# OriginalDataset\_FashionProductImagesSmall

* Dataset downloaded from Kaggle (https://www.kaggle.com/datasets/paramaggarwal/fashion-product-images-small)
  + images (folder)
    - 44,441 cloth images of men, women, boys, girls, unisex.
  + styles.csv
    - csv file of the original dataset
    - Features: id, gender, masterCategory, subcategory, articleType, baseColour, season, year, usage, productDisplayName
  + EDAOnMaleDS.ipynb
    - Original csv file is erroneous because 22 rows (6044, 6569, 7399, 7939, 9026, 10264, 10427, 10905, 11373, 11945, 14112, 14532, 15076, 29906, 31625, 33020, 35748, 35962, 37770, 38105, 38275, 38404) contain comma separated values in productDisplayName hence causing 1 extra column. After skipping bad lines df has 44,424 entries (44,447 - 22 -1).
    - The original csv file has 44,447 rows but images are 44,441
    - Df has men, boys, women, girls, unisex images.
    - Has too many features, we only require id, articleType/subCategory, usage