Submitted By : Mrinal Bhan

DSAI - 211020428

```
!pip install pycocotools
```

```
Requirement already satisfied: pycocotools in /opt/conda/lib/python3.10/site-packages (2.0.7)
Requirement already satisfied: matplotlib>=2.1.0 in /opt/conda/lib/python3.10/site-packages (from pycocotools) (3.7.2)
Requirement already satisfied: numpy in /opt/conda/lib/python3.10/site-packages (from pycocotools) (1.23.5)
Requirement already satisfied: contourpy>=1.0.1 in /opt/conda/lib/python3.10/site-packages (from matplotlib>=2.1.0->pycocotools) (1.1.0)
Requirement already satisfied: cycler>=0.10 in /opt/conda/lib/python3.10/site-packages (from matplotlib>=2.1.0->pycocotools) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in /opt/conda/lib/python3.10/site-packages (from matplotlib>=2.1.0->pycocotools) (4.40.0
Requirement already satisfied: kiwisolver>=1.0.1 in /opt/conda/lib/python3.10/site-packages (from matplotlib>=2.1.0->pycocotools) (1.4.4)
Requirement already satisfied: packaging>=20.0 in /opt/conda/lib/python3.10/site-packages (from matplotlib>=2.1.0->pycocotools) (21.3)
Requirement already satisfied: pillow>=6.2.0 in /opt/conda/lib/python3.10/site-packages (from matplotlib>=2.1.0->pycocotools) (9.5.0)
Requirement already satisfied: pyparsing<3.1,>=2.3.1 in /opt/conda/lib/python3.10/site-packages (from matplotlib>=2.1.0->pycocotools) (3.
Requirement already satisfied: python-dateutil>=2.7 in /opt/conda/lib/python3.10/site-packages (from matplotlib>=2.1.0->pycocotools) (2.8
Requirement already satisfied: six>=1.5 in /opt/conda/lib/python3.10/site-packages (from python-dateutil>=2.7->matplotlib>=2.1.0->pycocot
```

```
import os
from pycocotools.coco import COCO

#Set paths to annotation files

annotations_dir = '/kaggle/input/coco-dataset/coco2017/annotations/'
train_ann_file = os.path.join('/kaggle/input/coco-2017-dataset/coco2017/annotations/instances_train2017.json')
val_ann_file = os.path.join('/kaggle/input/coco-2017-dataset/coco2017/annotations/instances_val2017.json')

# Initialize COCO instances for train and validation sets
train_coco = COCO(train_ann_file)
val_coco = COCO(val_ann_file)
```

```
loading annotations into memory...
Done (t=21.34s)
creating index...
index created!
loading annotations into memory...
Done (t=0.87s)
creating index...
index created!
```

```
train_image_ids = train_coco.getImgIds()
val_image_ids = val_coco.getImgIds()

# Initialize dictionaries to store image file names and URLs
train_images = {}
val_images = {}

# Initialize counter for removed images
removed_images = 0

# Iterate over train image IDs and extract file names and URLs
for img_id in train_image_ids:
    # Check if the image has annotations
    if len(train_coco.imgToAnns[img_id]) == 0:
        removed_images += 1
        continue
    img = train_coco.loadImgs(img_id)[0]
    file_name = img['file_name']
    url = img['coco_url']
    train_images[img_id] = {'file_name': file_name, 'url': url}

# Iterate over validation image IDs and extract file names and URLs
for img_id in val_image_ids:
    # Check if the image has annotations
    if len(val_coco.imgToAnns[img_id]) == 0:
        removed_images += 1
        continue
    img = val_coco.loadImgs(img_id)[0]
    file_name = img['file_name']
    url = img['coco_url']
    val_images[img_id] = {'file_name': file_name, 'url': url}

# Print number of removed images
print(f"{removed_images} images were removed because they had no annotations.")
```

```
1069 images were removed because they had no annotations.
```

```
min_annotations = 3

# Get image ids with number of annotations
```

```
# Get image ids with number of annotations
img_ids = train_coco.getImgIds()
img_id_to_ann_count = {img_id: len(train_coco.getAnnIds(img_id)) for img_id in img_ids}

# Filter out images with too few annotations
filtered_img_ids = [img_id for img_id, ann_count in img_id_to_ann_count.items() if ann_count >= min_annotations]

# Update train_coco to only contain images with enough annotations
train_coco.imgs = {img_id: train_coco.imgs[img_id] for img_id in filtered_img_ids}

# Print number of removed images
removed_images = len(img_ids) - len(filtered_img_ids)
print(f"{removed_images} images were removed because they had fewer than {min_annotations} annotations.")
```

```
36305 images were removed because they had fewer than 3 annotations.
```

```

```