

Modelado de datos

Descripción

Las empresas generan grandes cantidades de datos diariamente. Este curso enseña a analizar esos datos para comprender a los clientes en profundidad. Aprenderás a formular hipótesis, generar estadísticas descriptivas y escribir consultas SQL avanzadas, incluso trabajando con tipos de datos como JSON. Además, explorarás análisis geoespacial y de texto, y técnicas avanzadas como el perfilado y la automatización. Al concluir, dominarás técnicas SQL para escenarios empresariales y analizarás datos con perspectiva profesional.

Prerrequisitos

- Nociones de lenguaje de programación: Python o R.

Contenido temático

1. Entendiendo y Describiendo Datos
2. Fundamentos de SQL para análisis
3. SQL para Preparación de Datos
4. Funciones de Agregación para Análisis de Datos
5. Funciones de Ventana para Análisis de Datos
6. Importación y Exportación de Datos
7. Análisis Utilizando Tipos de Datos Complejos

Evaluación

- 30 Asistencias
- 40 Prácticas
- 30 Trabajo final

Referencias

- T. Blanchard, Data Science for Marketing Analytics: A practical guide to forming a killer marketing strategy through data analysis with Python, 2nd Edition, 2019.
- B. Lubanovic, Introducing Python: Modern Computing in Simple Packages, 2019.
- W. McKinney, Python for Data Analysis: Data Wrangling with Pandas, Numpy, and Jupyter, 2022.

Guía de Aprendizaje

ITESO, Departamento de Economía, Administración y Mercadología Clave de la Asignatura: ITE3793

Periodo Escolar: Otoño 2024 Ordinario

No. de Créditos BCD y TIE y dedicación en horas: Créditos: 8.

Horas TIE: 48 horas

Asignatura: Modelado de datos

Grado y Nombre del Profesor: 918864 LUIS MIGUEL ESCOBAR VEGA

Horario: Sáb 08:30 - 12:30 Espacio: V-406

Propósito general de la asignatura

La asignatura se centra en el diseño estratégico organizacional para la recolección, actualización y almacenamiento de datos. Destaca el papel fundamental del analista de negocios y clientes, quien se integra activamente en los equipos de proyectos. Su misión es identificar, explorar y perfeccionar sistemas y fuentes de datos alineados con los objetivos organizacionales. Esta labor se realiza siempre respetando rigurosos estándares de ética y legalidad.

Descripción del proyecto formativo

1. Entendiendo y Describiendo Datos

Este tópico se enfoca en la comprensión fundamental de los datos con los que trabajamos. Incluye aprender a reconocer las estructuras de datos, su procedencia, naturaleza y calidad. Además, se abordan técnicas estadísticas y visuales básicas para describir y resumir conjuntos de datos, permitiendo obtener una vista preliminar de la información y detectar posibles anomalías o tendencias.

2. Fundamentos de SQL para análisis

En este apartado, se introduce a SQL (Structured Query Language), el lenguaje de consulta estructurada usado para gestionar y recuperar datos de bases de datos relacionales. Se cubren conceptos básicos como selección, filtrado, ordenación y unión de datos, ofreciendo una base sólida para análisis más complejos.

3. SQL para Preparación de Datos

Más allá de las consultas básicas, aquí se exploran técnicas avanzadas de SQL para limpiar, transformar y preparar datos para análisis. Esto puede incluir normalización de datos, manejo de valores faltantes, conversión de tipos de datos, y otras operaciones esenciales antes de cualquier análisis en profundidad.

4. Funciones de Agregación para Análisis de Datos

Este tópico aborda cómo consolidar o agrupar datos en conjuntos más pequeños y manejables utilizando funciones de agregación en SQL, como SUM, AVG, COUNT, entre otras. Estas funciones son cruciales para obtener resúmenes y métricas de conjuntos de datos grandes.

5. Funciones de Ventana para Análisis de Datos

Las funciones de ventana en SQL permiten realizar cálculos a través de un conjunto específico de filas, denominado “ventana”, que están relacionadas con la fila actual dentro de una consulta. Son esenciales para operaciones como cálculos acumulativos, rankings o comparaciones de datos con períodos anteriores.

6. Importación y Exportación de Datos

Este módulo aborda las técnicas y herramientas para introducir datos en una base de datos (importación) o extraer datos de una base de datos para su uso en otros programas o formatos (exportación). Se enfoca en garantizar la integridad de los datos durante estos procesos y en manejar diferentes formatos de datos.

7. Análisis Utilizando Tipos de Datos Complejos

A diferencia de los datos simples como números o cadenas, los datos complejos pueden incluir estructuras como arrays, mapas o registros. En este tópico, se exploran técnicas y herramientas para analizar estos tipos de datos, permitiendo extraer insights más profundos y gestionar información en formatos más ricos y variados.

Competencias por desarrollar / Propósitos específicos

Al finalizar esta asignatura, el estudiante será capaz de evidenciar avances en las siguientes competencias:

- Visión sistémica y estratégica: analizar las necesidades de una empresa desde una perspectiva compleja e identificar las prioridades desde un enfoque estratégico, de forma que pueda fundamentar una propuesta de proyectos para el desarrollo empresarial.
- Pensamiento creativo e innovador: identificar y fundamentar alternativas que permitan el desarrollo de la empresa a intervenir y/o la solución de problemas o procesos críticos en su operación.
- Compromiso ético: demostrar que los aprendizajes y las actividades realizadas durante el curso se operan en respeto y apoyo mutuos, en aprendizaje colaborativo y el diálogo de saberes y que las propuestas de proyectos integran los propósitos y valores de la responsabilidad social empresarial.
- Aprendizaje crítico y reflexivo: identificar y discriminar los elementos más significativos de la forma cómo aprenden y desarrollan las competencias profesionales.
- Planeación: establecer objetivos de acuerdo con un diagnóstico sistémico, definir una estrategia de intervención y organizar etapas y acciones en el tiempo para el logro de dichos objetivos.

Competencias previas del estudiante

Para poder inscribirse en esta asignatura, el estudiante debe evidenciar ser capaz de:

- Nociones de lenguaje de programación: Python o R.
- Investigar: buscar y procesar información procedente de fuentes diversas, de manera sistemática, coherente y consistente para comprender una realidad y plantear recomendaciones.
- Analizar y sintetizar: distinguir y contrastar los contenidos teórico-conceptuales, así como los empíricos, que permitan entender y caracterizar la información de la empresa necesaria para el desarrollo de su proyecto y su justificación.

Gestión de la calidad del proyecto

Autor (es): Dr. Luis Escobar

Fecha: Segunda versión: Otoño 2024

Periodicidad de la revisión: Semestral

Normas del curso

- La comunicación con los alumnos se hace mediante el correo institucional del alumno.
- Las reglas del curso obedecen a la normativa ITESO:
 - Normativa propia de posgrados, disponible en la página web de posgrados.iteso.mx/admision,
 - Normativa general para alumnos de ITESO, disponible en la página web de ITESO normativa universitaria.

Calificación de la asignatura

Elementos por calificar

- Asistencias 30%
- Prácticas 40%
- Trabajo final 30%

Referencias documentales

Bibliografía básica y complementarias, artículos, vínculos de internet y casos de estudio

- W. McKinney, Python for Data Analysis: Data Wrangling with Pandas, Numpy, and Jupyter, 2022.
- B. Lubanovic, Introducing Python: Modern Computing in Simple Packages, 2019.
- T. Blanchard, Data Science for Marketing Analytics: A practical guide to forming a killer marketing strategy through data analysis with Python, 2nd Edition, 2019.

Otros materiales de apoyo (artículos, vínculos de internet, casos de estudio):

- <https://colab.research.google.com/>
- <https://quickref.me/python>
- <https://github.com/limspiga/data-modeling-fall-24>