

**INSTITUTO POLITÉCNICO NACIONAL
ESCUELA SUPERIOR DE CÓMPUTO
MINERÍA DE DATOS**

Semestre escolar: 2026-1 (Agosto-Diciembre 2025)

Grupo: _____

Nombre: _____

Descripción

Los datos se recopilaron para una investigación que se llevó a cabo como parte del programa de Investigación Ecológica a Largo Plazo de la Estación Palmer, Antártida, que fue apoyado por subvenciones a través de la Fundación Nacional de Ciencias, Oficina de Programas Polares (NSF-OPP).

Autores del conjunto de datos:

Gorman KB, Williams TD & Fraser WR (2014). Palmer Penguins Dataset for EDA. Disponible en: url=<https://www.kaggle.com/ds/5489301>. DOI={10.34740/KAGGLE/DS/5489301}.

Enlace de acceso a los datos de Palmer Penguins Dataset for EDA

<https://www.kaggle.com/datasets/satyajeetrai/palmer-penguins-dataset-for-eda>

Generar un reporte con el BIRT del KNIME que considere los siguientes aspectos:

- 1) Portada de Presentación del curso, autores, UA, etc. Práctica 1. Aspectos descriptivos de los pingüinos de Palmer.
- 2) Investigar sobre el contexto de este conjunto de datos, agregando los tipos de pingüinos, las islas en donde viven.
- 3) **Descripción del conjunto de datos:** cantidad de elementos, describir cada atributo en la siguiente tabla:

No	Nombre	Tipo	Significado	Dominio
1				
2				

- 4) **Para todo el conjunto de datos** realizar lo siguiente:
 - a) **Análisis y tratamiento de datos.** Analizar cada uno de los atributos y en caso de ser necesario realizar el tratamiento correspondiente.
 - b) **Descripción de atributos.** Describir cada uno de los atributos mediante gráficas o tablas cruzadas o de frecuencias. Agregar el título del o los atributos.
 - c) **Consultas.** Realizar tres consultas para describir la información de todo el conjunto de datos. Agregar el enunciado de la consulta.
 - d) **Correlación.** Realice pruebas de correlación y explique el significado del coeficiente obtenido

e) Agrupamiento.

(1) Clúster. Al menos en dos ocasiones, aplicar el algoritmo *K-Means* para crear grupos de los atributos que usted elija y realice la descripción de los grupos resultantes. Usted elija la cantidad de grupos y explique la valoración de los mismos.

(2) Contenedores (Binnners). Seleccione dos atributos y realice binnners de igual amplitud para cada uno de ellos. Realice consultas que describan los datos en cada categoría de datos.

5) Para cada tipo de pingüino realizar:

a) Portada con el nombre del pingüino y los siguientes datos: nombre científico, imagen del pingüino, características, cantidad de elementos de este tipo en el estudio.

b) Realizar consultas, gráficas, correlaciones y tablas que describan las características de tal pingüino.

6) Agregar Conclusiones.

7) Referencias bibliográficas.