**INSTITUTO POLITÉCNICO NACIONAL**

ESCUELA SUPERIOR DE CÓMPUTO

LICENCIATURA EN CIENCIA DE DATOS

UNIDAD DE APRENDIZAJE

DESARROLLO DE APLICACIONES PARA ANÁLISIS DE DATOS

PRÁCTICA 3

NOMBRE DE LOS ALUMNOS:

DE LUNA OCAMPO YANINA

PROFESOR:

ITURIEL FLORES

GRUPO:

4CDM1

FECHA:

04/05/2022

En esta práctica pusimos a prueba el uso de pandas y DataFrames con algunas sus funciones aprendidas durante la clase.

Los pasos que se siguieron fueron los siguientes:

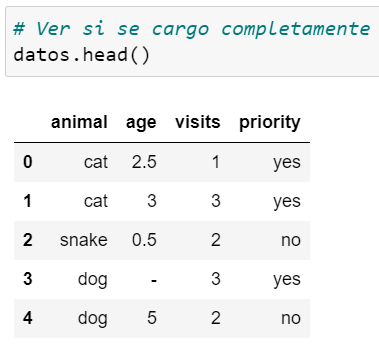
1. Primero importamos las librerías que utilizaremos durante toda la práctica. El de “*csv”* se utilizó para poder importar el, como dice el nombre, csv brindado en la práctica. El “*describe*” se usó para una función del DataFrame que se explicará adelante y por último “*pandas*”, que nos ayuda al análisis de datos y proporcionando estructuras de datos flexibles.



1. Leemos el csv para poder pasarlo posteriormente al DataFrame.

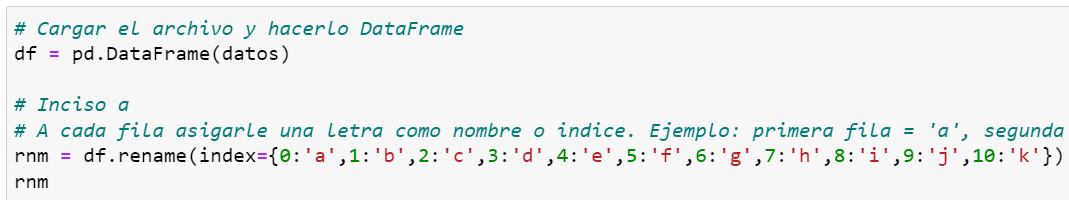


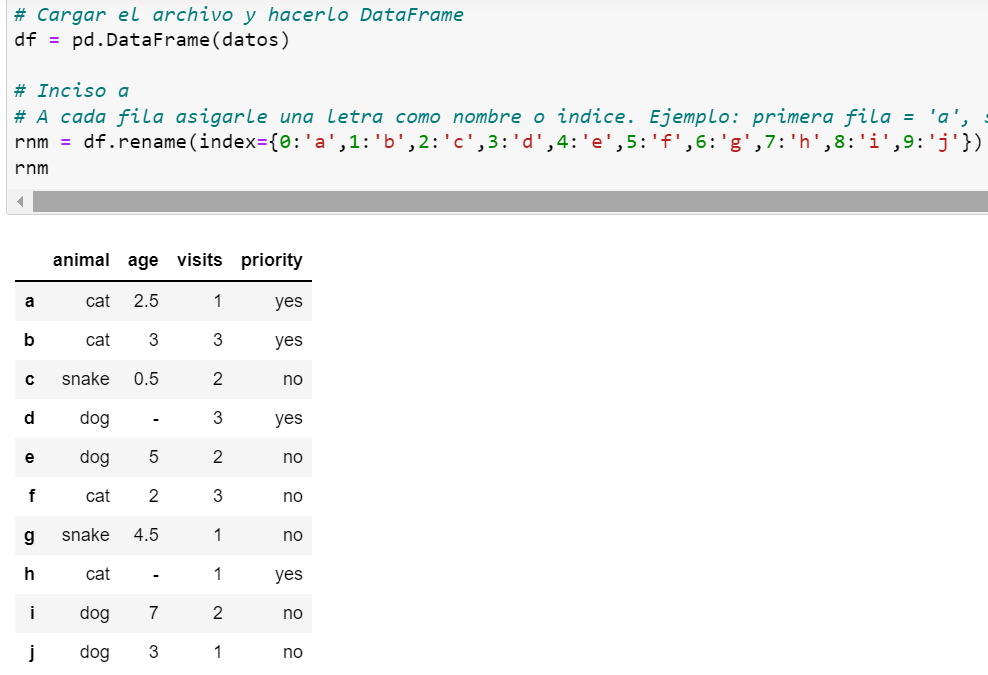
1. Vemos si los datos se cargaron correctamente, para esto utilizamos el comando “*head*” y poder visualizar las primeros 5 columnas de todo el DataFrame.



1. En la siguiente instrucción se realizaron dos pasos, el primero fue convertir el csv importado, a DataFrame como se solicitó.

La primera instrucción empezando la práctica es cambiar los index a las primeras letras del abecedario hasta que se terminen las filas del DataFrame, en este caso fue hasta j. Aquí es importante recalcar que al momento de cambiarle los index, cambié el nombre del conjunto.



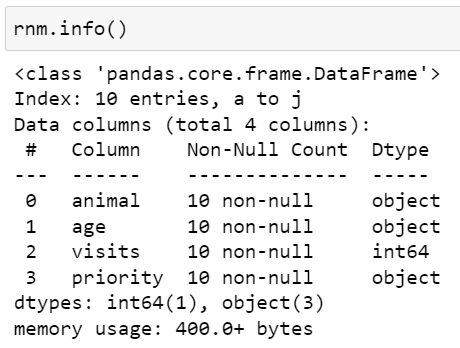


1. Mostramos la información básica del DataFrame, junto con: incluido el número de filas, los nombres de las columnas, el número y el tipo de valores en cada columna.

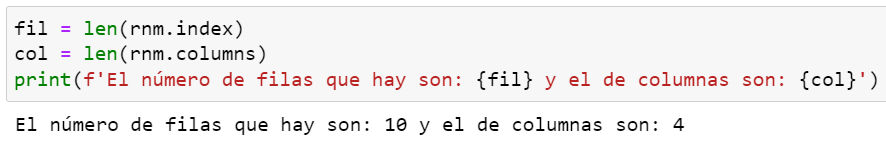
# Aquí es en donde se utiliza la librería de pydoc para obtener este, nos brinda como se ve en la imagen: desviación estándar, media, cuartiles, mínimo, máximo, etc.



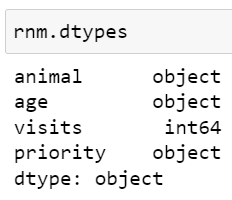
# Con .info obtuvimos el nombre de las columnas y el tipo de datos que se almacena en estas.



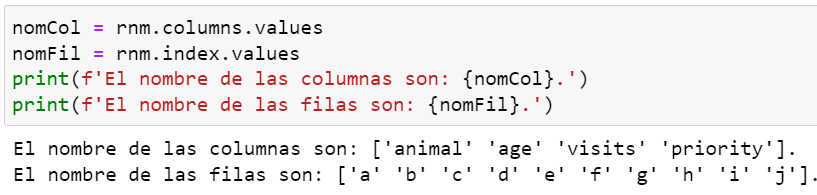
# Obtenemos cuántas filas y columnas son, con un format las unimos e imprimimos los valores.



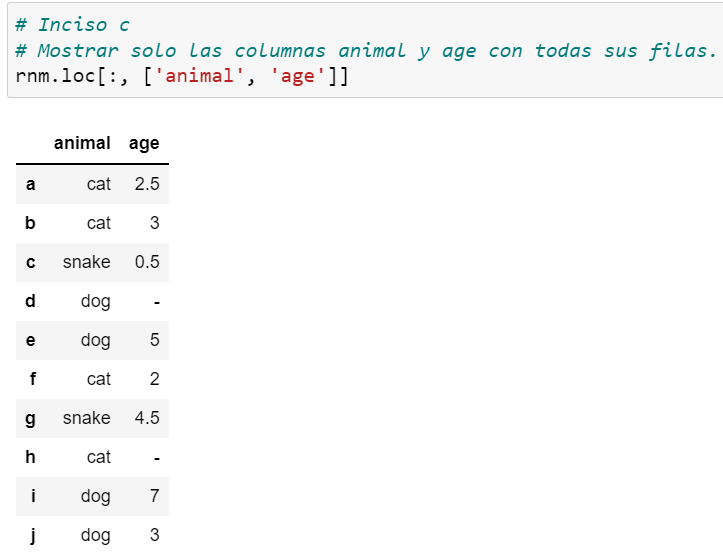
# Aquí, con dtypes vemos el tipo de datos que tiene cada columna, de la misma forma como pudimos ver en el .info.



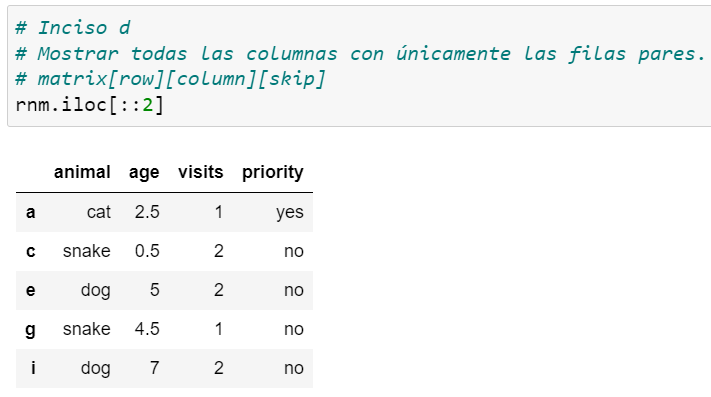
1. Para obtener el nombre de columnas y filas se utiliza .columns.values e index.values, nuevamente con un format unimos las respuestas y las mostramos.



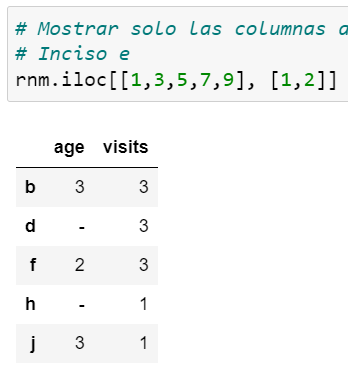
1. Siguiendo con las instrucciones, solo mostramos las columnas “animal” y ”age” con todas sus filas. Para esto usamos .loc[] con los parámetros estipulados.



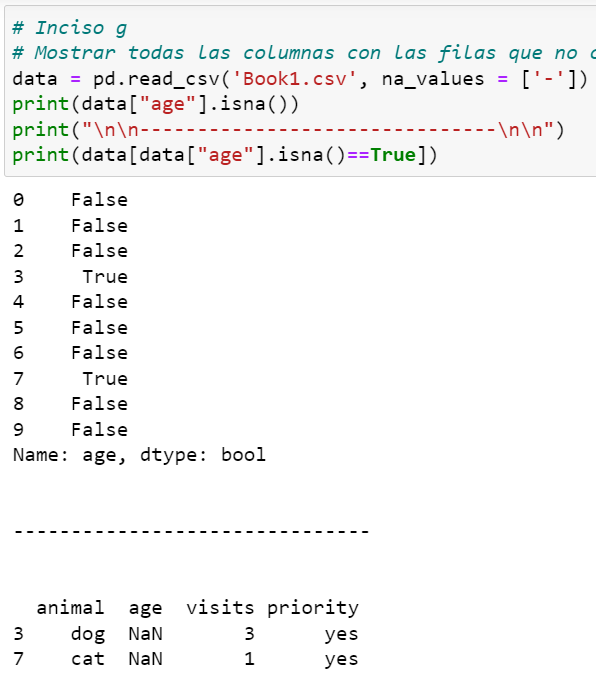
1. Mostramos las columnas con las filas pares, para esto se usa la función .iloc[] con sus parámetros correctamente estipulados.



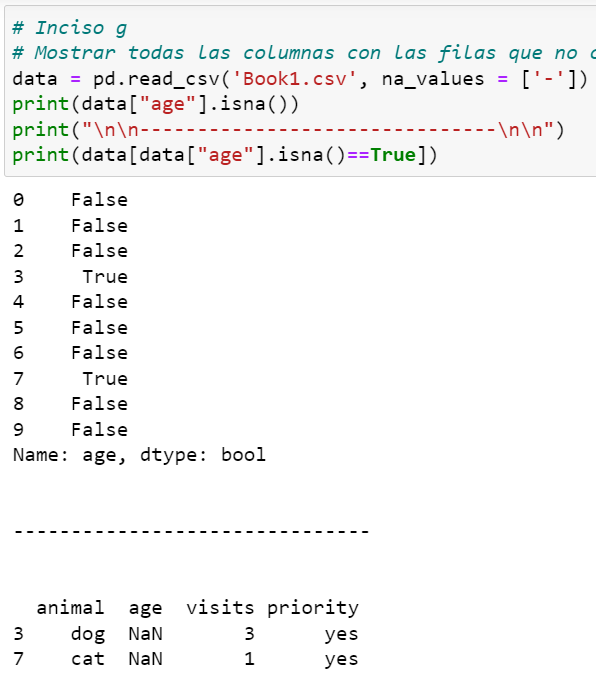
1. Mostramos de igual forma como arriba, solo las columnas age y visits pero únicamente con las filas impares.



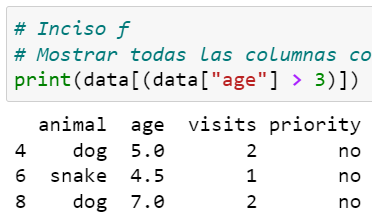
1. Mostrar todas las columnas con las filas que no contengan un dato para edad. Ponemos los na\_values = “-”, ya que no lo toma como entero, si no como string.



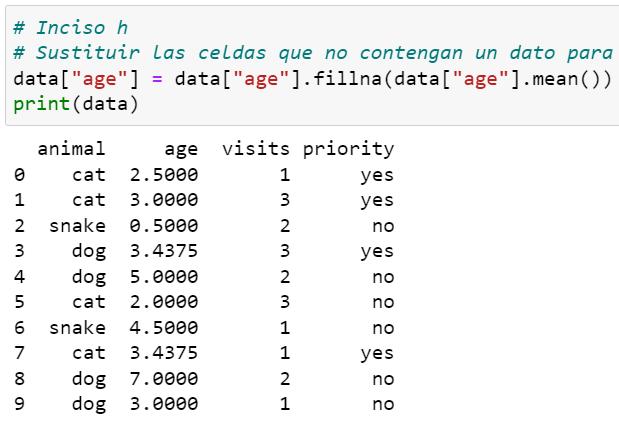
# Podemos imprimirlos así o solo las columnas que tiene los valores nulos.



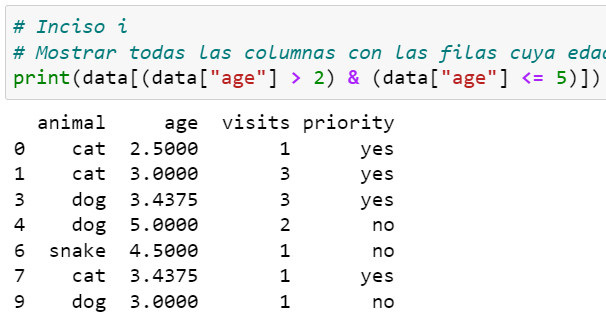
1. Mostrar todas las columnas con las filas cuya edad sea mayor a 3. Accedemos a la fila y comparamos, ya quitando los nulos para que no tenga problema.



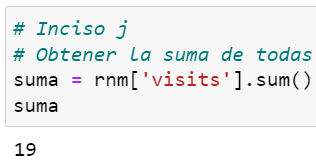
1. Sustituir las celdas que no contengan un dato para edad con el promedio del resto de edades. Para esto usamos la función. mean()



1. Mostrar todas las columnas con las filas cuya edad este sea mayor a 2 y menor o igual 5. Usamos lo mismo para este pero con “&” para poder comparar las dos.



1. Sacamos la suma de todas las visitas con la función .sum de la columna del dataset de “visits”.

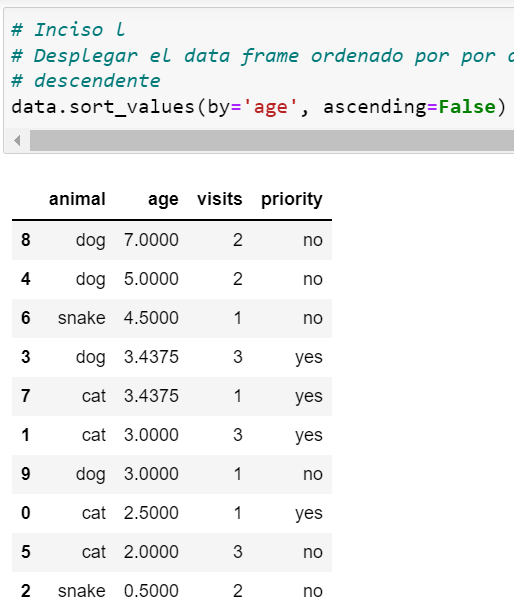


1. Calcular el promedio de la edad de cada tipo diferente de animal. Tip: Investigar y usar el método "groupby".



1. Desplegar el data frame ordenado por por dos criterios, primero por edad en orden descendente y después por visitas en orden ascendente. Tip: Investigar y usar el método "sort\_values".

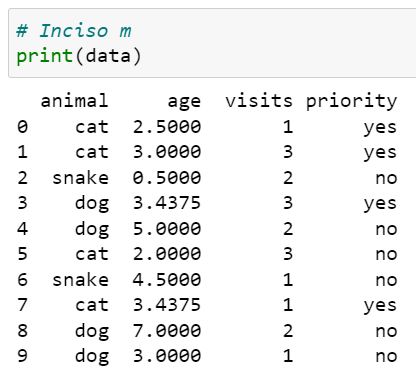
# descendente



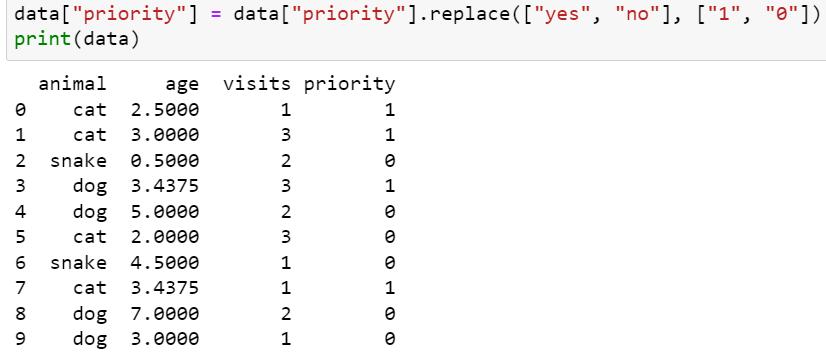
# ascendente



1. Imprimimos los datos para la siguiente instrucción y ver el cambio realizado.



# Reemplazar sí y no en la columna de prioridad con su valor booleano correspondiente. Utilizamos el .replace múltiple para poder cambiar dichos valores.



1. Reemplazar las celdas de la columna de animales por su respectivo nombre en español. Volvemos a utilizar el .replace múltiple para lograr realizar esta instrucción.

