This A3_qhuangak_20548333_code.py file is used to cluster 5011 images.

It includes four processes:

1. The first process import the images and transform the images into a 4 dimension list:
1) Read images
2) Resize the images to 224*224*3
3) Convert to from BGR to RGB
4) Store each images into a list(images[ ]) with size (number of images)*224*224*3.

2. The second process is extract the feature from the images:
In this process, we use ResNet50 model in keras to extract the feature. After that, reshape the list from 4 dimension to 2 dimension(the same as 1d with 5011 numbers), and the size of this list is (number of images)*7*7*2048.
After reshape step, use PCA to compress the data, it will reduce the time of next process.

3. The third process is use K-means model to cluster images:
1) train the K-means model from sklearn with 11 clusters by the result of the inertia's trend
2) Use the training model to predict the feature we got from last process
3) Get the predict label of each images

4. The last process is output the result as a .csv file.
This process use a dict to store all the cluster with their corresponding images, and output it as the given format.