

PROGRAMA DE ASIGNATURA

Unidad Académica: Facultad de Ciencias Económicas	
ASIGNATURA: INTRODUCCION A LA CIENCIA DE DATOS CON PYTHON	REQUISITOS DE CORRELATIVIDAD: Estadística I
CATEDRA: CARO	REQUIERE CURSADA: Si
TIPO: Electiva	UBICACIÓN EN LA CARRERA: OCTAVO
DICTADO: Normal	MODALIDAD: Presencial
CARRERA: LICENCIATURA EN ECONOMÍA (2009)	
SEMESTRE DE CURSADO: SEGUNDO	CARGA HORARIA: 84 HORAS
CARGA HORARIA TEORICA: 42 HORAS	CARGA HORARIA PRÁCTICA: 42 HORAS

FUNDAMENTACIÓN Y OBJETIVOS

La ciencia de datos es un campo interdisciplinario que se enfoca en extraer conocimiento a partir de conjuntos de datos combinando habilidades en estadística, matemática, programación y conocimientos del dominio específico. Se trata de un enfoque sistemático para analizar grandes volúmenes de datos con el fin de describir patrones, tendencias, relaciones, conocimientos útiles para la toma de decisiones y la generación de valor en diferentes ámbitos.

En esencia, la ciencia de datos se compone de varias etapas, que incluyen la recolección de los datos, la limpieza y preparación de los mismos, el análisis exploratorio, la modelización y la interpretación de resultados. Estas etapas pueden implicar el uso de diversas técnicas y herramientas, como algoritmos de aprendizaje automático, minería de datos, visualización de datos y estadísticas.

El objetivo final de la ciencia de datos es generar información que pueda ser utilizada para resolver problemas, hacer predicciones, optimizar procesos y tomar decisiones informadas en una amplia gama de áreas, como negocios, ciencia, salud, gobierno, entre otros. Con el crecimiento exponencial de la cantidad de datos disponibles en la era digital, la ciencia de datos se ha vuelto cada vez más relevante y demandada en la actualidad.

Con este curso se pretende brindar una visión general y equilibrada entre los fundamentos y el conocimiento práctico de los principales conceptos de la inteligencia de negocios (business intelligence) o de la analítica gubernamental (government analytics) que permita a los estudiantes participar y desarrollar proyectos mediante los cuales se puedan utilizar los datos que posee una organización o el estado, para transformarlos en información valiosa que permita ayudar a resolver problema de negocio o de políticas públicas y aportar a la toma de decisiones.

En la actualidad, se generan enormes cantidades de datos económicos provenientes de diversas fuentes, como transacciones financieras, encuestas, redes sociales y registros gubernamentales. La capacidad para recopilar, procesar y analizar estos datos de manera efectiva es crucial para obtener información valiosa sobre el comportamiento económico.

En un mercado laboral cada vez más competitivo, contar con habilidades en ciencia de datos brinda una ventaja diferencial a los profesionales de las ciencias económicas.

Python ofrece una plataforma versátil y potente para la manipulación y análisis de datos, además de ser un software libre, lo que permite a los estudiantes de nuestras carreras explorar patrones

complejos, identificar tendencias y desarrollar modelos predictivos para informar políticas, pronósticos y estrategias empresariales.

El curso tiene como objetivo general:

Proporcionar a los y las estudiantes los conocimientos, habilidades y herramientas necesarias para aplicar con éxito la ciencia de datos en el contexto de las ciencias económicas.

PROGRAMA ANALÍTICO

Unidad 1: Ciencia de datos

Objetivos Específicos:

- Comprender el significado de la ciencia de datos y su importancia en diversos campos.
- Familiarizarse con la terminología y los conceptos básicos de la analítica de datos
- Identificar los roles comunes dentro del proceso de ciencia de datos y sus responsabilidades.
- Entender las diferentes fases del proceso de ciencia de datos, desde la obtención de datos hasta la presentación de resultados
- Aprender sobre la importancia de la obtención de datos de calidad y la ingeniería de características para el éxito de un proyecto de ciencia de datos.
- Entender la importancia de la disponibilidad de datos para las organizaciones y proyectos de ciencia de datos.
- Reconocer la importancia de la gobernanza de datos en la gestión y el uso responsable de la información.
- Comprender los desafíos éticos asociados con el uso de datos en proyectos de ciencia de datos.

Contenidos:

- Ciencia de datos: definición y su importancia en la actualidad. Aplicaciones prácticas en el campo de las ciencias económicas.
- Introducción a la analítica de datos (*Data Analytics*): terminología y conceptos fundamentales, categorías de *Data Analytics*: descriptivo, predictivo, prescriptivo.
- El proceso de data science: roles comunes, datos, obtención de datos, curación e ingeniería de características (feature engineering).
- *Big data*: introducción y nociones básicas. Identificación de grandes volúmenes de datos
- Arquitectura de datos. *Business Intelligence*.
- Disponibilidad de datos
- Gobernanza de los datos
- Ética en la ciencia de datos.
- Software de programación en ciencia de datos

Bibliografía:

Gorelik, A. (2019). *The enterprise big data lake: Delivering the promise of big data and data science*. O'Reilly Media.

Unidad 2: Base de datos

Objetivos específicos:

- Identificar y distinguir entre datos estructurados y no estructurados.
- Aprender técnicas específicas para el tratamiento de cada tipo de dato.
- Reconocer la importancia de la limpieza y la integración de datos para garantizar la calidad y la coherencia de los conjuntos de datos.
- Entender los principios de la ingeniería de características y cómo crear nuevas características relevantes a partir de los datos existentes para mejorar el rendimiento de los modelos de aprendizaje automático.
- Comprender los principios fundamentales del manejo de bases de datos.
- Dominar conceptos básicos de programación en Python y trabajar con scripts y notebooks.

Contenidos:

- Tipos de datos (estructurados y no estructurados) y su tratamiento.
- Limpieza e integración de datos. Técnicas de curación de datos e ingeniería de características (feature engineering).
- Estandarización de procesos
- Principios en el manejo de base de datos
- Bases de datos
- Uso de Python. Librerías

Bibliografía:

Vander Plas (2016) *Python Data Science Handbook* " O'Reilly Media, Inc.". <https://jakevdp.github.io/PythonDataScienceHandbook/00.00-preface.html>

Unidad 3: Visualización de datos

Objetivos Específicos:

- Comprender por qué la visualización de datos es fundamental en el proceso de análisis, facilitando la detección de patrones, tendencias y relaciones.
- Reconocer cómo las visualizaciones mejoran la capacidad de comunicar resultados y conclusiones de manera efectiva.
- Identificar diferentes tipos de gráficos y visualizaciones y comprender cuándo es apropiado utilizar cada uno.
- Explorar una variedad de bibliotecas de visualización de datos en Python.
- Practicar la creación de narrativas visuales que guíen al espectador a través de los datos y las conclusiones.

Contenidos:

- Importancia de la visualización de datos en la exploración, análisis y comunicación de la información.
- Conceptos básicos del diseño visual, selección de colores, formas y estilos para representar los datos.
- Tipos de gráficos y visualizaciones. Interpretación.

- Galería de gráficos en Python
- comunicación de visualizaciones y storytelling (narrativa visual).
- Uso de Python. Librerías

Bibliografía:

Vander Plas (2016) *Python Data Science Handbook* " O'Reilly Media, Inc.". <https://jakevdp.github.io/PythonDataScienceHandbook/00.00-preface.html>

Wilke, C. O. (2019). *Fundamentals of data visualization: a primer on making informative and compelling figures*. O'Reilly Media. Disponible en <https://clauswilke.com/dataviz/index.html>

Unidad 4: Análisis de datos

Objetivos Específicos:

- Integrar herramientas de ciencia de datos de manera eficiente en un flujo de trabajo coherente para un adecuado análisis de datos.
- Garantizar la calidad y la reproducibilidad de los análisis de datos

Contenidos:

- Análisis exploratorio de datos. Su integración con la visualización de datos en entornos con Python.
- Documentación clara y completa de los procesos de análisis exploratorio de datos, incluyendo descripciones de las herramientas utilizadas, los pasos del análisis y los resultados obtenidos.
- Tableros de indicadores.
- Introducción a las herramientas de aprendizaje automático.

Bibliografía:

Vander Plas (2016) *Python Data Science Handbook* " O'Reilly Media, Inc.". <https://jakevdp.github.io/PythonDataScienceHandbook/00.00-preface.html>

Unidad 5: Introducción al aprendizaje automático

Objetivos Específicos:

- Entender la diferencia entre aprendizaje supervisado y no supervisado, así como sus aplicaciones.
- Explorar algoritmos de aprendizaje supervisado y no supervisado.
- Implementar estos algoritmos en conjuntos de datos para comprender su interpretación y el ajuste de parámetros.

Contenidos:

- Introducción.
- Aprendizaje Supervisado y no supervisado. Algoritmos.
- Análisis Supervisado: Regresión Logística

- Análisis No supervisado: Análisis de Clúster
- Uso de Python. Librería scikit learn

Bibliografía:

Vander Plas (2016) *Python Data Science Handbook* " O'Reilly Media, Inc.". <https://jakevdp.github.io/PythonDataScienceHandbook/00.00-preface.html>

Albon, C (2018) *Machine Learning with Python Cookbook* " O'Reilly Media, Inc.". <https://www.oreilly.com/library/view/machine-learning-with/9781491989371/>

METODOLOGIA DE ENSEÑANZA Y APRENDIZAJE

Los contenidos se desarrollan a través de clases teóricas – practicas. Se utilizan los gabinetes de computación como aula taller para trabajar con Python.

Se trabaja con bases de datos reales y/o simuladas con la finalidad de capacitar a los y las estudiantes en la manipulación de los datos a fin de obtener la base de datos adecuada para su procesamiento e interpretación. Para las bases de datos simuladas, se podrá utilizar chat basados en lenguajes de inteligencia artificial (chat GPT u otros).

Con la finalidad de realizar un buen seguimiento de las clases los y las estudiantes contarán con comandos y procesamientos en Python para que no sea una condición saber programar, sino que vayan adquiriendo habilidades a medida que se avanza en el cursado.

A través del aula Moodle se subirán materiales, bases de datos, autoevaluaciones, entre otros.

TIPO DE FORMACION PRACTICA

Se trabaja con proyectos donde se aplican herramientas e ciencias de datos a bases de datos con uso de software de libre acceso y con énfasis en la interpretación y visualización de los datos.

EVALUACIÓN

Evaluaciones parciales: 2

Trabajo práctico: 1

Recuperatorio: 1

CRITERIOS DE EVALUACIÓN

Los parciales tienen contenidos teóricos-conceptuales y prácticos. El recuperatorio tendrá un nivel de dificultad similar al de los parciales.

El trabajo práctico consiste en elaborar un proyecto de intervención donde se aplicarán los conceptos aprendidos en la asignatura a una base de datos.

CONDICIONES DE REGULARIDAD Y/O PROMOCION

Para regularizar la materia se debe aprobar dos parciales con opción a un recuperatorio.

Se puede acceder al sistema de promoción directa: aprobando 2 (dos parciales) y presentando el trabajo integrador correspondiente al proyecto de intervención.

MODALIDAD DE EXAMEN FINAL

Estudiantes regulares: trabajo integrador escrito

Estudiantes libres: examen integral escrito

CRONOLOGIA DE ACTIVIDADES DE LA ASIGNATURA

Unidades	Semana
1	1-2
2	3 -5
3	6 - 8
4	9 - 11
5	12 - 14

PLAN DE INTEGRACION CON OTRAS ASIGNATURAS

Se relaciona con Estadística I, Estadística II y Métodos cuantitativos para la toma de decisiones.

LECTURAS EXIGIDAS

Albon, C (2018) *Machine Learning with Python Cookbook* " O'Reilly Media, Inc.". <https://www.oreilly.com/library/view/machine-learning-with/9781491989371/>

Gorelik, A. (2019). *The enterprise big data lake: Delivering the promise of big data and data science*. O'Reilly Media.

Vander Plas (2016) *Python Data Science Handbook* " O'Reilly Media, Inc.". <https://jakevdp.github.io/PythonDataScienceHandbook/00.00-preface.html>

Wilke, C. O. (2019). *Fundamentals of data visualization: a primer on making informative and compelling figures*. O'Reilly Media. Disponible en <https://clauswilke.com/dataviz/index.html>

LECTURAS RECOMEN DADAS

McKinney, W. (2022). *Python for data analysis*. " O'Reilly Media, Inc.". <https://wesmckinney.com/book/>

Anderson, D., Sweeney, D., Williams, T., Camm, J., & Cochran, J. (2019). *Estadística para negocios y Economía*. Decimotercera edición. México: Cengage Learning.

Gobernanza Digital. Publicación de las Naciones Unidas. Disponible en: <https://repositorio.cepal.org/server/api/core/bitstreams/6a12e389-3dcb-4cba-830a-99f038835423/content>

Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and Tensor Flow: Concepts, Tools, and Techniques to Build Intelligent Systems* (2nded.). O'ReillyMedia

Turrell, A. (2024) Coding for Economist- Disponible en <https://aeturrell.github.io/coding-for-economists/intro.html>

Fiol, J (2023) clases de Python del Instituto Balseiro. Disponible en: <https://fiolj.github.io/data/ClasesdePython.pdf>

Turrell, A. (2024) Python for data science. Disponible en: <https://aeturrell.github.io/python4DS/welcome.html>