



Proyecto 2. Data Science:
Análisis Exploratorio

Kristen Brandt 171482
Oliver Graf 17190
Estuardo Ureta 17010

Situación Problemática:

La evaluación de la edad ósea se realiza con frecuencia en pacientes pediátricos para evaluar el crecimiento y diagnosticar y controlar una multitud de trastornos endocrinos y síndromes pediátricos. Durante décadas, la determinación de la madurez ósea se ha basado en una evaluación visual del desarrollo esquelético de la mano y muñeca. Actualmente, se cuenta con herramientas digitales como, por ejemplo, BoneXpert, que permiten eliminar el error humano y son más eficientes. Dichas herramientas utilizan radiografías de la mano del paciente para hacer el diagnóstico.

Nuestro grupo cuenta con una base de datos de aproximadamente 9 gigabytes de radiografías de manos y se busca poner a prueba diferentes algoritmos predictivos, para ver si se pueden utilizar en el campo de la misma manera que BoneXpert.

Problema científico:

Con las tecnologías de imagenología digital, se busca realizar y se han realizado múltiples intentos para desarrollar técnicas de procesamiento de imágenes que extraigan automáticamente las características morfológicas clave de la osificación en los huesos para proporcionar un resultado más efectivo y objetivo, con el fin de facilitar la evaluación de la madurez ósea. El desarrollo de este tipo de aplicaciones para diagnóstico médico, que utilizan imágenes digitales para proporcionar diagnósticos, es de suma importancia para aquellas regiones que cuentan con escasos recursos, porque utilizando la tecnología de los celulares inteligentes se puede llevar estas soluciones hasta las personas más remotas y necesitadas, fácil y económicamente. El reto está en elevar lo más posible la fidelidad de estos algoritmos, no con el objetivo de reemplazar el criterio de los médicos profesionales, sino complementarlo y volverlo más eficiente y accesible.

Objetivos:

General:

1. Generar un análisis exploratorio que ayude a determinar de forma automatizada la edad ósea de un infante utilizando la radiografía de las manos.

Específicos:

1. Crear 8 subgrupos de edad y mostrar un ejemplo de ambos géneros para cada categoría para tener una idea de lo que es importante y también la diversidad en los datos.
2. Hallar los elementos clave que lo pueden llevar a comprender lo que está causando el problema.

Descripción de los datos:

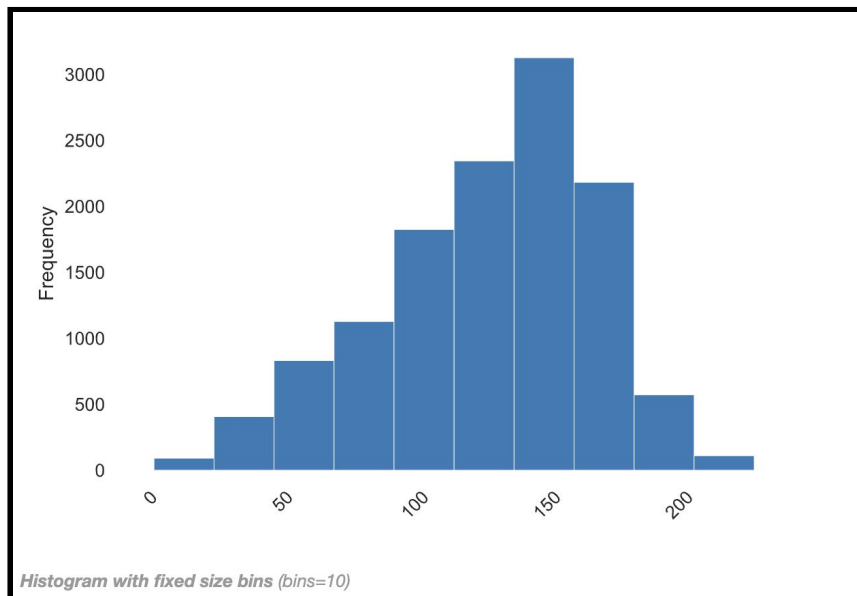
Number of variables	3	NUM	2
Number of observations	12611	BOOL	1
Missing cells	0		
Missing cells (%)	0.0%		
Duplicate rows	0		
Duplicate rows (%)	0.0%		
Total size in memory	209.5 KiB		
Average record size in memory	17.0 B		

Overview de los datos

Hay tres variables en los datos id, boneage y male.

- id: es el identificador único, este es el que nos permite juntarlo con la imagen, es una variable numérica
- boneage: edad osea calculada para el paciente, este esta en meses, es una variable numérica
- male: es un booleano, nos dice si el paciente es mujer o hombre.

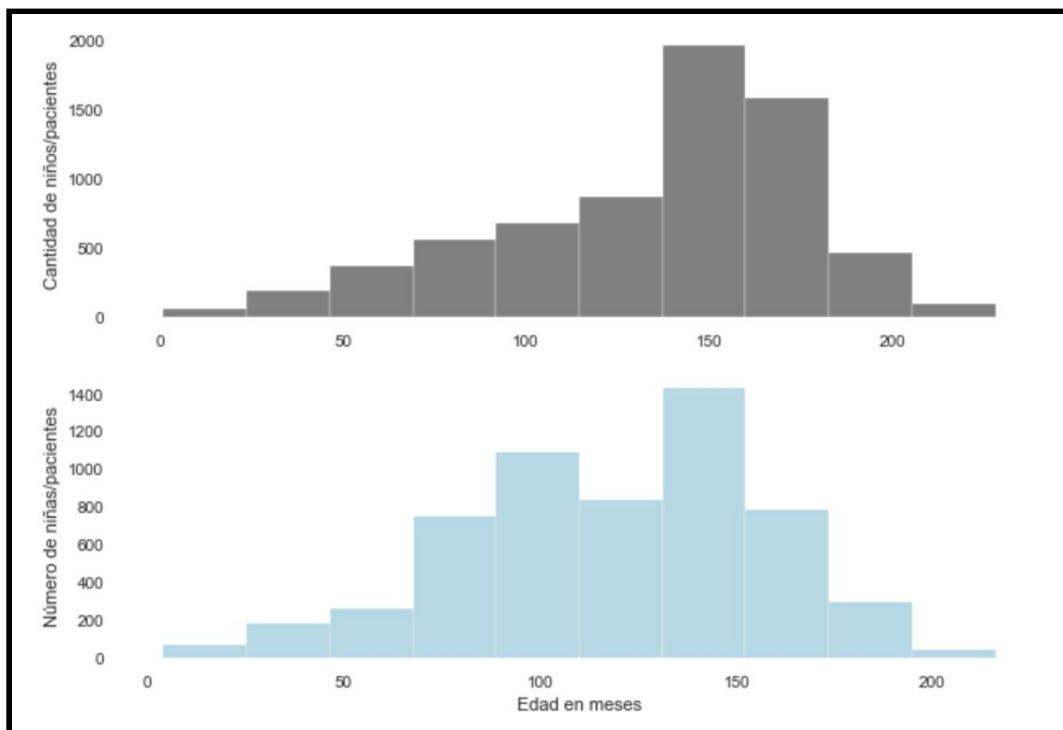
Análisis Exploratorio:



Histograma de la variable boneage

Minimum	1	Standard deviation	41.1820214
5-th percentile	50	Coefficient of variation (CV)	0.3234509759
Q1	96	Kurtosis	-0.2373008857
median	132	Mean	127.3207517
Q3	156	Median Absolute Deviation (MAD)	26
95-th percentile	186	Skewness	-0.4749626722
Maximum	228	Sum	1605642
Range	227	Variance	1695.958887
Interquartile range (IQR)	60		

Estadística descriptiva de la variable boneage



Histogramas de boneage de niños y niñas por separado

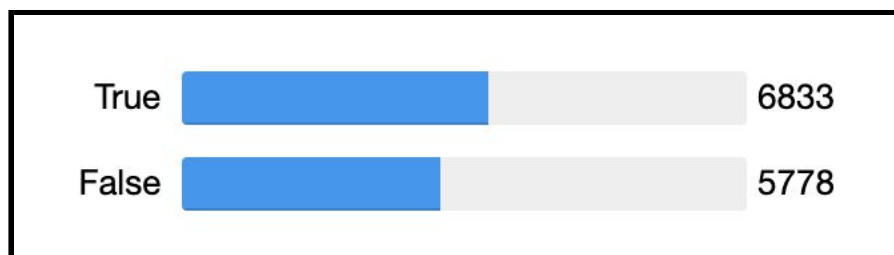
Se crearon una serie de histogramas para ver de qué manera están distribuidos los datos. Con el objetivo de ver normalidad en los mismos y corregir si esta fuera la acción correcta. Podemos observar, en los histogramas, que el dataset presenta una normalidad relativamente correcta o aceptable. Se debe mencionar que la frecuencia hace referencia a la cantidad de muestras (niños), mientras que el eje x a la edad ósea en meses (bonage). Así mismo se puede notar que un porcentaje alto de niños reside en el rango de 100 y 150 meses de edad.

Minimum	1	Standard deviation	42.14316183
5-th percentile	54	Coefficient of variation (CV)	0.3114709363
Q1	108	Kurtosis	0.04707907023
median	150	Mean	135.3036733
Q3	162	Median Absolute Deviation (MAD)	18
95-th percentile	192	Skewness	-0.7291214277
Maximum	228	Sum	924530
Range	227	Variance	1776.046089
Interquartile range (IQR)	54		

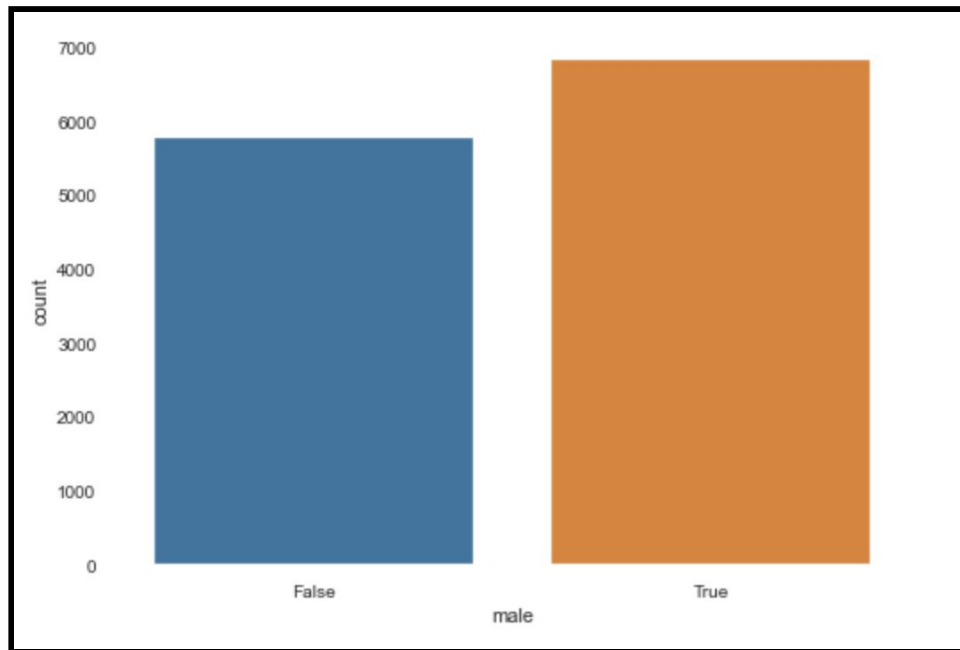
Estadística descriptiva boneage para males

Minimum	4	Standard deviation	37.9081119
5-th percentile	50	Coefficient of variation (CV)	0.3215815762
Q1	94	Kurtosis	-0.1877732542
median	120	Mean	117.8802354
Q3	144	Median Absolute Deviation (MAD)	26
95-th percentile	180	Skewness	-0.3165917273
Maximum	216	Sum	681112
Range	212	Variance	1437.024948
Interquartile range (IQR)	50		

Estadística descriptiva boneage para females



Frequency table de la variable male

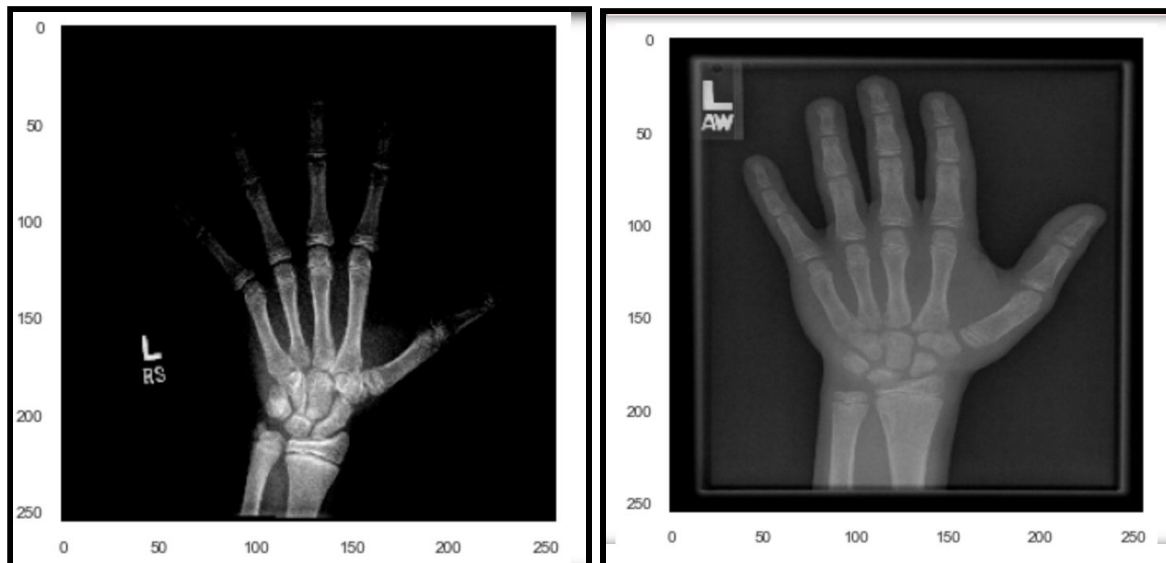


Pacientes hombres versus mujeres: las mujeres son false y hombres son true



Ejemplo de una de las imágenes antes de procesarla

Las imágenes de la radiografía contienen una L o una R esto indica de qué mano del paciente es la radiografía. Las otras letras que están presentes en la radiografía son las iniciales del doctor encargado de la radiografía.



Ejemplo de unas de las imágenes después de procesarlas

Con eso, concluye la limpieza de la data. Se empleó una forma de ajustar las imágenes. Creando un DataImageGenerator con tensorflow. Los rasgos morfológicos que representan la madurez esquelética se han visto obstaculizados por incapacidad para gestionar la gran variabilidad en el desarrollo y la osificación de los múltiples huesos de la mano y/o la muñeca.

En un intento de superar estas dificultades, se espera que estos análisis y análisis futuros ayuden o planten una semilla para el desarrollo de técnicas automatizadas de esta índole.

Conclusiones:

- Se logró transformar las imágenes para que todas cumplieran con un estándar de tamaño y color.
- La distribución de la edad en el dataset se puede considerar normal, para hombres y mujeres.