Su generosidad y constante aliento merecen un lugar de honor en la culminación de este gran logro.

Por último, dedico este trabajo a mi propio esfuerzo. Al compromiso, la paciencia y las incontables horas y noches de dedicación que le entregué a mi carrera. Me agradezco por no haberme rendido y por demostrarme que soy capaz de cumplir lo que me propongo.

Agradecimientos

Quiero expresar mi más profundo agradecimiento a mi director, el profesor Iván David Ortiz Pineda. Aunque no tuve la fortuna de coincidir con usted en un aula de clases durante mi carrera, encontrarlo como director de este proyecto fue un apoyo invaluable. Gracias por impulsarme a salir de mi zona de confort, por retarme a aprender cosas nuevas y por darme la confianza necesaria para creer en mis propias capacidades.

A Liz, mi primera jefa. Gracias por tus invaluable enseñanzas, por darme ánimo en los momentos en que más lo necesité y por abrirme las puertas de este gran proyecto. Tu recomendación y respaldo fueron el punto de partida de este logro.

A la Escuela de Estudios Industriales y Empresariales de la Universidad Industrial de Santander, por haberme brindado mi primera oportunidad laboral y por ser la institución que me formó integralmente como ingeniera industrial.

A mis amigos de la universidad: Andrés, Sara, Ronald y Juan David. Gracias por hacer de mi vida universitaria una experiencia hermosa y llevadera. Su compañía, sus risas y su apoyo incondicional fueron una parte fundamental de este camino.

A Julián, por ser mi recordatorio constante de que soy capaz de alcanzar todo lo que me propongo. Gracias por hacerme sentir segura, fuerte y capaz en cada paso de este proceso.

A Camila, mi mejor amiga y hermana, por no soltar mi mano en los momentos más difíciles y por hacer de este proceso un camino más llevadero con su presencia. Y a Edwin, por su apoyo incondicional, por creer siempre en mis capacidades y darme la fuerza necesaria para alcanzar este logro.

Tabla de Contenido

	Pág.
Introducción	12
1. Objetivos.....	15
1.1 Objetivo General	15
1.2 Objetivos Específicos.....	15
1.3 Cumplimiento de los objetivos.....	16
2. Marco Referencial	18
2.1 Marco de antecedentes	18
2.2 Marco teórico	22
2.2.1 Enseñanza de la estadística.....	22
2.2.2 Analítica de datos y profesiones basadas en datos	23
2.2.3 Competencias analíticas y herramientas computacionales	25
2.2.4 Fundamentos pedagógicos para el diseño de laboratorios	27
3. Desarrollo metodológico.....	29
3.1 Fase 1: Diagnóstico inicial	29
3.1.1 Descripción de la asignatura.....	29
3.1.2 Entrevistas a docentes del área	30
3.1.2.1 Diseño del instrumento.....	31
3.1.2.2 Selección de participantes	32
3.1.2.3 Análisis de las entrevistas.....	32

3.1.3	Estudio de percepción de los estudiantes.....	36
3.1.3.1	Diseño del estudio	37
3.1.3.2	Análisis de resultados de la encuesta de percepción.....	38
3.2	Fase 2: Revisión de experiencias y buenas prácticas.....	46
3.2.1	Exploración de literatura en didáctica de la estadística y analítica de datos.....	46
3.2.2	Identificación de buenas prácticas pedagógicas	47
3.2.3	Transición hacia nuevas herramientas tecnológicas	48
3.3	Fase 3: Diseño de los materiales tipo laboratorio.....	49
3.3.1	Selección y caracterización de las bases de datos reales	49
3.3.2	Definición de la Estructura Pedagógica de los Laboratorios	51
3.3.3	Contenido y justificación de los laboratorios	52
3.3.4	Criterios y rúbrica de evaluación.....	56
3.3.5	Validación interna y control de calidad de los materiales	57
3.4	Fase 4: Implementación piloto y validación de los resultados	58
3.4.1	Aplicación piloto y percepción por sesión.....	58
3.4.1.1	Resultados Laboratorio 1.....	60
3.4.1.2	Resultados Laboratorio 2.....	62
3.4.1.3	Resultados laboratorio 3 y valoración general.....	65
3.4.2	Evaluación de impacto del aprendizaje	68
3.4.2.1	Diseño y estructura del instrumento de evaluación	69
3.4.2.2	Resultados del diagnóstico inicial (Pre-test).....	70

3.4.2.3	Resultados del diagnóstico final (Post-test).....	73
3.4.3	Evaluación integral de proyectos finales	76
3.5	Fase 5: Consolidación y documentación de resultados	78
3.5.1	Ajuste y optimización de los materiales didácticos	79
3.5.2	Organización del repositorio digital	79
3.5.3	Diseño del instructivo de implementación para docentes.....	80
4.	Conclusiones	81
5.	Recomendaciones.....	84
	Referencias Bibliográficas	86

Lista de Tablas

	Pág.
TABLA 1. <i>CUMPLIMIENTO DE OBJETIVOS</i>	16
TABLA 2. <i>CONJUNTOS DE DATOS PÚBLICOS SELECCIONADOS PARA LOS LABORATORIOS</i>	51
TABLA 3. <i>ESTRUCTURA DE LAS GUÍAS DE LOS LABORATORIOS PROPUESTOS</i>	52
TABLA 4. <i>ESTRUCTURA DE LAS GUÍAS DE LOS LABORATORIOS PROPUESTOS</i>	57
TABLA 5. <i>FRECUENCIAS DE PERCEPCIÓN ESTUDIANTEL SOBRE LA APLICACIÓN DEL LABORATORIO 1</i>	60
TABLA 6. <i>FRECUENCIAS DE PERCEPCIÓN ESTUDIANTEL SOBRE LA APLICACIÓN DEL LABORATORIO 2</i>	62
TABLA 7. <i>FRECUENCIAS DE PERCEPCIÓN SOBRE LA APLICACIÓN DEL LABORATORIO 3 Y CIERRE</i>	65
TABLA 8. <i>RESULTADOS DE LA DIMENSIÓN 1 EN EL DIAGNÓSTICO INICIAL (PRE-TEST)</i>	70
TABLA 9. <i>RESULTADOS DE LA DIMENSIÓN 1 EN LA EVALUACIÓN DE SALIDA (POST-TEST)</i>	73

Lista de Figuras

	Pág.
FIGURA 1. <i>FRECUENCIA DE ACTIVIDADES PRÁCTICAS GUIADAS EN COMPUTADOR EN ESTADÍSTICA I</i>	40
FIGURA 2. <i>NIVEL ACTUAL PERCIBIDO EN COMPETENCIAS DIGITALES PARA DATOS</i>	42
FIGURA 3. <i>SOFTWARE O HERRAMIENTAS UTILIZADAS DURANTE O DESPUÉS DE CURSAR ESTADÍSTICA I.</i>	43
FIGURA 4. <i>DISPOSICIÓN PARA REALIZAR TRABAJOS PRÁCTICOS CON PYTHON O R CON GUÍAS PASO A PASO.</i>	44
FIGURA 5. <i>NIVEL DE COMODIDAD DE LOS ESTUDIANTES USANDO EXCEL PARA EL ANÁLISIS DE DATOS.</i>	45

Lista de Apéndices

Apéndice A: Entrevista semiestructurada aplicada a docentes del área.

Apéndice B: Instrumento del estudio de percepción estructurado para estudiantes de la asignatura.

Apéndice C: Guías de los laboratorios

Apéndice D: Rúbricas de evaluación

Apéndice E: Muestras de laboratorios realizados

Apéndice F: Encuestas y respuestas de la retroalimentación aplicada

Apéndice G: Evaluación aplicada

Apéndice H: Repositorio Digital

Apéndice I: Presentación Instructivo Docente

Resumen

Título: Integración de la analítica de datos en la enseñanza de Estadística I mediante el diseño de laboratorios computacionales orientados al desarrollo de competencias*

Autor: Paula Juliana Hernández Herrera**

Palabras Clave: Analítica de datos, estadística descriptiva, laboratorios computacionales, herramientas computacionales, competencias analíticas

Descripción: Este proyecto de grado, desarrollado en modalidad de práctica en docencia, tiene como objetivo diseñar e implementar laboratorios computacionales que integren la analítica de datos en la asignatura Estadística I del programa de Ingeniería Industrial de la Universidad Industrial de Santander. La propuesta surgió de la necesidad de fortalecer el aprendizaje aplicado y reducir la brecha entre la formación teórica tradicional y el desarrollo de competencias digitales requeridas en el contexto de la Industria 4.0, donde el análisis e interpretación de datos se ha convertido en habilidades fundamentales para el ejercicio profesional del ingeniero industrial.

La metodología se estructura en cinco fases: diagnóstico y análisis curricular, revisión de experiencias y buenas prácticas, diseño de laboratorios, implementación piloto y validación de resultados. Los materiales desarrollados incorporan herramientas como Python, Excel y Power BI, junto con el uso de conjuntos de datos reales provenientes de fuentes abiertas, permitiendo a los estudiantes aplicar conceptos estadísticos en contextos cercanos al ejercicio profesional mediante actividades guiadas y progresivas.

La implementación piloto evidenció una alta aceptación de la estrategia por parte de los estudiantes y mostró mejoras en la interpretación de resultados estadísticos, la selección de visualizaciones adecuadas y el uso de herramientas computacionales para el análisis de datos. Asimismo, el proyecto permitió consolidar un repositorio de materiales didácticos, conformado por guías de laboratorio, solucionarios, rúbricas de evaluación e instructivos para docentes, proporcionando una estrategia pedagógica replicable que fortalece la articulación entre la teoría y la práctica y contribuye al desarrollo de competencias analíticas en la formación del ingeniero industrial.

* Trabajo de Grado

** Facultad de Ingeniería Fisicomecánicas. Escuela de Estudios Industriales y Empresariales. Director: Iván David Ortiz Pineda. Candidato a Doctor en Ciencias de la Computación.

Abstract

Title: Integration of Data Analytics in the Teaching of Statistics I through the Design of Competency-Oriented Computational Laboratories*

Author: Paula Juliana Hernández Herrera**

Key Words: Data analytics, descriptive statistics, computational laboratories, computational tools, analytical skills

Description: This degree project was developed as a teaching practice and aimed to design, implement, and validate computational laboratories that integrate data analytics into the Statistics I course of the Industrial Engineering program at the Universidad Industrial de Santander. The project was motivated by the need to strengthen applied learning and reduce the gap between traditional theoretical teaching and the digital skills required in the context of Industry 4.0, where data analysis and interpretation have become essential skills for industrial engineers.

The methodology was organized into five phases: initial diagnosis and curriculum analysis, review of previous experiences and best practices, laboratory design, pilot implementation and validation, and documentation of the developed materials. The laboratories included tools such as Python, Excel, and Power BI, together with real datasets from open data sources. This approach allowed students to apply statistical concepts in practical situations through guided activities.

The pilot implementation showed positive results. Students improved their ability to interpret statistical results, select appropriate data visualizations, and use computational tools for data analysis. In addition, the project produced a digital repository with laboratory guides, answer keys, assessment rubrics, and teaching instructions. These materials provide a teaching strategy that can be adapted by other instructors and contribute to strengthening analytical and digital skills in the education of industrial engineering students.

* Degree Work

** Faculty of Physical-Mechanical Engineering. School of Industrial and Business Studies. Director: Iván David Ortiz Pineda. Ph.D.(c) Computer Science.

Introducción

La transformación digital impulsada por la Industria 4.0 ha redefinido el perfil profesional del ingeniero industrial, quien actualmente debe contar con competencias analíticas y tecnológicas que le permitan interpretar datos, automatizar procesos, comunicar resultados y tomar decisiones fundamentadas en evidencia. Tecnologías como la inteligencia artificial, el Internet de las Cosas, el Big Data y la analítica de datos han generado nuevas exigencias en los entornos productivos, lo que implica que la formación universitaria debe responder no solo con fundamentos conceptuales, sino también con experiencias prácticas orientadas al uso de herramientas computacionales.

En este contexto, Estadística I es una asignatura base del programa de Ingeniería Industrial de la Universidad Industrial de Santander, cuyos contenidos se orientan principalmente hacia la estadística descriptiva, el manejo básico de datos y los fundamentos de probabilidad para el análisis estadístico inferencial posterior. Aunque estos elementos son esenciales para la formación del estudiante, en la enseñanza tradicional de asignaturas cuantitativas persiste una brecha entre el dominio teórico de los contenidos y su aplicación mediante herramientas computacionales.

El planteamiento del problema se centra, por tanto, en la limitada articulación entre los conceptos estadísticos abordados en el aula y el desarrollo de competencias digitales requeridas en el contexto profesional actual. Esta situación puede dificultar que los estudiantes trasladen los conocimientos adquiridos a escenarios reales de análisis, donde se requiere procesar información, identificar patrones, construir visualizaciones y formular conclusiones útiles para la toma de decisiones. Esta necesidad se evidencia en el diagnóstico inicial del proyecto, desarrollado a partir de la revisión curricular, entrevistas a docentes del área y un estudio de percepción aplicado a estudiantes que cursaron previamente la asignatura.

Como antecedente del proyecto, se reconocen experiencias académicas previas orientadas a la incorporación de analítica de datos y herramientas digitales en procesos formativos, tanto en asignaturas del programa de Ingeniería Industrial como en otros contextos educativos. Estos trabajos evidencian la pertinencia de usar software, casos de estudio, visualización de datos y metodologías activas para favorecer el aprendizaje aplicado. No obstante, también muestran retos asociados a la accesibilidad tecnológica, la necesidad de guías estructuradas y la importancia de adaptar las herramientas al nivel de los estudiantes. En este sentido, persiste la necesidad de diseñar estrategias pedagógicas que integren la analítica de datos en cursos base como Estadística I, sin reemplazar su fundamentación teórica.

Por esta razón, el presente trabajo de grado, desarrollado en modalidad de práctica en docencia, tiene como propósito diseñar e implementar laboratorios computacionales que integren temáticas de analítica de datos en la asignatura Estadística I. Estos laboratorios buscan fortalecer el aprendizaje aplicado de los conceptos estadísticos mediante el uso de herramientas computacionales, conjuntos de datos reales y actividades guiadas que permitan a los estudiantes pasar de la comprensión teórica a la interpretación práctica de la información.

Metodológicamente, el proyecto se desarrolla a partir de un diagnóstico curricular y pedagógico, una revisión de experiencias y buenas prácticas, el diseño de materiales tipo laboratorio, una implementación piloto y la validación de los resultados obtenidos. De esta manera, el trabajo no solo busca producir guías o actividades aisladas, sino consolidar un repositorio estructurado de laboratorios y conjuntos de datos que pueda ser utilizado, adaptado y fortalecido por los docentes del programa.

En este sentido, el proyecto aporta una estrategia didáctica replicable que articula la teoría estadística con la práctica profesional, contribuyendo al cierre de la brecha entre la formación académica y las competencias digitales demandadas por el entorno industrial actual. Asimismo, la propuesta busca generar recursos didácticos que apoyen tanto al estudiante en su proceso de aprendizaje como al docente en la incorporación gradual de tecnologías digitales en la asignatura.

1. Objetivos

1.1 Objetivo General

Desarrollar materiales tipo laboratorio, basados en herramientas computacionales, que integren temáticas de analítica de datos en la asignatura de Estadística I, con el propósito de fortalecer el aprendizaje aplicado de los conceptos estadísticos y el uso de software de análisis estadístico.

1.2 Objetivos Específicos

- Analizar el contenido de la asignatura de Estadística I en relación con los conceptos y competencias asociadas a la analítica de datos, identificando oportunidades de integración curricular.
- Revisar experiencias y buenas prácticas de diseño de materiales didácticos y laboratorios de analítica de datos en programas afines para fundamentar el diseño pedagógico del proyecto.
- Diseñar materiales tipo laboratorio, con base en contextos y conjuntos de datos reales (o sintéticos realistas), que permitan a los estudiantes aplicar herramientas como Python, R, Excel o Power BI en el análisis, visualización y modelado de datos.
- Implementar y validar los materiales diseñados mediante la aplicación piloto con estudiantes de la asignatura, evaluando su pertinencia, usabilidad y contribución al desarrollo de competencias en analítica de datos.
- Consolidar y documentar los resultados del proceso, incluyendo el repositorio de laboratorios y el banco de datos sugeridos, para su uso y adaptación por parte de otros docentes del programa.

1.3 Cumplimiento de los objetivos

En la Tabla 1 se presenta la relación de cada objetivo específico desarrollado en marco de este trabajo de grado, junto con las secciones del presente documento donde se evidencia su cumplimiento.

Tabla 1.

Cumplimiento de objetivos

Objetivo específico	Cumplimiento
Analizar el contenido de la asignatura de Estadística I en relación con los conceptos y competencias asociadas a la analítica de datos, identificando oportunidades de integración curricular.	Sección 3.1: Fase 1: Diagnóstico inicial
Revisar experiencias y buenas prácticas de diseño de materiales didácticos y laboratorios de analítica de datos en programas afines para fundamentar el diseño pedagógico del proyecto.	Sección 2.1: Marco de antecedentes Sección 3.2: Fase 2: Revisión de experiencias y buenas prácticas
Diseñar materiales tipo laboratorio, con base en contextos y conjuntos de datos reales (o sintéticos realistas), que permitan a los estudiantes aplicar herramientas como Python, R, Excel o Power BI en el análisis, visualización y modelado de datos.	Sección 3.3: Fase 3: Diseño de los materiales tipo laboratorio Apéndice C: Guías de los laboratorios Apéndice D: Rúbricas de evaluación
Implementar y validar los materiales diseñados mediante la aplicación piloto con estudiantes de la asignatura, evaluando su pertinencia, usabilidad y contribución al desarrollo de competencias en analítica de datos.	Sección 3.4: Fase 4: Implementación piloto y validación de los resultados Apéndice E: Muestras de laboratorios realizados Apéndice F: Encuestas y respuestas Apéndice G: Evaluación aplicada

Consolidar y documentar los resultados del proceso, incluyendo el repositorio de laboratorios y el banco de datos sugeridos, para su uso y adaptación por parte de otros docentes del programa.

Sección 3.5: Fase 5: Consolidación
y documentación de resultados

Apéndice H: Enlace y estructura
del Repositorio Digital

Apéndice I: Presentación
Instructivo Docente

2. Marco Referencial

El marco referencial del presente proyecto reúne los antecedentes y fundamentos teóricos que sustentan el diseño de laboratorios computacionales orientados a la integración de la analítica de datos en la asignatura Estadística I. En primer lugar, se presentan trabajos previos relacionados con el uso de herramientas digitales, analítica de datos, visualización de información y estrategias didácticas aplicadas en contextos académicos, con el fin de identificar aportes, limitaciones y oportunidades para la propuesta. Posteriormente, se desarrollan los conceptos centrales que fundamentan el proyecto, entre ellos la estadística descriptiva, los perfiles profesionales orientados a datos, los enfoques de la analítica de datos y las estrategias pedagógicas asociadas al aprendizaje aplicado. De esta manera, este apartado permite establecer la base conceptual y metodológica sobre la cual se estructura la propuesta de laboratorios computacionales para fortalecer el aprendizaje práctico y el desarrollo de competencias analíticas en los estudiantes.

2.1 Marco de antecedentes

Revisando proyectos de grado desarrollados por esta misma línea se puede destacar el trabajo realizado por Castro y Puerto (2025), en donde se hizo el diseño e implementación de casos de estudio con técnicas de analítica de datos para las asignaturas de Logística Integral e Investigación de Mercados, utilizando el software Orange Data Mining. Sin embargo, en este trabajo no se consideró la limitante técnica asociada al software, el cual, por su alto consumo de recursos, presentó dificultades de ejecución en equipos de cómputo con capacidades inferiores, restringiendo así el acceso y la experiencia práctica de los estudiantes.

Este trabajo es pertinente en el marco del proyecto porque sienta un precedente metodológico sólido sobre la integración de la analítica de datos en las bases del ingeniero industrial y sus hallazgos sobre las barreras técnicas son un insumo fundamental. Este proyecto

parte de esta base para proponer una solución o alternativa tecnológica que supere dicha limitación, garantizando accesibilidad, mediante el uso de herramientas como Python, R, Excel y Power BI.

En una línea complementaria, se encuentra el trabajo desarrollado por Vásquez (2023) titulado “*Diseño, desarrollo e implementación de un sistema de información web para la gestión y analítica de datos de los recursos electrónicos disponibles en la Biblioteca Virtual de la Universidad Industrial de Santander*”, el cual se enfocó en la construcción de una plataforma para automatizar la recolección, consolidación y análisis de datos de uso de recursos electrónicos. En su desarrollo se incorporaron tecnologías web (HTML, CSS, JavaScript, PHP y MySQL) junto con estándares como COUNTER y el protocolo SUSHI, permitiendo generar métricas de desempeño y visualizaciones dinámicas a través de librerías como Chart.js.

Aunque este trabajo no se sitúa directamente en un escenario de aula, constituye un antecedente significativo en tanto demuestra el valor del análisis de datos para apoyar decisiones institucionales y facilitar procesos de interpretación mediante visualizaciones accesibles. En contraste con el presente proyecto, su principal diferencia radica en que se orienta a un objetivo administrativo-operativo, sin incorporar una estrategia didáctica asociada al desarrollo de competencias analíticas en estudiantes; no obstante, aporta un referente técnico útil sobre procesamiento, estructuración de indicadores y comunicación visual de resultados.

De forma coherente con la intención de integrar herramientas digitales en contextos académicos, se identifica el trabajo “*Estrategia didáctica para mejora del proceso de enseñanza y de aprendizaje en la asignatura de sistemas de información mediante el uso de Power Platform*”, desarrollado por Vidal y Blanco (2025), el cual propone una estrategia para fortalecer la enseñanza mediante la incorporación de herramientas como Power BI, Power Apps y Power Automate,

articuladas con metodologías activas como el aprendizaje basado en proyectos, la gamificación y modelos instruccionales como ADDIE.

Este antecedente es particularmente relevante para el presente proyecto, ya que evidencia que el uso de plataformas tecnológicas no solo incrementa la motivación y participación estudiantil, sino que también posibilita el fortalecimiento de competencias técnicas aplicadas y el desarrollo de habilidades útiles para el entorno laboral. Aunque el área de aplicación corresponde a otra asignatura, su valor reside en la estructuración pedagógica y metodológica de una estrategia didáctica basada en el uso de herramientas de análisis y automatización, lo cual constituye un insumo significativo para orientar el diseño de laboratorios computacionales enfocados en Estadística I.

Por su parte, un antecedente nacional que contribuye a ampliar el enfoque de analítica en educación es el trabajo de Hernández (2024) titulado “*Modelo de dominio específico para análisis y minería de datos educativos*” (Universidad Nacional de Colombia), en el cual se propone un modelo conceptual para apoyar la aplicación de técnicas de minería de datos en contextos educativos, integrando componentes como la preparación de datos, la representación del dominio, la minería y la visualización/interacción. El aporte principal de esta propuesta se centra en destacar que el análisis de datos en educación requiere incorporar conocimiento del dominio, por ejemplo, normativas y estructuras académicas, para lograr interpretaciones válidas y útiles.

No obstante, su alcance se orienta mayormente al análisis institucional y al aprovechamiento de datos educativos como insumo para diagnóstico o gestión, sin profundizar en estrategias didácticas concretas para el aprendizaje de estadística mediante actividades prácticas. Aun así, este antecedente resulta pertinente para el presente proyecto porque reafirma la importancia de adaptar procesos analíticos a contextos formativos específicos, principio que se

retoma en esta investigación al proponer laboratorios contextualizados para la asignatura de Estadística I.

Finalmente, como referencia ampliada, se considera el trabajo de Bedoya (2021) titulado *“Desarrollo de una plataforma de análisis de datos para el estudio de la brecha digital aprovechando el ecosistema de datos abiertos de Colombia”* (UNAD), donde se desarrolló una solución basada en Tableau para analizar datos provenientes de la Gran Encuesta TIC 2017, con el propósito de identificar patrones de acceso y uso de tecnologías a partir de variables socioeconómicas y demográficas.

Este antecedente es valioso en tanto evidencia la aplicabilidad de herramientas de visualización para explorar datos reales y comunicar resultados de forma comprensible, permitiendo reconocer la importancia de la exploración gráfica como apoyo interpretativo en escenarios de análisis. Sin embargo, su enfoque se concentra en una herramienta específica y no plantea un componente pedagógico asociado al desarrollo de competencias en estudiantes, aspecto que el presente proyecto busca fortalecer mediante la integración de laboratorios computacionales estructurados con objetivos de aprendizaje y evidencias evaluables en Estadística I.

En síntesis, los antecedentes revisados evidencian que el uso de analítica de datos y herramientas digitales se ha establecido como una estrategia pertinente tanto para fortalecer procesos académicos como para apoyar decisiones organizacionales; no obstante, también se identifican vacíos asociados a la accesibilidad tecnológica, la limitada incorporación pedagógica en algunos trabajos y la necesidad de fortalecer la enseñanza de competencias analíticas desde cursos base.

En este marco, el presente proyecto se diferencia al proponer el diseño de laboratorios computacionales aplicados a Estadística I, orientados a integrar analítica de datos con una

estrategia didáctica clara, buscando cerrar la brecha entre la formación conceptual y el desarrollo de habilidades prácticas necesarias para el análisis y la interpretación de datos en escenarios reales.

2.2 Marco teórico

El marco reúne los fundamentos conceptuales y pedagógicos que sustentan la propuesta de integrar laboratorios computacionales en la enseñanza de la asignatura Estadística I. En primer lugar, se aborda el papel de la estadística como base para el desarrollo del pensamiento estadístico y de las competencias necesarias para el análisis e interpretación de datos. Posteriormente, se presenta la evolución hacia la analítica de datos y su importancia en la formación del ingeniero industrial, considerando exigencias del entorno profesional y el uso de herramientas computacionales. Finalmente, se desarrollan los principales fundamentos pedagógicos que respaldan el diseño de laboratorios computacionales como una estrategia orientada al aprendizaje activo.

2.2.1 Enseñanza de la estadística

Diversos autores han señalado que el propósito de la educación estadística no consiste únicamente en enseñar técnicas de cálculo de medidas, sino en promover formas de pensamiento que permitan comprender la variabilidad, identificar patrones, reconocer comportamientos relevantes en los datos e interpretar los resultados dentro de un contexto determinado. Garfield y Ben-Zvi (2008) sostienen que el aprendizaje de la estadística debe favorecer el desarrollo del pensamiento estadístico, entendido como la capacidad para utilizar conceptos estadísticos al analizar información, explicar resultados y tomar decisiones fundamentadas.

Dentro de este proceso formativo, la estadística descriptiva desempeña un papel fundamental al ser la primera aproximación al análisis de datos. Su objetivo principal consiste en organizar, resumir, y comprender el comportamiento general de un conjunto de datos, mediante

tablas, medidas numéricas y gráficos, antes de aplicar procedimientos más complejos. En cursos introductorios, su valor no está únicamente en calcular medidas, sino en aportar elementos para identificar tendencias, detectar valores atípicos, reconocer relaciones entre variables y formular interpretaciones iniciales.

En esa misma línea, Tukey (1977), bajo el enfoque de análisis exploratorio de datos (*Exploratory Data Analysis*), afirma que el propósito inicial del trabajo con datos es “mirar los datos para descubrir lo que parecen decir”, enfatizando en la importancia de comprender la información antes de formar explicaciones o inferencias. En este sentido, el primer paso de cualquier proceso analítico consiste en observar cuidadosamente los datos antes de aplicar modelos o procedimientos complejos.

Este enfoque resulta especialmente pertinente en cursos introductorios, como Estadística I, donde el objetivo no es aplicar modelos estadísticos avanzados, sino fortalecer la capacidad del estudiante para la interpretación de la información y comprender el significado de los resultados.

2.2.2 Analítica de datos y profesiones basadas en datos

Actualmente, las empresas e instituciones producen grandes volúmenes de información provenientes de procesos productivos, sistemas de información, plataformas digitales o dispositivos conectados, lo que ha incrementado la necesidad de profesionales capaces de convertir dichos datos en conocimiento útil para el negocio. En ese contexto, aunque la estadística sigue desempeñando un papel fundamental, su aplicación se complementa con herramientas computacionales, métodos de procesamiento de datos y tecnologías que superan métodos tradicionales.

Como resultado de esta evolución, surge la analítica de datos, entendida como el uso de un conjunto de métodos, procesos y herramientas orientadas en transformar datos en información útil

para apoyar la toma de decisiones mediante evidencia. Davenport y Harris (2007) señalan que las organizaciones que fundamentan sus decisiones en procesos de análisis de datos obtienen una ventaja competitiva. Desde esta perspectiva, la analítica de datos no consiste únicamente en utilizar software especializado, sino en formular preguntas relevantes, seleccionar indicadores apropiados, interpretar resultados y traducirlos en conclusiones dentro del contexto del problema.

Uno de los métodos que mejor representa este proceso es el CRISP-DM (Cross Industry Standard Process for Data Mining), propuesto por Chapman et al. (2000), el cuál organiza el desarrollo de un proyecto de datos en seis fases: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación y despliegue. Aunque este modelo fue propuesto inicialmente para los proyectos de minería de datos, actualmente sigue siendo un marco metodológico ampliamente utilizado por su carácter iterativo y la importancia del conocimiento del contexto.

Dentro de este proceso, el preprocesamiento de los datos es una etapa crítica. Diversos estudios coinciden en que una parte considerable del tiempo de un proyecto se dedica a tareas de selección, limpieza, integración y transformación de datos, las cuales condicionan la calidad de los análisis posteriores (Chapman et al., 2000). Estas actividades conforman un campo conocido como ingeniería de datos, cuyo propósito consiste en garantizar que la información sea consistente, completa y adecuada para responder las preguntas planteadas. En ese sentido, la ingeniería de datos no reemplaza el análisis estadístico, sino que crea las condiciones para que dicho análisis produzca resultados favorables.

Por otra parte, la ciencia de datos amplía aún más este panorama al integrar conocimientos provenientes de la estadística, las ciencias de la computación y el dominio del problema. Donoho (2017) señala que la ciencia de datos puede entenderse como la disciplina dedicada a extraer

conocimiento a partir de datos mediante la combinación de métodos estadísticos, algoritmos computacionales y comprensión del contexto. El autor relaciona esta disciplina con las ideas planteadas por Tukey, quien proponía una visión más amplia del análisis de datos como una “ciencia aún no reconocida”, como propósito era el “aprendizaje a partir de los datos”.

En el contexto de la formación de un ingeniero industrial, la integración con estas disciplinas de profesionales en datos adquiere importancia debido a que el ejercicio profesional demanda con mayor frecuencia competencias relacionadas con el análisis de información, construcción de indicadores y visualización de resultados.

2.2.3 Competencias analíticas y herramientas computacionales

En relación con la necesidad de la integración de ingeniería industrial con las profesiones basadas en datos, diversos organismos han destacado la importancia de fortalecer las competencias digitales y analíticas en la formación universitaria. El foro económico mundial, en el “*The Future of Jobs Report 2025*” identifica el pensamiento analítico, la alfabetización tecnológica y el manejo de datos como algunas de las competencias con mayor demanda para los próximos años, debido al crecimiento de la automatización y la inteligencia artificial (Leopold et al., 2025).

Sin embargo, las exigencias actuales del entorno laboral requieren que estas capacidades sean complementadas con el uso de herramientas computacionales que permitan procesar datos de gran volumen, automatizar procedimientos repetitivos y generar resúmenes gráficos que faciliten la interpretación de resultados. De esta manera, se puede identificar la tecnología como un medio para fortalecer el desarrollo de las competencias analíticas.

Entre las herramientas más utilizadas actualmente para el análisis de datos se encuentran programas como Microsoft Excel, Python, R y Microsoft Power BI, cada una con sus características particulares que responden una necesidad dentro del proceso analítico. Excel es una

herramienta ampliamente utilizada por su facilidad para organizar información, realizar cálculos básicos y construir visualizaciones sencillas. Por otra parte, lenguajes de programación como Python y R han adquirido relevancia por su capacidad de automatizar procesos de análisis, manipular grandes volúmenes de datos y desarrollar modelos reproducibles.

De manera complementaria, herramientas de inteligencia de negocios como Power BI han fortalecido la importancia de la visualización interactiva dentro de los procesos de análisis. La representación gráfica de la información es una etapa esencial para comprender el comportamiento de los datos y comunicar los resultados de manera efectiva a diferentes audiencias. Few (2013) señala que una visualización adecuada permite revelar relaciones, tendencias y comportamientos que difícilmente podrían identificarse mediante resúmenes numéricos, facilitando la interpretación y el proceso de toma de decisiones.

La comunicación de resultados constituye igualmente una competencia indispensable dentro del análisis de datos, considerando que un proceso analítico pierde utilidad cuando sus hallazgos no logran transmitirse de forma clara a los encargados de tomar decisiones. Bajo esta perspectiva surge el concepto de *data storytelling*, entendido como la integración de datos, visualizaciones y narrativa para construir mensajes comprensibles y orientados a una audiencia específica. Knaflitz (2015) sostiene que una visualización adquiere verdadero valor cuando se acompaña con una narrativa que explique el significado de los resultados y sus implicaciones para el problema analizado.

No obstante, la incorporación de estas herramientas dentro de asignaturas introductorias debe realizarse de manera gradual y coherente con los objetivos de aprendizaje del curso. El propósito no es formar especialistas en programación o inteligencia de negocios, sino en

proporcionar experiencias que permitan aplicar conceptos estadísticos mediante el uso de herramientas usadas en el ejercicio profesional.

2.2.4 Fundamentos pedagógicos para el diseño de laboratorios

La evidencia disponible en educación superior muestra que la simple incorporación de software no garantiza mejores resultados de aprendizaje, por el contrario, los beneficios dependen de que las actividades propuestas correspondan a objetivos de aprendizaje y promuevan la participación del estudiante. En ese sentido, el diseño de laboratorios computacionales requiere una fundamentación pedagógica que trascienda el uso de la tecnología como recurso a una estrategia que favorezca la comprensión de los conceptos mediante prácticas contextualizadas.

Uno de los enfoques que sustenta este tipo de estrategias es el aprendizaje activo, el cuál plantea que los estudiantes aprenden con mayor profundidad cuando participan de manera directa en actividades que requieren analizar información, resolver problemas, discutir resultados y reflexionar sobre los conceptos abordados, en vez de recibirlos de forma pasiva. Prince (2004) define el aprendizaje activo como cualquier metodología de enseñanza que involucre al estudiante en el proceso de aprendizaje mediante actividades que favorezcan el análisis, la síntesis y la evaluación del conocimiento.

Otro referente ampliamente reconocido en educación corresponde a la alineación constructivista desarrollada por Biggs y Tang (2011). El modelo sostiene que la calidad del aprendizaje depende de la coherencia que exista entre los resultados de aprendizaje, las actividades del curso y los mecanismos de evaluación utilizados por el docente. En consecuencia, las actividades prácticas deben ser diseñadas de manera que permitan al estudiante demostrar las competencias que posteriormente serán evaluadas, evitando la desarticulación entre lo enseñado, lo que se practica y lo que se evalúa.

Esta perspectiva resulta coherente con el aprendizaje basado en competencias, el cual orienta la formación hacia el desarrollo integrado de conocimientos, habilidades y actitudes que puedan transferirse a situaciones reales. Tobón (2013) señala que la formación por competencias busca que el estudiante sea capaz de unificar diferentes saberes para resolver problemas propios de su contexto profesional, superando modelos educativos centrados exclusivamente en la adquisición de contenidos.

Dentro de esta misma orientación se encuentra el aprendizaje basado en proyectos, metodología que promueve el aprendizaje mediante el desarrollo de problemas auténticos y significativos para los estudiantes. Thomas (2000) plantea que el aprendizaje basado en proyectos favorece la integración de conocimientos provenientes de diferentes disciplinas y fortalece actividades como el trabajo colaborativo, la autonomía y la resolución de problemas. En cursos relacionados con el análisis de datos, permite que los estudiantes desarrollen procesos completos de exploración, tratamiento e interpretación de la información, aproximándose a situaciones que se pueda enfrentar en su práctica profesional.

A partir de estos fundamentos, los laboratorios computacionales pueden entenderse como espacios estructurados de aprendizaje donde convergen los principios del aprendizaje activo y la formación por competencias. Su propósito no consiste únicamente en enseñar el funcionamiento de un software, sino en ofrecer experiencias estructuradas que permitan a los estudiantes aplicar conceptos estadísticos en contextos cercanos a la realidad, fortaleciendo tanto el razonamiento, la interpretación de resultados y el uso de herramientas digitales.

3. Desarrollo metodológico

La metodología del presente proyecto se estructuró en cinco fases orientadas al diagnóstico, diseño, construcción, validación y documentación de materiales computacionales para la asignatura Estadística I. Cada fase permitió avanzar de manera progresiva desde el análisis de las necesidades curriculares y pedagógicas del curso hasta la elaboración de laboratorios enfocados en fortalecer el aprendizaje aplicado de los conceptos estadísticos y el desarrollo de competencias analíticas y digitales en los estudiantes.

3.1 Fase 1: Diagnóstico inicial

Se presenta un diagnóstico del estado actual de la asignatura Estadística I en Ingeniería Industrial (UIS), construido a partir de información recolectada mediante aspectos cualitativos y cuantitativos. Para ello, se describen las condiciones curriculares y académicas del curso y se analizan las percepciones de los actores que participan en su desarrollo, con el fin de identificar fortalezas, limitaciones y necesidades que orienten la propuesta del proyecto. En los apartados se aborda, primero, la caracterización del programa y su dinámica de enseñanza; luego, los hallazgos de las entrevistas semiestructuradas aplicadas a docentes del área; y finalmente, la perspectiva estudiantil obtenida mediante una encuesta.

3.1.1 Descripción de la asignatura

La asignatura Estadística I (código 29003) forma parte del plan de estudios del programa de Ingeniería Industrial (Plan 11) de la Universidad Industrial de Santander y se ubica en la malla curricular en el quinto nivel del proceso formativo del estudiante, actuando como curso base para el desarrollo posterior de materias de la línea de métodos cuantitativos del programa. En la malla se observa su articulación con asignaturas de fundamentación matemática previas y su continuidad

hacia cursos estadísticos, lo cual confirma su carácter estructurante dentro de la línea cuantitativa del plan de estudios.

De acuerdo con el syllabus oficial, Estadística I tiene una intensidad de 3 créditos académicos, con 4 horas teóricas de acompañamiento docente y 5 horas de trabajo independiente por parte de los estudiantes semanales y sin componente práctico formal asignado, lo que refuerza su orientación principalmente hacia lo conceptual dentro del plan de formación. La asignatura está planeada para ser desarrollada en cinco unidades temáticas, cubriendo contenidos de estadística descriptiva, probabilidad y conteo, variables aleatorias, modelos teóricos de variables aleatorias, y distribuciones bivariantes y multivariantes.

En coherencia con ello, el propósito de la asignatura se centra en brindar al estudiante los conocimientos necesarios para aplicar técnicas estadísticas básicas tanto en investigación científica como en el ejercicio profesional; adicionalmente, busca que el estudiante adquiera herramientas para procesar información de manera descriptiva a partir de datos recolectados en procesos industriales o administrativos, así como realizar pronósticos a corto y mediano plazo utilizando probabilidades. Este propósito también incorpora el desarrollo de habilidades intelectuales y de pensamiento analítico mediante la adquisición de conocimientos, valores y actitudes propios del campo estadístico.

3.1.2 Entrevistas a docentes del área

Con el propósito de complementar la caracterización curricular de la asignatura Estadística I y profundizar en su estado actual, se aplicaron entrevistas semiestructuradas a docentes del área, la cual se encuentra como Apéndice A. Este componente cualitativo permite comprender no solo el diseño formal de la asignatura, sino también las prácticas pedagógicas actuales, herramientas

utilizadas, percepciones sobre competencias estudiantiles y las oportunidades de mejora identificadas por quienes orientan el curso.

3.1.2.1 Diseño del instrumento. Se aplicó una entrevista semiestructurada con preguntas abiertas organizadas en ejes temáticos que facilitaron la comparación de las respuestas de los docentes que participaron. Los ejes abordados fueron:

- Enfoque general de enseñanza de la asignatura
- Valoración de la relevancia del programa actual
- Integración de datos reales y contextos industriales en las actividades
- Nivel de alfabetización digital en los estudiantes
- Herramientas computacionales empleadas
- Percepción sobre la relación entre Estadística I y analítica de datos
- Barreras y oportunidades de fortalecimiento

Esos ejes temáticos se establecieron para poder caracterizar los enfoques de enseñanza más utilizados por los profesores, su experiencia sobre el uso de herramientas computacionales como herramienta de apoyo a la clase y su percepción sobre la relevancia del uso de estos recursos, así como limitaciones y oportunidades actuales.

3.1.2.2 Selección de participantes. Las entrevistas se realizaron de forma presencial a tres docentes planta que orientan de manera recurrente la asignatura en el programa de Ingeniería Industrial. La elección de este grupo se fundamenta en primer lugar, en su representatividad directa, dado que son quienes implementan el curso y poseen una visión integral del desarrollo de contenidos, evaluaciones y dificultades recurrentes. En segundo lugar, al tratarse de docentes de planta, aportan continuidad curricular y conocimiento acumulado sobre la evolución de la asignatura. Finalmente, para el alcance del presente diagnóstico (de carácter cualitativo y orientado a identificar tendencias y necesidades pedagógicas), la muestra permitió alcanzar información relevante, ya que las entrevistas convergieron en patrones claros sobre enfoque teórico-práctico, competencias logradas, limitaciones tecnológicas y oportunidades de innovación.

3.1.2.3 Análisis de las entrevistas. En relación con el enfoque pedagógico, se evidencia una orientación mixta con predominio de la fundamentación conceptual; los docentes resaltan la importancia de preservar el componente teórico, especialmente en probabilidad y modelos de incertidumbre. En este sentido, un profesor afirma que su enseñanza mantiene siempre un núcleo conceptual: “la parte teórica siempre ha estado presente en todos mis cursos”, mientras que la práctica basada en toma directa de datos reales es limitada en el día a día del curso. De forma consistente, otro docente describe su asignatura como una combinación de teoría y práctica, aunque reconoce que el peso mayor recae en lo conceptual: “con un componente más fuerte a lo teórico que a lo práctico”. A su vez, el tercer docente puntualiza que la práctica se concentra con mayor claridad en estadística descriptiva, mientras que probabilidad conserva un estilo más abstracto, indicando que “el corte estadístico-descriptivo intento que sea muy práctico, pero probabilidad ya es un poco más teórico”. En síntesis, Estadística I es concebida como asignatura base con prácticas relevantes, pero subordinadas a la consolidación teórica.

Por otra parte, sobre la relevancia del programa actual, los tres docentes coinciden en que los contenidos son adecuados para el propósito formativo de Estadística I. Se resalta que la asignatura constituye el soporte de la línea estadística y de otras materias cuantitativas del plan de estudios. Un docente sostiene que los contenidos “funcionan como un correcto abrebocas para toda la línea estadística que llegue en adelante”, dado que proveen bases para cursos posteriores de estadística y analítica de datos. En la misma dirección, otro profesor enfatiza el carácter estructural del curso al afirmar que los modelos teóricos vistos son “el fundamento de todo lo que tiene que ver con la estadística”. No obstante, surge un aspecto clave para este proyecto: aunque el contenido es pertinente, se identifica una brecha tecnológica frente a las exigencias actuales, pues se señala que los planes de estudio permanecen en una estadística clásica mientras hoy se requieren competencias nuevas asociadas a herramientas computacionales, IA y Big Data. Este hallazgo no cuestiona los fundamentos del programa, sino que evidencia la necesidad de fortalecer su dimensión aplicada con soporte digital.

En cuanto al uso de datos reales y contextos industriales, los docentes reportan una práctica dual. En el trabajo cotidiano predominan datos simulados, utilizados para controlar didácticamente el comportamiento estadístico que se desea ilustrar. Un profesor explica que, en general, “simulo los datos buscando un comportamiento específico” para evidenciar resultados y gráficos en clase. Sin embargo, el trabajo con datos reales aparece con fuerza en proyectos finales, donde los estudiantes buscan y analizan bases abiertas. Un docente indica que utiliza repositorios como Kaggle, UCI Machine Learning y datos institucionales como Banco de la República o DANE. Complementariamente, otro profesor describe el uso de la plataforma datos.gov.co para que los estudiantes “se enfrenten a lo que son los datos reales con todo el ruido que está allá afuera”, destacando el valor formativo de analizar información auténtica y no solo sintética. En

consecuencia, existe una intención clara de aplicación, aunque concentrada en momentos específicos del semestre.

Respecto a las competencias estadísticas desarrolladas, los docentes coinciden en que la estadística descriptiva es el componente más sólido en el estudiantado. Se reconoce que es una temática familiar para ellos y la que mejor aplican durante el curso; uno de los docentes afirma que “la estadística descriptiva es lo que ellos más utilizan y más se les facilita”. A pesar de ello, también se señala que la comprensión y el análisis es lo que más se les dificulta, pues falta mayor capacidad interpretativa sobre el sentido de las medidas descriptivas en contextos reales. En contraste, los principales obstáculos se concentran en probabilidad y en razonamiento estadístico de mayor abstracción. Un docente enfatiza que las “variables aleatorias multidimensionales para ellos es bastante complejo”, lo cual refleja barreras conceptuales persistentes en la transición hacia modelos probabilísticos avanzados. Este resultado sugiere un punto crítico donde los apoyos computacionales pueden facilitar comprensión conceptual y aplicación.

En relación con la alfabetización digital estudiantil, los tres docentes reportan niveles heterogéneos, con tendencia general a competencias bajas en herramientas básicas. Se afirma que los grupos “son bastante heterogéneos hay algunos que vienen bien preparados y otros que no”, y en promedio el nivel se ubica “intermedio o bajo”. Esta situación implica que la integración de software en el curso encuentra una barrera estructural, pues no todos los estudiantes cuentan con habilidades previas que les permitan avanzar al mismo ritmo. En consecuencia, los laboratorios computacionales propuestos en este proyecto se justifican como una estrategia de nivelación progresiva, enfocada en interpretación y uso aplicado más que en exigencias técnicas iniciales. Además, con el fin de facilitar la apropiación de las herramientas digitales y reducir las dificultades de entrada, se plantea la elaboración de videos introductorios por cada software, en los cuales se

presenten: una explicación inicial del entorno y sus funciones básicas, el desarrollo guiado de un ejemplo completo, y orientaciones para que posteriormente los estudiantes puedan replicar y aplicar lo aprendido de manera autónoma en los laboratorios. De este modo, se busca asegurar una transición gradual desde la comprensión operativa mínima hacia el análisis estadístico aplicado con apoyo computacional.

Finalmente, sobre las herramientas digitales y la relación con analítica de datos, se evidencia que actualmente sí se utilizan herramientas de software, aunque de forma no completamente sistemática. Un docente menciona que emplea principalmente “R, Power BI y Python, todos hemos migrado a Python”, mostrando una transición hacia herramientas más abiertas y vigentes en el campo analítico. Sin embargo, también se expresa precaución frente a una integración directa de analítica de datos en Estadística I; en particular, un docente sostiene que el curso debe mantener su rol de base teórica y que “la estadística uno tiene que seguir siendo la fundamentación teórica”. Esto revela que la oportunidad de mejora no es reemplazar contenidos, sino articularlos con experiencias computacionales que complementen el aprendizaje sin perder su esencia formativa.

En conjunto, las tres entrevistas muestran que Estadística I posee un programa funcional y una base teórica sólida, pero enfrenta retos asociados al bajo nivel digital promedio del estudiantado, la concentración del proyecto aplicado hacia el final del semestre y la integración parcial de herramientas computacionales. Adicionalmente, aunque hay consenso en la importancia de preservar fundamentos, se reconoce la necesidad de fortalecer la conexión con competencias tecnológicas que demanda el contexto profesional actual. Estos hallazgos sustentan la pertinencia del presente proyecto, orientado al diseño de laboratorios computacionales que integren analítica de datos de manera gradual, accesible y coherente con el enfoque conceptual de la asignatura.

3.1.3 Estudio de percepción de los estudiantes

Con el fin de complementar el diagnóstico realizado sobre el estado actual de la asignatura Estadística I y sustentar el diseño de los laboratorios computacionales, se planteó un estudio de percepción dirigido a estudiantes que ya hayan cursado la asignatura. Este estudio busca identificar la experiencia formativa de los estudiantes en relación con la aplicabilidad de los contenidos, el uso de herramientas digitales durante el curso, el nivel de desarrollo de competencias asociadas al análisis y la visualización de datos, así como su disposición frente a la implementación de laboratorios computacionales integrados con analítica de datos.

El estudio se desarrolló bajo un enfoque cuantitativo de tipo descriptivo, mediante la aplicación de un instrumento de recolección de información basado en encuesta estructurada, que se encuentra en el Apéndice B. La encuesta fue diseñada con preguntas cerradas y escalas tipo Likert (1 a 5), complementadas con preguntas de selección múltiple y un espacio abierto para comentarios, con el propósito de recolectar información estandarizada que permita caracterizar la percepción de los estudiantes y comparar tendencias entre diferentes semestres académicos.

El instrumento de medición corresponde a una encuesta estructurada y anónima aplicada a estudiantes que cursaron Estadística I, cuyo propósito es identificar su percepción sobre la asignatura y la pertinencia de integrar laboratorios computacionales con analítica de datos. La encuesta incluye preguntas de caracterización (edad, género, semestre actual y semestre en el que cursó la asignatura), una valoración general de la calidad de la enseñanza (escala 1 a 5), e ítems sobre aplicabilidad de los contenidos, uso de casos reales y realización de actividades prácticas guiadas en computador.

Adicionalmente, incorpora una autoevaluación del manejo de herramientas digitales antes del curso y del nivel actual en competencias específicas (manejo de Excel, limpieza de datos,

visualización e interpretación de resultados), así como una pregunta de selección múltiple sobre herramientas utilizadas (Excel, Python, R, Power BI, Minitab u otras). Finalmente, se incluyen ítems tipo Likert (1 a 5) para medir el grado de acuerdo y disposición frente a la implementación de laboratorios computacionales, el interés por trabajar con Python o R mediante guías paso a paso y la comodidad usando Excel para análisis básico de datos.

3.1.3.1 Diseño del estudio. La población objetivo estuvo conformada por $N = 779$ estudiantes matriculados entre sexto y décimo semestre, niveles en los cuales la asignatura Estadística I constituye un requisito académico previo. El tamaño de muestra se estimó empleando la fórmula de estimación de una proporción, considerando que la mayor parte de la encuesta de percepción eran variables de naturaleza categórica, y bajo la asunción de un nivel de confianza del 95% ($Z_{\alpha/2} = 1.96$), un margen de error de 5 puntos porcentuales sobre las proporciones ($\varepsilon = 0.05$) y máxima variabilidad poblacional ($p = q = 0.5$).

Primero se calcula el tamaño muestral para población infinita,

$$n_0 = \frac{Z^2 pq}{e^2},$$

obteniendo un tamaño muestral para población infinita de $n_0 = 384,16$. Posteriormente, considerando la naturaleza de la población definida, se realiza el ajuste para población finita,

$$n = \frac{n_0}{1 + \frac{n_0 - 1}{N}},$$

obteniendo un tamaño muestral de $n = 257,4$ el cual se aproxima al entero superior, en consecuencia, el tamaño de muestra requerido es de 258 estudiantes.

Para el desarrollo del estudio se establecieron criterios que garanticen la pertinencia de los participantes, considerando el objetivo del diagnóstico. Como criterios de inclusión se consideraron estudiantes de ingeniería industrial que:

- se encuentren activos matriculados entre 6° y 10° semestre,
- hayan cursado y aprobado la asignatura Estadística I,
- acepten participar voluntariamente en la encuesta bajo condiciones de anonimato y uso académico de la información.

Además, se contempla como criterio de exclusión el eliminar todos los registros que presenten respuestas incompletas o inconsistentes, que no permitan su análisis apropiado.

3.1.3.2 Análisis de resultados de la encuesta de percepción. Para la aplicación de la encuesta se utilizó un muestreo no probabilístico de tipo intencional o por voluntarios, dado que la selección de los participantes se realizó con base en condiciones académicas previamente definidas. La encuesta fue distribuida en grupos académicos de WhatsApp y aplicada en salones de asignaturas de niveles superiores del programa de Ingeniería Industrial. Este procedimiento permitió garantizar que los encuestados cumplieran con los criterios de inclusión establecidos, asegurando así la pertinencia y coherencia de las respuestas con los objetivos del diagnóstico.

Una vez aplicada la encuesta a 258 estudiantes que habían cursado previamente la asignatura Estadística I, cuyas respuestas se encuentran consolidadas en el Apéndice B, se procedió al análisis descriptivo de las variables más relevantes para el diagnóstico del estado actual del curso y la pertinencia de integrar laboratorios computacionales orientados a la analítica de datos. Para el tratamiento de las preguntas medidas mediante escala tipo Likert se emplearon frecuencias absolutas y porcentajes por categoría de respuesta, considerando que este tipo de escala permite identificar la distribución de las percepciones de los estudiantes en cada nivel de valoración.

En primer lugar, se analizó la percepción general de los estudiantes frente a la calidad de la enseñanza recibida en Estadística I. Los resultados muestran una valoración mayoritariamente

positiva, dado que 111 estudiantes (43.02%), asignaron la calificación máxima de 5, mientras que 103 estudiantes (39.92%), seleccionaron el nivel 4. Por su parte, 30 estudiantes (11.63%) seleccionaron el nivel 3, y únicamente 14 estudiantes (5.43%) se ubicaron en los niveles 1 y 2. Este resultado indica que la asignatura cuenta con una percepción favorable en términos de enseñanza, aunque ello no excluye la existencia de oportunidades de fortalecimiento en su componente aplicado.

Respecto a la aplicabilidad de los contenidos de Estadística I en problemas reales, 190 estudiantes (73.64%) consideraron que los contenidos sí resultaron aplicables. No obstante, 57 estudiantes (22.09%) señalaron que solo algunas temáticas fueron aplicables, mientras que 11 estudiantes (4.26%) indicaron que no percibieron dicha aplicabilidad. Este comportamiento evidencia que, aunque la mayoría reconoce la utilidad de los contenidos, existe un grupo importante de estudiantes que percibe una aplicación parcial, lo cual refuerza la necesidad de fortalecer la conexión entre los conceptos estadísticos y situaciones prácticas cercanas al contexto profesional.

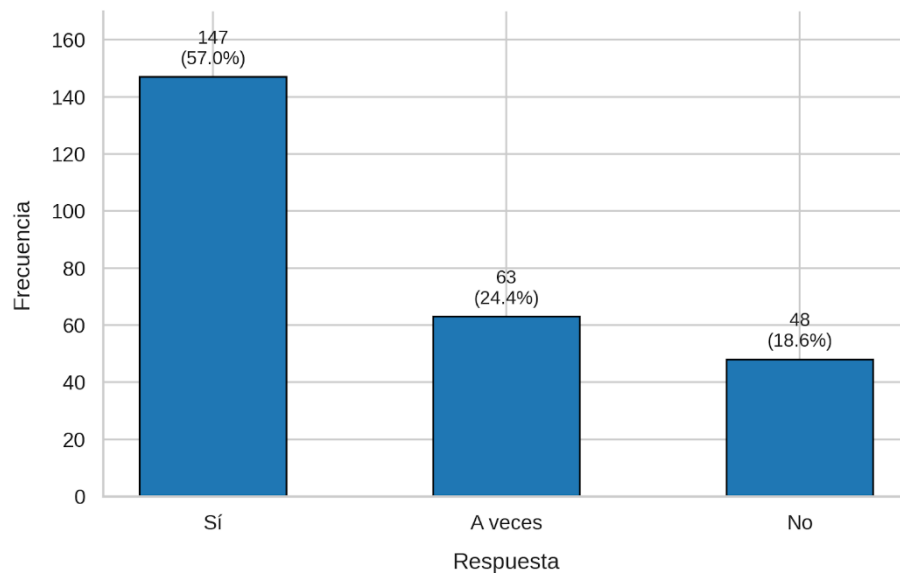
De manera complementaria, frente al uso de casos de estudio o ejemplos de la vida real en clase, 150 estudiantes (58.14%) respondieron afirmativamente, 81 estudiantes (31.40%) indicaron que estos se utilizaron solo algunas veces, y 27 estudiantes (10.47%) señalaron que no se usaron. Estos resultados muestran que sí existe una incorporación de ejemplos aplicados dentro de la asignatura, pero no de forma completamente homogénea para todos los estudiantes. En consecuencia, se identifica una oportunidad para estructurar actividades prácticas más sistemáticas que permitan trabajar con datos y contextos reales durante el desarrollo del curso.

En cuanto a la experiencia previa con actividades prácticas guiadas en computador, tales como laboratorios, talleres, proyectos finales o notebooks, 147 estudiantes (56.98%) manifestaron

haber recibido este tipo de actividades. Sin embargo, 63 estudiantes (24.42%) indicaron que las recibieron solo algunas veces, y 48 estudiantes (18.60%) señalaron que no tuvieron esta experiencia. Como se observa en la Figura 1, aunque más de la mitad de los encuestados reportó haber realizado actividades prácticas guiadas, el 43.02% restante se ubicó entre las opciones “A veces” y “No”, lo que evidencia que el componente práctico no ha sido uniforme en todos los casos.

Figura 1.

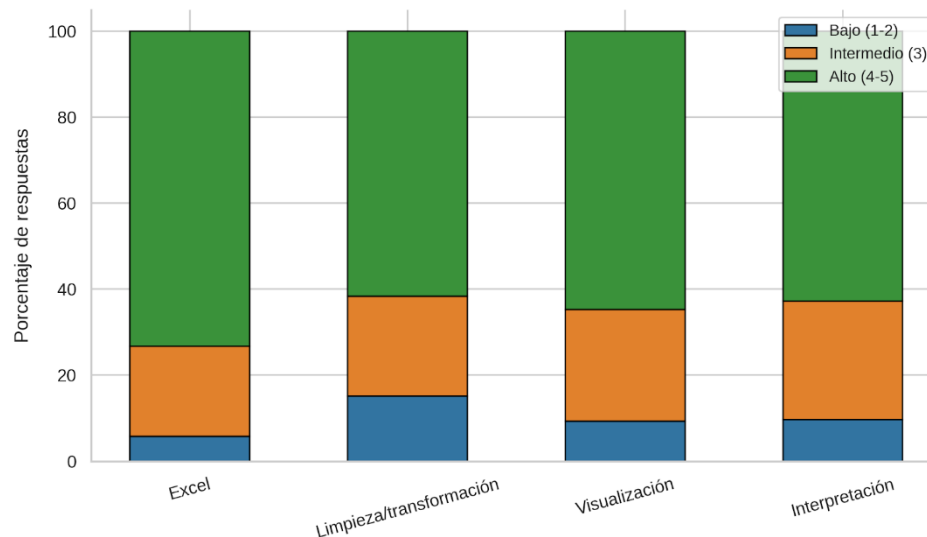
Frecuencia de actividades prácticas guiadas en computador en Estadística I



Posteriormente, se analizó el nivel de manejo que los estudiantes manifestaron tener antes de cursar Estadística I frente a herramientas digitales para datos, como Excel, Python, R, Power BI, Minitab u otras. Los resultados evidencian una distribución heterogénea: 17 estudiantes (6.59%) se ubicaron en el nivel 1, 51 estudiantes (19.77%), en el nivel 2, 86 estudiantes (33.33%) en el nivel 3, 52 estudiantes (20.16%) en el nivel 4 y 52 estudiantes (20.16%) en el nivel 5. En términos generales, el 26.36% de los encuestados se ubicó en niveles bajos, el 33.33% en un nivel intermedio y el 40.31% en niveles altos. Esta distribución confirma el hallazgo de las encuestas de los docentes, los estudiantes no ingresan a la asignatura con un nivel homogéneo de alfabetización

digital, por lo que los laboratorios deben diseñarse de manera gradual, con instrucciones claras y acompañamiento suficiente.

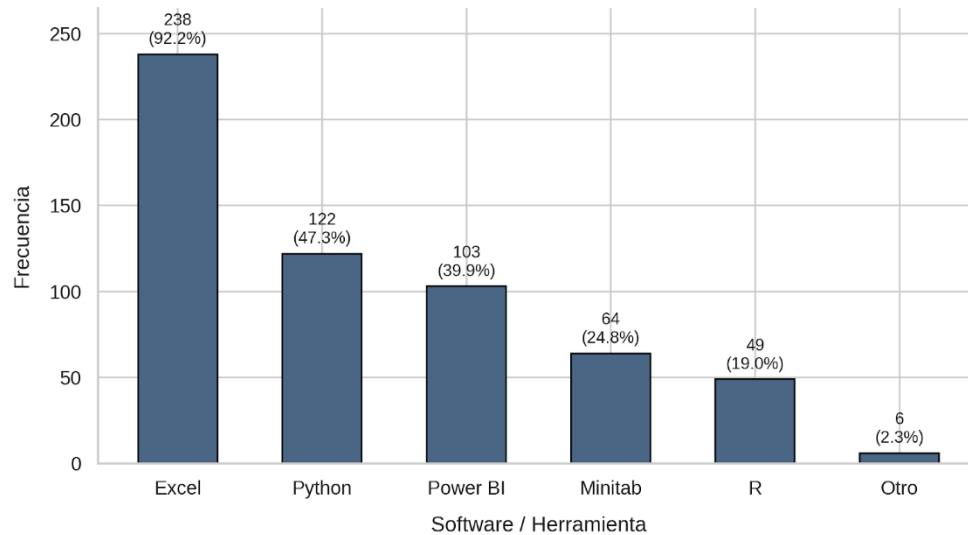
En la competencia de limpieza y transformación de datos, 159 estudiantes (61.63%) se ubicaron en los niveles 4 y 5; 60 estudiantes (23.26%) seleccionaron el nivel 3; y 39 estudiantes (15.12%) se ubicaron en los niveles bajos. En visualización de datos, 167 estudiantes (64.73%) se ubicaron en los niveles superiores; 67 estudiantes (25.97%) en el nivel intermedio, y 24 estudiantes (9.30%) en niveles bajos. Finalmente, en la interpretación de resultados generados por software, 162 estudiantes (62.79%) se ubicaron en los niveles 4 y 5, 71 estudiantes (27.52%) en el nivel 3 y 25 estudiantes (9.69%) en niveles bajos. Como se presenta en la Figura 2, en términos generales estas competencias muestran una base favorable, pero también evidencian la necesidad de fortalecer habilidades asociadas al tratamiento, visualización e interpretación de datos. Además, el considerar que un 40% aproximado de estudiantes que no llegan con un nivel alto de competencias digitales es un factor que se debe considerar en el diseño de actividades que involucren TIC en el aula.

Figura 2.*Nivel actual percibido en competencias digitales para datos*

En relación con las herramientas utilizadas durante o después de cursar Estadística I, se identificó que Excel fue la herramienta más frecuente, con 238 menciones (92.25%) de los encuestados. Le siguieron Python con un 47.29% de uso, Power BI con un 39.92%, Minitab con un 24.81%; y en menor medida el lenguaje de programación R con 49 menciones (18.99%). Además, 6 estudiantes (2.33%) indicaron el uso de otras herramientas. Es importante señalar que esta pregunta permitía seleccionar más de una opción, por lo que cada herramienta se considera una variable binaria aislada para el análisis de frecuencias. Como se observa en la Figura 3, Excel se consolida como la herramienta de mayor reconocimiento y familiaridad entre los estudiantes, lo cual evidencia que existe una base inicial para el trabajo con datos; sin embargo, también se observa una presencia significativa de herramientas como Python y Power BI, asociadas a procesos de análisis, visualización y comunicación de información.

Figura 3.

Software o herramientas utilizadas durante o después de cursar Estadística I.



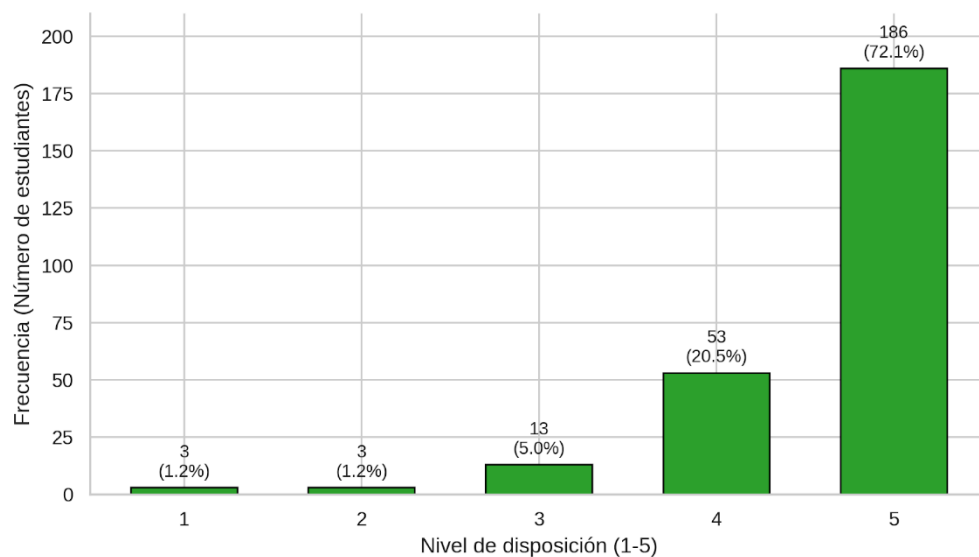
Posteriormente, se evaluó que tan de acuerdo están los estudiantes frente a la afirmación: “Integrar laboratorios computacionales en Estadística I mejoraría mi aprendizaje”. Los resultados muestran una postura favorable por parte de los encuestados, dado que 190 estudiantes (73.64%) manifestaron el nivel máximo de acuerdo, y 52 estudiantes (20.16%) seleccionaron el nivel 4, lo que indica que más del 90% de los encuestados en los niveles superiores de la escala. Por otra parte, 9 estudiantes (3.49%) optaron por la opción intermedia, y únicamente 7 estudiantes (2.71%) se ubicaron en los niveles de desacuerdo 1 y 2.

En la misma línea, se consultó sobre la disposición de los alumnos para realizar trabajos prácticos empleando lenguajes de programación como Python o R, bajo la condición de contar con guías didácticas estructuradas paso a paso en el curso. Ante este planteamiento, se registró una recepción positiva, donde 186 estudiantes (72.09%) manifestaron estar totalmente dispuestos y 53 estudiantes (20.54%) marcaron el nivel 4. De esta manera, se evidenció una alta apertura hacia la incorporación de herramientas de programación en la asignatura. Por otro lado, 13 estudiantes

(5.04%) se mantuvieron en una postura neutral, mientras que solo 6 estudiantes (2.33%) se inclinaron por los niveles de disposición baja 1 y 2. Este comportamiento (ver **Figura 4**) es fundamental para el proyecto, pues demuestra que el temor técnico o la resistencia natural hacia entornos de programación se mitiga por completo si el proceso pedagógico está mediado por una instrucciones claras, secuenciales y guiadas.

Figura 4.

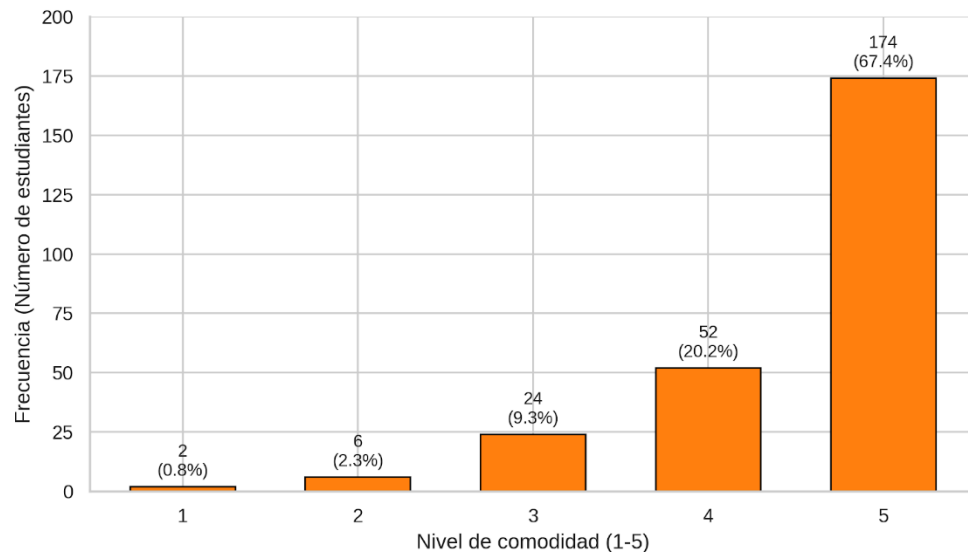
Disposición para realizar trabajos prácticos con Python o R con guías paso a paso.



Finalmente, se examinó el nivel de comodidad autopercibido por los encuestados respecto al uso de la herramienta Microsoft Excel para la ejecución de análisis básicos de datos (ver Figura 5). Los resultados muestran que el 87.60% de los participantes se ubica en los niveles altos de confianza con la herramienta de hojas de cálculo. Por su parte, 24 estudiantes (9.30%) manifestaron un nivel de comodidad intermedio, y únicamente 8 estudiantes (3.10%) indicaron sentirse incómodos en los niveles 1 y 2.

Figura 5.

Nivel de comodidad de los estudiantes usando Excel para el análisis de datos.



A pesar de esta marcada familiaridad instrumental con las hojas de cálculo tradicionales, los requerimientos metodológicos planteados en este proyecto (enfocados en el uso de contextos y conjuntos de datos reales) exigieron replantear el alcance de las herramientas convencionales de aula. Al trabajar con bases de datos reales y de gran escala, el software de escritorio tradicional suele presentar limitaciones operativas para el procesamiento eficiente de grandes volúmenes de información. Esta restricción técnica se alinea directamente con los precedentes identificados en el marco de antecedentes, como fue la investigación documentada por Sánchez y Puerto (2025), donde resaltan que la incorporación de técnicas analíticas en el aula puede verse severamente limitada por el factor de accesibilidad tecnológica, especialmente cuando el software seleccionado demanda un alto consumo de recursos e introduce dificultades de ejecución en los equipos de cómputo personales de los estudiantes.

Por consiguiente, el cruce de este antecedente con los resultados del diagnóstico inicial planteó una triple necesidad de cara a las siguientes fases del proyecto: en primer lugar, la urgencia de migrar hacia herramientas computacionales capaces de gestionar bases de datos de un volumen

intermedio-alto, superando el enfoque de las hojas de cálculo tradicionales que, si bien son familiares para los estudiantes, la literatura especializada señala como herramientas limitadas y simplistas frente a la complejidad, dimensionalidad y automatización que exigen los problemas reales de la industria (Provost y Fawcett, 2013); en segundo lugar, la oportunidad de utilizar la alta disposición manifestada por los estudiantes para interactuar con lenguajes de programación, la cual se encuentra directamente condicionada a la existencia de un acompañamiento estructurado mediante guías paso a paso; y, en tercer lugar, se evidencia la necesidad seleccionar alternativas tecnológicas que superen las barreras de hardware local.

De este modo, el diagnóstico justifica la transición hacia entornos de procesamiento más robustos, accesibles y guiados, garantizando una experiencia de aprendizaje equitativa, metodológica y adaptada a las exigencias informáticas de la Industria 4.0.

3.2 Fase 2: Revisión de experiencias y buenas prácticas

En concordancia con los objetivos establecidos, la Fase 2 consistió en la exploración, revisión y análisis de la literatura académica y de buenas prácticas relacionadas con la enseñanza de estadística y la analítica de datos en programas de ingeniería. El propósito de esta revisión fue identificar los enfoques pedagógicos y los requerimientos tecnológicos documentados en casos de éxito, para establecer los lineamientos base que orientarían la posterior estructuración de los laboratorios computacionales.

3.2.1 Exploración de literatura en didáctica de la estadística y analítica de datos

El análisis documental evidenció una preocupación generalizada en la educación superior respecto a la enseñanza tradicional de la estadística. Autores como Cobb (2007) señalan que los currículos tradicionales suelen estar sobrecargados de fórmulas y teoría probabilística, generando un "currículo ptolemaico" que aleja al estudiante de la verdadera naturaleza del análisis de datos.

Para contrarrestar esto, la literatura reciente resalta la necesidad de orientar las asignaturas cuantitativas hacia el desarrollo de la capacidad de "pensar con datos" (think with data), integrando conceptos de computación y modelado desde los primeros semestres (Hardin et al., 2015).

Esta transición se fundamenta teóricamente en la perspectiva de Tukey (1977), quien enfatiza que el análisis exploratorio requiere observar los datos críticamente antes de imponer modelos rígidos. Al trasladar este enfoque a la ingeniería industrial contemporánea, Davenport y Harris (2007) confirman que el perfil profesional demanda competencias para formular preguntas relevantes y traducir resultados matemáticos en decisiones de negocio apoyadas en evidencia empírica.

3.2.2 Identificación de buenas prácticas pedagógicas

De la revisión de referentes pedagógicos aplicados a disciplinas STEM (Ciencia, Tecnología, Ingeniería y Matemáticas), se identificó que el aprendizaje activo y el uso de laboratorios prácticos superan significativamente a las clases teóricas tradicionales. El metaanálisis desarrollado por Freeman et al. (2014) demostró que la integración de actividades prácticas estructuradas (como los laboratorios computacionales) no solo incrementa el desempeño académico, sino que reduce sustancialmente las tasas de reprobación en asignaturas de ingeniería y matemáticas.

Bajo la perspectiva del diseño curricular, se adoptó el principio de alineación constructiva de Biggs y Tang (2011), asegurando que los laboratorios no sean un anexo aislado, sino una actividad coherente con los resultados de aprendizaje esperados. Asimismo, la revisión de buenas prácticas resalta la comunicación de resultados como una competencia indispensable. Knaflitz (2015) establece que el *Data Storytelling* es el puente entre el rigor técnico y la aplicación práctica,

exigiendo que las actividades de laboratorio obliguen al estudiante a articular narrativas visuales coherentes.

De manera complementaria, la Asociación Estadounidense de Estadística (American Statistical Association), mediante las recomendaciones del informe GAISE College Report (Carver et al. 2016), propone que la enseñanza de la estadística en educación superior debe orientarse hacia el aprendizaje activo, el uso de datos reales, el desarrollo del pensamiento estadístico y la integración de herramientas tecnológicas que permitan al estudiante explorar, analizar e interpretar información en contextos auténticos. Estas recomendaciones enfatizan que el aprendizaje mejora cuando los estudiantes participan en experiencias donde los conceptos estadísticos se aplican a problemas cercanos a su realidad, en lugar de limitarse a resolver ejercicios descontextualizados o centrados exclusivamente en procedimientos algebraicos.

3.2.3 Transición hacia nuevas herramientas tecnológicas

Finalmente, la revisión metodológica permitió definir la estructura técnica requerida para los laboratorios. Siguiendo el estándar analítico de la industria, se revisó la metodología CRISP-DM (Chapman et al., 2000), la cual establece que la comprensión del dominio y la preparación de los datos son etapas críticas.

Para lograr esta inmersión realista en el aula, Saltz y Stanton (2017) recomiendan abandonar las hojas de cálculo convencionales al enseñar analítica, promoviendo en su lugar la inmersión temprana en entornos de programación y software de inteligencia de negocios. Este enfoque exige trabajar exclusivamente con bases de datos crudas y reales, evitando los "datos de juguete" (toy datasets), para simular el ruido y la complejidad de los entornos industriales. Estos requerimientos documentales conformaron el marco de referencia que justificó la selección

tecnológica (Python y Power BI) y el diseño de la arquitectura didáctica ejecutada en la Fase 3 del proyecto.

3.3 Fase 3: Diseño de los materiales tipo laboratorio

Una vez consolidados los requerimientos prácticos del diagnóstico inicial (Fase 1) y tomando como base los lineamientos y buenas prácticas identificados en la revisión de la literatura (Fase 2), se procedió con el desarrollo de la Fase 3 del proyecto. Esta etapa se centró en el diseño y la estructuración de los laboratorios computacionales, lo cual abarcó la selección de bases de datos reales, la definición de una estructura pedagógica para las guías y el desarrollo de los contenidos técnicos específicos para cada sesión práctica.

3.3.1 Selección y caracterización de las bases de datos reales

Para el desarrollo de las prácticas se definió el uso de bases de datos reales extraídas del portal de datos abiertos del Gobierno de Colombia (www.datos.gov.co). Esta elección responde a la necesidad de que el estudiante se enfrente a escenarios reales de ingeniería y análisis de información, superando los entornos controlados o datos ajustados para facilitar su análisis. Como señalan Provost y Fawcett (2013), el desarrollo del pensamiento analítico orientado a datos (*data-analytic thinking*) requiere que el estudiante interactúe con la complejidad y la naturaleza multifactorial de los datos del entorno real, entendiendo que la información no se presenta de forma estructurada en los contextos organizacionales.

Para la selección del banco de datos, se aplicaron los siguientes criterios técnicos:

- **Volumen de información:** Conjuntos de datos con un número elevado de registros que justificaran el uso de herramientas especializadas, demostrando las limitaciones del procesamiento manual tradicional.

- Tipos de variables mixtas: Bases de datos que integraran variables cualitativas (nominales y ordinales) y cuantitativas (discretas y continuas) para abordar tanto la estadística descriptiva como los modelos probabilísticos.
- Presencia de valores inconsistentes o faltantes: Datos reales que requirieran un proceso previo de limpieza, permitiendo desarrollar habilidades prácticas en la preparación de la información.
- Pertinencia del contexto: Temáticas del entorno social, público o de salud que facilitaran la interpretación de los resultados estadísticos en problemas aplicados.

Es importante destacar que los repositorios públicos oficiales seleccionados poseen originalmente una alta dimensionalidad y una cantidad considerable de variables administrativas, técnicas y de registro propias de cada entidad gubernamental. Por lo tanto, como parte del diseño metodológico de este proyecto, se realizó un proceso de delimitación y selección dirigida de variables. Esto permitió acotar el análisis a los objetivos específicos de aprendizaje de la asignatura Estadística I, evitando la dispersión cognitiva del estudiante en el uso de la herramienta informática, a la vez que estimula de manera directa la limpieza de datos al obligar al alumno a identificar, aislar y depurar únicamente los campos de interés bajo criterios lógicos, conservando el volumen masivo y el comportamiento real de la información.

Bajo estos criterios, se seleccionaron tres conjuntos de datos institucionales para el desarrollo de los laboratorios, como se puede observar en la Tabla 2.

Tabla 2.*Conjuntos de datos públicos seleccionados para los laboratorios.*

Conjunto de datos	Dimensión	Justificación
Casos positivos de COVID-19	6,39 millones de registros con 23 variables	Es un conjunto de datos de alta dimensionalidad, de un contexto conocido por lo estudiantes en los que se pueden analizar y visualizar indicadores. Ideal para el tema de medidas descriptivas y visualizaciones.
Accidentes de tránsito ocurridos en el municipio de Bucaramanga	39,2 mil registros distribuidos en 24 variables	Es un conjunto de datos de una temática cercana al contexto de la ciudad, ideal para calcular proporciones y probabilidades de un evento asociado a datos geográficos o características del siniestro.
Embarazo en adolescentes	9.513 registros y 51 variables	Es un conjunto de datos que contiene información de los padres y del control médico del nacimiento. Ideal para analizar distribuciones de los conjuntos de datos, ya que tiene varias variables numéricas, algunas con simetrías marcadas y otras con valores atípicos.

Para el diseño de los laboratorios, se decidió no aplicar ningún preprocesamiento de los datos, sino que los estudiantes descargaran los datos directamente del portal web. Esta decisión se fundamenta en que, al dedicar parte del laboratorio a estas actividades de preprocesamiento, eliminación de valores nulos y atípicos, se permite reforzar estas habilidades en los estudiantes.

3.3.2 Definición de la Estructura Pedagógica de los Laboratorios

Con el fin de estandarizar las prácticas y facilitar el aprendizaje autónomo de los estudiantes, se diseñó una estructura común para las guías de laboratorio. Esta arquitectura didáctica se fundamenta en el principio de alineación constructiva de Biggs y Tang (2011), el cual establece que, para lograr un aprendizaje profundo, debe existir una coherencia directa entre los objetivos declarados, las actividades técnicas que el estudiante ejecuta y las herramientas de

evaluación implementadas. En consecuencia, cada guía se compuso de cinco secciones, como se puede observar en la Tabla 3.

Tabla 3.

Estructura de las guías de los laboratorios propuestos.

Sección	Descripción
Propósito de aprendizaje	Define el objetivo conceptual de la sesión, relacionando el tema estadístico con el uso de la herramienta informática.
Competencias a desarrollar	Detalla las habilidades de análisis, visualización o modelado que el estudiante consolidará al finalizar la práctica.
Caso de estudio	Presenta el problema del entorno real que contextualiza el análisis de los datos y le da sentido al ejercicio.
Instrucciones paso a paso	Describe secuencialmente el procedimiento técnico para cargar los datos, limpiarlos, realizar los cálculos o programar las funciones necesarias.
Preguntas de análisis e interpretación	Serie de preguntas diseñadas para que el estudiante interprete el significado de los resultados y construya conclusiones aplicadas.

En cada uno de estos laboratorios se propuso de forma adicional un entregable para verificar el entendimiento de los estudiantes sobre los contenidos dados. Estas actividades son las que permiten a los estudiantes poner en práctica las herramientas y las técnicas ampliadas en los laboratorios. Algunos entregables consistieron en replicar el diseño de un dashboard con otro conjunto de datos público o el responder unas preguntas de análisis después de seguir el laboratorio de forma autónoma.

3.3.3 Contenido y justificación de los laboratorios

De acuerdo con las necesidades de formación identificadas y las actividades planeadas, el repositorio se estructuró en tres laboratorios computacionales que avanzan de acuerdo con los contenidos de la asignatura durante el semestre y de forma secuencial desde la descripción visual

de datos hasta el modelado probabilístico avanzado. El contenido completo de estas guías didácticas se encuentra en el Apéndice C.

- Laboratorio 1: Diseño de dashboards descriptivos en Power BI

Caso de estudio: Análisis descriptivo de casos positivos de COVID-19 en Colombia.

Herramienta: Power BI (Power Query y expresiones DAX).

Contenido temático: estadística descriptiva, medidas numéricas y representaciones gráficas.

Justificación: Este primer laboratorio introduce al estudiante en la gestión de bases de datos reales y con un gran volumen de información. Se seleccionó Power BI debido a su capacidad para conectarse a fuentes de datos externas y por la potencia de Power Query, el cual permite realizar de forma gráfica la limpieza de datos. Asimismo, la herramienta permite introducir el modelado lógico mediante expresiones DAX básicos como COUNTROWS (para conteo de registros), AVERAGE y MEDIAN (para medidas de tendencia central), y la función condicional SWITCH (para la agrupación de la variable edad en rangos). El laboratorio cierra con un componente de Data Storytelling, donde el estudiante aprende a estructurar un tablero con filtros interactivos y a redactar un informe dirigido a tomadores de decisiones.

Aplicación práctica: Esta sesión se desarrolla bajo una metodología de aprendizaje progresivo y transferencia de conocimiento. Durante la jornada presencial en el aula de cómputo, el docente guía a los estudiantes en la construcción conjunta del dashboard principal utilizando el conjunto de datos de COVID-19. Posteriormente, como actividad de aplicación autónoma y entregable calificable, cada grupo de trabajo debe seleccionar, descargar y limpiar una base de datos completamente nueva desde el portal Datos.gov.co

para diseñar de forma independiente un segundo tablero interactivo que responda a una problemática social o pública de su propio interés, replicando las competencias técnicas adquiridas. Cabe resaltar que, dentro de la dinámica de esta práctica, el dominio instrumental de la herramienta informática se considera un elemento secundario; la prioridad fundamental de la sesión radica en la capacidad crítica del estudiante para procesar la información, identificar patrones analíticos abstractos, extraer conclusiones estadísticamente válidas y comunicarlas de forma clara a públicos no especializados.

- Laboratorio 2: Aplicación del teorema de Bayes en Colab

Caso de estudio: Análisis probabilístico de factores asociados a accidentes de tránsito en el municipio de Bucaramanga.

Herramienta: Python en el entorno virtual Google Colab.

Contenido temático: método empírico de cálculo de probabilidad y teorema de Bayes

Justificación técnica: En esta sesión se realiza la transición formal hacia los lenguajes de programación analítica utilizando Python. La elección de Google Colab como entorno de desarrollo responde a la necesidad de procesar el dataset de siniestralidad vial sin generar conflictos de compatibilidad o rendimiento en los equipos de los estudiantes, al ejecutar los scripts directamente en la nube. El contenido técnico se enfoca en la manipulación de dataframes para calcular frecuencias absolutas y relativas. El estudiante aprende a construir condiciones lógicas sobre las variables seleccionadas utilizando funciones de Python. Esto permite estimar empíricamente probabilidades simples, conjuntas y condicionales, aplicando finalmente el Teorema de Bayes para actualizar probabilidades posteriores ante nueva evidencia y contrastar los resultados matemáticos con los datos observados en el caso real.

Aplicación práctica: La dinámica de este laboratorio exige que los estudiantes completen de manera lógica la sintaxis del código para terminar de definir las variables probabilísticas y estructurar el cálculo de probabilidades condicionales. Con el propósito de mitigar el riesgo de copia genérica y fomentar el análisis situado en ingeniería, el docente asigna de manera exclusiva un barrio específico del municipio de Bucaramanga a cada equipo de trabajo. Los estudiantes se ven obligados a realizar un filtrado condicional personalizado del dataset general para estimar las probabilidades del Teorema de Bayes.

- Laboratorio 3: Modelos de probabilidad aplicados en Colab

Caso de estudio: Análisis de variables numéricas asociadas al embarazo en adolescentes

Herramienta: Python en el entorno virtual Google Colab.

Contenido temático: modelos teóricos de probabilidad (variables discretas y continuas)

Justificación técnica: El último laboratorio consolida el uso de la programación para el modelado probabilístico de variables discretas y continuas utilizando datos reales de salud pública. La guía incluye un proceso inicial de depuración donde el estudiante debe identificar y tratar valores especiales o inconsistentes sin eliminar filas completas, preservando la mayor cantidad de registros dentro del conjunto de datos. La justificación técnica de este laboratorio radica en el uso de la librería científica *scipy.stats*, mediante la cual se aplican de forma directa cuatro modelos de probabilidad: la distribución Normal (para analizar el peso y la talla del neonato), la distribución Uniforme (para evaluar la edad de la madre), la distribución Binomial (aplicada al cálculo de probabilidades en muestras de bajo peso o madres menores de edad) y la distribución Exponencial (para modelar la diferencia de edad entre el padre y la madre). El diseño incluye la automatización de un diccionario de respuestas que recopila las interpretaciones lógicas del estudiante,

facilitando la reproducibilidad del análisis estadístico y la entrega estructurada de resultados.

Aplicación práctica: Este laboratorio interactivo está diseñado para ejecutarse de manera guiada y secuencial dentro del notebook de Google Colab. Los estudiantes avanzan ejecutando los bloques de código destinados al cálculo matemático y generación gráfica de las distribuciones (Normal, Uniforme, Binomial y Exponencial). A medida que interpretan el comportamiento de cada modelo estadístico, el entorno virtual ejecuta una función en Python que guarda de manera automatizada las respuestas analíticas redactadas por el alumno dentro de un diccionario estructurado de datos. Al finalizar la práctica, el mismo código consolida y exporta estas respuestas en un archivo Word, sirviendo como el entregable final de la sesión y garantizando un proceso de evaluación transparente que mide la capacidad de argumentación crítica del estudiante.

3.3.4 Criterios y rúbrica de evaluación

En concordancia con el enfoque de competencias y el principio de alineación constructiva, el proceso de evaluación de los laboratorios computacionales se diseñó mediante un sistema de rúbricas estandarizadas, las cuales se encuentran recopiladas en el Apéndice D. Cada instrumento evalúa el desempeño estudiantil en una escala de 1 a 5 (donde 1 representa un nivel muy bajo y 5 un nivel excelente), distribuyendo los porcentajes de calificación en criterios específicos que responden a los objetivos técnicos y conceptuales de cada sesión, como se presenta en la Tabla 4. Estas evaluaciones son aplicadas sobre las actividades entregables solicitadas a los estudiantes como complemento a lo trabajado en el laboratorio en clase.

Tabla 4.*Estructura de las guías de los laboratorios propuestos.*

Laboratorio	Énfasis	Ítems de evaluación
Power BI	Medir la capacidad de transformación y comunicación descriptiva del estudiante	Selección y justificación de la base de datos (10%)
		Preparación y comprensión de los datos (20%)
		Construcción de indicadores descriptivos (20%)
		Diseño y calidad del dashboard (10%)
		Interpretación y análisis de resultados (20%)
		Data Storytelling y comunicación (20%)
Teorema de Bayes en Colab	Calificar la transición hacia la programación analítica aplicada a probabilidades	Preparación y comprensión de la base de datos (20%)
		Definición de eventos probabilísticos a partir de las variables (20%)
		Cálculo de probabilidades simples y condicionales (25%)
		Interpretación y análisis de resultados (30%)
		Presentación y organización del trabajo entregado (5%)
Modelos de Probabilidad en Colab	Evaluar el rigor matemático y el modelado probabilístico automatizado	Exploración e interpretación inicial de los datos a través de histogramas (20%)
		Aplicación de modelos de probabilidad a partir de las variables mediante la librería <code>scipy.stats</code> (20%)
		Interpretación de probabilidades y comparación entre resultados teóricos y observados (25%)
		Análisis crítico e interpretación de resultados (30%)
		Presentación y organización del trabajo entregado (5%)

3.3.5 Validación interna y control de calidad de los materiales

Como actividad de cierre para la Fase 3, se ejecutó un proceso de validación interna y control de calidad sobre el repositorio de laboratorios computacionales, esto se desarrolló una semana antes de la aplicación de cada sesión práctica en el aula. La dinámica consistió en presentar el material didáctico y la idea pedagógica al director del proyecto; posteriormente, se realizaba la ejecución conjunta de las herramientas, donde el director evaluaba el funcionamiento en tiempo

real y realizaba las correcciones técnicas o metodológicas que consideraba pertinentes para optimizar la guía antes de su implementación con los estudiantes.

Durante estas revisiones, el director evaluó el funcionamiento en tiempo real para verificar la sintaxis técnica de las expresiones lógicas en Power BI (fórmulas DAX) y la correcta estructuración de los scripts de Python en Google Colab, corrigiendo cualquier error de manera inmediata. Asimismo, se comprobó el procesamiento de los conjuntos de datos masivos seleccionados, validando que estas plataformas virtuales superaran eficazmente las restricciones de hardware local identificadas en los antecedentes de la asignatura para garantizar un entorno de trabajo fluido. Tras implementar de forma oportuna las correcciones y sugerencias metodológicas acordadas en cada sesión de revisión, los materiales didácticos obtuvieron el aval definitivo por parte del director quedando plenamente consolidados para iniciar la fase de implementación piloto con los estudiantes de la asignatura.

3.4 Fase 4: Implementación piloto y validación de los resultados

De acuerdo con los objetivos metodológicos establecidos en el plan, la Fase 4 se centró en la aplicación práctica de los laboratorios diseñados y en la evaluación de su impacto real en el aula de clase. Esta fase comprendió tres actividades centrales: el desarrollo presencial de los laboratorios con estudiantes de la asignatura Estadística I, la medición del nivel de adquisición de competencias mediante una evaluación diagnóstica aplicada antes y después de las prácticas (Pre-test y Post-test), y la verificación de la transferencia autónoma de conocimiento a través de la revisión de los proyectos finales del curso.

3.4.1 Aplicación piloto y percepción por sesión

La implementación piloto de los tres laboratorios computacionales se llevó a cabo en el salón de clase de la Escuela de Estudios Industriales y Empresariales con los estudiantes

matriculados el grupo B1 de Estadística I, en el primer semestre del 2026. Las sesiones se desarrollaron de forma presencial durante el horario habitual del curso, establecido los miércoles y viernes de 10:00 a. m. a 12:00 m.

El cronograma se distribuyó de la siguiente manera: el Laboratorio 1 (Power BI) se aplicó el viernes 27 de febrero, el Laboratorio 2 (Teorema de Bayes en Colab) se ejecutó el viernes 17 de abril y el Laboratorio 3 (Modelos de probabilidad en Colab) se desarrolló el miércoles 20 de mayo. Cada sesión se desarrolló bajo un modelo de transferencia gradual de conocimiento, las clases iniciaban con una demostración técnica guiada por la autora del proyecto para contextualizar el caso de estudio, dando paso inmediato a una fase de trabajo autónomo donde los estudiantes debían procesar los datos, resolver los requerimientos técnicos y argumentar sus propios análisis.

Al finalizar cada una de las jornadas, se aplicó a los estudiantes una breve encuesta de percepción con el propósito de realizar una retroalimentación sobre la claridad de las instrucciones paso a paso, la usabilidad de las herramientas informáticas y las principales dificultades técnicas encontradas durante cada sesión. Este instrumento permitió registrar la experiencia cualitativa de los alumnos por clase y sirvió como mecanismo de control para evaluar la pertinencia del material. Para estandarizar esta medición, se diseñó un instrumento de evaluación estructurado bajo una escala tipo Likert de cinco niveles (donde 1 representa total desacuerdo y 5 total acuerdo), compuesto por ítems cerrados y preguntas abiertas orientadas a la retroalimentación cualitativa. Los formatos de las encuestas se encuentran adjuntos en el Apéndice F.

3.4.1.1 Resultados Laboratorio 1. Luego de la ejecución de la primera sesión práctica, orientada al diseño de tableros descriptivos y visualización de datos en Power BI, se recolectaron los proyectos desarrollados de manera autónoma por los equipos de trabajo. Algunas muestras se encuentran adjuntas en el Apéndice E.

Posteriormente, con el objetivo de evaluar la experiencia de aprendizaje, se aplicó el instrumento de percepción individual, logrando la participación de 29 estudiantes, el formato de este instrumento y sus respuestas se encuentran consolidados en el Apéndice F. La **¡Error! No se encuentra el origen de la referencia.** presenta la proporción de favorabilidad agrupando los niveles de valoración superiores (4: De acuerdo y 5: Totalmente de acuerdo), categorizados en las cuatro dimensiones de evaluación del instrumento.

Tabla 5.

Frecuencias de percepción estudiantil sobre la aplicación del Laboratorio 1.

Dimensión	Ítem evaluado (N=29)	Frecuencia Positiva	Porcentaje de Favorabilidad
Comprensión conceptual	Comprendí adecuadamente el cálculo e interpretación de medidas de tendencia central (media, mediana, moda).	28	96.55%
	Puedo diferenciar entre variables numéricas y categóricas al analizar una base de datos.	28	96.55%
	Logré interpretar correctamente los gráficos construidos en Power BI.	29	100%
	Me siento capaz de explicar con mis propias palabras los resultados obtenidos en un dashboard.	28	96.55%
Desarrollo de habilidades de analítica de datos	El laboratorio fortaleció mi capacidad para analizar datos de manera estructurada.	29	100%
	Aprendí a seleccionar el tipo de gráfico más adecuado según el objetivo del análisis.	29	100%
	Me siento más preparado para trabajar con bases de datos reales.	23	79.31%
	El uso de Power BI facilitó la comprensión de los conceptos estadísticos.	29	100%

Experiencia con el laboratorio	Las instrucciones del laboratorio fueron claras y comprensibles.	27	93.1%
	El caso de estudio facilitó la aplicación práctica de los conceptos vistos en clase.	29	100%
	El nivel de dificultad del laboratorio fue adecuado.	27	93.1%
	La actividad me permitió relacionar la teoría con situaciones reales.	29	100%
Data Storytelling y comunicación	Aprendí a comunicar resultados de análisis de datos de manera más clara y estructurada.	26	89.66%
	La construcción del dashboard me ayudó a organizar la información con un mensaje central.	27	93.1%
	Me siento más seguro al presentar resultados basados en datos.	28	96.55%

Nota. Los porcentajes reflejan la sumatoria de las frecuencias absolutas en los niveles 4 (De acuerdo) y 5 (Totalmente de acuerdo) de la escala Likert.

Como se observa en la **¡Error! No se encuentra el origen de la referencia.**, los resultados evidencian una percepción mayormente positiva por parte de los estudiantes de lo trabajado en el laboratorio, obteniendo una proporción de favorabilidad mayor al 79% en todos los ítems evaluados. Se podrían destacar positivamente la percepción de los estudiantes en la interpretación de los gráficos y datos de forma estructurada, la capacidad de seleccionar un gráfico según el objetivo del análisis, y la capacidad de relacionar la teoría en situaciones reales.

De los ítems evaluados, se puede observar que hay oportunidades de mejora, considerando que aproximadamente el 20% de los encuestados no se sienten preparados para trabajar con bases de datos reales. Este aspecto podría ser reforzado al complementar este laboratorio con algún proyecto o actividad de clase donde se le dé la oportunidad a los estudiantes de diseñar su propia visualización de datos a partir de una base de datos de libre elección.

De forma post hoc, se calculó el alfa de Cronbach sobre los bloques de las preguntas tipo Likert para analizar la consistencia interna de la encuesta. Se obtuvo una confiabilidad excelente ($\alpha = 0.937$), lo que evidencia una alta consistencia interna entre las preguntas del instrumento.

Estos resultados se encuentran respaldados por la retroalimentación obtenida en las preguntas abiertas. Al consultar sobre el mayor aporte del laboratorio, los estudiantes destacaron reiteradamente el interés por aprender una herramienta que facilita el manejo de grandes volúmenes de datos, y la capacidad de tomar decisiones a partir de interfaces interactivas. Como oportunidad de mejora para futuras iteraciones, algunas respuestas sugirieron contar con un monitor de apoyo adicional durante la sesión práctica para agilizar la resolución de dudas técnicas individuales, observación que resulta pertinente para la estandarización del instructivo docente.

3.4.1.2 Resultados Laboratorio 2. Luego de la ejecución del segundo laboratorio, enfocado en el modelado de probabilidades simples, conjuntas y la aplicación del Teorema de Bayes en el entorno de Google Colab, se recopilaron las evidencias del código estructurado y las respuestas de análisis realizados por los estudiantes a partir de los datos de accidentalidad vial. Las muestras representativas de estos proyectos en Python se encuentran alojadas en el Apéndice E.

Para evaluar la experiencia pedagógica de esta transición hacia lenguajes de programación, se aplicó la encuesta de percepción individual a los 23 estudiantes asistentes. El formato de la encuesta y sus respuestas se consolidan en el Apéndice F. La **¡Error! No se encuentra el origen de la referencia.** expone la proporción de favorabilidad del grupo en las cinco dimensiones evaluadas, agrupando los niveles altos de la escala Likert (4: De acuerdo y 5: Totalmente de acuerdo).

Tabla 6.

Frecuencias de percepción estudiantil sobre la aplicación del Laboratorio 2

Dimensión	Ítem evaluado (N=23)	Frecuencia positiva	Porcentaje de favorabilidad
Aplicación práctica de los	El laboratorio me permitió aplicar los conceptos de probabilidad vistos previamente en clase.	23	100%

contenidos vistos en clase	La actividad me ayudó a relacionar el Teorema de Bayes con una situación real de análisis de datos.	22	95.65%
	El caso de accidentes de tránsito facilitó la aplicación práctica de probabilidades simples y condicionales.	22	95.65%
	El laboratorio me permitió reforzar lo aprendido en clase mediante el uso de datos reales.	22	95.65%
	La actividad contribuyó a conectar la teoría trabajada en clase con el análisis de información en un contexto real.	21	91.3%
Desarrollo de habilidades de análisis de datos	Aprendí a cargar, revisar y explorar una base de datos en Google Colab.	23	100%
	El laboratorio fortaleció mi capacidad para analizar datos reales de manera estructurada.	22	95.65%
	Logré utilizar condiciones lógicas para filtrar registros y calcular probabilidades.	21	91.3%
	El uso de Python facilitó el desarrollo de los cálculos requeridos en el laboratorio.	21	91.3%
	Me siento más preparado para trabajar con bases de datos reales en actividades estadísticas.	19	82.61%
Experiencia con el laboratorio	Las instrucciones del laboratorio fueron claras y comprensibles.	21	91.3%
	El desarrollo paso a paso en Google Colab facilitó la realización de la actividad.	21	91.3%
	El nivel de dificultad del laboratorio fue adecuado para el curso de Estadística I.	21	91.3%
	Los recursos proporcionados fueron útiles para completar el laboratorio.	21	91.3%
Interpretación y análisis de resultados	La actividad me permitió formular conclusiones basadas en evidencia obtenida de los datos.	23	100%
	El simulador visual interactivo contribuyó a comprender mejor el comportamiento de los resultados.	22	95.65%
	Pude comparar probabilidades para identificar posibles relaciones entre las variables analizadas.	21	91.3%
	El laboratorio me ayudó a interpretar resultados obtenidos a partir de probabilidades.	20	86.96%
Comunicación de resultados	El laboratorio me ayudó a organizar los resultados para responder preguntas de análisis.	23	100%

Considero que este tipo de actividades mejora mi aprendizaje en Estadística I.	23	100.00%
Aprendí a comunicar resultados probabilísticos de manera más clara y estructurada.	20	86.96%
Me siento más seguro al explicar resultados obtenidos a partir de datos.	20	86.96%

Nota. Los porcentajes reflejan la sumatoria de las frecuencias absolutas en los niveles 4 (De acuerdo) y 5 (Totalmente de acuerdo) de la escala Likert.

El análisis cuantitativo de la Tabla 6 demuestra que la transición hacia la programación en Python fue altamente exitosa y alcanzó los objetivos establecidos. En términos generales, se obtuvo un alto porcentaje favorable, llegando a superar el 80% de respuestas positivas en todos los ítems de evaluación. Como resultados positivos a destacar, los estudiantes perciben que lograron trasladar los conceptos matemáticos de probabilidad al entorno de la programación, permitiendo relacionar mejor estos conceptos con situaciones reales. Además, se puede observar una valoración positiva de las habilidades de ingeniería de datos como cargar, exportar o filtrar utilizando condiciones lógicas.

Como posibles aspectos de mejora para estos resultados, los ítems con los indicadores más bajos son los relacionados a trabajo con datos reales, punto en común con el laboratorio 1, y los relacionados a comunicar resultados probabilísticos. Esto puede indicar que, durante la lectura inicial del laboratorio se aconsejaría a los docentes hacer especial énfasis en cómo se comunica un resultado de naturaleza probabilística en un informe técnico.

Igualmente, se aplicó un análisis de la coherencia interna del instrumento, mediante un alfa de Cronbach. Los 22 ítems obtuvieron un valor de $\alpha = 0.964$, lo que evidencia una excelente confiabilidad y una elevada consistencia interna del cuestionario.

La recopilación de respuestas cualitativas destacó que para los alumnos es importante la parte práctica y la integración de tecnologías, mencionando que "aprender a usar estas nuevas

herramientas se torna necesario para cualquier campo" y que facilitó "poner en práctica lo que vemos en clase". Por su parte, dentro de los aspectos por mejorar, algunos estudiantes sugirieron ampliar el tiempo de entrega y profundizar en la explicación de la sintaxis del código para conocer otras funciones operativas en Python.

3.4.1.3 Resultados laboratorio 3 y valoración general. Para evaluar el cierre del ciclo de prácticas, enfocado en los modelos de probabilidad mediante el uso de Python en Google Colab, se aplicó un tercer cuestionario a los 27 estudiantes asistentes. Adicionalmente, al tratarse del último laboratorio del semestre, el instrumento incorporó una dimensión de valoración final para medir el impacto global de la estrategia pedagógica. El formato físico y la consolidación de los resultados de esta sesión se encuentran disponibles en el Apéndice F.

Para evaluar el cierre del ciclo de prácticas, enfocado en los modelos de probabilidad mediante el uso de Python en Google Colab, se recopilaron los códigos y análisis desarrollados de manera autónoma por los equipos de trabajo. Las muestras representativas de estos entregables se encuentran adjuntas en el Apéndice E. Adicionalmente, se aplicó un tercer cuestionario a los 27 estudiantes asistentes para evaluar la experiencia durante la sesión y al tratarse del último laboratorio del semestre, medir el impacto de la estrategia pedagógica. La síntesis de los resultados favorables de esta encuesta es presentada en la Tabla 7. El formato de esta encuesta y la consolidación de sus respectivas respuestas se encuentran en el Apéndice F.

Tabla 7.

Frecuencias de percepción sobre la aplicación del Laboratorio 3 y cierre

Dimensión	Ítem evaluado (N=27)	Frecuencia Positiva	Porcentaje de Favorabilidad
	La actividad contribuyó a conectar la teoría con la información en un contexto real.	27	100.00%

Aplicación práctica de los contenidos	El caso facilitó la comprensión práctica de las distribuciones de probabilidad.	26	96.30%
	El laboratorio me permitió reforzar lo aprendido en clase mediante datos reales.	26	96.30%
	El laboratorio me permitió aplicar los conceptos de modelos de probabilidad.	26	96.30%
	La actividad me ayudó a relacionar la teoría estadística con una situación real.	25	92.59%
Desarrollo de habilidades de análisis	Logré interpretar histogramas y reconocer comportamientos en los datos.	26	96.30%
	Me siento preparado para trabajar con bases de datos reales en estadística.	26	96.30%
	El laboratorio fortaleció mi capacidad para analizar datos de manera estructurada.	26	96.30%
	El uso de Python facilitó el desarrollo de los cálculos probabilísticos requeridos.	25	92.59%
	Aprendí a explorar y depurar bases de datos utilizando Google Colab.	24	88.89%
Experiencia con el laboratorio	Los recursos proporcionados fueron útiles para completar el laboratorio.	27	100.00%
	El desarrollo paso a paso en Google Colab facilitó la realización de la actividad.	26	96.30%
	El nivel de dificultad fue adecuado para el curso de Estadística I.	25	92.59%
	Las instrucciones del laboratorio fueron claras y comprensibles.	24	88.89%
	La metodología favoreció mi participación durante la actividad.	24	88.89%
Interpretación y análisis	Los gráficos facilitaron la comprensión de las distribuciones de probabilidad.	26	96.30%
	La actividad me permitió formular conclusiones basadas en evidencia.	26	96.30%
	El laboratorio me ayudó a interpretar resultados de modelos probabilísticos.	26	96.30%
	Comprendí cómo los modelos probabilísticos se aproximan a datos reales.	25	92.59%
	Pude comparar resultados teóricos y observados para analizar el comportamiento.	25	92.59%
Comunicación de resultados	Aprendí a comunicar resultados probabilísticos de manera clara y estructurada.	26	96.30%
	Los laboratorios fortalecieron mi interés por la estadística aplicada.	26	96.30%
	Me siento más seguro al explicar resultados obtenidos a partir de datos.	25	92.59%

Valoración final del proyecto	El laboratorio me ayudó a organizar resultados para responder preguntas.	25	92.59%
	Considero que estas actividades mejoran mi aprendizaje en Estadística I.	25	92.59%
	El uso de datos reales hizo que las actividades fueran más significativas.	26	96.30%
	Google Colab y Power BI fueron herramientas adecuadas para desarrollar actividades.	26	96.30%
	Considero que este tipo de estrategias deberían seguir implementándose.	26	96.30%
	Los laboratorios contribuyeron positivamente a mi proceso de aprendizaje.	26	96.30%
	Los laboratorios computacionales facilitaron mi comprensión de los temas vistos.	26	96.30%

Nota. Los porcentajes reflejan la sumatoria de las frecuencias absolutas en los niveles 4 y 5 de la escala Likert.

Los resultados cuantitativos de este tercer laboratorio, presentados en la Tabla 7, evidencian la correcta aplicación del modelado probabilístico por parte de los estudiantes y respaldan la viabilidad de la metodología implementada. En la dimensión de aplicación práctica y habilidades, el 100% de la clase asimiló la conexión entre la teoría matemática y los contextos de ingeniería, resaltando que un 96.30% logró interpretar de manera exitosa los histogramas y gráficos generados para reconocer el comportamiento de las distribuciones de probabilidad.

A nivel de experiencia e interpretación, el 100% de los alumnos manifestó que los recursos entregados en la sesión fueron útiles, permitiendo que el 96.30% lograra formular conclusiones lógicas con base en los resultados probabilísticos que el código de Python guardaba de forma automática.

Igual que en los apartados anteriores, la consistencia interna del cuestionario fue evaluada mediante el coeficiente alfa de Cronbach. El instrumento, conformado por 30 ítems, obtuvo un

valor de $\alpha = 0.983$, lo que evidencia una excelente confiabilidad y una elevada consistencia interna entre los reactivos.

Por otro lado, la valoración final del proyecto arroja un consenso frente a la viabilidad de la estrategia, un 96.30% de los estudiantes (26 de 27) considera que la integración de Power BI, Python y datos reales hizo que la asignatura fuera más significativa y contribuyó positivamente a su proceso de aprendizaje. El mismo porcentaje concluye firmemente que esta metodología orientada a la analítica de datos debería continuar implementándose de forma permanente en el curso de Estadística I.

En la parte cualitativa, los estudiantes reconocieron como mayor aporte el haber comprendido visualmente cómo los datos reales se distribuyen en distintos modelos matemáticos. Al consultarles qué práctica les resultó más significativa en el semestre, las respuestas mostraron una preferencia equilibrada: algunos señalaron Power BI por permitir "manejar datos fácilmente a pesar de que sean masivos", mientras otros destacaron Google Colab y el entorno de programación porque facilitaron los cálculos probabilísticos y el desarrollo del proyecto final de la asignatura. Estas retroalimentaciones validan que la transición técnica desde tableros interactivos hacia lenguajes de programación resultó pertinente para atender diferentes estilos de aprendizaje y necesidades operativas.

3.4.2 Evaluación de impacto del aprendizaje

Como mecanismo central para validar la contribución real de los laboratorios en el desarrollo de competencias analíticas, se diseñó e implementó un sistema de evaluación diagnóstica de tipo longitudinal, estructurado en dos momentos de medición: un Pre-test, aplicado al inicio de la primera clase antes de cualquier intervención informática, y un Post-test, administrado al finalizar la ejecución del tercer y último laboratorio. Este enfoque comparativo

permitió aislar y medir la ganancia cognitiva y procedimental del estudiantado, mejora asociada a la intervención piloto.

3.4.2.1 Diseño y estructura del instrumento de evaluación. Para garantizar la objetividad de la medición, se construyó un cuestionario estandarizado orientado a evaluar el nivel de alfabetización de datos (*Data Literacy*) y el pensamiento estadístico en contextos aplicados a la ingeniería. El instrumento, cuyo formato completo se encuentra disponible para consulta en el Apéndice G, se estructuró bajo dos dimensiones fundamentales de competencia:

- Dimensión 1. Interpretación y razonamiento estadístico: Compuesta por un bloque de ocho preguntas cerradas (de selección única, selección múltiple y falso/verdadero) apoyadas en gráficos y escenarios descriptivos. El objetivo de este componente fue evaluar la capacidad técnica del estudiante para extraer conclusiones válidas, identificar tendencias, detectar valores atípicos y seleccionar las visualizaciones más adecuadas según el tipo de variable, superando la simple lectura numérica para alcanzar un nivel de inferencia argumentada.
- Dimensión 2. Planeación analítica: Consistente en una pregunta abierta estructurada como un caso de estudio (análisis de atención en entidades públicas). Este ítem fue diseñado para evaluar las habilidades de orden superior, exigiendo al alumno proponer una arquitectura de solución que definiera qué variables priorizar, qué indicadores calcular y qué esquema de *Data Storytelling* aplicar para respaldar la toma de decisiones gerenciales.

La aplicación controlada de este instrumento en sus fases de Pre-test y Post-test proporcionó la base de datos cuantitativa y cualitativa necesaria para evaluar de forma objetiva la transferencia de competencias, cuyos resultados comparativos se detallan en el siguiente apartado.

3.4.2.2 Resultados del diagnóstico inicial (Pre-test). La aplicación del instrumento de evaluación en su fase diagnóstica (Pre-test) se llevó a cabo al inicio del semestre con la participación de 28 estudiantes, previo a cualquier intervención con herramientas de programación o inteligencia de negocios. Es relevante indicar que esta prueba fue aplicada después de tres semanas después de haber iniciado clases, con la temática de estadística descriptiva ya abordada en clase.

Para el análisis de la Dimensión 1 (Interpretación y razonamiento estadístico), se tabularon las frecuencias absolutas de las opciones seleccionadas en las ocho preguntas cerradas, clasificando las respuestas correctas. La **¡Error! No se encuentra el origen de la referencia.** consolida estos resultados y presenta el porcentaje de acierto del grupo.

Tabla 8.

Resultados de la Dimensión 1 en el diagnóstico inicial (Pre-test)

Ítem / Competencia evaluada	Respuestas correctas (N)	Porcentaje de acierto
1. Identificación de concentración en gráfico de barras	28	100.00%
2. Conclusión sobre participación porcentual comparativa	27	96.43%
3. Interpretación prudente ante valores de alta cercanía	22	78.57%
4. Selección de gráfico para variables cuantitativas	2	7.14%
5. Lectura de distribución de frecuencias en histogramas	28	100.00%
6. Identificación de concentración mayoritaria por grupos	28	100.00%
7. Detección de valores atípicos (outliers) en un conjunto	21	75.00%
8. Identificación de tendencias de crecimiento temporal	27	96.43%
Total	183	81.70%

Nota. N = 28 estudiantes.

El análisis cuantitativo del diagnóstico inicial revela una brecha significativa en las competencias de los estudiantes. Por un lado, se evidencia un desempeño sobresaliente en tareas de lectura visual directa: la totalidad del grupo (100%) logró interpretar correctamente gráficos de barras e histogramas sencillos (ítems 1, 5 y 6), y más del 96% identificó tendencias de crecimiento y participaciones porcentuales (ítems 2 y 8). Sin embargo, al transitar hacia habilidades de inferencia estadística, el rendimiento disminuye (75% al 78% en los ítems 3 y 7).

El hallazgo más crítico se presentó en el ítem 4, orientado a evaluar el dominio teórico para la selección de visualizaciones (determinar que un diagrama de dispersión es el adecuado para evaluar la relación entre dos variables cuantitativas). En este punto, únicamente el 7.14% de la clase (2 estudiantes) respondió correctamente. Este déficit evidencia que los alumnos ingresan con capacidades para leer valores explícitos, pero carecen de los fundamentos técnicos para estructurar un análisis de datos desde cero.

Este comportamiento se corroboró al evaluar la Dimensión 2 (Planeación analítica), mediante la resolución de un caso abierto sobre la gestión de servicios en una entidad pública. El análisis cualitativo de las respuestas manuscritas evidenció un nivel de alfabetización de datos incipiente. Aunque la mayoría de los estudiantes logró identificar las variables base del problema (tiempos de atención, número de usuarios y estado del servicio) y sugirió el cálculo de métricas descriptivas tradicionales como el promedio, se detectó una carencia significativa en la estructuración del flujo de trabajo analítico.

Entre las falencias más recurrentes observadas en esta etapa diagnóstica, destacan:

- Desconexión con la toma de decisiones: Los alumnos propusieron cálculos estadísticos por inercia matemática, pero sin articular cómo esa información resolvería el problema de gestión o ayudaría a priorizar los recursos de la entidad.
- Confusión instrumental en la visualización: En concordancia con el error del ítem 4, diversos estudiantes propusieron representaciones gráficas inadecuadas para la naturaleza de los datos planteados en el caso.
- Falta de delimitación del alcance: Una proporción notable de estudiantes propuso priorizar y calcular indicadores basados en el "nivel de satisfacción" o los "costos operacionales", variables que no existían en el conjunto de datos planteado. Esto evidenció una dificultad inicial para acotar el análisis estrictamente a la información disponible.

Estos resultados demostraron que, previo a la intervención con los laboratorios computacionales, los estudiantes poseían nociones aisladas de lectura gráfica, pero presentaban marcadas dificultades para planear y ejecutar una estrategia de analítica y *Data Storytelling* coherente con los requerimientos de la ingeniería industrial.

3.4.2.3 Resultados del diagnóstico final (Post-test). Para medir la evolución de las competencias analíticas y evaluar el impacto real de la intervención pedagógica, se aplicó la segunda fase del instrumento diagnóstico (Post-test) al finalizar el ciclo de los tres laboratorios computacionales. En esta etapa participaron 24 estudiantes, una leve variación frente a la muestra inicial propia de la dinámica de asistencia académica, pero una muestra representativa para la validación del estudio. El procesamiento de la Dimensión 1 (Interpretación y razonamiento estadístico) mantuvo la estructura de validación cerrada. La **¡Error! No se encuentra el origen de la referencia.** expone las frecuencias observadas y los porcentajes de acierto consolidados tras la intervención.

Tabla 9.

Resultados de la Dimensión 1 en la evaluación de salida (Post-test)

Ítem / Competencia evaluada	Respuestas correctas (N)	Porcentaje de acierto
1. Identificación de concentración en gráfico de barras	24	100.00%
2. Conclusión sobre participación porcentual comparativa	24	100.00%
3. Interpretación prudente ante valores de alta cercanía	22	91.67%
4. Selección de gráfico para variables cuantitativas	15	62.50%
5. Lectura de distribución de frecuencias en histogramas	24	100.00%
6. Identificación de concentración mayoritaria por grupos	24	100.00%
7. Detección de valores atípicos (outliers) en un conjunto	21	87.50%
8. Identificación de tendencias de crecimiento temporal	24	100.00%
Total	178	92.71%

Nota. N = 24 estudiantes.

El análisis comparativo entre la evaluación inicial (Pre-test) y la evaluación final (Post-test) evidenció una mejora en todas las métricas analizadas. La proporción global de respuestas correctas aumentó del 81.7% en el Pre-test al 92.7% en el Post-test. La comparación de proporciones mediante una prueba Z mostró que esta diferencia fue estadísticamente significativa ($p = 0.0015$), lo que permitió rechazar la hipótesis nula de igualdad de proporciones. La diferencia absoluta fue de 11.0 puntos porcentuales, con un intervalo de confianza del 95% de 4.7% a 17.3%.

Al analizar a profundidad los ítems, las competencias de lectura visual directa (ítems 1, 2, 5, 6 y 8) alcanzaron un nivel alto, consolidando un 100% de acierto, lo que permite suponer que la interacción constante con tableros interactivos (dashboards) y gráficos programados en Python ayudó a afianzar la seguridad de los estudiantes frente a la representación de datos.

El mayor impacto cuantitativo se observó en la reducción de la brecha conceptual relacionada con la selección de visualizaciones. En el ítem 4, correspondiente a la elección de un diagrama para el análisis de variables cuantitativas, la proporción de respuestas correctas aumentó del 7.14% en el Pre-test a un 62.50% en el Post-test. La comparación entre ambas proporciones mediante la prueba exacta de Fisher evidenció una diferencia estadísticamente significativa ($p < 0.001$), lo que permitió rechazar la hipótesis nula de igualdad de proporciones.

Asimismo, la capacidad de los estudiantes para interpretar escenarios con valores cercanos (ítem 3) aumentó del 78.57% al 91.67%, y la destreza para detectar valores atípicos (ítem 7) ascendió al 87.50%. Estas variaciones validan empíricamente que el aprendizaje basado en el procesamiento de bases de datos reales optimizó el razonamiento crítico del grupo.

Esta ganancia cognitiva se revalida mediante el análisis cualitativo de las respuestas manuscritas correspondientes a la Dimensión 2 (Planeación analítica). Al contrastar las propuestas de los estudiantes en el Post-test frente a las del diagnóstico inicial, se evidenció una maduración técnica en la estructuración de soluciones. Entre los hallazgos cualitativos más relevantes, sustentados en las propias respuestas de los estudiantes, destacan:

- Apropiación de vocabulario estadístico avanzado: A diferencia del Pre-test, donde las respuestas se limitaban a sugerir el cálculo de "promedios", en el Post-test los estudiantes demostraron un dominio de conceptos estadísticos robustos. Se observaron propuestas que incluían calcular explícitamente "el resumen de los 5 números, así como la extracción de "cuartiles, media, mediana, varianza, coeficiente de variación y curtosis para tener una información más completa y reducir incertidumbre".
- Estructuración de flujos de análisis de datos: Los estudiantes superaron la visión unidimensional de los datos y propusieron arquitecturas de análisis por fases. Un estudiante especificó: "Primero realizaría un análisis univariado para cada una... luego realizaría un análisis bivariado entre el estado de los servicios y el número de usuarios, lo que demuestra una asimilación del flujo de trabajo algorítmico inculcado durante el modelado en Python. Otro estudiante propuso realizar "estadísticos de frecuencia y luego análisis multivariado", reflejando un nivel de abstracción propio de la analítica avanzada.
- Coherencia instrumental en la visualización: Se corrigió la confusión inicial sobre el uso de herramientas gráficas. Los alumnos lograron asociar correctamente la naturaleza de la variable con su representación visual óptima, sugiriendo el uso de "histogramas para ver el panorama de la distribución de los tiempos" (variables continuas) y "gráficos de barras para el estado del servicio" (variables categóricas). Varios estudiantes propusieron

específicamente el uso del "diagrama de cajas y bigotes" para analizar la varianza del tiempo de atención.

- Alineación del modelado con la toma de decisiones: Quizás el avance más representativo fue la capacidad de conectar la estadística con el contexto de la ingeniería. Los estudiantes argumentaron que el cálculo de probabilidades y medidas de dispersión servía para medir la "eficiencia operativa" y justificar decisiones. Un estudiante planteó una relación causal clara para la toma de decisiones: "relacionaría X_2 y X_3 considerando que, si el estado de los servicios es malo, el número de usuarios baja", e incluso se propuso el uso del "modelo binomial para interpretar porcentajes de tiempos de atención y estado del servicio".

En conclusión, los resultados longitudinales confirman que el diseño e implementación de los laboratorios computacionales aportó de forma positiva en la apropiación de conceptos estadísticos en los estudiantes. Se logró una transición desde un nivel de alfabetización de datos incipiente hacia un pensamiento analítico aplicable para solucionar problemas.

3.4.3 Evaluación integral de proyectos finales

Como mecanismo de cierre para medir la apropiación real y sostenida de las competencias analíticas, se asistió a las sustentaciones orales de los proyectos finales de la asignatura. En el proyecto de clase los estudiantes debían tomar un conjunto de datos público de una entidad colombiana, y realizar un análisis estadístico desde una perspectiva descriptiva y validando modelos de probabilidad para algunas variables de interés.

Este proceso de evaluación se realizó en un trabajo conjunto con el docente titular, con el propósito fundamental de observar en tiempo real la defensa argumentativa de los estudiantes y verificar si lograban transferir las enseñanzas instrumentales y teóricas de los laboratorios hacia la resolución de un problema estadístico de diseño propio.

Durante estas sustentaciones, se buscaba evidenciar el nivel de madurez analítica de los grupos y su capacidad para tomar decisiones sobre los datos sin un sistema de aprendizaje guiado. La capacidad del estudiante para aplicar herramientas tecnológicas en escenarios libres y no estructurados constituye la evidencia principal del aprendizaje autónomo. Según referentes en psicología educativa y pedagogía activa (Zimmerman, 2002; Mentz & Oosthuizen, 2016), el aprendizaje autónomo se consolida cuando el individuo transita de una instrucción guiada hacia la autorregulación, asumiendo el control sobre qué recursos informáticos e instrumentales utilizar para resolver un problema complejo. Esta transferencia de control fue precisamente el comportamiento documentado durante la presentación de los proyectos.

Para garantizar una medición objetiva y ágil de esta competencia en tiempo real, se diseñó e implementó un instrumento de evaluación estructurado mediante Microsoft Forms. La rúbrica permitió cuantificar el desempeño de los 11 grupos de trabajo bajo cuatro dimensiones en una escala de 1 a 5: uso de Power BI, aplicación de código en Google Colab/Python, calidad de las visualizaciones e interpretación de los resultados.

El procesamiento de los datos recolectados arrojó resultados altamente positivos, demostrando una adopción tecnológica contundente por parte del estudiantado:

- Uso de Power BI: 9 de los 11 grupos evaluados (81.8%) alcanzaron la calificación máxima (Nivel 5), presentando dashboards completos, organizados e interactivos que facilitaron el análisis integral de sus bases de datos. Un grupo alcanzó un desempeño adecuado (Nivel 4), y únicamente un grupo optó por prescindir de esta herramienta (Nivel 1).
- Modelado en Google Colab / Python: Todos los grupos incorporaron código en sus entregas, evidenciando la apropiación de esta herramienta durante el desarrollo del curso. El 54.5% (6 grupos) logró el Nivel 5, utilizando correctamente Python para el análisis, la

visualización y la interpretación matemática; por su parte, 4 grupos (36.4%) implementaron análisis correspondientes al Nivel 4, y solo un grupo se ubicó en el Nivel 3 con cálculos básicos.

- Calidad de las visualizaciones: En términos de comunicación visual de datos (*Data Storytelling*), 9 grupos (81.8%) lograron el Nivel 5, generando representaciones gráficas claras, variadas y de pertinentes para argumentar sus hallazgos, mientras que los 2 grupos restantes obtuvieron un sólido Nivel 4.
- Interpretación y conclusiones: En concordancia con los resultados observados en el Post-test, los grupos evidenciaron un buen nivel de razonamiento estadístico en la interpretación de los resultados. El 81.8% de los grupos (9 de 11) alcanzaron el Nivel 5, demostrando la capacidad teórica para conclusiones fundamentadas en el análisis de la evidencia y trascendiendo la descripción de estadísticas descriptivas básicas. Los 2 grupos restantes (18.2%) interpretaron adecuadamente sus resultados ubicándose en el Nivel 4.

Estos resultados reafirman el impacto de la intervención pedagógica. La evaluación en tiempo real de las sustentaciones demuestra que los laboratorios computacionales funcionaron como un ecosistema de entrenamiento. El estudiante logró apropiar el conocimiento técnico e instrumental impartido, transformándolo en un aprendizaje autónomo, crítico y plenamente alineado con las competencias analíticas exigidas en el perfil del ingeniero industrial contemporáneo.

3.5 Fase 5: Consolidación y documentación de resultados

Tras la culminación de la fase de implementación piloto y el análisis exhaustivo de los resultados de aprendizaje, la última etapa de la metodología se centró en garantizar la

sostenibilidad, transferibilidad y replicabilidad del proyecto. El objetivo de esta fase fue estructurar los productos finales de tal manera que puedan ser adoptados por el equipo docente de la Escuela de Estudios Industriales y Empresariales en futuros ciclos académicos. Para el cumplimiento de este propósito, la consolidación se ejecutó a través de tres acciones fundamentales:

3.5.1 Ajuste y optimización de los materiales didácticos

Como primer paso de la fase de consolidación, y tomando como insumo principal la retroalimentación obtenida durante la implementación piloto, se llevó a cabo un proceso de revisión y mejora de las guías de laboratorio. El objetivo central fue refinar el contenido técnico y pedagógico a partir de la experiencia directa adquirida al dirigir las clases, garantizando que los materiales finales fueran completamente funcionales y transferibles al resto de los docentes.

Las observaciones realizadas por los estudiantes a través de las encuestas de percepción permitieron identificar oportunidades de mejora en la claridad de las prácticas, lo que derivó en ajustes específicos como la ampliación del nivel de detalle en la explicación de la sintaxis de programación en Python para el segundo laboratorio, la recalibración de los tiempos de ejecución y la reestructuración de las instrucciones metodológicas en un solo párrafo descriptivo continuo, garantizando un flujo de lectura lógico que facilita la comprensión del paso a paso.

3.5.2 Organización del repositorio digital

Para asegurar la transferibilidad del proyecto, se estructuró un repositorio digital en Google Drive que contiene la totalidad de los recursos construidos durante la ejecución de este proyecto. Este entorno garantiza el acceso inmediato a todo el material que el equipo docente requiere para la ejecución de los laboratorios. El drive incluye:

- Guías completas y detalladas para la implementación paso a paso de cada laboratorio en el aula.

- Documentos de resolución (solucionarios) para el profesorado.
- Rúbricas de evaluación estandarizadas para medir las competencias adquiridas en cada sesión.

Es importante destacar que el repositorio no almacena los conjuntos de datos reales de forma local debido a su alto volumen y peso informático. En su lugar, las guías metodológicas proporcionan los enlaces de acceso directo al portal oficial de datos abiertos del Gobierno de Colombia (*datos.gov.co*). Desde allí, los usuarios deben descargar la información estrictamente en formato .csv. Esta decisión técnica obedece a que el gran tamaño de estas bases de datos supera la capacidad de procesamiento de herramientas tradicionales de hojas de cálculo (como Microsoft Excel), haciendo obligatorio el uso del formato CSV para garantizar la descarga, lectura y procesamiento íntegro de la totalidad de los registros mediante Power BI o Python.

Asimismo, todo el material, especialmente las guías y casos de estudio, se encuentran de manera dual: en formato PDF, listos para ser distribuidos e implementados de forma directa por los profesores, y en formato Word, con el fin de otorgar flexibilidad por si el docente requiere realizar ajustes, modificaciones o personalizaciones según las dinámicas de su grupo (Ver Apéndice H: Enlace y estructura del Repositorio Digital).

3.5.3 Diseño del instructivo de implementación para docentes

Con el objetivo de facilitar la adopción institucional de estos nuevos materiales por parte de los profesores que orienten la asignatura, se diseñó una presentación visual e interactiva a modo de instructivo. Este recurso funciona como una herramienta de inducción rápida que explica de manera estructurada cómo está organizado el repositorio en Google Drive, qué contiene cada uno de los archivos suministrados y cómo pueden articularse dentro del flujo normal de la clase de Estadística I. Esta estrategia de comunicación gráfica busca agilizar la apropiación tecnológica por parte del cuerpo docente (Ver Apéndice I: Presentación Instructivo Docente).

4. Conclusiones

El desarrollo de este trabajo de grado permitió responder a la necesidad de fortalecer la articulación de la fundamentación teórica de la asignatura Estadística I con las competencias digitales relacionadas con la analítica de datos. A partir del diseño, implementación y validación de laboratorios computacionales, fue posible integrar herramientas de uso frecuente en el contexto profesional dentro del proceso de enseñanza, manteniendo la coherencia con los objetivos de aprendizaje de la asignatura y dando cumplimiento a los objetivos planteados en el proyecto. En este sentido, se destacan las siguientes conclusiones:

- El diagnóstico inicial confirmó que, si bien la asignatura posee una base conceptual sólida, los estudiantes presentan niveles heterogéneos de alfabetización digital al iniciar el curso. No obstante, la encuesta de percepción evidenció una alta disposición para aprender herramientas como Python, siempre y cuando el proceso estuviera mediado por guías estructuradas y actividades desarrolladas paso a paso. Este resultado sustentó el diseño metodológico de los laboratorios y demostró la importancia de incorporar estrategias de acompañamiento que faciliten la transición gradual hacia el uso de herramientas computacionales.
- La revisión de experiencias y buenas prácticas permitió fundamentar el diseño pedagógico de los laboratorios y evidenció la necesidad de incorporar herramientas computacionales ampliamente utilizadas en el análisis de datos. La adopción de Power BI y Google Colab (Python) facilitó el trabajo de bases de datos reales de gran tamaño, permitiendo desarrollar actividades acordes con las dinámicas actuales de la analítica de datos y favoreciendo un entorno de aprendizaje más cercano al contexto profesional.

- La implementación piloto permitió evidenciar mejoras en la capacidad de los estudiantes para interpretar información estadística y seleccionar representaciones gráficas adecuadas. La evaluación aplicada mostró un incremento en el porcentaje de respuestas correctas, pasando de un 7.14% en el diagnóstico inicial a un 62.50% tras la ejecución de los 3 laboratorios. Asimismo, el análisis cualitativo permitió observar avances en la interpretación de resultados, el uso de terminología propia del análisis de datos y la construcción de explicaciones fundamentadas para apoyar la toma de decisiones.
- La evaluación de los proyectos finales evidenció que los laboratorios funcionaron eficazmente para promover la aplicación autónoma de los conocimientos adquiridos. La mayoría de los grupos desarrolló dashboards interactivos con altos niveles de desempeño y más de la mitad incorporó de manera autónoma código en Python dentro de sus proyectos. Estos resultados ratifican que la estrategia propuesta favoreció no solo el aprendizaje del uso de herramientas computacionales, sino también la integración de competencias relacionadas con la visualización, interpretación y comunicación de resultados.
- El uso de conjuntos de datos provenientes del portal de Datos Abiertos de Colombia permitió contextualizar el aprendizaje de la estadística mediante problemáticas de interés nacional y regional, como la accidentalidad vial, la vigilancia epidemiológica del COVID-19 y el embarazo adolescente. Esta aproximación contribuyó a fortalecer la relación entre los conceptos estadísticos desarrollados en la asignatura y su aplicación en escenarios reales, favoreciendo una comprensión más amplia del papel que desempeña el análisis de datos en la ingeniería industrial.

- Finalmente, el proyecto aporta a la Escuela de Estudios Industriales y Empresariales de la Universidad Industrial de Santander un conjunto de recursos didácticos organizados para facilitar la implementación de esta metodología en futuras versiones de la asignatura. El repositorio digital, las guías de laboratorio, los solucionarios, las rúbricas de evaluación y el instructivo para docentes constituyen materiales que favorecen la continuidad, adaptación y fortalecimiento de la estrategia, promoviendo la incorporación gradual de herramientas computacionales en la enseñanza de la estadística.

Si bien los resultados obtenidos evidencian la pertinencia de la estrategia propuesta, estos corresponden a una implementación piloto desarrollada en un contexto académico específico. En consecuencia, futuras aplicaciones permitirán consolidar la evidencia sobre su impacto y fortalecer progresivamente los materiales diseñados.

5. Recomendaciones

A partir de la experiencia obtenida durante el diseño, validación y ejecución de la fase piloto de los laboratorios computacionales, se formulan las siguientes recomendaciones orientadas a optimizar la implementación futura de esta metodología en la asignatura Estadística I:

- Debido a que los laboratorios utilizan conjuntos de datos reales de gran tamaño, especialmente aquellos descargados desde el portal de Datos Abiertos de Colombia, se recomienda solicitar a los estudiantes la descarga previa de los archivos requeridos antes de cada sesión práctica. Esta medida permitirá optimizar el tiempo disponible en el aula y concentrar las actividades en el análisis, tratamiento e interpretación de los datos, reduciendo el tiempo destinado a procesos de descarga e importación de información.
- Durante la implementación piloto se identificó que no todos los estudiantes disponen de equipos de cómputo con las condiciones necesarias para desarrollar las actividades propuestas. Por esta razón, se recomienda gestionar con anticipación la disponibilidad de las salas de informática institucionales, garantizando igualdad de condiciones para todos los participantes y facilitando el acompañamiento permanente por parte del docente durante el desarrollo de los laboratorios.
- Considerando los resultados obtenidos durante la implementación piloto y la aceptación manifestada por los estudiantes, se recomienda evaluar la incorporación progresiva de esta estrategia en asignaturas posteriores de la línea estadística, como Estadística II y Estadística III. La continuidad en el uso de herramientas como Python y Power BI permitiría fortalecer de manera gradual las competencias analíticas y digitales del estudiante a lo largo de su proceso de formación.

- Se recomienda realizar nuevas implementaciones de los laboratorios con diferentes cohortes de estudiantes, con el fin de evaluar la estabilidad de los resultados obtenidos y generar evidencia adicional sobre el impacto de esta estrategia en el aprendizaje. Asimismo, futuras investigaciones podrían incorporar diseños comparativos entre grupos con metodologías tradicionales y grupos que utilicen laboratorios computacionales, permitiendo medir con mayor precisión los efectos de la propuesta sobre el desarrollo de competencias analíticas.

Referencias Bibliográficas

- An Introduction to Data Science. (s. f.). Google Books.
<https://books.google.com.co/books?id=XSekDgAAQBAJ&lpq=PP1&hl=es&pg=PP1#v=onepage&q&f=false>
- Bedoya, J. M. (2021). Desarrollo de una plataforma de análisis de datos para el estudio de la brecha digital aprovechando el ecosistema de datos abiertos de Colombia. [Proyecto aplicado]. Repositorio Institucional UNAD. <https://repository.unad.edu.co/handle/10596/42610>
- Biggs, J., & Tang, C. (2011). Teaching for quality learning at university (4th ed.). McGraw-Hill Education/Open University Press.
- Carver, R., Everson, M., Gabrosek, J., Horton, N., Lock, R., Mocko, M., ... & Wood, B. (2016). Guidelines for assessment and instruction in statistics education (GAISE) college report 2016.
- Castro Sánchez, L. T., & Puerto Álvarez, E. V. (2025). Diseño de casos de estudio con integración de técnicas de analítica de datos Para la asignatura Logística Integral e Investigación de Mercados del Programa de Ingeniería Industrial de la Universidad Industrial de Santander. [Trabajo de Grado para Optar el título de Ingeniería Industrial]. Universidad Industrial de Santander.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide. CRISP-DM Consortium.
- Cobb, GW (2007). El curso introductorio de estadística: ¿un currículo ptolemaico? Innovaciones tecnológicas en la educación estadística, 1(1). <http://dx.doi.org/10.5070/T511000028>
Recuperado de <https://escholarship.org/uc/item/6hb3k0nz>
- Davenport, T. H. (2006, January). Competing on analytics. Harvard Business Review.

- Davenport, T. H., & Harris, J. G. (2007). *Competing on analytics: The new science of winning*. Harvard Business Press.
- Donoho, D. (2017). 50 Years of Data Science. *Journal Of Computational and Graphical Statistics*, 26(4), 745-766. <https://doi.org/10.1080/10618600.2017.1384734>
- Few, S. (2013). *Information dashboard design: Displaying data for at-a-glance monitoring* (2nd ed.). Analytics Press.
- Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. Springer Science & Business Media.
- Hardin, J., Hoerl, R., Horton, NJ, Nolan, D., Baumer, B., Hall-Holt, O., ... Ward, MD (2015). Ciencia de datos en los planes de estudio de estadística: Preparando a los estudiantes para “pensar con datos”. *The American Statistician*, 69 (4), 343–353. <https://doi.org/10.1080/00031305.2015.1077729>
- Hernández, E. J. (2024). Modelo de dominio específico para análisis y minería de datos educativos. Recuperado de: <https://repositorio.unal.edu.co/handle/unal/85851>
- Inmon, W. H. (2005). *Building the data warehouse* (4th ed.). John Wiley & Sons.
- Leopold, T., Di Battista, A., Jativa, X., Sharma, S., Li, R., & Grayling, S. (2025). Future of jobs report 2025. En World Economic Forum.
- Mentz, E., & Oosthuizen, I. J. (2016). *Self-directed learning research : an imperative for transforming the educational landscape*. AOSIS.
- Nussbaumer Knaflic, C. (2015). *Storytelling with data: A data visualization guide for business professionals*. John Wiley & Sons.
- Prince, M. (2004). Does active learning work? A review of the research. *Journal of engineering education*, 93(3), 223-231.

- Provost, F., & Fawcett, T. (2013). Data science for business: What you need to know about data mining and data-analytic thinking. O'Reilly Media.
- S. Freeman, S.L. Eddy, M. McDonough, M.K. Smith, N. Okoroafor, H. Jordt, & M.P. Wenderoth, Active learning increases student performance in science, engineering, and mathematics, Proc. Natl. Acad. Sci. U.S.A. 111 (23) 8410-8415, <https://doi.org/10.1073/pnas.1319030111> (2014).
- Sebastian, V. G. J., & Daniel, B. H. C. (2025, 7 mayo). Estrategia didáctica para mejora del proceso de enseñanza y de aprendizaje en la asignatura de sistemas de información mediante el uso de Power Platform. <https://noesis.uis.edu.co/items/1aa5f9dc-0b89-4484-a314-f8750feea408>
- Thomas, J. W. (2000). A review of research on project-based learning. The Autodesk Foundation.
- Tobón, S. T. (2013). Formación integral y competencias. Ecoe ediciones.
- Tukey, J. W. (1977). Exploratory data analysis. Addison-Wesley.
- Vásquez Caldas, J. F. (2023). Práctica empresarial: Diseño, desarrollo e implementación de un sistema de información web para la gestión y analítica de datos De los recursos electrónicos disponibles en la biblioteca virtual de la Universidad Industrial de Santander. [Trabajo de Grado para Optar al Título de Ingeniero de Sistemas]. Universidad Industrial de Santander.
- Wiggins, G., & McTighe, J. (1998). Understanding by design. Association for Supervision and Curriculum Development (ASCD).
- Zimmerman, B. J. (2002). Becoming a Self-Regulated Learner: An Overview. *Theory into Practice*, 41(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2