**Facultad de Ingeniería y Ciencias  
Escuela de Informática y Telecomunicaciones**

**PROGRAMA DE ASIGNATURA**Data Science

1. **Identificación de la asignatura:**

| Nombre de la Asignatura: Data Science | |
| --- | --- |
| Códigos: CIT-3620 | Créditos: 6 |
| Duración: Semestral | Ubicación en el plan de estudios: Semestre 9 |
| Requisitos: CIT-2309 Big Data, CIT-2313 Inteligencia Artificial | |
| Sesiones cátedras semanales: 2 cátedras | |
| Sesiones de Ayudantía: 1 | |

1. **Descripción de la asignatura:**

Curso aplicado, orientado al desarrollo de proyectos interdisciplinarios en ciencia de datos, donde los/as estudiantes diseñan, ejecutan y comunican soluciones basadas en datos para problemas reales de distintas industrias. Se trabaja con estudios de caso de alto impacto, integrando habilidades técnicas con la comprensión del contexto de aplicación, la ética del uso de datos y la colaboración con perfiles no técnicos.

1. **Resultados de aprendizaje:**
2. Desarrolla soluciones de ciencia de datos, contextualizadas a problemas reales e interdisciplinarios.
3. Participa en equipos, colaborando con especialistas de otras áreas para entender problemas, formular hipótesis y comunicar resultados.
4. Aplica herramientas de ciencia de datos de manera estratégica, desde la recolección y limpieza de datos hasta el modelado y visualización.
5. Comunica hallazgos de forma efectiva a públicos técnicos y no técnicos mediante informes, dashboards y presentaciones.
6. **Unidades Temáticas:**

* **Unidad 1: Contexto y desafíos en ciencia de datos interdisciplinaria**
  + Rol del científico de datos en distintas industrias
  + Comunicación efectiva con perfiles no técnicos
  + Ética, sesgo y responsabilidad en la toma de decisiones basadas en datos
* **Unidad 2: Estructura del proyecto de ciencia de datos**
  + Formulación del problema y validación con stakeholders
  + Definición de métricas de éxito y restricciones prácticas
  + Diseño del pipeline de trabajo
* **Unidad 3: Estudios de caso por industria**  
  *Se desarrollarán ejemplos concretos y datasets para trabajo práctico, se sugieren temas a analizar:*
  + **Salud**: predicción de diagnósticos, optimización de recursos hospitalarios
  + **Finanzas**: segmentación de clientes, detección de fraude
  + **Retail/Marketing**: análisis de comportamiento, recomendadores
  + **Transporte y logística**: optimización de rutas, demanda y congestión
  + **Energía y medio ambiente**: consumo energético, análisis climático
  + **Educación**: predicción de deserción, personalización de contenidos
* **Unidad 4: Desarrollo del proyecto final**
  + Selección de dominio e identificación de problema real
  + Desarrollo iterativo en equipo (con entregas parciales)
  + Validación, visualización y presentación de resultados

1. **Descripción general del método de enseñanza:**

El curso se desarrolla como un laboratorio práctico con foco en trabajo en equipo y resolución de problemas reales. Cada grupo de estudiantes abordará un proyecto aplicado desde una perspectiva interdisciplinaria, en colaboración con expertos o usuarios de otras disciplinas (o mediante simulaciones guiadas).

Se trabajará con herramientas estándar de ciencia de datos (Python, Pandas, Scikit-learn, Streamlit, GitHub, etc.) y se incentivará el uso de recursos abiertos (datasets públicos, APIs gubernamentales, Kaggle, etc.). Las sesiones combinan clases cortas, talleres, trabajo guiado y presentaciones de avance.

1. **Descripción general de la modalidad de evaluación:**

La evaluación se centra en el desarrollo progresivo de un proyecto de ciencia de datos con entregas parciales durante el semestre, complementado con trabajos prácticos y dos pruebas solemnes. Se evalúan tanto las competencias técnicas como la capacidad de comunicar resultados a distintos públicos.

Entregas del proyecto final: El proyecto se evalúa en etapas a lo largo del semestre:

* Propuesta de proyecto y definición del problema
* Avance 1: exploración de datos + visualización
* Avance 2: modelado y validación
* Informe final + presentación oral

1. **Bibliografía Básica Obligatoria:**
2. Cathy O’Neil y Rachel Schutt. Doing Data Science. O’Reilly.
3. Jake VanderPlas. Python Data Science Handbook. O’Reilly.
4. Aurélien Géron. Hands-On Machine Learning. O’Reilly.
5. Eric Ma, Alice Zhao. Communicating Data with Python.
6. Documentación oficial: Pandas, Scikit-learn, Streamlit, FastAPI
7. Casos de estudio y datasets: [kaggle.com](https://kaggle.com), [data.gov](https://data.gov), [ourworldindata.org](https://ourworldindata.org)

**

Elaborado por: Jonathan Frez

Fecha revisión: Marzo de 2025

Fecha vigencia: Marzo de 2026