The motivation behind the creation of this dataset is to have a challenging Test set for the task of classifying fake and real human faces. Most of the available datasets on Kaggle are "Uniform" and doesn't present a good variance of face features, particularly for the "Fake" class. Moreover, the fake faces collected in this dataset are generated using the StyleGAN2, which present a harder challenge to classify them correctly even for the human eye. The real human faces in this dataset are gathered so that we have a fair representation of different features(Age, Sex, Makeup, Ethnicity, etc…) that may be encountered in a production setup.

Content

The images available in this dataset are in a JPEG format and of uniform size of 300x300.  
There "Fake" faces are collected from the website [thispersondoesnotexist.com](https://thispersondoesnotexist.com/).  
There "Real" faces images are collected through the API of the website [Unsplash](https://unsplash.com/) and then the faces are cropped out of using OpenCV library.

Total number of images: 1288  
Number of "Fake" faces: 700  
Number of "Real" faces: 589

The data.csv contains the images Id and the corresponding label.