

# PROGRAMA DE ASIGNATURA

## Prosecución de estudios en Economía Segundo Semestre 2021

<b>Asignatura</b>	<b>R &amp; Data Science</b>	
<b>Carrera</b>	<b>Ingeniería Comercial</b>	
<b>Código</b>	351495	
<b>Créditos</b>	7 SCT	
<b>Nivel</b>	8° semestre	
<b>Requisitos</b>	Econometría I y Econometría II	
<b>Categoría</b>	Electivo de especialidad	
<b>Área de conocimiento según OCDE</b>	Ciencias Sociales	
<b>Profesor</b>	Eduardo Jiménez	
<b>Correo electrónico</b>	Eduardo@datrics.cl	
<b>Horario</b>	Clases: Lunes y miércoles 8:00 am Ayudantía: Miércoles 9:00 am Atención Alumnos: Jueves 12 am (se requiere agendar una hora)	
<b>Ayudante</b>	Dylan Arce (dylan@datrics.cl)	
<b>ID Zoom</b>		
<b>Github</b>	<a href="https://github.com/edojimenez/r_datascience">https://github.com/edojimenez/r_datascience</a>	
<b>Descripción</b>	<b>Contribución al Perfil de Egreso</b> <b>Este curso es el curso electivo del área de economía del grado de Economía.</b> Antes de esta asignatura los estudiantes debiesen haber cursado las asignaturas de Estadística y Econometría II que cubren estadística descriptiva, probabilidad e inferencia, el modelo de regresión lineal tanto con datos de corte transversal como con datos temporales. El objetivo de este curso es que los estudiantes utilicen softwares de códigos abiertos, como el paquete R y su IDE R Studio. Además, profundizar el manejo de datos, la elaboración de visualizaciones, introducir los conceptos y prácticas relevantes de la ciencia de datos, al mismo tiempo introducir elementos propios de Machine learning.	
	<b>Resultado de aprendizaje general</b> Al finalizar el curso los alumnos deberán ser capaces de: <ul style="list-style-type: none"> <li>– Utilizar el software estadístico R, y su IDE Rstudio.</li> <li>– Manejo de vectores y bases de datos.</li> <li>– Regresiones e inferencia en R</li> <li>– Realizar reportes con visualizaciones interactivas</li> <li>– Incorporar conceptos relacionados con el Machine Learning</li> <li>– Manejar grandes bases de datos de manera eficiente</li> </ul>	
	<b>Resultados de aprendizaje específicos</b>	<b>Unidades temáticas</b>
	Introducción a R, conceptos importantes en Data Science.	I R y Data Science
	Importar bases de datos, formas de exploración transformación de las bases de datos y creación de nuevas variables, creación de funciones, loops.	II Manejo de bases de datos
	Aprender a hacer diferentes tipos de gráficos de acuerdos al tipo de programa. Realizar gráficos interactivos	III Visualizaciones

	Estimación de regresiones de corte transversal y series de tiempo. Test de hipótesis.	IV Regresiones en R
	Introducción a conceptos relevantes de ML: Aprendizaje Supervisado (Clasificación, Regresión) y No supervisado (Clustering, DR)	V Introducción a Machine Learning
	<p><b>Metodologías de enseñanza y de aprendizaje</b></p> <p>La asignatura tiene un enfoque principalmente práctico, combinado con ayudantías donde se harán ejercicios de aplicación. Estas actividades se complementarán con controles de lecturas y exposiciones prácticas sobre contenidos aplicados de la materia que ocurren en la industria.</p> <p>Dado la naturaleza del curso, se promoverá que, de forma paralela, los estudiantes realicen un trabajo propio donde puedan utilizar las herramientas aprendidas en diferentes escenarios, como diferentes bases de datos. Se alentará a los estudiantes que complementen las bases de datos que poseen con datos no estructurados (Por ejemplo, textos que provengan desde la web), de forma de ampliar sus análisis.</p> <p>Otro objetivo del curso es potenciar la visualización de los datos, buscando que los alumnos sean capaces de reconocer y realizar diferentes elementos visuales (gráficos, mapas, nubes de palabras, etc.) tanto de manera estática como interactiva. Además, revisar las principales herramientas disponibles en código libre para que puedan realizar trabajos colaborativos, de manera eficiente y automatizada.</p> <p>Finalmente, se hará una introducción a temas relativos a Machine learning, sus principales algoritmos y los tipos de problemas que pueden resolver</p> <p>En cuanto a las ayudantías, se enfocarán para hacer una instancia de trabajo asistido por el ayudante para resolver problemas con los códigos, así como asesorar a los estudiantes en formas de abordar el problema.</p> <p>Para aclarar dificultades con la materia del curso, se realizan consultas en horario de atención de los alumnos, a través de Intranet, por email, o solicitando una cita con el profesor.</p>	

	<p><b>Procedimientos de evaluación</b></p> <p>La evaluación del alumno se realizará a partir de la consideración de los siguientes aspectos:</p> <ul style="list-style-type: none"> <li>- Control de lectura</li> <li>- Exposición tema de actualidad</li> <li>- Problem Set</li> <li>- Trabajo final</li> </ul>	
	<p><b>Bibliografía básica</b></p> <ul style="list-style-type: none"> <li>- R y Data Science <ul style="list-style-type: none"> <li>▪ Varian, H. (2014). Big Data: New Tricks for Econometric. <i>Journal of Economic Perspectives</i>, [online] (Volume 28, Number 2), pp.3-28. Disponible en : <a href="https://pubs.aeaweb.org/doi/pdfplus/10.1257/jep.28.2.3">https://pubs.aeaweb.org/doi/pdfplus/10.1257/jep.28.2.3</a></li> <li>▪ Mason, H. (2019). dataists » A Taxonomy of Data Science. [online] D ataists.com. Disponible en: <a href="http://www.dataists.com/2010/09/a-taxonomy-of-data-science/">http://www.dataists.com/2010/09/a-taxonomy-of-data-science/</a></li> <li>▪ Harvard Business Review. (2019). <i>Data Scientist: The Sexiest Job of the 21st Century</i>. [online] Disponible en: <a href="https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century">https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century</a></li> <li>▪ Barocas, Solon and Selbst, Andrew D., Big Data's Disparate Impact (2016). 104 California Law Review 671 (2016). Disponible en : <a href="https://ssrn.com/abstract=2477899">https://ssrn.com/abstract=2477899</a> or <a href="http://dx.doi.org/10.2139/ssrn.2477899">http://dx.doi.org/10.2139/ssrn.2477899</a></li> </ul> </li> <li>- Manejo de Bases de datos y Visualización <ul style="list-style-type: none"> <li>▪ Wickham, H., &amp; Grolemund, G. (2017). R for data science. Beijing: OReilly. Disponible en: <a href="https://r4ds.had.co.nz/">https://r4ds.had.co.nz/</a>.</li> </ul> </li> <li>- Regresiones en R <ul style="list-style-type: none"> <li>▪ James, G. (2017). An introduction to statistical learning: With applications in R. New York: Springer. Capítulo 3.</li> </ul> </li> <li>- Exportación de resultados. <ul style="list-style-type: none"> <li>▪ Wickham, H., &amp; Grolemund, G. (2017). R for data science. Beijing: OReilly. Disponible en: <a href="https://r4ds.had.co.nz/">https://r4ds.had.co.nz/</a></li> </ul> </li> <li>- Introducción al Machine Learning</li> </ul>	

	<ul style="list-style-type: none"><li>▪ Kelleher, J. D., Namee, B. M., &amp; D'Arcy, A. (2015). Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies. MIT Press.</li><li>▪ Müller, A. C., &amp; Guido, S. (2016). Introduction to Machine Learning with Python: A Guide for Data Scientists. O'Reilly Media, Incorporated.</li><li>- Text Mining Silge, Julia, and David Robinson. 2016. "tidytext: Text Mining and Analysis Using Tidy Data Principles in R." JOSS 1 (3). The Open Journal. <a href="https://doi.org/10.21105/joss.00037">https://doi.org/10.21105/joss.00037</a></li></ul>	

## Apéndice

<b>Contenidos Detallados</b>	<b>Tema 1:</b> R y Data Science <ol style="list-style-type: none"> <li>1. Definición de la Ciencia de Datos</li> <li>2. Introducción al ambiente digital (API, Cloud, Big data, etc.)</li> <li>3. Programas de código abierto</li> <li>4. Introducción a R y R Studio</li> <li>5. Creación de R Project</li> </ol>
	<b>Tema 2:</b> Manejo de Bases de datos <ol style="list-style-type: none"> <li>1. Importar bases de datos (.xlsx, .txt, csv,.dta, etc)</li> </ol>

	2. Configuración de directorios y uso de proyectos 3. Limpiar datos, manejo de outliers y NA. 4. Creación de variables 5. Algoritmos para el análisis exploratorio en R <b>Tema 3:</b> Visualización 1. Propiedades de los gráficos y su correspondencia con el problema y los datos que se están analizando. 2. Creación de gráficos con diversos paquetes (ggplot, plotly) para gráficos univariados, multivariados y mapas. 3. Creación de Gráficos interactivos 4. Creación de Dashboards <b>Tema 4:</b> Regresiones en R 1. Mínimos cuadrados ordinarios 2. Logit, probit 3. Inferencia estadística <b>Tema 5:</b> Introducción al Machine Learning 1. Modelos supervisados 2. Modelos no supervisados 3. Introducción a herramientas de big data		
<b>Evaluaciones</b>	EVALUACIÓN	POND.	FECHAS TENTATIVAS
	Control de Lectura	20%	30 de noviembre
	Problem Set (2)	40%, (20% c/u)	3 de noviembre 14 de diciembre
	Proyecto Final	30% (15% proyecto, 15% presentación)	19 de enero
	Presentación tema de actualidad	10%	Entrega de tema – 6 de octubre Entrega de grabación de presentación – 20 de octubre.
	Evaluación Recuperativa. Consta de una prueba de contenido de todo el semestre.	Misma evaluación de la prueba faltante	Por Confirmar
	En Caso de inasistencia a alguna evaluación, se debe avisar con anticipación a los profesores. Este requerimiento será fundamental para acceder a la evaluación recuperativa, posterior autorización por parte de los profesores. Se considera copia cualquier tipo de similitud entre evaluaciones entre compañeros de una misma generación o generaciones pasadas. En el caso de presentarse esta situación, se procederá a evaluar con nota 1 a todos los involucrados y se le dará aviso al director de escuela, quien determinará el curso de acción.		