



UAI
UNIVERSIDAD ADOLFO IBÁÑEZ

UNIVERSIDAD ADOLFO IBÁÑEZ
Ingeniería Comercial
Sede Viña del Mar

Curso	:	Data Science para Economía y Negocios
Profesores	:	Javier Fernández y Esteban López
Ayudante	:	Carolina Molinare
Emails	:	javier.fernandezch@edu.uai.cl esteban.lopez@uai.cl mmolinare@alumnos.uai.cl
Oficina – Horario Consulta	:	Sala A-229 – TBD
Horario:	:	Martes 8:30 - 11:05
Año y Semestre	:	2020 - Primer Semestre

Última Actualización: 21.02.2020

INTRODUCCIÓN-PRESENTACION DEL CURSO

Vivimos en una época donde la cantidad de información y la velocidad con que ésta se actualiza crece exponencialmente. El manejo de bases de datos en la era de la información es cada vez más relevante tanto en el sector privado como en el sector público, y en consecuencia diferentes organismos han empezado a demandar trabajadores especializados en el manejo de datos.

La misión de este curso es entregar las bases de programación para el procesamiento y análisis de datos que apoyen a la toma de decisiones en la empresa, instituciones públicas y privadas, gobiernos, etc.

El objetivo fundamental de esta asignatura es desarrollar en los alumnos la capacidad de leer, escribir y pensar en un lenguaje de programación que les permita analizar datos para que la toma de decisiones sea mejor informada. Siguiendo la tendencia mundial en adopción y uso de software de “open-source” (acceso libre) que permita la reproductibilidad, este curso utilizará como base el software estadístico R con el cual los alumnos inscritos podrán manipular y analizar bases de datos. Para un buen desempeño en el curso se requieren conocimientos mínimos de computación (manejo de archivos). También se requieren nociones básicas de estadística descriptiva e inferencial.

OBJETIVOS DEL CURSO

Los objetivos generales de aprendizaje del curso son los siguientes:

- Desarrollar la capacidad de conceptualización abstracta de estructuras de datos y cómo estas se enlazan con metodologías para responder preguntas básicas que sirvan de apoyo a la toma de decisiones.
- Entender y aplicar la generación de información desde datos ‘crudos’.
- Aprender a trabajar datos de manera ordenada y sistemática con el fin de lograr generar proyectos que sean reproducibles.
- Introducir a los estudiantes al lenguaje estadístico R, entregando una experiencia tipo “aprender-haciendo” de las múltiples funcionalidades que tiene el software.

Al finalizar el curso se espera que la/el estudiante tenga una formación inicial, pero sólida de programación para manejo de bases de datos. Adicionalmente, se espera que los/las alumnos/as tengan una visión clara de las utilidades prácticas que tiene la programación para las problemáticas que afrontarán en su futura vida laboral.

COMPETENCIAS PARA EL PERFIL DE EGRESO

Esta asignatura contribuye a la formación de las siguientes competencias del perfil de egreso de Ingeniería Comercial en la Universidad Adolfo Ibáñez:

- 1.2 Identifica oportunidades de mejora o nuevas iniciativas que contribuyan al crecimiento de la organización y la sociedad.
- 2.1 Anticipa escenarios, identificando las variables asociadas al desafío y su posible comportamiento en el futuro.
- 2.2 Diseña soluciones para diferentes escenarios, sólidamente fundamentadas, identificando las consecuencias de cada alternativa.
- 2.3 Ejecuta la solución seleccionada, realizando un seguimiento en pos de asegurar su correcta implementación.
- 3.2 Transforma ideas en acción, mediante el diseño de un prototipo de solución que sea viable y tenga el potencial de crear valor.
- 4.1 Selecciona los datos pertinentes, para la toma de decisiones considerando las variables relevantes del entorno.
- 4.2 Analiza los datos, mediante la aplicación de modelos matemáticos y económicos y el uso de tecnología, para la construcción de información que le permita comprender el funcionamiento de los mercados y sus componentes.
- 4.3 Interpreta la información acorde a los parámetros relevantes de la disciplina y el contexto particular que sirva para la toma de decisiones en el ámbito económico y de la gestión de organizaciones nacionales e internacionales.
- 3.1 Realiza un análisis personal de sus habilidades, intereses y tiempo disponible, de manera responsable antes de asumir un compromiso.
- 3.2 Asume los compromisos de manera rigurosa y sistemática, identificando procesos, rutinas y estándares que favorezcan su adecuada concreción.
- 3.3. Se hace cargo de las consecuencias de sus actos o decisiones que lo afectan personalmente y/o a distintos actores del contexto organizacional, local o global.

METODOLOGÍA

Para cumplir estos objetivos del curso, se combinará la presentación de las bases teóricas, con actividades prácticas basadas en el manejo del software R. Dado que el curso consiste principalmente en entregar herramientas prácticas de programación a los alumnos, las clases se realizarán en laboratorio. Las clases consistirán en exposiciones teóricas y prácticas de diferentes paquetes del software estadístico R con el fin de que los alumnos comprenden las múltiples funcionalidades que puede tener el estadístico para luego aplicarlas a la resolución de problemas.

Dado que se hará uso intensivo de herramientas computacionales durante el curso, es preciso que cada alumno cuente con un laptop en el que pueda instalar el software R.

Las clases serán complementadas con tareas prácticas y un proyecto final.

CONTENIDOS

Unidad 1: Natación 101 para Data Science: Introducción a la Programación y Manejo de Datos

- Conceptualización de Data Science y Big Data
- Introducción a la programación en R
 - Programación Orientada al objeto vs. Orientada a la tarea
 - Sintaxis en R y tipología de objetos
 - Infraestructura de datos
 - Limpieza y Manipulación de datos en base a data.table

Aplicación 1:

- Objetivo: Utilizar información disponible a través de encuestas y censos con el objetivo de emparejar datos de fuentes de datos diferentes. De esta forma los estudiantes podrán identificar el potencial de mercado distribuido a lo largo de las grandes ciudades del país.
- Aprendizaje esperado: Los estudiantes podrán ser capaces de entender la estructura de una base de datos y, de forma simultánea, comprender el cómo el vincular información desde fuentes de información diferentes puede permitirle responder múltiples preguntas relacionadas a la actividad empresarial.

Unidad 2: Buscando la aguja en el pajar: Visualización de Datos

- Principios de Visualización:
 - Dos, Don'ts & Everything in between
 - Análisis exploratorio de datos desde lo visual
- Programación de Visualizaciones
 - Gráficos estáticos (base, y ggplot)
 - Gráficos dinámicos (base y plotly)
 - Introducción al mapeo de datos
 - Gráficos interactivos (base y leaflet)
- Aplicación 2:
 - Objetivo: Utilizar información disponible de tipo espacial para localizar la competencia, definir áreas de mercado e identificar potenciales lugares donde localizar una nueva empresa.

- Aprendizaje esperado: Los estudiantes podrán ser capaces de recoger información para definir las áreas de mercado y cotejar esa información con fuentes de datos secundarias y así resolver dudas empresariales como: proyecciones de demanda, niveles de competencia, entre otros.

Unidad 3: Obsesivo de los Datos VS. Teórico Conservador: El dilema actual de un buen Científico de Datos

- Introducción a los Métodos Estadísticos en Data Science
 - Análisis de Regresión
 - Validación Cruzada
 - Análisis de Cluster
 - Árboles de Decisión (CART)
- Aplicación 3:
 - Objetivo: Utilizar instrumental estadístico con el objetivo de analizar la información proveniente de encuestas y desarrollar indicadores que permitan comprender la relación entre variables observables.
 - Aprendizaje esperado: Los alumnos podrán ser capaces de entender tanto a nivel abstracto como práctico el cómo relacionar variables entre sí. De esta forma, los estudiantes podrán utilizar distintas fuentes de información para modelar procesos económicos y de toma de decisiones para los negocios.

EVALUACIÓN

El promedio (nota final) se forma con las notas de: **tareas, participación en clases, y un proyecto final**. La ponderación de cada actividad será la siguiente:

Evaluación	Ponderación	Fechas
Control 1	15%	Por definir
Control 2	10%	
Control 3	15%	
Participación en clases		No aplica
Entrega de productos en clase	12,5%	
Presentación del Tema de clase	12,5%	
Proyecto Final – Primera Entrega	10%	Por definir
Proyecto Final – Informe y Presentación	25%	Por definir

Habrán un total de 3 controles **individuales** que tienen como objetivo que el estudiante realice aplicaciones de los contenidos vistos en clases y tenga entrenamiento previo el proyecto final. Estos se realizarán en horario de clases en las fechas **(tentativas)** a continuación:

- Control 1: 7 de abril
- Control 2: 5 de mayo
- Control 3: 9 de junio

Si bien la asistencia a clases no es obligatoria, se evaluará la participación en clases de los estudiantes en función de su rendimiento en las actividades propuestas en clases. La “Entrega de productos en clase” consistirá en ejercicios de programación alineados al proyecto final que los estudiantes deberán resolver. Por otro lado la “Presentación del Tema de clase” corresponde a la presentación de la lectura de la semana por un alumno la que tiene por objetivo motivar la clase y el tema de esa semana.

Tanto el informe como la presentación del proyecto final consistirán en evaluaciones en grupos de 2 alumnos los cuales deberán ser capaces de, utilizando el software estadístico R, enfrentar diferentes problemas de programación que abarcarán desde la limpieza de bases de datos hasta el desarrollo de visualizaciones de bases de datos. En este curso no considera la posibilidad de eximición.

Condiciones de aprobación

Para aprobar el curso se debe tener una nota final (NF) igual o superior a 3,95. Aquellos alumnos que obtengan nota final (NF) inferior a 3,95, reprobarán con dicha nota. Adicionalmente alumnos que tengan menos de un 3 como promedio en Proyecto Final, reprobarán la asignatura automáticamente.

PROYECTO FINAL

La habilidad de emprender una investigación, preparar un reporte escrito y desarrollar una presentación oral puede generar grandes ventajas tanto en el sector privado como público. El mercado laboral valora de mejor forma a aquellos profesionales que tienen cierta experiencia en la solución de un problema, explicando su estatus y proponiendo una solución. Este mismo procedimiento será exigido en el trabajo de investigación desarrollado a lo largo del semestre. En este trabajo, los estudiantes utilizarán el enfoque de la ciencia de datos para comprender un problema, presentar una solución basada en herramientas estadísticas vistas a lo largo del curso y finalmente desarrollar una presentación vinculada a un reporte.

Objetivos:

1. Entender la estructura de las bases de datos y el cómo el paradigma de la ciencia de datos permite ayudarnos a responder preguntas aplicadas a fenómenos de la economía y negocios.
2. Desarrollar la capacidad del estudiante a entender la información contenida dentro de una base de datos, comprendiendo las limitantes que pueden tener estos para entender un problema.
3. Entregarle al estudiante múltiples herramientas para presentar información relevante para comprender un fenómeno, tanto a nivel de visualizaciones como de métodos estadísticos clásicos.
4. Preparar un informe escrito y una presentación oral.

Procedimiento:

La presentación final y el reporte consistirán en los siguientes ítems:

1. Introducción y descripción general de la base de datos elegida por el grupo utilizando estadísticos y visualizaciones enseñadas durante el curso.

2. Empleo de estrategia enseñada en el curso para responder a la problemática tratada en el Proyecto.
3. Desarrollo de una sección donde se describan los principales resultados encontrados en el análisis de datos, explicando los mecanismos y los limitantes de estos.

Evaluación:

El alumno será evaluado en base a (i) su claridad a la hora de realizar la presentación, (ii) creatividad con la cual la presentación fue hecha, (iii) el uso idóneo de las herramientas vistas a lo largo del semestre y (iv) la explicación clara de los mecanismos que explican los resultados de su estudio. Con el objetivo de que el proyecto de investigación de los estudiantes cumpla con ciertos mínimos, el profesor solicitará en el horario de consultas reuniones con los estudiantes para dar feedback.

NORMAS PARA EL BUEN FUNCIONAMIENTO DEL CURSO

- **Faltas Académicas:** En caso que el alumno sea sorprendido en actitudes y comportamientos poco éticos o dudosos antes, durante y después de cualquier actividad, evaluación y/o instancia de esta asignatura, se aplicarán las medidas disciplinarias del caso. Bastará para la aplicación de estas medidas que el profesor y/o los ayudantes tengan sospechas fundadas de la conducta del alumno. Respecto a las medidas disciplinarias, por favor considere lo siguiente:
 - Los alumnos que sean sorprendidos copiando, intentando copiar o dejándose copiar en una prueba, control o trabajo, serán evaluados con nota 1,0 en dicha evaluación sin derecho a apelación.
 - Por “copia” se entenderá cualquier medio ilícito para capturar información propia o de terceros sin que exista autorización para hacerlo (esto incluye, por ejemplo, mirar para el lado, simple posesión y/o uso de “ayuda memoria”, medios electrónicos como celulares y computadores, entre otros, utilizados para el traspaso de información ilícita.)
 - Si un alumno es sorprendido por segunda vez cometiendo una falta académica reprobará automáticamente la asignatura sin derecho a apelación.
 - En cualquiera de los casos descritos anteriormente, los antecedentes serán entregados al Director de Carrera, al Decano de Facultad y al Decano de Pregrado para que la Universidad tome las medidas disciplinarias adicionales que estime conveniente.
 - El listado anterior no es exhaustivo y, por tanto, se pueden incluir en el concepto de copia situaciones no contempladas en este apartado, pero que claramente se pueden considerar como faltas al honor y a la ética estudiantil.
- **Otros:** Cualquier situación que no esté contemplada en este programa pero que el alumno considere que puede interferir con su desempeño durante el semestre debe ser conversada oportunamente con el profesor, quien buscará la forma apropiada para solucionar eventuales problemas. Esto significa que si el alumnos tiene algún problema debe acercarse a conversar con el profesor inmediatamente presentado el problema y no una vez que aquel haya interferido con su desempeño.

BIBLIOGRAFÍA

Libros Base: (Lecturas requeridas)

Teoría:

- Big Data (BDWS) – Walter Sosa. Editorial Siglo Veintiuno. (Disponible en Biblioteca)
- Storytelling with Data: A Data Visualization Guide for Business Professionals – Cole Nussbaumer

Programación:

- R para Ciencia de Datos (R4DS) – link: <https://es.r4ds.hadley.nz>
- Introduction to Statistical Learning: with applications in R – link: <http://faculty.marshall.usc.edu/gareth-james/ISL/>

Recursos adicionales (opcionales):

- 1- “The art of R programming: tour of statistical software design” - Norman S. Matloff (Disponible en Biblioteca)
- 2- “R in action : Data analysis and graphics with R” - Robert Kabacoff (Disponible en Biblioteca)
- 3- “Introductory statistics with R” - Peter Dalgaard
- 4- “Political Analysis Using R” – James E. Monogan
- 5- “ggplot2: Elegant Graphics for Data Analysis” – Hadley Wickham
- 6- “Advanced R” – Hadley Wickham
- 7- “An Introduction to “R” for Spatial Analysis and Mapping” – Chris Brunsdon, Lex Comber