

GUIA DOCENTE

MÁSTER UNIVERSITARIO EN DATA ANALYTICS FOR BUSINESS

Edición 2

Curso 2024-2025

1. ASIGNATURA

- **Nombre:** Análisis exploratorio de datos
- **Tipo de asignatura:** Obligatoria
- **Trimestre:** PRIMERO
- **Créditos:** 3 ECTS
- **Idioma de docencia:** castellano
- **Coordinador de la asignatura:** Alexandra Abós Ortega
- **Profesor/es de la asignatura:** Alexandra Abós Ortega

2. PRESENTACIÓN DE LA ASIGNATURA

Objetivos de la asignatura

El objetivo principal de esta asignatura es proporcionar una introducción a los aspectos fundamentales del análisis exploratorio de datos. También se proporcionan las herramientas básicas para identificar y resolver los problemas más comunes que se encuentran durante esta fase de análisis y preproceso de datos.

El temario está organizado siguiendo las distintas etapas, desde la exploración inicial de los datos, su visualización, la corrección de valores anómalos y el preproceso para preparar los datos para su posterior análisis.

Nuestro compromiso con el impacto social y el bienestar planetario se traduce en contenidos formativos alineados con los Objetivos de Desarrollo Sostenible (ODS) previstos en la agenda 2030:



En la asignatura que nos ocupa, los ODS implicados son:

- ODS.4. Educación de Calidad
- ODS.8. Trabajo Decente y Crecimiento Económico

Contenidos

Esta asignatura está compuesta de 9 unidades didácticas, cuyo contenido se detalla a continuación.

1. Introducción al análisis exploratorio de datos. ¿Qué es y por qué es importante?
2. Visualización para el análisis exploratorio de datos.
 - 2.1. Gráficos más comunes y cuándo usarlos.
3. Análisis descriptivo
 - 3.1. Estadísticos descriptivos
 - 3.2. Gestión de los tipos de variables
4. Detección y gestión de datos nulos y atípicos.
 - 4.1. Detección de valores nulos
 - 4.2. Gestión de valores nulos según su tipo
 - 4.3. Detección de valores atípicos
 - 4.4. Corrección de valores atípicos
5. Conversión de variables categóricas a numéricas
 - 5.1. Tipos de datos categóricos
 - 5.2. Codificación de categorías: Label Encoding y One Hot Encoding
6. Correlaciones
 - 6.1. Correlación: definición e importancia
 - 6.2. Impacto de la presencia de variables correlacionadas en el análisis de datos
 - 6.3. Detección de correlación entre variables
 - 6.4. Gestión de variables correlacionadas
7. Análisis exploratorio de series temporales
 - 7.1. ¿Qué es una serie temporal?
 - 7.2. Visualización de series temporales
 - 7.3. Conversión de tipo de fecha
 - 7.4. Detección y gestión de datos nulos y atípicos en series temporales
 - 7.5. Suavizado de series temporales

8. Normalización y estandarización de datos
 - 8.1. Normalización y estandarización: conceptos básicos
 - 8.2. Técnicas de normalización y estandarización de datos

9. Balanceo de datos
 - 9.1. Datos desbalanceados: definición y problemas asociados.
 - 9.2. Métricas para evaluar datos desbalanceados
 - 9.3. Algoritmos de balanceo de datos: sobremuestreo y submuestreo, generación de muestras sintéticas.

La asignatura dentro del plan de estudios

*Esta asignatura obligatoria/optativa se emmarca dentro de la materia 1. **Análisis de Datos. Data Analytics** del plan de estudios. Se realiza durante el **primer trimestre** una vez se han adquirido los conocimientos de *Introducción a Python* e *Introducción a la Inteligencia Artificial*.*

Competencias/Resultados de aprendizaje

- RA1. Mat 1.1 Seleccionará la infraestructura necesaria para hacer frente a un proyecto que involucre datos masivos.
- RA3 Mat 1.2 Realizará un informe completo que incluya visualizaciones diversas sobre un conjunto de datos y que ayuden a la toma de decisiones.
- RA4. Mat 1.3 Extraerá información de modo visual a partir de datos masivos.
- RA5. Mat. 1.4 Identificará las diferencias fundamentales entre bases de datos relacionales y no relacionales.
- RA6. Mat 1.5 Analizará un conjunto de datos mediante análisis univariante y bivariante.
- RA6. Mat 1.6 Distinguirá correlaciones entre las diferentes dimensiones de un conjunto de datos.
- RA7. Mat 1.7 Realizará un plan de gobernanza de datos que minimice riesgos y costes.
- RA19. Mat 1.1 Utilizará python para generar visualizaciones adecuadas al tipo de datos que se estén trabajando en diversos sectores empresariales
- RA14. Mat 1.1 Planteará unas hipótesis sobre un conjunto de datos realizando un test de hipótesis que te permita saber si debes aceptar o rechazar la hipótesis nula.
- RA14. Mat 1.2 Propondrá un algoritmo para predecir la aceptación, por parte de clientes banco visionarios, de diferentes campañas.
- RA14. Mat 1.3 Construirá un algoritmo de clasificación de pacientes sanos vs enfermos.
- RA14. Mat 1.4 Diseña un algoritmo para la segmentación de clientes.
- RA14. Mat 1.5 Lista todas las variables que deberías tener en cuenta para optimizar los procesos de almacenaje de una empresa.
- RA14. Mat 1.6 Formulará nuevas aplicaciones que podría desarrollar en el ámbito empresarial o sectorial utilizando las herramientas más punteras de analítica de datos.

3.PLAN DE APRENDIZAJE DE LA ASIGNATURA

Metodología docente

El enfoque de esta asignatura es práctico, pero las sesiones contienen una primera parte de teoría, seguida de una parte práctica donde se plantean problemas comunes de las distintas fases del análisis exploratorio de datos. El alumnado tiene que identificar y resolver estas cuestiones mediante las funcionalidades estudiadas.

Horas de dedicación (horas lectivas+trabajo del alumno): 75

Evaluación

Esta asignatura se evalúa con un trabajo final en grupo, donde el alumnado tiene que aplicar los conceptos estudiados en cada unidad. El trabajo se tendrá que presentar de manera oral el último día del curso, junto con una memoria escrita.

Sistema de recuperación

A los estudiantes que suspendan el curso durante la evaluación ordinaria se les permitirá hacer un trabajo para poder aprobar la asignatura. Los estudiantes que aprueben obtendrán por defecto un 5 como nota final del curso. Si el curso vuelve a ser suspendido después de la repetición, los estudiantes tendrán que matricularse de nuevo en el curso del año siguiente.

Plagio

Plagio es utilizar el trabajo de otro y presentarlo como propio sin reconocer las fuentes de forma correcta. Todos los trabajos, informes o proyectos entregados por los estudiantes deben ser trabajos originales realizados por ellos mismos.

Actividades formativas

*Información sobre las sesiones.

Primera sesión	Unidad 1: Introducción al análisis exploratorio de datos Unidad 2: Visualización para el análisis exploratorio de datos.
Segunda sesión	Unidad 3: Análisis descriptivo Unidad 4: Detección y gestión de datos nulos y atípicos.
Tercera sesión	Unidad 5: Conversión de variables categóricas a numéricas Unidad 6: Correlaciones

Cuarta sesión	Unidad 7: Análisis exploratorio de series temporales
Quinta sesión	Unidad 8: Normalización y estandarización de datos Unidad 9: Balanceo de datos
Sexta sesión	Presentaciones orales trabajo final

3. PROFESORADO

Alexandra Abós es Data Scientist en la farmacéutica Sanofi. Doctora en Ingeniería Biomédica por la Universidad de Barcelona, también es profesora en el Máster y Grado de Data Science en la Universitat Oberta de Catalunya (UOC).

Cuenta con un título de Grado y Máster en Ingeniería Biomédica por la Universidad de Barcelona, y un Postgrado en Inteligencia Artificial (IA) con redes neuronales profundas (Deep learning) por la Universidad Politécnica de Cataluña (UPC). También es autora y/o coautora de múltiples artículos científicos ([Google Scholar](#)).

4. BIBLIOGRAFIA (*obligatoria/ recomendada*)

Python Crash Course, 3rd Edition. Eric Matthes. No Starch Press.
Python for Data Analysis. Data Wrangling with pandas, NumPy, and Jupyter. Wes McKinney. O'Reilly Media, Inc.
Python Data Science Handbook. Jake VanderPlas. O'Reilly Media, Inc.
Fundamentals of Data Visualization. Claus O. Wilke. O'Reilly Media, Inc.
Practical Python Data Wrangling and Data Quality. Susan E. McGregor. O'Reilly Media, Inc.