**Wrangle Report**

***Gathering:***

Wasn´t easy to me to import all pieces of data into my jupyter notebook, but with this Project I've learned a lot. Read_csv is one of the most useful tool we have in Python to import pieces of data, but the problem has showed up when I had to read the JSON file from a TXT. With Monday classes I've learned some different ways to get the data and I appreciated that a lot. To be honest, I have to continue practicing and learning about how to use API because I don't feel confident with that yet.

***Assesing and Cleaning:***

I have done some manually assessing but more programatic assesing of the data. At the bigining I have used ".head()" or ".sample()" to get information about how each dataframe was built. With ".info()" I could check if datatypes was correct (i.e. "timestamp" column was defined as object) and also which columns had null values and decide how would I handle them. I also recognised many string values as "None" or "a" and I changed for NaN (with np.nan) because is the correct way to manage nulls afterwards.

I had corrected two tidiness issues, melting some columns:

1. Dog stage was wrong created into the dataframe, because each stage (category) was consigned as a value (column), so I created a new column for the variable stage and I deleted all the other columns. Here is where I discovered some dirty data because there was some observations with 2 stages, so I finally consigned as NaN.
2. Into the image prediction dataframe there was also a building mistake. We got a column for each number of prediction, each breed prediction, each level of confidence and each boolean value if it was or not a dog. I created a column for each variable: number of prediction, dog stage, confidence of prediction and if it was a dog or not.

During the Journey I was discovering new dirty data to clean and it was all what I've done into the jupyter notebook.

Of course all cleaning efforts was done programatically.