Adrian Marinovich
Springboard Data Science Career Track
Capstone Project #1 - Data Wrangling
September 24, 2018

Project: Detection of smiles in images of faces

What kind of cleaning steps did you perform?

The primary data wrangling tasks were not cleaning, but rather were focused on writing scripts to fill the AWS S3 buckets with data, establishing a pipeline from the Jupyter notebook to the S3 buckets using Boto, and then using the lists of smile and non-smile image file names to select the corresponding images from the larger LFWcrop dataset of images.

How did you deal with missing values, if any?

Two file names in the smile list did not have matching image files, and were identified as extraneous .txt file names ('listt.txt', and 'SMILE_list.txt') which did not interfere with the correct matching of image file names.'

Were there outliers, and how did you handle them?

Images corresponding to the smile and non-smile lists were perused to check for validity of the smile/non-smile labels, and found to be consistent with the given labels.

Submit a link to the document.

https://github.com/adriatic13/springboard/blob/master/dsct_capstone1/Adrian_Marinovich_Cap1_smiles_data_wrangling.ipynb