Eugene Kim

DS-DC19

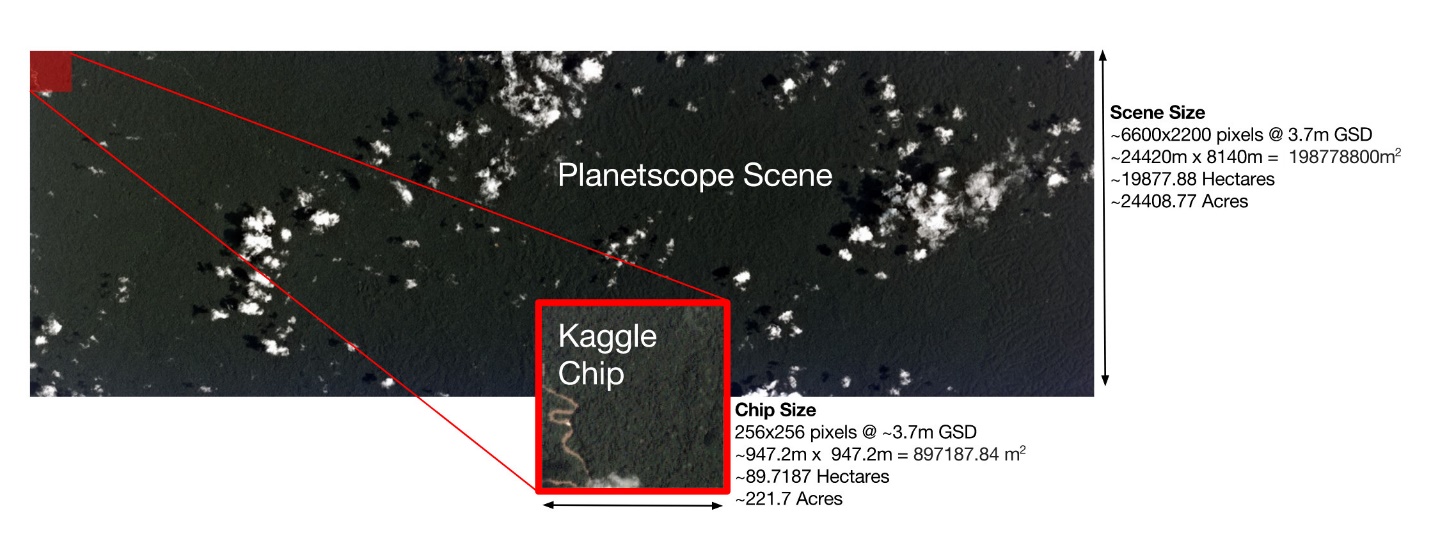
Project Outline for Classifying Satellite Images of Amazon Rainforest

## Overview

Planet has provided satellite images of the Amazon Rainforest that need to be classified by what is within the image. Images can be classified with multiple classifications depending on what can be seen in each image. Examples can be seen below



Images are divided by chips created from a specified Scene Size. Each image in the data set represents a chip. Overall the whole image displays approx. 20000 Hectares. More detailed specification can be seen in the image below



These are the more important classifications that I will be focusing on:

**Primary**: virgin rainforest; thick forestry not affected by human activity

**Agriculture**: Farms and human activity

**Roads**: Logging Roads

**Habitation:** Human Homes

**Selective Logging**: Areas of the forest which are cleared due to logging

## Hypothesis

When classifying images,

80% of images will be classified as primary

30% of images will be classified as Agriculture/roads/habitation/human activity

40% of images will be classified as Selective Logging

Of those classified as primary, 50% will include one of the other classification labels

## Risks and Assumptions

1. There are cloudy/hazy images, which are virtually useless when classifying what is in the forest.
2. There are many rare classifications that may or may not be impactful to what I would be looking for.
3. There are also a variety of other classifications that do not particularly add

## Goals

1. Create a machine learning algorithm that can classify images with computer vision!
2. Visualize the correlation of certain classifications with a heatmap
3. Create a predictive model that can classify an image if given labels.

## Success Metrics

1. Create a machine learning algorithm that can correctly classify at least one label per picture for over 50% of images provided
2. Determine the correlation between primary labeled images vs other labels