

## 1. Información del equipo pedagógico y horario atención a estudiantes

**Profesor:** Ignacio Sarmiento-Barbieri ([i.sarmiento@uniandes.edu.co](mailto:i.sarmiento@uniandes.edu.co))

- Horario Clase: Lunes y miércoles, de 3:30p.m. a 4:50p.m.
- Web del Curso: Bloque Neón
- Horario de atención a estudiantes: Hacer cita en <https://calendly.com/i-sarmiento/horarios-atencion-estudiantes>

**Asistente de docencia:** Julián David Rojas Aguilar ([j.rojas27@uniandes.edu.co](mailto:j.rojas27@uniandes.edu.co))

- Horario de atención a estudiantes: Hacer cita en [este enlace](#).

## 2. Descripción del curso

Este curso se sitúa en la intersección entre la Ciencia de Datos y la Economía Aplicada. Está dirigido a estudiantes interesados en abordar problemas empíricos contemporáneos que requieren el uso de técnicas modernas de análisis de datos, con un énfasis especial en la inferencia causal y el aprendizaje de máquinas.

El objetivo del curso es introducir a los estudiantes a los métodos fundamentales de la ciencia de datos y a las herramientas computacionales necesarias para aplicarlos rigurosamente a problemas económicos reales. Los contenidos combinan teoría, implementación computacional y aplicaciones, con un enfoque práctico centrado en proyectos.

Para cursarlo, es requisito haber aprobado Econometría I y II. Se asume familiaridad con conceptos de probabilidad, estadística, álgebra lineal, cálculo diferencial, así como con teoría económica básica y Microeconomía 3.

También se espera experiencia previa en análisis de datos con herramientas como Excel, Stata, R u otro software equivalente. El curso se dictará principalmente en R; tener conocimientos previos es recomendable, pero no indispensable. Estudiantes sin experiencia, pero con disposición y ganas de aprender, son bienvenidos. ¡Este curso (como todos) se aprende haciendo!

## 3. Resultados de aprendizaje

- **Distingue** entre predicción e inferencia causal, **aplicando** conceptos como parámetro de interés, método de estimación y valor estimado, **evaluando** supuestos de identificación en modelos de regresión lineal, **utilizando** datos **observacionales y experimentales, reales y simulados, en problemas económicos y de negocios aplicados**.

- **Implementa** técnicas de aprendizaje de máquinas, como Lasso y árboles de decisión, **evaluando** su capacidad predictiva mediante desempeño fuera de muestra y su utilidad para estimar efectos causales heterogéneos, **usando datos simulados y reales en aplicaciones económicas y de negocios**.
- **Analiza** datos espaciales representados por polígonos, **definiendo** estructuras de vecindad y **estimando** modelos económicos espaciales, como SAR y SEM, **para estudiar dependencia espacial e inferencia en aplicaciones económicas reales**.
- **Procesa y analiza** texto como dato utilizando técnicas de minería de texto y modelado de tópicos, **extrayendo** información relevante **para el análisis de problemas de negocios y el diseño de sistemas de recomendación**.
- **Colabora** de manera efectiva con su equipo para **desarrollar y presentar** análisis de datos aplicados, **comunicando** de forma clara, estructurada y profesional los hallazgos y recomendaciones, **en el marco de talleres, presentaciones y el proyecto final del curso**.

#### **4. Cronograma**

Duración	Tema	¿Qué estudiaremos?
~ 4 semanas	Regresión lineal, inferencia, predicción y Lasso	Revisaremos el modelo de regresión lineal desde las perspectivas de predicción e inferencia, discutiendo la diferencia entre parámetro de interés, método de estimación y valor estimado. Introduciremos el marco de inferencia causal y el rol de los experimentos. Estudiaremos cómo evaluar el desempeño de los modelos fuera de muestra y compararemos OLS con técnicas de regularización como Lasso, analizando sus ventajas y limitaciones en tareas predictivas y causales.

*Continúa en la siguiente página*

Duración	Tema	¿Qué estudiaremos?
~ 4 semanas	Árboles de decisión y árboles causales	Estudiaremos árboles de decisión como herramienta predictiva, analizando su construcción, poda y evaluación fuera de muestra. Luego exploraremos árboles causales para estimar efectos heterogéneos de tratamiento, destacando su estructura, principios de honestidad y su utilidad para responder preguntas de política económica.
~ 4 semanas	Datos espaciales y econometría espacial	Trabajaremos con datos espaciales representados mediante polígonos, como barrios o municipios, aprendiendo a definir estructuras de vecindad y a construir matrices de contigüidad. Exploraremos patrones espaciales mediante análisis exploratorio y estimaremos modelos econométricos espaciales como SAR y SEM para estudiar problemas económicos donde el espacio importa.
~ 3 semanas	Texto como datos y sistemas de recomendación	Exploraremos técnicas para procesar y analizar texto como dato, incluyendo limpieza, tokenización, minería de texto y modelado de tópicos. Aplicaremos estas herramientas al análisis de problemas de negocios, como segmentación de clientes y diseño de sistemas de recomendación, trabajando con datos reales extraídos de fuentes textuales.

## 5. Referencias

- Arbia, G. (2014). Primer for Spatial Econometrics (Vol. 230). Basingstoke: Palgrave Macmillan.
- Békés, G., & Kézdi, G. (2021). Data analysis for business, economics, and policy. Cambridge University Press.
- Berk, R. A. (2008). Statistical learning from a regression perspective (Vol. 14). New York: Springer.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An introduction to statistical learning (ISLR)
- Hernán MA & Robins JM (2020). Causal Inference: What If. Boca Raton: Chapman & Hall/CRC

- Patel, A. A. (2019). Hands-on unsupervised learning using Python: how to build applied machine learning solutions from unlabeled data. O'Reilly Media.
- Silge, J., Robinson, D., & Robinson, D. (2025). Text mining with R: A tidy approach. Boston (MA): O'reilly.

## 6. Metodología

El curso se desarrollará mediante clases presenciales. Durante las sesiones, el profesor explicará los conceptos teóricos y su implementación práctica utilizando **Jupyter Notebooks**, integrando código y análisis en tiempo real. Las clases combinarán teoría, ejemplos aplicados y ejercicios programados directamente en clase.

No se distribuirán diapositivas ni notas de clase. Se espera que los estudiantes tomen apuntes y programen en clase cuando se les solicite, siguiendo el ritmo de las explicaciones y resolviendo ejercicios en conjunto.

La participación en clase es un componente esencial del curso. Se fomentará un ambiente de diálogo, donde se espera que los estudiantes hagan preguntas, comenten y contribuyan a la discusión técnica. Para ello, es fundamental que lleguen preparados, habiendo leído el material asignado.

La evaluación del curso se basa en talleres grupales que deben ser entregados y presentados, así como en un proyecto final que consolida los aprendizajes. La participación activa también será evaluada.

## 7. Evaluaciones

**Cuadro 1.** Puntajes

	Puntaje Individual	Puntaje Total
Talleres grupales	10 %	40 %
Presentaciones	15 %	30 %
Proyecto Final		20 %
Participación		10 %
Total		100 %

- **Talleres grupales ( $4 \times 10\% = 40\%$ ):** Ejercicios prácticos en grupos para aplicar los métodos discutidos en clase. Cada taller incluirá entregables específicos (código, slides, etc.) los detalles específicos estarán disponibles en Bloque Neón.
- **Presentaciones ( $2 \times 15\% = 30\%$ ):** Dos exposiciones por equipo seleccionadas al azar entre los talleres. Se evaluará la capacidad de comunicar de forma clara y profesional los resultados.

- **Proyecto final (20 %)**: Trabajo grupal que simula un encargo de consultoría en ciencia de datos. Los requisitos y criterios de evaluación estarán disponibles en Bloque Neón.
- **Participación en clase (10 %)**: Se evaluará la asistencia, preparación previa y la participación activa en discusiones, preguntas, resolución de ejercicios y trabajo en clase.

### **Sistema de aproximación de notas definitiva**

Las calificaciones definitivas de las materias serán numéricas de uno cinco (1,50) a cinco (5,00), en unidades, décimas y centésimas. La calificación aprobatoria mínima será de tres (3,0). En este curso se aproximarán la nota a la centésima más cercana. Por ejemplo, si el cálculo del cómputo es 3.245, la nota final se aproximarán a 3.25; si el resultado del cálculo es 2.994 la nota final será de 2.99

### **Excusas**

Los estudiantes que no presenten las actividades y evaluaciones del curso en la fecha establecida previamente recibirán una calificación de cero (0), a menos que justifiquen su ausencia ante el profesor dentro de un término no superior a ocho (8) días hábiles. Para excusas validas ver Artículo 45 del [Reglamento General de Estudiantes de Pregrado \(RGEPr\)](#)

## **8. Asistencias**

Se espera que los estudiantes asistan a todas las clases. Si un estudiante se ausenta en más del 20 % de las sesiones programadas, se aplicará una penalización del 1 % sobre la nota final del curso por cada inasistencia adicional.

## **9. Políticas generales de los cursos de Economía y fechas importantes**

Los estudiantes deben consultar este enlace, donde se encuentran las reglas sobre asistencia a clase, excusas válidas, fraude académico y faltas disciplinarias, reclamos, políticas de bienestar (ajustes razonables y MAAD), política sobre el uso de herramientas de inteligencia artificial generativa y fechas importantes del semestre: <https://economia.uniandes.edu.co/programas/politicas-generales-para-cursos-ofrecidos>