



Unidad de Aprendizaje:

Minería de Datos

Profesor(a):

Dra. Fabiola Ocampo Botello

Práctica 1. Aspectos descriptivos de los pingüinos de Palmer.



Grupo: 6BM2

Autores:

Pérez Núñez Rúben Gabriel

Rodríguez Hernández María José

Sánchez Butrón Eduardo

Fecha de Entrega: 12/03/2025

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}.



Palmer Penguins Dataset for EDA

El conjunto de datos Palmer Penguins proviene del Programa de Investigación Ecológica a Largo Plazo de la Estación Palmer (LTER Palmer Station), un proyecto respaldado por la Fundación Nacional de Ciencias (NSF-OPP) de los Estados Unidos, cuyo propósito es monitorear la ecología de especies antárticas y comprender las variaciones ambientales en la región.

El dataset contiene observaciones de tres especies de pingüinos pertenecientes al género *Pygoscelis*, recolectadas entre los años 2007 y 2009 en distintas islas del Archipiélago Palmer, ubicado al oeste de la Península Antártica. Cada registro representa un individuo identificado con su especie, sexo, año de observación y medidas morfológicas específicas, tales como la longitud y profundidad del pico, la longitud de la aleta y la masa corporal.

Especies de pingüinos:

- Pingüino Adelie

Es una especie característica de la Antártida y sus islas circundantes. Los datos del conjunto incluyen 152 individuos de esta especie, reconocida por su distribución amplia en zonas heladas y su dieta basada principalmente en kril.

- Pingüino Gentoo

Presente en regiones subantárticas y áreas de la península, es la especie más grande del conjunto, con 124 individuos registrados. Su rasgo más notable es la franja blanca que se extiende sobre la cabeza y su mayor masa corporal en comparación con las otras especies.

- Pingüino Chinstrap

Habita en diversas islas subantárticas y la península Antártica. En el conjunto se registran 68 ejemplares, que se distinguen por la delgada franja negra que recorre su mentón, la cual da origen a su nombre común.

Islas de observación:

Las mediciones se realizaron en tres islas principales del Archipiélago Palmer:

- Isla Biscoe:

Una de las islas donde se observaron las tres especies de pingüinos. Su diversidad ecológica la convierte en un punto clave del estudio.

- Isla Dream:

Ubicada al norte del archipiélago, es un sitio donde se registraron principalmente ejemplares de las especies Adelie y Chinstrap.

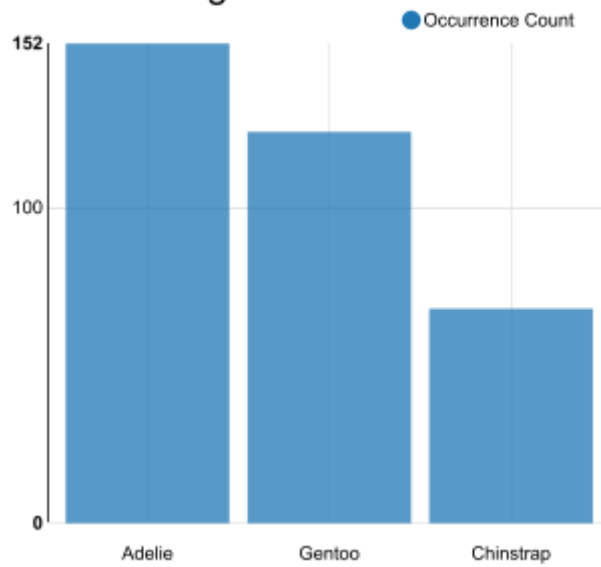
- Isla Torgersen:

También parte del Archipiélago Palmer, es una de las islas de menor tamaño y fue utilizada para observaciones repetidas de individuos Adelie.

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



Gráfica Pingüinos





Descripción del Conjunto de Datos

El conjunto de datos Palmer Penguins está formado por 344 registros y nueve atributos que describen las características morfológicas y contextuales de tres especies distintas de pingüinos observados en la región de Palmer, Antártida.

Cada registro representa un individuo, que se identifican mediante un número único y acompañado de variables tanto categóricas (especie, isla, sexo, año) como numéricas continuas (mediciones morfométricas). Las unidades de medida utilizadas para los valores de longitud y profundidad fueron en milímetros (mm) y la masa corporal fueron en gramos (g).

Row ID	Nombre	Tipo	Descripción	Dominio
Row0	ID	Number (Integer)	Identificador único para cada pingüino.	(0 - 343)
Row1	species	String	Especie de pingüino.	"Adelie", "Chinstrap", "Gentoo"
Row2	island	String	Isla donde se observó el pingüino.	"Biscoe", "Dream", "Torgersen"
Row3	Bill_Length	Number(Double)	Longitud del pico del pingüino.	(mm)
Row4	Bill_Depth	Number(Double)	Profundidad del pico del pingüino.	(mm)
Row5	Flipper_Length	Number(Double)	Longitud de la aleta del pingüino.	(mm)
Row6	Body mass	Number(Double)	Masa corporal del pingüino.	(gr)
Row7	sex	String	Sexo del pingüino.	"Male", "Female"
Row8	year	Number (Integer)	Año en que se realizó la observación.	2007, 2008, 2009.



Análisis y Tratamiento de Datos

Antes de aplicar cualquier técnica de Minería de datos se necesita revisar y preparar la base de datos. Comenzamos revisando la base para detectar valores faltantes asegurándonos de que todos los registros tuvieran información completa; validamos también los tipos de datos, dejando las variables numéricas con formato integer y double y las categóricas como string, para así evitar errores en los cálculos posteriores y que los análisis fueran confiables.

Durante la exploración se identificaron algunos valores faltantes en los atributos sex, bill_length_mm, bill_depth_mm, flipper_length_mm y body_mass_g, 2 en cada atributo numérico y 11 para el atributo categórico sexo

Row ID	Column	No. missings
species	species	0
island	island	0
bill_length_mm	bill_length_mm	2
bill_depth_mm	bill_depth_mm	2
flipper_length_mm	flipper_length_mm	2
body_mass_g	body_mass_g	2
sex	sex	11

El tratamiento de los valores faltantes se realizó de dos formas distintas:

Para las variables numéricas, se decidió eliminar los registros con datos faltantes, ya que imputar estos valores podría generar medidas irreales que no pertenecen a un pingüino real. Esta decisión se apoyó en la idea de Han, Kamber y Pei (2012), quienes mencionan que cuando la cantidad de datos faltantes es reducida, es válido ignorar las tuplas afectadas para mantener la coherencia del conjunto.

Para la variable sexo, en cambio, se aplicó un proceso de imputación por moda, pero calculada dentro de cada especie e Isla. Esto se hizo porque los pingüinos de una misma especie suelen tener proporciones de machos y hembras similares y también puede que el entorno en el que se encuentren repercuta en su desarrollo o alimentación, por lo que reemplazar los valores faltantes con el sexo más frecuente dentro de cada grupo es una forma más lógica y representativa. Esta decisión se basa en Larose & Larose (2015), quienes proponen reemplazar los valores faltantes por valores imputados considerando las características de cada registro y su contexto.

Fuente: Diapositivas realizadas por la Dra. Fabiola Ocampo Botello Microsoft teams canal 2 "Tratamiento de Datos"

Después del tratamiento de los valores faltantes, se revisaron los datos nuevamente mediante el nodo Statistics y para verificar que las variables numéricas ya no tuvieran vacíos, que por lo que podemos apreciar ya no hay ningún valor missing en nuestro conjunto de datos:

Row ID	Column	No. missings
species	species	0
island	island	0
bill_length_mm	bill_length_mm	0
bill_depth_mm	bill_depth_mm	0
flipper_length_mm	flipper_length_mm	0
body_mass_g	body_mass_g	0
sex	sex	0






Descripción de atributos.

Para conocer mejor las características del conjunto de datos se realizó un análisis descriptivo de los principales atributos, tanto numéricos como categóricos; este análisis se apoyó en tablas de frecuencias, tablas cruzadas y gráficas de barras generadas con los nodos Statistics, GroupBy, Bar Chart (JavaScript), Histogram, etc., dentro de KNIME.

Atributos Categóricos

Aquí se analizaron los atributos categóricos con el objetivo de conocer cuántos registros hay de cada tipo y si existe alguna relación entre las especies(species), las islas(island) y el sexo (sex) de los pingüinos.

Row ID	Column	No. missings	Histogram
species	species	0	
island	island	0	
sex	sex	0	

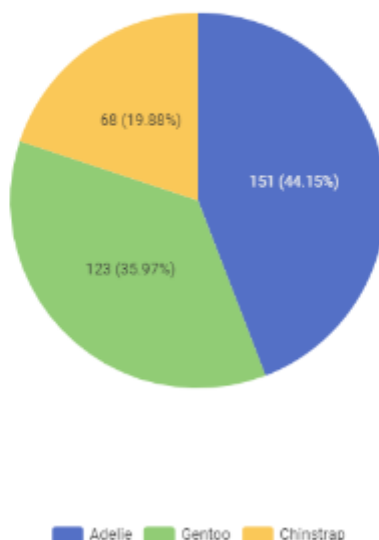
Row ID	species	Count (species)	Relative Frequency (species)	island	Count (island)	Relative Frequency (island)	sex	Count (sex)	Relative Frequency (sex)
Row0	Adelie	151	0.4415204	Biscoe	167	0.4883040	male	173	0.5058479
Row1	Gentoo	123	0.3596491	Dream	124	0.3625730	female	169	0.4941520
Row2	Chinstrap	68	0.1988304	Torgersen	51	0.1491228			

"species"

En el gráfico de pastel se puede ver que la especie Adelie es la que más aparece en el conjunto de datos, con 151 registros, lo que equivale al 44.15% del total; le sigue la especie Gentoo con 123 individuos (35.97%), y por último Chinstrap con 68 (19.88%). Esto nos muestra que la mayoría de las observaciones corresponden a pingüinos Adelie, lo que podría deberse a que es una de las especies más comunes en las islas donde se tomaron los datos.



Ocurrencia de Especie



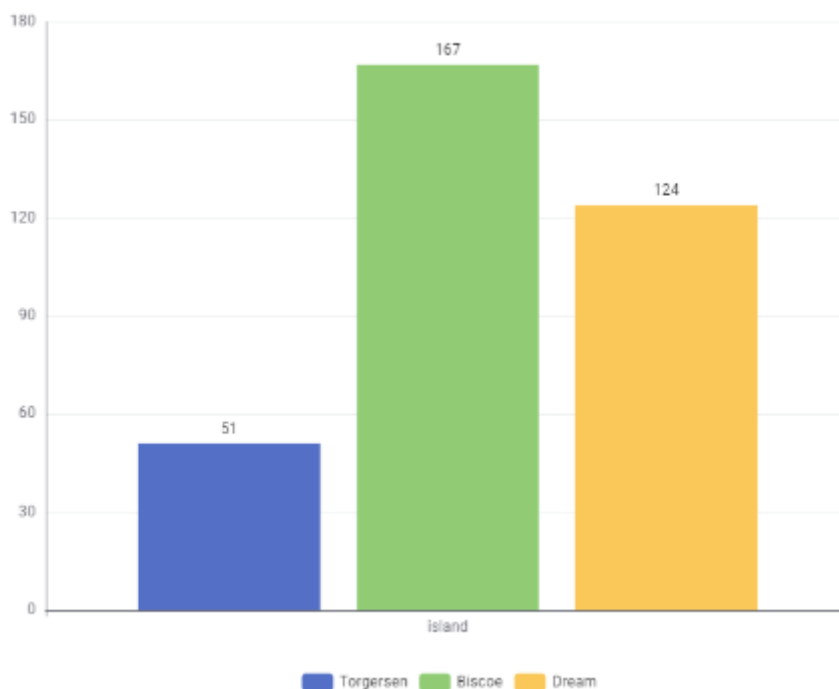
"island"

Se puede ver que la Isla Biscoe tiene el mayor número de registros con 167 observaciones (48.8%), seguida por Dream con 124 (36.3%) y Torgersen con 51 (14.9%); o sea que casi la mitad de los pingüinos estudiados se encuentran en Biscoe, siendo la isla más representada en el dataset.

Cuando se cruza la información de especie e isla los resultados reflejan que cada especie tiende a habitar en áreas distintas; los Adelie son más versátiles encontrándose en las tres islas (Biscoe, Dream y Torgersen), mientras que Chinstrap y Gentoo parecen adaptarse a islas concretas: isla Dream e isla Biscoe respectivamente, lo que se podría atribuir a una relación con factores ambientales o alimenticios de cada lugar.



Islas donde los Pingüinos fueron Observados



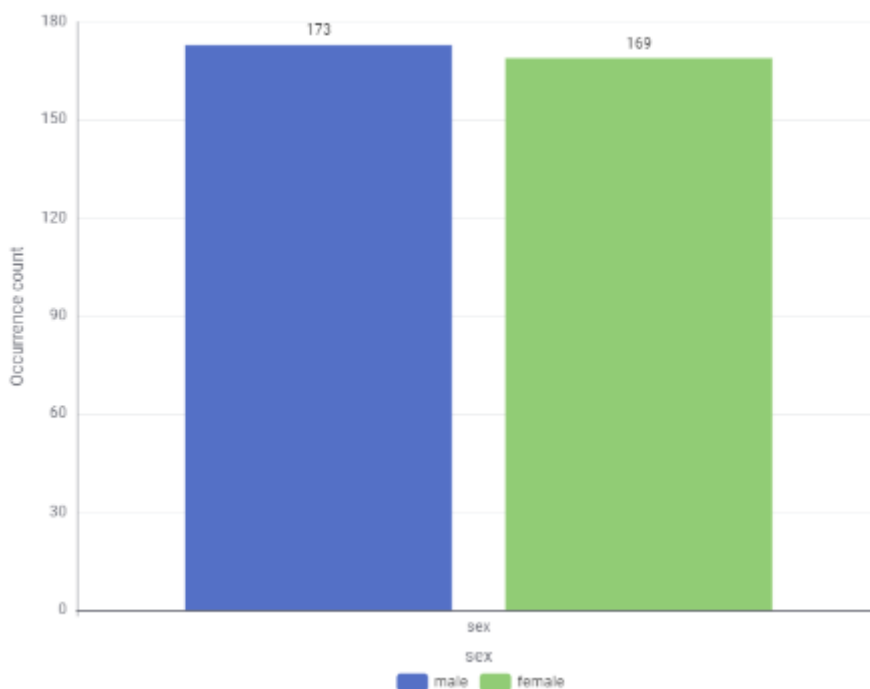
Row ID	species	island	Count*(id)
Row0	Adelie	Biscoe	44
Row1	Adelie	Dream	56
Row2	Adelie	Torgersen	51
Row3	Chinstrap	Dream	68
Row4	Gentoo	Biscoe	123

"sex"

El atributo sex muestra una distribución balanceada hay 173 machos (50.6%) y 169 hembras (49.4%); al cruzar los atributos especie y sexo, también se puede observar el mismo comportamiento de equilibrio entre machos y hembras en todas las especies, lo que nos sugiere que los datos fueron recolectados de forma uniforme y sin sesgo por género.



Ocurrencia de Sexo de los Pingüinos



Row ID	species	sex	Count*(id)
Row0	Adelie	female	77
Row1	Adelie	male	74
Row2	Chinstrap	female	34
Row3	Chinstrap	male	34
Row4	Gentoo	female	58
Row5	Gentoo	male	65

Atributos Numéricos

En esta parte el objetivo fue ver cómo cambian los valores numéricos de longitud del pico (bill_length_mm), profundidad del pico (bill_depth_mm), longitud de la aleta (flipper_length_mm) y masa corporal (body_mass_g) según la especie y entender si hay diferencias notables entre ellas dependiendo de sus diferencias morfológicas.

Row ID	Std. deviation	Variance	Skewness	Kurtosis	Histogram
id	99.18	9835.68	0.01	-1.2	





Row ID	Std. deviation	Variance	Skewness	Kurtosis	Histogram
bill_length_mm	5.46	29.81	0.06	-0.88	
bill_depth_mm	1.98	3.9	-0.15	-0.91	
flipper_length_mm	14.07	197.74	0.35	-0.99	
body_mass_g	801.96	643131.08	0.48	-0.72	
year	0.82	0.67	-0.06	-1.51	

"bill_length_mm"

Los valores de la longitud del pico en general van desde 32.1 hasta 59.6 mm; donde los Adelie tiene los picos más cortos con 38.79 mm, mientras que Chinstrap y Gentoo son más grandes, con promedios de 48.83 mm y 47.50 mm respectivamente; esto nos muestra que los pingüinos Gentoo y Chinstrap suelen tener picos más largos que los Adelie, probablemente por diferencias en su alimentación o hábitat. En el histograma se puede observar una distribución bastante uniforme.

Row ID	species	Mean(bill_length_mm)	Min*(bill_length_mm)	Max*(bill_length_mm)
Row0	Adelie	38.79139072847681	32.1	46
Row1	Chinstrap	48.83382352941176	40.9	58
Row2	Gentoo	47.50487804878049	40.9	59.6

"bill_depth_mm"

La profundidad del pico tiene valores de entre 13.1 y 21.5 mm. Aquí el patrón cambia un poco porque los Adelie y Chinstrap tienen picos más profundos, con medias de 18.34 mm y 18.42 mm, mientras que los Gentoo tienen picos más delgados, con 14.98 mm de promedio, o sea que aunque los Gentoo tienen picos



largos, no son tan robustos como los de las otras especies. También en este atributo en el histograma se observa una distribución uniforme.

Row ID	species	Mean(bill_depth_mm)	Min*(bill_depth_mm)	Max*(bill_depth_mm)
Row0	Adelie	18.346357615894046	15.5	21.5
Row1	Chinstrap	18.42058823529412	16.4	20.8
Row2	Gentoo	14.982113821138212	13.1	17.3

"flipper_length_mm"

Los datos de la aleta varían entre los 172 y 231 mm, y aquí sí hay una diferencia más clara; los Gentoo tienen las aletas más largas (217.18 mm), luego los Chinstrap (195.82 mm) y al final los Adelie (189.95 mm); esto se puede relacionar con el tamaño y la habilidad para nadar, ya que los Gentoo son conocidos por ser los más rápidos en el agua. El histograma muestra una forma cercana a una distribución normal, sin grandes concentraciones en un solo valor.

Row ID	species	Mean(flipper_length_mm)	Min*(flipper_length_mm)	Max*(flipper_length_mm)
Row0	Adelie	189.95364238410593	172	210
Row1	Chinstrap	195.82352941176472	178	212
Row2	Gentoo	217.18699186991873	203	231

"body_mass_g"

La masa corporal es la variable más variable del conjunto con valores que van desde 2700 g hasta 6300 g; los Gentoo también destacan aquí con un peso promedio de 5076 g, mucho más alto que el de los Adelie (3700 g) y los Chinstrap (3733 g). Tomando en cuenta lo que hemos visto con los demás atributos el peso solo termina de confirmar que los Gentoo son la especie más grande en general. En el histograma se ve una distribución algo simétrica, aunque con una pequeña inclinación hacia la derecha por algunos pingüinos más pesados.

Row ID	species	Mean(body_mass_g)	Min*(body_mass_g)	Max*(body_mass_g)
Row0	Adelie	3700.662251655629	2850	4775
Row1	Chinstrap	3733.088235294118	2700	4800
Row2	Gentoo	5076.016260162601	3950	6300

En resumen todas las variables numéricas muestran diferencias claras entre especies; los Gentoo tienden a tener los valores más altos en casi todas las mediciones, mientras que los Adelie son más pequeños en comparación. También se puede llegar a concluir que las especies más grandes de Chinstrap o Adelie pueden llegar a parecer con los más pequeños de Gentoo, lo que no cambia es el hecho que los Gentoo tienen el pico menos profundo, por lo que eso delataría si realmente es un Gentoo o no.

Si nos fijamos en los valores de skewness y kurtosis, se nota que las distribuciones no están muy sesgadas, por lo que los datos parecen bastante limpios y sin valores atípicos, o sea que esto indica que el conjunto es confiable y se puede seguir con análisis más avanzados.



Consultas

Promedio de masa corporal por especie y su sexo

En esta consulta se calculó el promedio de la masa corporal (body_mass_g) de cada especie según su sexo. Donde se observa que en la especie Adelie, las hembras tienen un peso promedio de 3,385 g, mientras que los machos alcanzan alrededor de 4,029 g; con los Chinstrap, las hembras pesan 3,527 g y los machos 3,938 g en promedio; en los Gentoo, la diferencia es más grande, las hembras pesan 4,679 g, mientras que los machos llegan hasta 5,429 g en promedio.

Todos los resultados nos muestra que en todas las especies los machos pesan más que las hembras, lo que ya nos esperabamos por la naturaleza de sus diferencias biológicas.

Row ID	species	sex	Mean(body_mass_g)
Row0	Adelie	female	3385.064935064935
Row1	Adelie	male	4029.0540540540537
Row2	Chinstrap	female	3527.205882352941
Row3	Chinstrap	male	3938.970588235294
Row4	Gentoo	female	4679.741379310347
Row5	Gentoo	male	5429.615384615385

Promedio de longitud y profundidad del pico por especie e isla

En esta consulta se analizaron la longitud y la profundidad del pico promedio de los pingüinos según la isla donde habitan, con la idea de ver si el tipo de entorno o la alimentación disponible influye en la forma del pico.

En las tres islas donde vive el pingüino Adelie (Biscoe, Dream y Torgersen), las medidas del pico son muy parecidas: la longitud ronda los 38.5–38.9 mm y la profundidad entre 18.2–18.4 mm.

Esto sugiere que, para esta especie, el tipo de isla no cambia mucho la morfología del pico, probablemente porque tienen una alimentación similar (básicamente kril y pequeños crustáceos) sin importar la ubicación.

Los Chinstrap, que solo están en la isla Dream, tienen picos más largos (48.8 mm) y con una profundidad de 18.4 mm. Este tipo de pico más alargado podría ser una adaptación para capturar presas más móviles o pequeñas, como peces o kril disperso.

Por otro lado, los Gentoo, que habitan únicamente en Biscoe, tienen un pico también largo (47.5 mm) pero mucho más plano (14.9 mm de profundidad). Esto encaja con lo que se sabe de su alimentación, ya que suelen bucear más profundo y cazar peces, por lo que un pico más fino y largo les resulta más útil.

En general, se puede decir que sí existe una relación entre la isla y las dimensiones del pico, aunque esta relación parece más influenciada por la especie dominante en cada isla que por el ambiente físico en sí.

Row ID	species	island	Mean(bill_length_mm)	Mean(bill_depth_mm)
Row0	Adelie	Biscoe	38.975	18.37045454545455
Row1	Adelie	Dream	38.5017857142857	18.25178571428571
Row2	Adelie	Torgersen	38.95098039215686	18.429411764705883
Row3	Chinstrap	Dream	48.83382352941176	18.42058823529412
Row4	Gentoo	Biscoe	47.50487804878049	14.982113821138212



¿Cuál es la especie con mayor longitud de aleta promedio?

Por último, se calculó la longitud promedio de la aleta (flipper_length_mm) para cada especie. Los resultados fueron que en promedio los Adelie tienen una medida de aleta de 189.95; las de los Chinstrap 195.82 mm y las de los Gentoo median 217.18 mm. Lo que se traduce a que la especie Gentoo tiene la aleta más larga de todas, seguida por Chinstrap y finalmente Adelie; este resultado tiene sentido porque los Gentoo también fueron los más pesados, y una mayor longitud de aleta les permite desplazarse adecuadamente en el agua.

Row ID	species	Mean(flipper_length_mm)
Row0	Adelie	189.95364238410593
Row1	Chinstrap	195.82352941176472
Row2	Gentoo	217.18699186991873



Correlaciones

Row ID	First column name	Second column name	Correlation value	p value	Degrees of freedom
Row0	bill_length_mm	bill_depth_mm	-0.235052870355	1.119662196135E	340
Row1	bill_length_mm	flipper_length_mm	0.656181340746	0	340
Row2	bill_length_mm	body_mass_g	0.595109824437	0	340
Row3	bill_depth_mm	flipper_length_mm	-0.583851216465	1.2327337668077	340
Row4	bill_depth_mm	body_mass_g	-0.471915621186	2.2769412725334	340
Row5	flipper_length_mm	body_mass_g	0.871201767306	0	340

En esta parte se observa qué tanto se relacionan las variables numéricas entre sí; al revisar los resultados de correlación se puede ver que hay relaciones tanto positivas como negativas, lo que nos indica que algunas variables crecen juntas y otras van al revés.

Por ejemplo, entre `bill_length_mm` y `bill_depth_mm` la correlación fue -0.23, lo que nos quiere decir que cuando el pico es más largo normalmente es menos profundo. Esto tiene sentido porque los Gentoo tienen el pico más alargado y delgado y los Adelie lo tienen más corto pero más ancho.

Entre `bill_length_mm` y `flipper_length_mm` la correlación es 0.65, y con `body_mass_g` es de 0.59, o sea, que a mayor longitud del pico también aumentan la aleta y el peso del pingüino. Esto muestra que las especies más grandes (como los Gentoo) tienden a tener picos más largos, aletas más largas y también pesan más.

Por otro lado, `bill_depth_mm` tiene correlación negativa con `flipper_length_mm` (-0.58) y con `body_mass_g` (-0.47), lo que significa que los que tienen picos más profundos tienden a ser más pequeños y con aletas más cortas, como los Adelie.

La correlación más alta fue entre `flipper_length_mm` y `body_mass_g` (0.87), lo que indica que los pingüinos con aletas más largas también son los que más pesan; o sea que el tamaño corporal está muy relacionado con la longitud de la aleta, lo cual es lógico porque los pingüinos grandes necesitan aletas más largas para poder nadar mejor.

En general los datos demuestran que las características físicas del pingüino están bastante conectadas, los más grandes tienen picos largos, más masa y aletas más extensas (Gentoo), mientras que los más chicos tienen picos más profundos y cuerpos más ligeros (Adelie). Todo esto concuerda con las diferencias entre las especies y cómo cada una se adapta al tipo de alimento que tiene disponible en su isla.



Agrupamiento por K-Means

Para esta parte se usó la técnica de K-means con el objetivo de agrupar a los pingüinos según sus características físicas, principalmente la combinación de los atributos que tuvieron correlaciones más destacables en el punto anterior, la longitud y profundidad del pico, la longitud de la aleta y la masa corporal. Antes de aplicar el modelo, se normalizaron los datos con Min-Max porque las medidas estaban en diferentes escalas (milímetros y gramos) y además los datos no tenían una distribución normal, ya que algunas variables estaban un poco sesgadas hacia un lado (por ejemplo, la masa corporal y la longitud de la aleta estaban inclinadas hacia la derecha y el profundo de pico hacia la izquierda) y según Larose & Larose (2015) esta es una de las razones por las que no se puede usar la técnica de z-score.

Fuente: Diapositivas realizadas por la Dra. Fabiola Ocampo Botello Microsoft teams canal 2 "Tratamiento de Datos"

El valor de K=3 se estableció porque ya se conoce que existen tres especies distintas en el conjunto de datos (Adelie, Chinstrap y Gentoo). El objetivo era comprobar si el algoritmo podía reconocer y agrupar correctamente a las especies basándose solo en las variables numéricas, sin utilizar la etiqueta de especie.

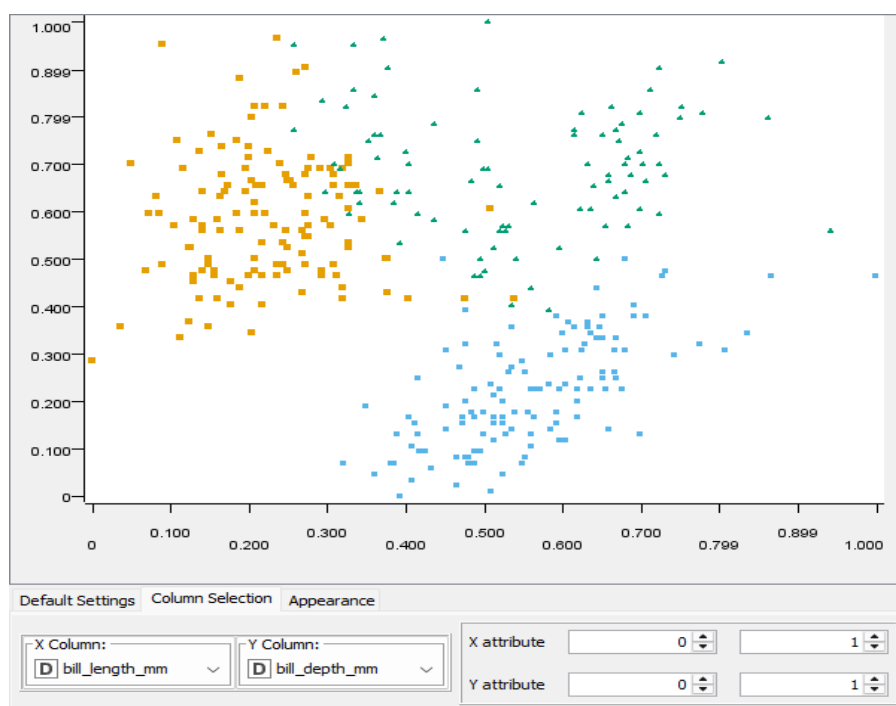
Row ID	Mean Silhouette Coefficient
cluster_0	0.40166927929653307
cluster_2	0.28300050666553855
cluster_1	0.5824990290924982
Overall	0.4351289310961915

Una vez ejecutado el modelo, el índice de silueta promedio fue de 0.435, lo que indica que la separación de los grupos es moderadamente buena. En detalle, el cluster 1 tuvo el mejor agrupamiento con 0.582, el cluster 0 alcanzó 0.402, y el cluster 2 fue el más difuso con 0.283, lo que sugiere que algunos ejemplares de ese grupo se parecen a los de otros clusters.

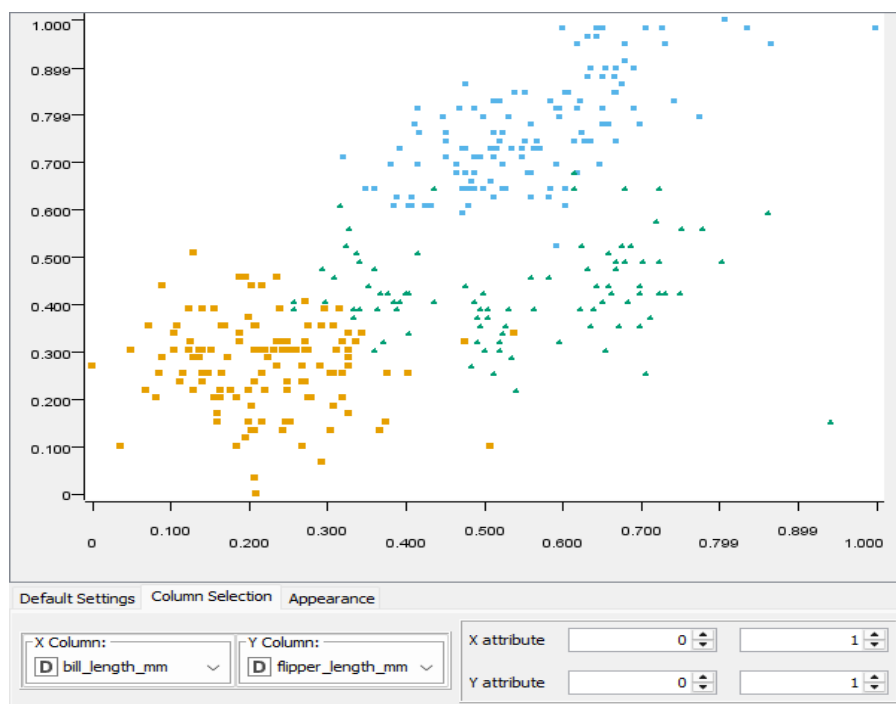
Row ID	Cluster	bill_length_ (Mean)	bill_depth_ (Mean)	flipper_len (Mean)	body_mass (Mean)	species (Unique concatenated with count*)	species (Count*)
Row0	cluster_0	0.224602272	0.584356398	0.269994703	0.233886718	Adelie(120), Chinstrap(8)	128
Row1	cluster_1	0.560177383	0.224061169	0.765881218	0.660004516	Gentoo(123)	123
Row2	cluster_2	0.542537462	0.687728937	0.426895138	0.346688034	Adelie(31), Chinstrap(60)	91

En la tabla de descripción de ejemplares, se observa que:

- El Cluster 1 agrupa principalmente a los Gentoo (123 ejemplares), que se caracterizan por tener el mayor promedio en masa corporal y longitud de aleta, confirmando que son los pingüinos más grandes.
- El Cluster 0 contiene en su mayoría Adelie (120) y algunos Chinstrap (8), con los valores más bajos en casi todas las variables, representando así a los pingüinos más pequeños.
- El Cluster 2 agrupa Adelie (31) y Chinstrap (60) con valores intermedios, lo que muestra que estas dos especies tienen características morfológicas muy parecidas, por eso su separación fue menos clara.



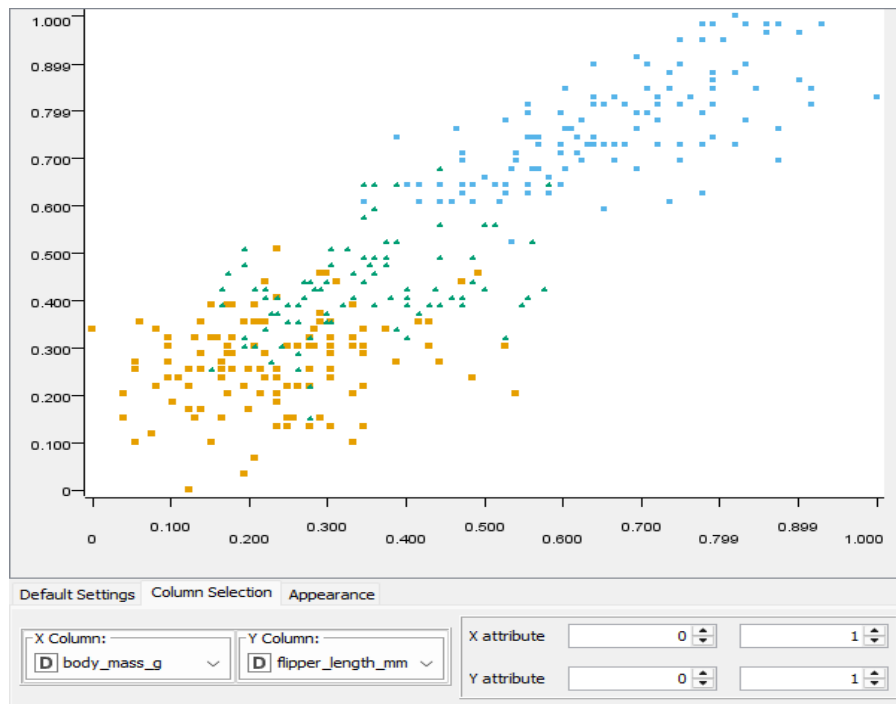
En esta primera gráfica se muestra la relación entre la longitud y la profundidad del pico, se nota una tendencia inversa que en el punto anterior de correlación vimos que conforme el pico es más largo, tiende a ser más delgado. Aquí se puede distinguir que los pingüinos Gentoo, representados en la parte inferior derecha, tienen picos largos pero poco profundos, mientras que los Adelie aparecen en la parte superior izquierda, con picos cortos pero más robustos. Los Chinstrap se ubican en medio, mezclándose un poco con ambos, pero aún se puede identificar su grupo; lo que hace sentido porque si pensamos en que si forma del pico cambia entre especies es por el tipo de alimento que consumen, ya que cada una está adaptada a su dieta.



Aquí se compara la longitud del pico con la longitud de la aleta, y visualmente sí se pueden distinguir los grupos. Los Gentoo siguen apareciendo en la parte con valores más altos en ambas variables (picos y aletas).



largas); los Adelie tienen picos y aletas más pequeñas, y los Chinstrap otra vez están justo en medio pero hay un poco de traslape entre ellos.



Aquí cuando el pingüino pesa más, también suele tener las aletas más largas según las correlaciones que se analizaron anteriormente, y a pesar de que fue de las que su valor era más alto visualmente en la nube de puntos se ve que los Chinstrap quedan sumamente empalmados con los Adelie, por lo que deben de ser otras variables que se tomaron en cuenta para estos nuevos agrupamientos; Gentoo sigue ocupando la parte más alta del gráfico (porque son los más grandes).

En general las tres gráficas nos permiten ver que las especies se diferencian bastante bien por su tamaño y proporciones corporales; los Gentoo son los de mayor tamaño, los Adelie los más pequeños y los Chinstrap los intermedios.



Contenedores Binnners

En este apartado se hizo la distribución de los datos usando contenedores o binnners, con la idea de ver en qué rangos se concentran más los pingüinos según el largo de sus aletas y su masa corporal. Esto ayuda a entender qué tan variados son los tamaños y si hay grupos más comunes que otros.

Binner del atributo Longitud de aleta

Row ID	count
(183.8,195.6]	131
(195.6,207.4]	59
(207.4,219.2]	84
(219.2,231]	43
[172,183.8]	25

Row ID	species	flipper_length_mm [Binned]	Count*(id)
Row0	Adelie	(183.8,195.6]	101
Row1	Adelie	(195.6,207.4]	26
Row2	Adelie	(207.4,219.2]	2
Row3	Adelie	[172,183.8]	22
Row4	Chinstrap	(183.8,195.6]	30
Row5	Chinstrap	(195.6,207.4]	31
Row6	Chinstrap	(207.4,219.2]	4
Row7	Chinstrap	[172,183.8]	3
Row8	Gentoo	(195.6,207.4]	2
Row9	Gentoo	(207.4,219.2]	78
Row10	Gentoo	(219.2,231]	43

Cuando se ve la distribución general del largo de las aletas se nota que casi todos los pingüinos se agrupan en los valores de en medio, entre 183.8 y 195.6 mm, o sea que la mayoría tienen aletas medianas, y conforme aumenta el tamaño va habiendo menos individuos. Los que tienen las aletas más largas, entre 219.2 y 231 mm, son pocos; en general los datos se concentran en el centro, lo que nos dice que no hay muchos pingüinos ni muy chicos ni muy grandes.

Pero cuando se ve esa misma variable separada por especie, las cosas cambian, ahí si se nota claramente que los Adelie son los que tienen las aletas más cortas, porque casi todos se concentran en los primeros rangos; los Chinstrap están un poco más repartidos, más o menos en los valores intermedios, y los Gentoo son los que tienen las aletas más largas. En esta parte se entiende que la forma de la distribución general está muy influenciada por la cantidad de Adelie en el conjunto, pero en realidad cada especie tiene su propio rango. Los Gentoo, por ejemplo, sobresalen mucho en la parte alta porque son más grandes, y eso se refleja en sus aletas.

Binner con el atributo Masa Corporal

Row ID	count
(3,420,4,140]	124
(4,140,4,860]	83
(4,860,5,580]	55
(5,580,6,300]	22



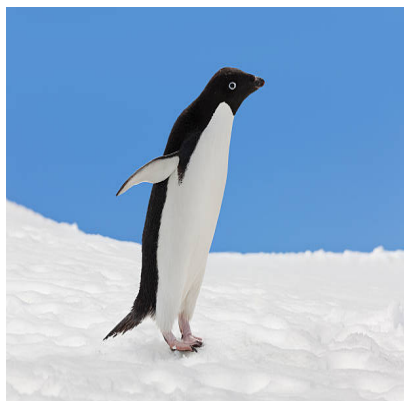
Row ID	count
[2,700,3,420]	58

Row ID	species	sex	body_mass_g [Binned]	Count*(id)
Row0	Adelie	female	(3,420,4,140]	35
Row1	Adelie	female	(4,140,4,860]	1
Row2	Adelie	female	[2,700,3,420]	41
Row3	Adelie	male	(3,420,4,140]	43
Row4	Adelie	male	(4,140,4,860]	29
Row5	Adelie	male	[2,700,3,420]	2
Row6	Chinstrap	female	(3,420,4,140]	21
Row7	Chinstrap	female	(4,140,4,860]	1
Row8	Chinstrap	female	[2,700,3,420]	12
Row9	Chinstrap	male	(3,420,4,140]	23
Row10	Chinstrap	male	(4,140,4,860]	8
Row11	Chinstrap	male	[2,700,3,420]	3
Row12	Gentoo	female	(3,420,4,140]	1
Row13	Gentoo	female	(4,140,4,860]	41
Row14	Gentoo	female	(4,860,5,580]	16
Row15	Gentoo	male	(3,420,4,140]	1
Row16	Gentoo	male	(4,140,4,860]	3
Row17	Gentoo	male	(4,860,5,580]	39
Row18	Gentoo	male	(5,580,6,300]	22

con la masa corporal pasa algo parecido, su distribución general la mayoría de los pingüinos pesan entre 3420 y 4140 gramos, o sea, más o menos entre tres y cuatro kilos; conforme sube el peso, hay menos individuos, y los más pesados, arriba de los 5500 gramos, son muy muy pocos. A simple vista parece que casi todos pesan lo mismo, pero eso cambia cuando se separan por especie y sexo.

En la comparación por especie y sexo se nota mucho que los Gentoo son los más pesados, con muchos individuos entre los 4860 y 6300 gramos; los Adelie se concentran más abajo, entre 3420 y 4140 gramos, y los Chinstrap están a la mitad, con pesos intermedios. También se puede ver que los machos casi siempre pesan más que las hembras, sobre todo en los Gentoo, donde la diferencia es muy evidente al ver que la mayoría de hembras están en más cantidad en los pesos intermedios y los machos se encuentran más en los intervalos de mayor peso; y con en los Adelie también hay diferencia, pero no tan marcada.

Pingüino Adelie(Adelia)



Nombre científico:
Pygoscelis adeliae

Características:

- Son pingüinos de tamaño medio-pequeño para la Antártida, con plumaje negro en la cabeza y la espalda, blanco en el frontal, y un anillo blanco alrededor del ojo que los distingue.

- Se alimentan principalmente de kril, peces y calamares, y viven en colonias densas en costas rocosas sin hielo.

Cantidad en el estudio:

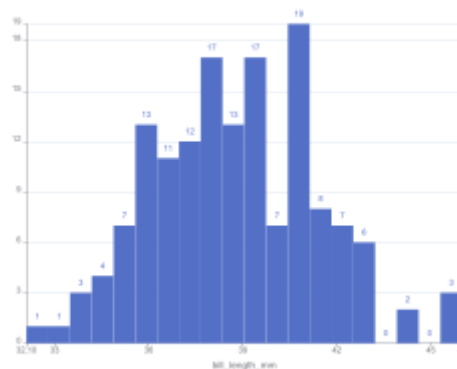
152 individuos de esta especie según tu conteo.

Longitud de Pico

Los valores de la longitud del pico en los pingüinos Adelia varían entre aproximadamente 32 mm y 46 mm.

La mayoría de los individuos tienen una longitud del pico en el rango de 37 mm a 42 mm, lo que sugiere que esta especie posee picos relativamente cortos en comparación con las otras.

Largo de Pico de los Adelia

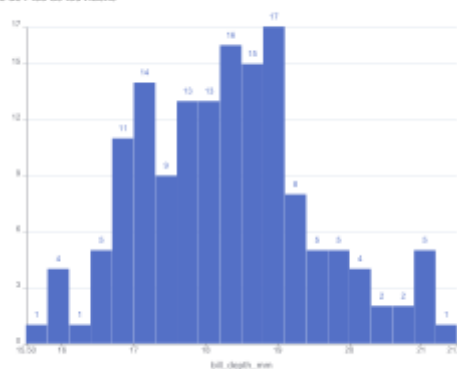


Ancho de Pico

Los datos muestran que la profundidad del pico de los pingüinos Adelia oscila entre 15 mm y 21 mm.

La mayor parte de los individuos tiene un pico con una profundidad promedio cercana a 18 mm, lo que indica que esta es una característica estable dentro de la especie.

Ancho de Pico de los Adelia

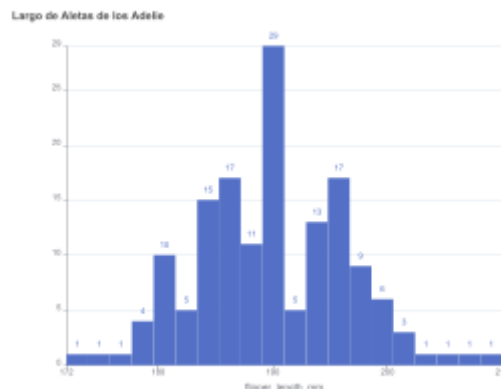




Largo de Aleta

La longitud de la aleta en los pingüinos Adelia se encuentra entre 172 mm y 210 mm.

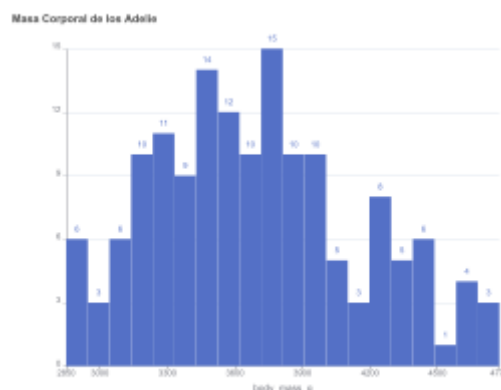
La mayoría de los individuos tiene aletas con una longitud de 180 mm a 195 mm, mostrando una distribución concentrada en ese rango



Masa Corporal

La masa corporal de los pingüinos Adelia se encuentra en un rango entre 2,700 g y 4,200 g.

La distribución de pesos muestra que la mayoría de los individuos pesa alrededor de 3,500 g a 3,800 g



Row ID	sex	Mean(bill_length)	Mean(bill_depth)	Mean(flipper_length)	Mean(body_mass)
Row0	female	37.3	17.66	187.77	3385.07
Row1	male	40.36	19.08	192.23	4029.06

Row ID	island	Mean(bill_length)	Mean(bill_depth)	Mean(flipper_length)	Mean(body_mass)
Row0	Biscoe	38.98	18.38	188.8	3709.66
Row1	Dream	38.51	18.26	189.74	3688.4
Row2	Torgersen	38.96	18.43	191.2	3706.38

En general los machos tienen un largo de pico de 40.36 mm, ancho de 19.08 mm, aletas de 192.23 mm y una masa corporal de 4029.06 g, mientras que las hembras tienen valores menores 37.3 mm de largo de pico, 17.66 mm de ancho, 187.77 mm de aletas y 3385.07 g de masa; esto confirma que los machos son físicamente más grandes, lo cual es algo normal en varias especies de pingüinos.

Cuando se comparan los Adelia por isla, las diferencias no son tan grandes, pero se nota que los de la isla Torgersen tienden a tener las aletas y los picos un poco más largos, con promedios de 191.2 mm en aletas y 38.96 mm de largo de pico. En cambio, los de Dream y Biscoe son apenas más pequeños, aunque el peso promedio es muy parecido entre las tres islas (alrededor de 3700 g).

Pingüino Chinstrap(Barbijo)



Nombre científico:
Pygoscelis antarcticus

Características:

- Esta especie se caracteriza por tener una fina banda negra que va de oreja a oreja pasando bajo el mentón ("barbijo"), lo que le da el nombre común.

- Su peso adulto está en el rango de aproximadamente 3.2 a 5.5 kg, y se distribuyen en islas sub-antárticas y península antártica.

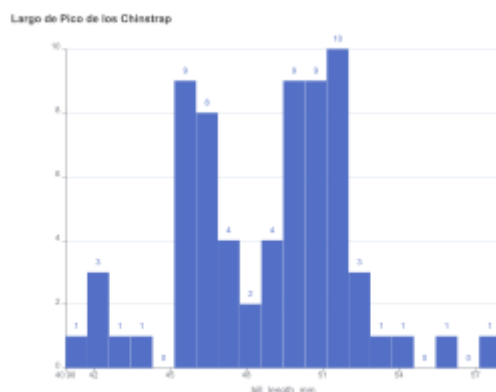
Cantidad en el estudio:

68 individuos de esta especie según tu conteo.

Longitud de Pico

La longitud del pico de los pingüinos Barbijo se encuentra entre 38 mm y 51 mm.

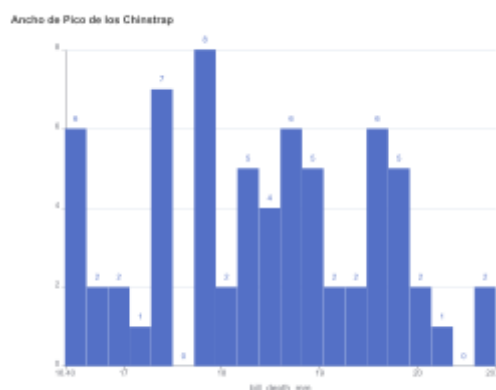
A diferencia de los Adelia, los pingüinos Barbijo tienen picos más largos en promedio, con la mayoría de los individuos en el rango de 42 mm a 47 mm



Ancho de Pico

La profundidad del pico de esta especie varía entre 16 mm y 21 mm.

Se observa que la mayoría de los pingüinos Barbijo tienen picos con una profundidad media de 18 mm a 19 mm, lo cual es similar a los pingüinos Adelia, pero con mayor dispersión en los valores.

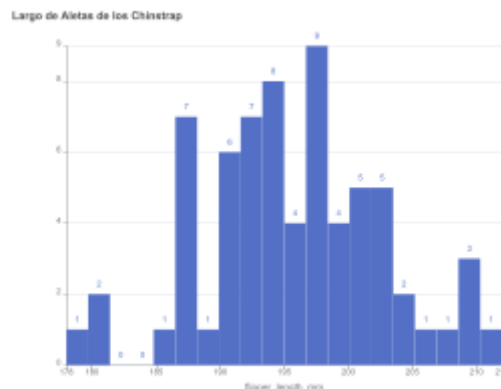




Largo de Aleta

En cuanto a la longitud de la aleta, los pingüinos Barbijo tienen una medida promedio de 185 mm a 205 mm, con pocos individuos que superan los 210 mm.

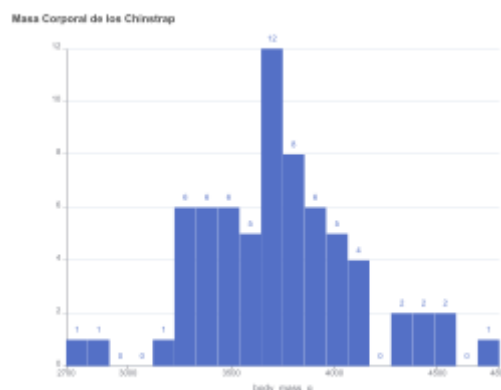
Esto sugiere que, en términos de aletas, los pingüinos Barbijo tienen medidas más similares a los Adelia que a los Gentoo



Masa Corporal

Los pingüinos Barbijo tienen masas corporales que oscilan entre 3,000 g y 4,600 g.

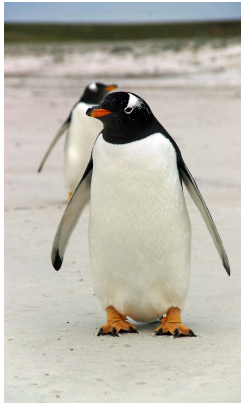
En comparación con los Adelia, estos pingüinos suelen ser un poco más pesados, con una media de 4,000 g, lo que indica que, aunque son de tamaño similar, pueden tener una mayor proporción de masa muscular o grasa.



Row ID	sex	Mean(bill_length)	Mean(bill_depth)	Mean(flipper_length)	Mean(body_mass)
Row0	female	46.58	17.59	191.74	3527.21
Row1	male	51.1	19.26	199.92	3938.98

En general, igual que en las otras especies, los machos Chinstrap son más grandes y pesados, mostrando un dimorfismo sexual evidente, aunque las diferencias no son tan grandes como en los Gentoo, pero sí lo suficiente para distinguirlos fácilmente por tamaño.

Pingüino Gentoo(Papúa)



Nombre científico:
Pygoscelis papua

Características:

- Los Gentoo son los más grandes de los tres estudiados. Tienen un pico de color rojo-anaranjado, pies color durazno, y una distintiva banda blanca que cruza su frente.
- Son también conocidos por nadar rápido y bucear profundo para alimentarse de kril, peces y calamares.

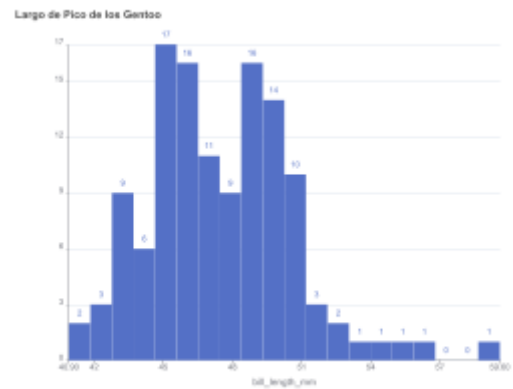
Cantidad en el estudio:

124 individuos de esta especie según tu conteo.

Longitud de Pico

Los pingüinos Papúa tienen los picos más largos entre las tres especies, con valores que oscilan entre 40 mm y 60 mm.

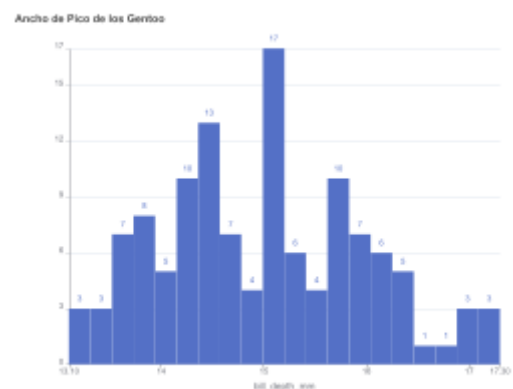
Se observa que la mayoría de los individuos tiene una longitud de pico en el rango de 45 mm a 55 mm, lo que los diferencia claramente de las otras dos especies.



Ancho de Pico

La profundidad del pico en los pingüinos Papúa es más variable, con valores entre 13 mm y 17 mm.

Se observa que estos pingüinos tienen picos más alargados pero menos profundos en comparación con los Adelia y Barbijo.

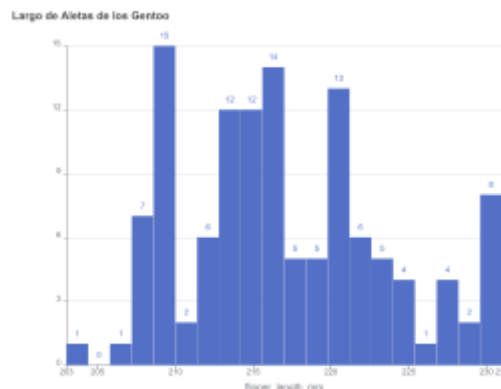




Largo de Aleta

La longitud de la aleta en los pingüinos Papúa se encuentra en un rango de 200 mm a 230 mm, siendo considerablemente mayor que en las otras dos especies.

La mayoría de los individuos presenta aletas con una longitud de 210 mm a 220 mm, lo que sugiere que esta característica es clave en su diferenciación morfológica.



Masa Corporal

La masa corporal de los pingüinos Papúa varía ampliamente entre 4,000 g y 6,000 g, lo que los convierte en la especie más grande de las tres.

La mayoría de los individuos pesa alrededor de 4,500 g a 5,300 g, con algunos casos extremos que superan los 5,800 g.



Row ID	sex	Mean(bill_length)	Mean(bill_depth)	Mean(flipper_length)	Mean(body_mass)
Row0	female	45.57	14.24	212.71	4679.75
Row1	male	49.24	15.65	221.19	5429.62

Las diferencias entre machos y hembras en los pingüinos Gentoo son bastante claras, los machos son más grandes en todas las medidas, tienen el pico más largo y ancho, las aletas más largas y también pesan más.

Y en general, el análisis muestra que los Gentoo tienen cuerpos más grandes, picos más desarrollados y aletas más largas que las demás especies, la única característica en donde es más pequeño es en el ancho del pico.