



MIA

Magíster en
Inteligencia Artificial

Introducción a Data Science

Programa del Curso

Clase 1



MIA

Magíster en
Inteligencia Artificial

IMT3860 – Introducción a Data Science
Prof. Nicolás Alvarado - nfalvarado@mat.uc.cl
Prof. Felipe Gutiérrez - figutier@uc.cl
Autora: Paula Aguirre A. - paaguirr@ing.puc.cl



AGENDA DE HOY

01

Presentaciones

02

Programa del Curso

03

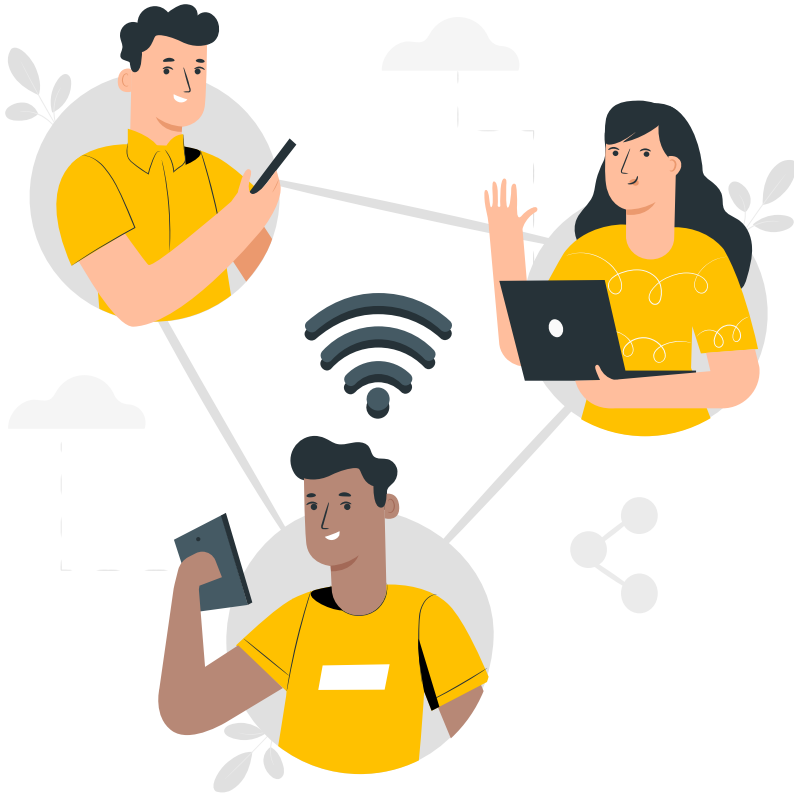
Introducción a Ciencia
de Datos

04

Herramientas Prácticas
de Programación

+ BREAK 20:00 (10-15')





01

Presentaciones



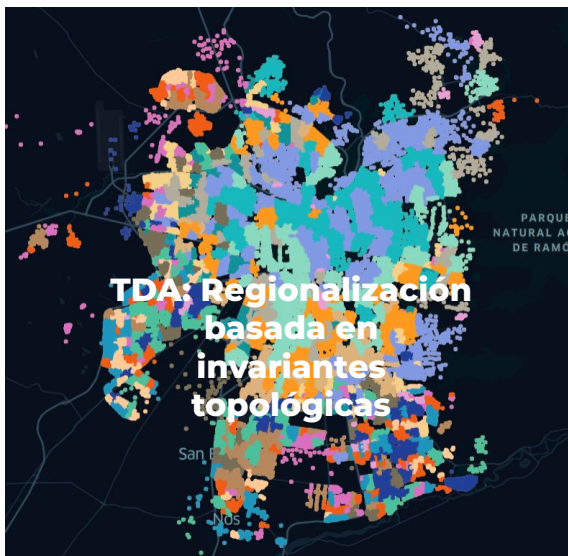
Equipo Docente

Profesor



Felipe Gutiérrez- figutier@uc.cl

- Planificador Urbano
- Magister en Desarrollo Urbano y en Procesamiento y Gestión de información
- Alumno de Doctorado en Ciencias de la Ingeniería
- Alumno investigador del Laboratorio de Inteligencia Artificial UC.
- Investigador del Instituto Milenio Fundamento de los Datos (IMFD)

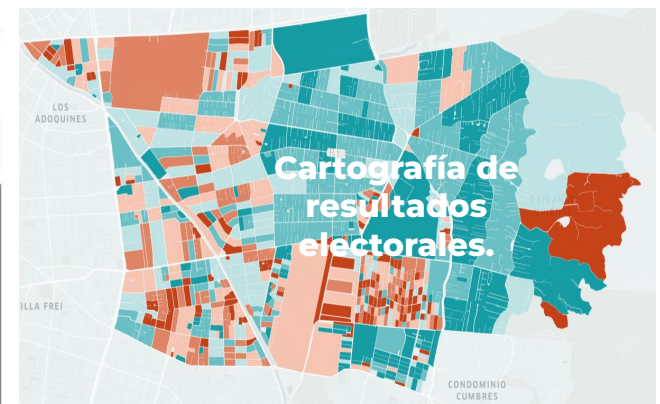


Inédito plan piloto en Ñuble: Monitorean aguas servidas para detección temprana del Covid-19

Publicado: Viernes, 26 de Junio de 2020 a las 14:52hrs. Autor: [Cooperativa.cl](#)

- La herramienta permite obtener información anticipada en grupos sociales.
- Según estudios europeos, test de heces fecales permite detectar el virus al tercer día del contagio.

Recuerda nuestra campaña **"Todos podemos hacer más"**, si tienes una iniciativa solidaria que quieras comunicar, llámanos al **2236-48002** o visita nuestro [sitio](#) si quieres hacer tu aporte



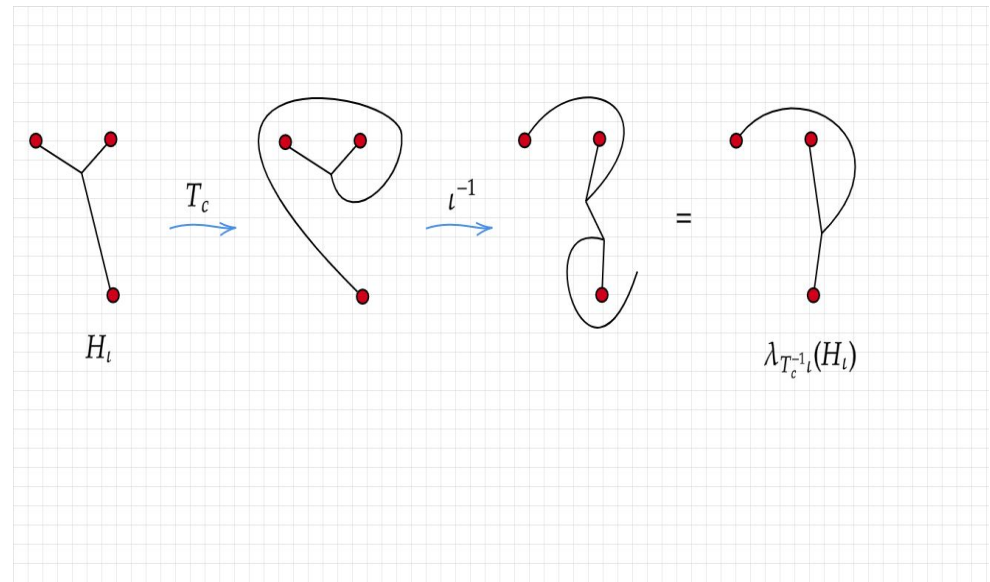
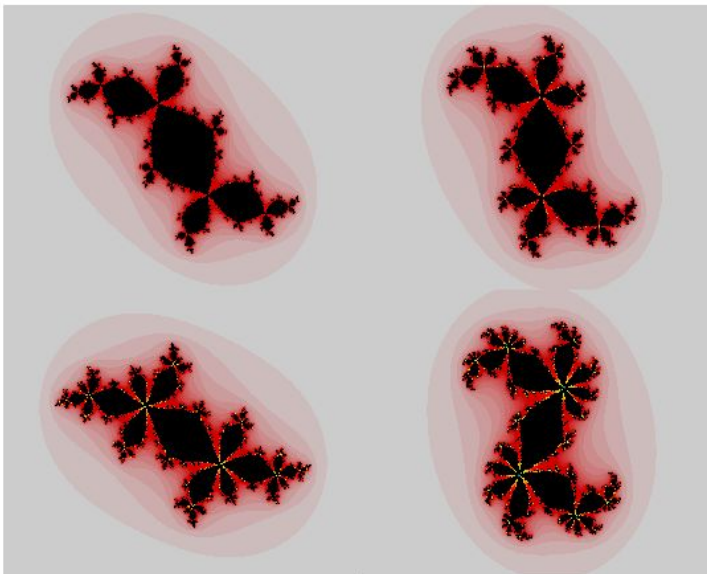
Equipo Docente

Profesor



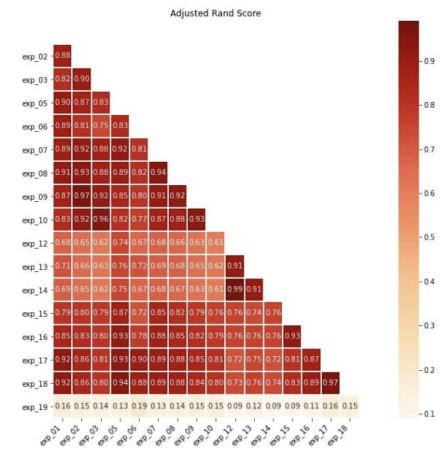
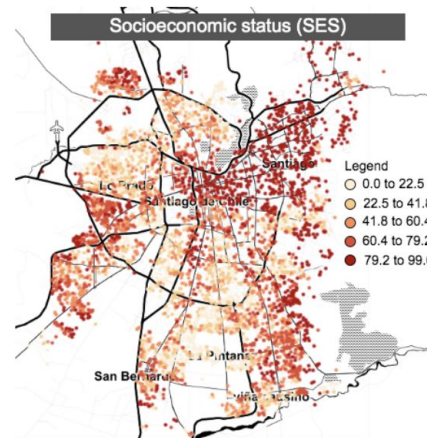
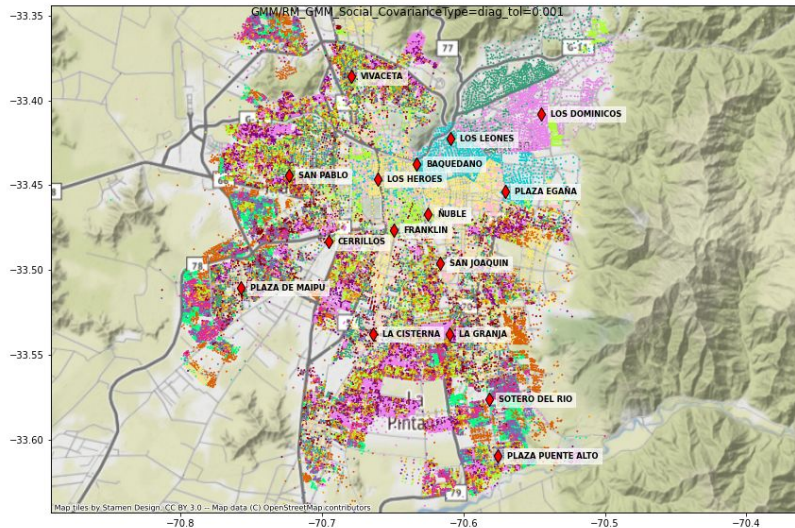
Nicolás Alvarado - nfalvarado@mat.uc.cl

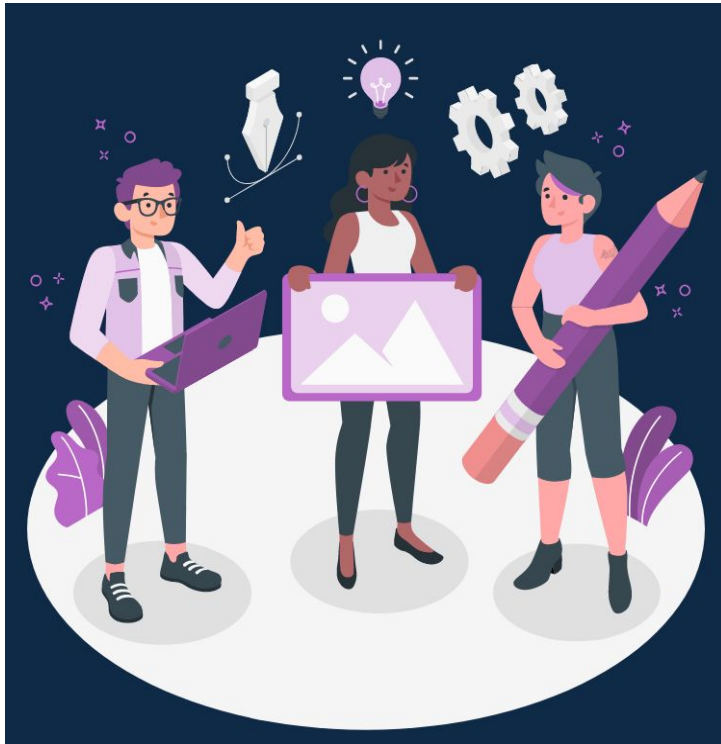
- Magíster en Matemática.
- Profesor de la Escuela de Ingeniería y Facultad de Matemáticas UC
- Alumno de Doctorado en Ciencias de la Computación.
- Alumno investigador del Laboratorio de Inteligencia Artificial UC.
- Investigador del Instituto Milenio Fundamento de los Datos (IMFD)



Ayudante

- Arquitecta y Profesora de Educación Básica
- Alumna de Magíster en Ciencias de la Ingeniería, área Ciencia de la Computación UC
- Ayudante en cursos del Departamento de Ciencias de la Computación UC
- Alumna de investigación en el Laboratorio de Inteligencia Artificial UC.
- Alumna de investigación en el Instituto Milenio Fundamento de los Datos (IMFD)





02

Programa del Curso



Descripción del Curso y Objetivos

Este curso **introduce** a los alumnos al campo de la Ciencia de Datos.

Estudiaremos en **amplitud**, más que en profundidad, sus **principios básicos** y las **principales técnicas y herramientas** utilizadas:

- ❖ Recolección, integración, transformación y almacenamiento de datos.
- ❖ Inferencia estadística.
- ❖ Análisis exploratorio de datos.
- ❖ Análisis descriptivo y predictivo.
- ❖ Visualización de datos.
- ❖ Herramientas prácticas (Python, Pandas, entre otros).
- ❖ Problemas éticos y de privacidad emergentes en Ciencia de Datos.

Descripción del Curso y Objetivos

Este curso **introduce** a los alumnos al campo de la Ciencia de Datos.

Estudiaremos en **amplitud**, más que en profundidad, sus **principios básicos** y las **principales técnicas y herramientas** utilizadas:

Bimestre 2	Bimestre 3	Bimestre 4	Bimestre 5	Bimestre 6	Bimestre 7
<ul style="list-style-type: none">• EPG4001 Aprendizaje Supervisado• EPG4002 Aprendizaje no Supervisado• FIL2000 Ética aplicada a Inteligencia Artificial	<ul style="list-style-type: none">• EPG4003 Métodos Probabilísticos• IMT3870 Computación de Alto Rendimiento• INF3812 Deep Learning	<ul style="list-style-type: none">• INF3813 Deep Learning Avanzado Optativos <ul style="list-style-type: none">• INF3822 Proyecto Aplicado I o• INF3250 Seguridad Computacional o• EPG4008 Análisis Estadístico con Datos Faltantes• INF3803 Minería de Procesos o• INF3863 Sistemas Recomendadores o• EPG4005 Métodos Bayesianos	<ul style="list-style-type: none">• INF3862 Plataformas de Big Data Optativos <ul style="list-style-type: none">• INF3823 Proyecto Aplicado II o• INF3260 Procesos de Negocio o• EPG4007 Análisis de Tiempos de Falla• INF3801 Minería de Medios Sociales o• EPG4004 Series de Tiempo o• EPG4006 Modelos Lineales Generalizados	<ul style="list-style-type: none">• INF3841 Recuperación de Información• INF3842 Visualización de Información Optativos <ul style="list-style-type: none">• INF3821 Aplicaciones o• INF3570 Aspectos Legales de las TI o• EPG4009 Análisis Multivariado	<ul style="list-style-type: none">• INF3851 Gobernanza de Datos• INF4980 Actividad de Graduación Optativos <ul style="list-style-type: none">• INF3852 Gestión de Proyectos o• INF3280 Gestión Estratégica o• EPG4010 Métodos Bayesianos Avanzados
Bimestre 1					Bimestre 8
					<ul style="list-style-type: none">• INF4980 Actividad de Graduación

Estrategias Metodológicas

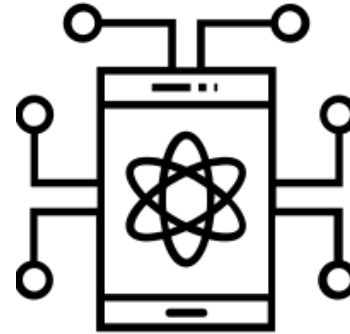
- ❖ Clases Expositivas.
- ❖ Talleres Prácticos.
- ❖ Lecturas y actividades de discusión.

Estrategias Evaluativas



Talleres prácticos

- 30%
- Individuales
- 3 Talleres prácticos durante el curso
- Trabajo en clase y entrega dentro de la semana siguiente.
- Se eliminará el taller con peor calificación



Proyecto de Aplicación

- 40%
- Grupal (3 personas)
- Avances + presentación + reporte.

Política de atrasos: Se descuenta 0.5 pto/hora.

Para atrasos con causa justificada, comunicarse con el equipo docente.

Calendario de Evaluaciones

Clase	Día	Fecha	Contenidos	Actividad evaluada
1	Jueves	24-mar	Introducción a Data Science. Proceso Data Science Introducción a herramientas prácticas	
2	Jueves	31-mar	Inferencia estadística	Taller 1: Inferencia estadística
3	Jueves	7-abr	ETL - Presentación proyecto del curso	
4	Miércoles	13-abr	Proceso ETL - EDA	
5	Jueves	21-abr	Transformación - Visualización de Datos	Taller 2: ETL-EDA Envío problema proyecto (Canvas, ventanilla abierta)
6	Jueves	28-abr	EDA - Algoritmos básicos para análisis descriptivo	
7	Jueves	5-may	Algoritmos básicos para análisis predictivo	Taller 4: análisis descriptivo y predictivo
8	Jueves	12-may	Algoritmos básicos para análisis predictivo - Ethics of Data Science	
9	Jueves	19-may	Proyecto	Presentación proyectos

Proyecto de Aplicación

- ◆ **Objetivo:** desarrollar todo el proceso de ciencia de datos para responder preguntas de interés y accionables en un ámbito a elección.

- ◆ **Componentes del proyecto:**
 - Planteamiento del problema.
 - Recolección de datos.
 - Limpieza y de datos.
 - Análisis exploratorio.
 - Preparación de datos.
 - Modelamiento y análisis descriptivo/predictivo.
 - Comunicación y reporte de resultados.

- ◆ **Formato:**
 - Avances parciales sin evaluación.
 - Presentación (19 de mayo 2022)
 - Reporte escrito + códigos (19 de Mayo 2022, fecha provisoria).

- ◆ El enunciado detallado del proyecto se presentará y discutirá en dos clases más.

Bibliografía

Mínima:

- *“Doing Data Science, Straight Talk from the Frontline”*.
Cathy O’Neil and Rachel Schutt, O’Reilly Media, 2013.
- *“Data Science for Business: What you need to know about data mining and data-analytic thinking”*.
Foster Provost and Tom Fawcett. O’Reilly Media, 2013.

Complementaria:

- *“Python for data analysis: Data wrangling with Pandas, NumPy, and IPython”*.
Wes McKinney, O’Reilly Media, 2017
- *“Python Data Science Handbook”: Essential Tools For Working With Data”*.
Jake VanderPlas, O’Reilly Media, 2016
- Blogs y artículos de revistas científicas, prensa, etc.



03

**Aspectos
Prácticos**



Comunicación y Herramientas Computacionales

- ❖ Todos los contenidos, anuncios y material del curso se canalizará a través de Canvas y correo electrónico (al menos hasta que todos tengan acceso a Canvas)
- ❖ Utilizaremos software de código abierto y lenguaje de programación en Python.
- ❖ Todo tipo de dudas, preguntas y comentarios bienvenidos!!! Utilizar correo electrónico, foros de Canvas, preguntar en clases, etc.

¿Preguntas?