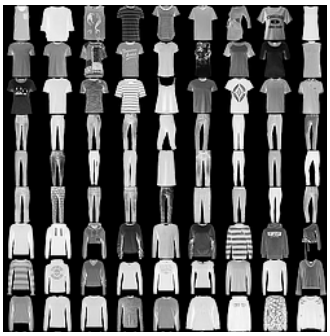
Instructions

This challenge is inspired by one of the most well known fashion image datasets: Fashion MNIST. Created by Zalando, it contains thousands of labeled clothing images. You can find all the documentation about the dataset here.



The original dataset contains 10 different labels:

0 T-shirt/top
1 Trouser
2 Pullover
3 Dress
4 Coat
5 Sandal
6 Shirt
7 Sneaker
8 Bag
9 Ankle boot

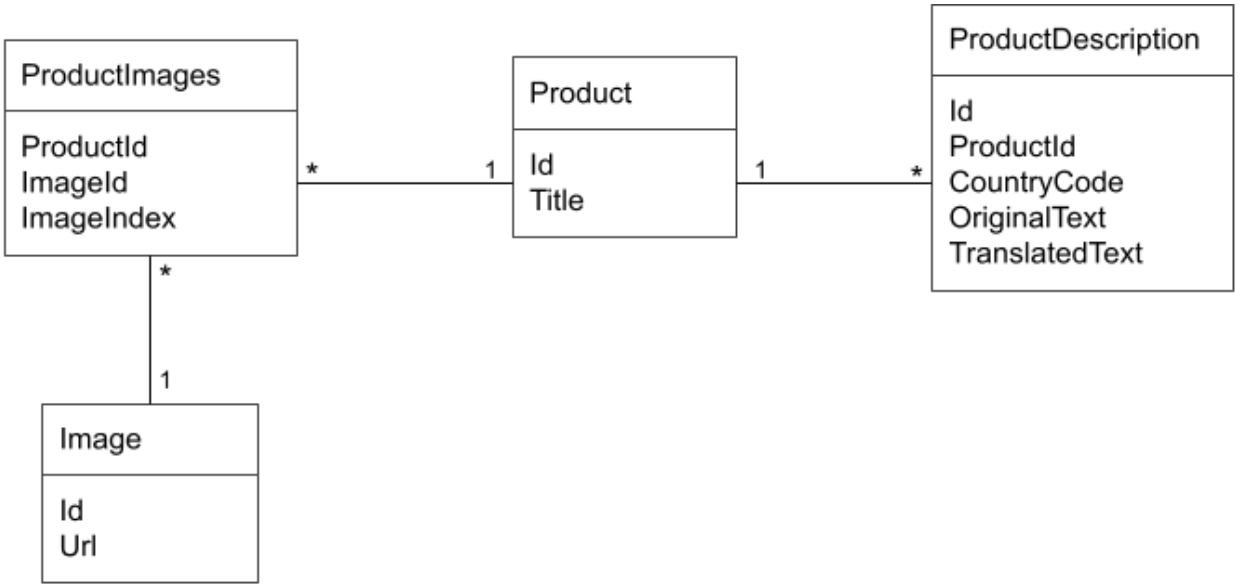
In this case, we want you to group the original labels in 5 new labels with the following mapping between them:

Upper part: T-shirt/top + Pullover + Coat + Shirt
Bottom part: Trouser
One piece: Dress
Footwear: Sandal + Sneaker + Ankle boot
Bags: Bag

The goal of this case is to understand the dataset and experiment with deep learning techniques to train a classifier for the given grouped classes (Upper part, Bottom Part, One piece, Footwear and Bags).

What aspects are highly valued when evaluating the exercise?

1. Data understanding and exploratory data analysis.
2. Deep learning experimentation and model evaluation with the proper metrics. *Please note that experimentation and originality are valued higher than pure metric scores.*
3. Design a CI/CD and deployment plan to serve the final model in a cloud based production environment. *This plan should include MLOps tools and KPIs monitoring and we don't expect you to implement any code here.*
4. Let's imagine that now you have to load the product images from our database along with some additional text features like the product description provided by the retailers. The data is stored in an SQL relational database (postgresql) with the following schema:



Write a query (either in django ORM or in SQL) to extract, for every existing product, the following fields:

- Product.Title
- Image.Url for the images with the ImageIndex = 0.
  ImageIndex field states the priority order of images of a certain product. So for a given ProductId, the image with ImageIndex = 0 would be the most relevant image for that product
- ProductDescription.TranslatedText if exists, else ProductDescription.OriginalText for ProductDescriptions in CountryCode = 'us'

The challenge must be completed using Python. Our personal recommendation is to use Jupyter Notebook and Keras/PyTorch for the initial experimentation. Additional

information about loading and using the dataset for different libraries can be found in the original dataset repository.

Your answer

You should share with us a link to your solution uploaded in a source control repository (e.g. github or gitlab) with the following:

- Source code for the points 1 and 2.
- Markdown file with your answer to points 3 and 4.
- Readme file with the documentation of your repo and references to any external source used.

There is no right way to do this. We are interested in the choices that you make, how you justify them, and your development process.

If you have any questions or comments, don’t hesitate to contact us. **Good luck!**