

Introducción al trabajo con Datos

(30 horas, 6 créditos)

Especialización en Economía

Opción Ciencia de Datos

Edición 2024

Leonardo Barceló

Descripción

Este curso está diseñado para introducir a los estudiantes al campo de la ciencia de datos, proporcionando una comprensión fundamental de los tipos de datos, las herramientas de procesamiento de datos y las metodologías para convertir datos en información valiosa. Los estudiantes aprenderán a trabajar con diversas herramientas de software y lenguajes de programación utilizados en la ciencia de datos.

Objetivos del Curso

Dentro de los objetivos del curso se encuentran:

- Comprender los conceptos básicos de los datos y su almacenamiento.
- Distinguir entre diferentes tipos y formatos de datos.
- Utilizar herramientas y lenguajes de programación para el procesamiento de datos.
- Aplicar técnicas de análisis de datos para extraer información valiosa.
- Introducirse en el manejo de datos no estructurados y el uso de APIs.

Contenido del Curso

Introducción a la Ciencia de Datos

- Definición y campo de aplicación
 - Definición e identificación de disciplinas entorno a la ciencia de Datos
 - Concepto de Dato e Información ○ Proceso de Extracción de conocimiento
 - Aplicaciones en la industria ○ Herramientas y Técnicas
 - Conceptos de Big Data y Análisis de Datos

- Datos estructurados y no estructurados
- Ética y privacidad
- Importancia de los datos en el mundo actual
 - Contexto actual
 - Innovación y Desarrollo futuro

Bases de Datos Estructuradas

- Introducción a las Bases de Datos Estructuradas
 - Definición y características de las bases de datos estructuradas.
 - Importancia y aplicaciones en el mundo real
 - Comparación con bases de datos no estructuradas y semiestructuradas.
- Modelado de Datos
 - Conceptos de modelado relacional
 - Modelo Entidad Relación
 - Modelos Relacionales
 - Conceptos de Normalización de datos
- Conceptos de SQL
 - Introducción a SQL
 - Operaciones CRUD
 - Consultas Básicas
 - Consultas Avanzadas
- Conceptos y diseño de bases de datos relacionales
 - Conceptos de RDBMS
 - Definición
 - Integridad
 - Seguridad
 - Transaccionalidad
 - Disponibilidad

Fundamentos de Programación para Procesamiento de Datos

- Conceptos Básicos de Programación
 - Variables y tipos de datos
 - Operadores y expresiones
 - Entrada y salida de datos
- Estructuras de Control
 - Condiciones (if, else if, else)
 - Bucles (for, while, do-while)
 - Control de flujo (break, continue, return)
- Estructuras de Datos Básicas
 - Arreglos o arrays
 - Cadenas de caracteres (strings)
 - Listas y diccionarios (según el lenguaje de programación)
 - DataFrames
- Funciones y Procedimientos
 - Definición y llamada de funciones
 - Parámetros y argumentos
 - Valor de retorno de las funciones

- Recursividad
- Herramientas básicas de Visualización
- Fundamentos y Conceptos de Programación Orientada a Objetos (OOP)
 - Conceptos básicos: clase, objeto, método y propiedad
 - Encapsulamiento, herencia, y polimorfismo
 - Constructores y destructores
- Lenguajes para procesamiento de datos
 - Introducción a R y RStudio
 - Introducción a Python
 - Utilización de librerías de procesamiento de datos y algoritmos
 - Uso de librerías: pandas y numpy para análisis de datos
- Introducción a MySQL
- Introducción a herramientas de desarrollo
 - Visual Studio Code
 - Jupyter notebooks
 - Google Colab
 - DBeaver

Introducción a la gestión de código con GitHub

- Introducción a control de versiones
- Uso básico de GitHub: repositorios, commits, branches, pull requests

Evaluación

La evaluación se realizará en base a un total de 100 puntos que estarán distribuidos de la siguiente forma:

- 20 puntos: Prueba al finalizar el curso
- 80 puntos: Trabajo obligatorio a ser realizado en equipo

Bibliografía

"Data Modeling Essentials" por Graeme Simsion y Graham Witt

Este libro ofrece una sólida introducción a las técnicas de modelado de datos, desde los fundamentos hasta conceptos avanzados.

"SQL for Data Scientists: A Beginner's Guide for Building Datasets for Analysis" por Renee M. P. Teate

Este libro está diseñado específicamente para científicos de datos que necesitan aprender SQL para manejar y analizar datos.

"Learning SQL" por Alan Beaulieu

Un libro introductorio que enseña SQL desde lo básico hasta conceptos más avanzados, ideal para principiantes.

"Python for Data Analysis" por Wes McKinney

Enfocado en el manejo de datos con pandas y NumPy, esencial para la ciencia de datos en Python.

"Data Science for Business" por Foster Provost y Tom Fawcett

Explica los principios fundamentales de la ciencia de datos y su aplicación en el ámbito empresarial.

"The Data Science Handbook" por Field Cady

Una guía completa que abarca todas las áreas fundamentales de la ciencia de datos.

"An Introduction to Statistical Learning" por Gareth James, Daniela Witten, Trevor Hastie, y Robert Tibshirani

Introduce los métodos estadísticos utilizados en la ciencia de datos, con ejemplos y aplicaciones prácticas.

"Practical Statistics for Data Scientists" por Peter Bruce y Andrew Bruce Explica conceptos estadísticos clave con ejemplos y aplicaciones en ciencia de datos.

Recursos en línea

SQLBolt

Un recurso en línea interactivo para aprender SQL básico y avanzado a través de lecciones y ejercicios prácticos.

"SQLZoo"

Una plataforma interactiva para aprender SQL en línea con ejercicios prácticos que cubren desde conceptos básicos hasta avanzados.

"Data Science from Scratch" por Joel Grus (disponible en O'Reilly Media) Un libro que enseña los fundamentos de la ciencia de datos y cómo implementarlos desde cero utilizando Python.

"Python Data Science Handbook" por Jake VanderPlas

Un recurso excelente para aprender a utilizar las bibliotecas de Python esenciales en la ciencia de datos como NumPy, Pandas, Matplotlib y Scikit-Learn.

"Think Stats: Probability and Statistics for Programmers" por Allen B. Downey. Ofrece una introducción a la estadística y la probabilidad desde la perspectiva de la programación.

"Introduction to Data Science" por JeVrey Stanton

Proporciona una visión general de las principales técnicas y conceptos en la ciencia de datos.

"Mining of Massive Datasets" por Jure Leskovec, Anand Rajaraman, y JeV Ullman. Este libro, disponible gratuitamente por sus autores, se centra en los algoritmos necesarios para trabajar con grandes conjuntos de datos.