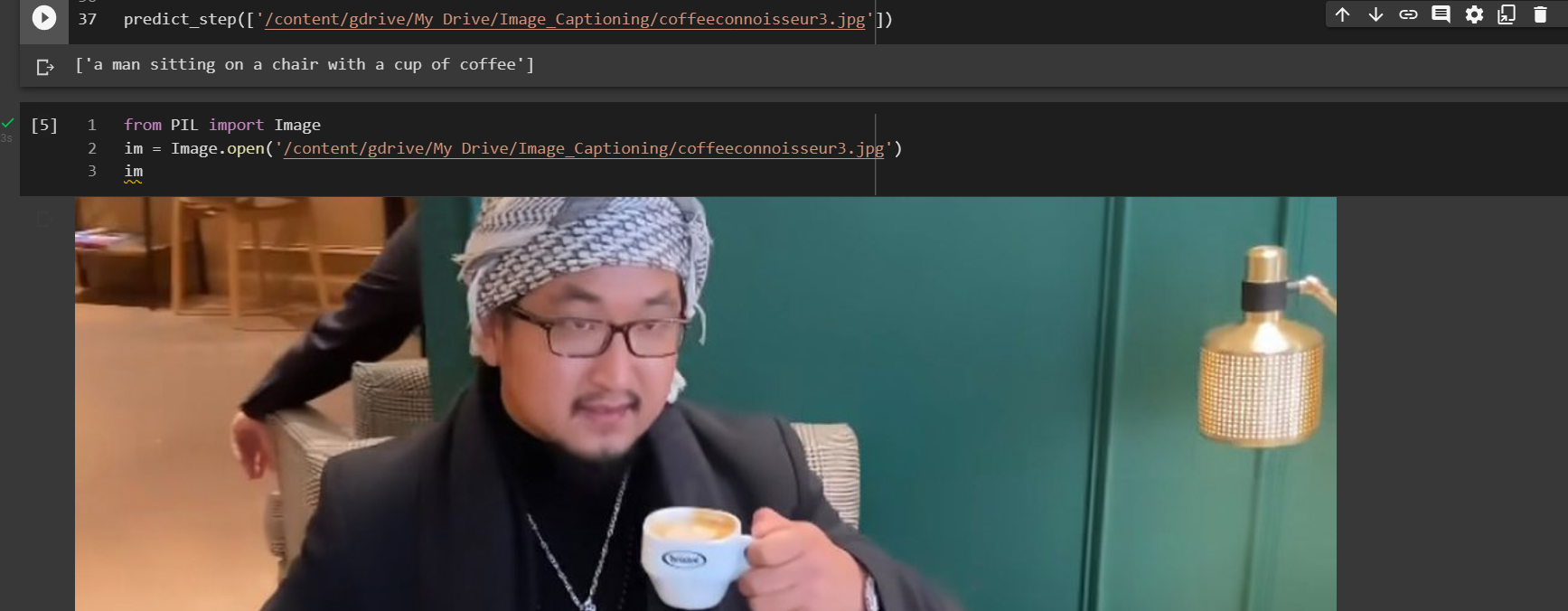
Image Captioning: Pretrained Model

1. We used Vision API for analyzing images for different features such as object detection, label detection, text extraction, etc. In addition, Vision API offers a best-label guess feature that assigns a caption to an image based on the best guess it can make from the input image.
2. However, as mentioned in the Google Cloud Vision API documentation, the best guess label does not correctly identify the image in many cases. To counter this problem I looked at pre-trained models image captioning deep learning models.
3. Hugging face has a lot of pre-trained models which can be used for image captioning. Of them, I selected the [nlpconnect/vit-gpt2-image-captioning · Hugging Face](https://huggingface.co/nlpconnect/vit-gpt2-image-captioning) model.
4. Using it, and mounting the drive to load the images, I was able to get the captions generated for images as seen below:

Image:



Caption Generated:



1. The code for the same can be found in my drive folder:
   1. Single Image: <https://colab.research.google.com/drive/1aqnVBUWcduYDC7QLlxhh6OeLazTmlMYS?usp=share_link>
   2. Mulitple Images: <https://colab.research.google.com/drive/1OxXViT3Ao7SKt1XrDJ3K3nHeHpRAN8kH?usp=share_link>