

Parcial #1: Penguin Species Classification

Instrucciones:

- Fecha de publicación: 06 de Marzo de 2025
- Fecha de entrega: 13 de Marzo 2025
- Medio de entrega: Sustentación en clase.

Conjunto de datos: Contiene datos sobre los pingüinos del archipiélago Palmer (Antártida).

<https://www.kaggle.com/datasets/parulpandey/palmer-archipelago-antarctica-penguin-data/data>

Ejercicio:

- ° Realizar un análisis exploratorio al conjunto de datos entregado, para ello debe crear diferentes preguntas e hipótesis a resolver en los datos partiendo de un problema que cada grupo debe plantear.
- ° Después de realizar el análisis exploratorio, debe hacer el preprocesamiento de los datos según como considere: limpieza, transformación, reducción de datos o discretización de los datos.
- ° Debe realizar un dashboard en PowerBI o Tableau para exponer visualmente las conclusiones del análisis exploratorio.

Entregables:

- Análisis exploratorio (10 puntos).
- Preprocesamiento de los datos (10 puntos).
- Visualización de datos con PowerBI o Tableau (20 puntos).
- Calidad de las visualizaciones y justificación (10 puntos).

Tema: Penguin Species Classification.

Penguin Species Classification: <https://github.com/dg2c4/Penguin-Species-Classification>

Problema:

¿Cuál es la mejor combinación de características para clasificar a las especies de los pingüinos?.

Dado que se tiene múltiples características como el peso, longitud de pico, profundidad de pico, longitud de aleta, entre otros, se busca identificar qué combinación de variables permite diferenciar mejor a las especies. Para resolver el problema, se debe realizar un análisis exploratorio, un preprocesamiento de datos y un Dashboard en power BI.

Objetivo:

Desarrollar un análisis completo que permita extraer datos sobre los pingüinos del archipiélago Palmer (Antártida), utilizando herramientas de visualización de datos para presentar los resultados de manera efectiva. El trabajo se divide en tres componentes principales:

Análisis Exploratorio:

- Formulación de preguntas de investigación.
- Creación de hipótesis basadas en un problema específico.
- Análisis inicial del conjunto de datos seleccionado.

Fundación Universitaria Konrad Lorenz.
Visualización De Datos Electiva-II. Parcial-1. Penguin Species Classification.
Estudiante: David Gutierrez Chaves Cod: 506222728.

Preprocesamiento de Datos:

- Limpieza de datos.
- Transformación de variables.
- Reducción de datos cuando sea necesario.
- Discretización de datos según corresponda.

Visualización:

- Creación de un dashboard en PowerBI o Tableau.
- Presentación visual de las conclusiones del análisis.
- Exposición clara de los hallazgos del análisis exploratorio.

Systems engineering:

User: David Gutierrez Chaves

Code: 506222728

Subject: Visualización De Datos Electiva-II

Institution: Fundación Universitaria Konrad Lorenz

Institutional email: david.gutierrec@konradlorenz.edu.co

Created By: <https://github.com/dg2c4>