



Coordinación de  
**Educación Abierta y a Distancia**  
VICERRECTORADO ACADÉMICO



---

## PROGRAMACIÓN 2

### Actividad Autónoma 6

**Unidad 3:** Análisis de la Información

**Tema 2:** Visualización y Análisis Exploratorio de Datos

---



FACULTAD DE  
Ingeniería

Nombres:

Fecha:

Carrera: Ciencia de Datos

Periodo académico:

Semestre:



### Objetivo de la actividad:

El objetivo de este proyecto es que el estudiante aplique el flujo completo de un proyecto de ciencia de datos: desde la selección, preprocessamiento y visualización de un dataset real (no Iris ni Penguins), hasta la implementación de modelos de clasificación y análisis de resultados. Se busca que el estudiante desarrolle habilidades para explorar, transformar, visualizar y modelar datos, justificando cada decisión técnica y comunicando los hallazgos de forma profesional y reproducible.

### Recursos o temas que debe haber estudiado antes de hacer la actividad:

- Fundamentos de Pandas y NumPy: manipulación de datos y arrays.
- Preprocesamiento: manejo de valores nulos y outliers.
- Visualización básica con Matplotlib y/o Seaborn.
- Codificación de variables categóricas: `.astype()`, `pd.get_dummies()`.
- Normalización/estandarización manual con Pandas o NumPy.
- Implementación básica de modelos de clasificación/regresión con scikit-learn.
- Uso de Jupyter Notebook o Google Colab.

**Formato de entrega:** Jupyter Notebook , a la nube de GitHub Classroom o SIGEA

**Importante:** El notebook debe tener un encabezado con los nombres de todos los integrantes y una sección final de reflexiones grupales.

### Instrucciones:

#### 1. Selección y Justificación del Dataset (Trabajo colaborativo)

- Elijan en grupo un dataset público de Kaggle (NO usar Iris ni Penguins). ✓
- **Justifiquen** la selección del dataset: ¿Por qué es relevante? ¿Qué preguntas de investigación pueden responder con estos datos?
- Formulen al menos **dos preguntas de investigación** o hipótesis que guiarán el análisis.

#### 2. Preprocesamiento y Análisis Exploratorio

- Realicen en grupo la limpieza de datos (nulos, outliers) y documente cada decisión. ✓



- Transformen y codifiquen variables categóricas usando `.as-type()` o `pd.get_dummies()`.
- Estandaricen o normalicen variables numéricas.
- Realicen visualizaciones exploratorias que permitan entender el dataset **y justifiquen sus decisiones.**
- **Respondan en equipo:** ¿Qué problemas encontraron en el dataset? ¿Cómo los resolvieron?

### 3. Modelado y Comparación de Resultados

- Seleccionen una variable objetivo categórica y **justifiquen por qué es interesante.**
- Dividan los datos en **entrenamiento** y prueba.
- Implementen en grupo **dos modelos de clasificación** distintos (ej: Logistic Regression, DecisionTree, SVM, Random Forest).
- **Analicen el desempeño usando** métricas (`accuracy`, matriz de confusión, precision, recall).

Agregarle una pequeña descripción a cada gráfico de cada modelo. 4 son de cada uno ya están implementados.

- Agreguen visualizaciones comparativas entre los modelos.
- **Discutan:** ¿Qué modelo funcionó mejor y por qué?

### 4. Análisis colaborativo y Comunicación de resultados

- **Redacten una sección de conclusiones grupales:** ¿Qué aprendieron en el proceso? ¿Cuáles fueron las principales dificultades? ¿Qué harían diferente?
- **Respondan:** ¿Qué técnicas de preprocessamiento o visualización fueron más útiles y por qué?
- **Comenten posibles mejoras y futuras líneas de investigación con estos datos.**

### 5. Presentación y trabajo en equipo

- Organíicense para que cada integrante aporte en el notebook (cada sección debe tener una breve firma indicando quién la desarrolló o coordinó).
- El notebook debe estar limpio, con celdas de texto explicativo y código bien comentado.



### Reto Adicional (+1 punto extra grupal)

Investiguen y apliquen en grupo una técnica de selección de características (feature selection). Analicen cómo cambia el desempeño de los modelos y discutan sus resultados como grupo.

#### Bibliografía:

- McKinney, W. (2022). Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter (3rd ed.). O'Reilly Media.
- Géron, A. (2022). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow (3rd ed.). O'Reilly Media.
- Documentación oficial de Pandas, NumPy, Seaborn, Matplotlib y Scikit-learn.



## Rúbrica de evaluación

<b>Componente de aprendizaje:</b>	Autónom o	X	Contacto con el Docente	
<b>Nombre de la Unidad:</b>	Unidad 3: Análisis de la Información			
<b>Resultado(s) de aprendizaje:</b>	<ul style="list-style-type: none"> <li>-Aplicar técnicas de limpieza, transformación y preprocesamiento de datos utilizando Pandas y NumPy en datasets reales.</li> <li>Comprender y emplear métodos de codificación de variables categóricas, estandarización y normalización de datos, así como visualización exploratoria.</li> <li>Implementar y analizar modelos matemáticos básicos de clasificación y regresión con scikit-learn.</li> <li>Desarrollar soluciones computacionales documentadas y reproducibles en Python, interpretando críticamente los resultados obtenidos.</li> <li>Trabajar en equipo para resolver preguntas de investigación y comunicar hallazgos de forma efectiva.</li> </ul>			
<b>Nombre de la Actividad:</b>	Preprocesamiento de Datos con Pandas			

Criterios de Evaluación	Escala de Valoración						
	Excelente (10 - 9,1)	Bueno (9 - 8,1)	Satisfactorio (8 - 7)	Necesita mejorar (6,9 - 0,1)	No entrega (0)	Puntaje	Comentarios (SIGEA)
<b>1. Selección, justificación y preguntas de investigación</b>	Justificación sólida, preguntas claras y relevantes , participación de todos los integrantes.	Justificación adecuada, preguntas presentes.	Justificación superficial, preguntas poco claras.	Justificación mínima o confusa, preguntas ausentes.	No presenta.		
<b>2. Preprocesamiento y exploración</b>	Técnicas completas, bien justificadas, visualizaciones claras y colaborativas.	Técnicas aplicadas correctamente, documentación aceptable.	Técnicas básicas, documentación incompleta.	Errores o ausencias notables.	No realiza.		
<b>3. Implementación y comparación de modelos</b>	Ambos modelos bien implementados, análisis comparativo.	Ambos modelos presentes, análisis.	Un solo modelo o análisis superficial.	Errores graves en modelos o métricas.	No presenta.		



	ivo profundo, uso adecuado o de métricas y visualizaciones.	s adecuado.					
<b>4. Reflexión y comunicación de resultados grupales</b>	Conclusiones profundas, reflexión colaborativa, comunicación clara y profesional.	Reflexión grupal adecuada.	Reflexión superficial o individual.	Sin reflexión grupal.	No presentada.		
<b>5. Presentación, limpieza y trabajo en equipo</b>	Notebook estructurado, roles claros, firmas de integrantes, excelente presentación.	Presentación adecuada, participación grupal aceptable.	Presentación desordenada, poca participación grupal visible.	Notebook desorganizado, sin evidencia de trabajo grupal.	No entrega.		

Puntaje  
total

Los criterios 4 y 5 están alineados a los ejes de formación del Modelo Educativo UNACH "Introspección y Prospectiva" y responden principalmente a dos de los siguientes ejes:

1. Ambiente;
2. Autonomía y adaptabilidad;
3. Comunicación;
4. Desarrollo humano;
5. Ética y valores;
6. Emprendimiento;
7. Inter y multidisciplinariedad;
8. Innovación;
9. Inclusión e interculturalidad;
10. Investigación;
11. Impacto social;
12. Tecnologías.