**ML Intern 1st Call Question**

You are provided with a dataset which contains images of scatter plots and correlations corresponding to each scatter plot.

* **Download the dataset and train a model**. The folder contains 150,000 images.
* Downloading the dataset (images and correlation values) from this link:
* <https://drive.google.com/file/d/1OqZJW8WUeUi2XrzJisJBHR4Er1lPowmG/view?usp=sharing>

Given these images,

1. **Separate the images into training and test sets. Predict the correlation between x and y values.**

To get acquainted with how it works, you can play the game here first:<http://guessthecorrelation.com/>

**Note: You do not need to conduct a detailed analysis for this task. Just training a model will suffice. we don’t ask for perfect predicting results, and the key point is to show the training and predicting structure.**

1. **Please share your code.** It can be a Github link, Jupyter notebook or any other format of your preference.
2. **After training the model, please answer the following questions.**
   1. How many images will you use for training?
   2. How many images will you use for testing? Why?
   3. How do you know when to stop training?
   4. If there are a few images of dirty data within this dataset that may not be correlated with the rest, how would you identify them?
   5. What kind of loss function and last-layer activation will you use? Why?