**Lab Report 1**

For this lab report we used the Flickr30 dataset provided by Mayanka in the documentation. The category we choose was Sports, which meant we had to scrub the Flickr30 dataset for pictures which made sense to our category. We went through the captions and searched for the keywords using the Standford Core NLP tokenization, the words were: baseball, football and basketball. The pictures which had those 3 keywords became our 3 datasets. We did basic tokenization and lemmatization of the image captions finding these 3 keywords and saved the images to separate locations for the image processing and SIFT feature extraction as well as doing image statistics (those can be found in the Github REPO). We did this for a small portion of the image captions. Once we calculated the image statistics we went ahead and computed the feature extraction. Feature extraction was done using the SIFT framework. While we attempted doing SIFT feature extraction through the code provided by Mayanka, I never got my version of SIFT to successfully extract features from our image dataset similar to test data, however I cannot say the same for my team members. In the Github repo you will find our versions of the SIFT code, as well as our versions of the lemmatization code based of the StandfordCore NLP provided in the source code for this class.