For the project, I looked into the datasets available in the UCI Machine Learning Repository. I picked the Image Segmentation dataset hoping that I can get most of machine learning.

**My understanding of the dataset:**

*About Image Segmentation:*

Image segmentation is considered to be one of the fundamental problems for computer vision. A primary goal of image segmentation is to partition an image into regions of coherent properties so that each region corresponds to an object or area of interest. In general, objects in outdoor scenes can be divided into two categories, namely, unstructured objects (e.g., sky, roads, trees, grass, etc.) and structured objects (e.g., cars, buildings, people, etc.).

*Data Set:*

The dataset contains various detections and mostly seem to be processed data. The data is provided for six different classes of images:

* Brickface
* Sky
* Foliage
* Cement
* Window
* Path
* Grass

The following attributes are recorded/calculated for each image:

1. Region-centroid-column
2. Region-centroid-row
3. Region-pixel-count
4. Short-line-density-5
5. Short-line-density-2
6. Vedge-mean
7. Vedge-sd
8. Hedge-mean
9. Hedge-sd
10. Intensity-mean
11. Rawred-mean
12. Rawblue-mean
13. Rawgreen-mean
14. Exred-mean
15. Exblue-mean
16. Exgreen-mean
17. Value-mean
18. Saturation-mean
19. Hue-mean

There is also a test data set available for testing the designed model.

As I look into the data, I can question the data as following:

1. Do the color detection help mostly in deciding the class of the image?
2. Do the hedge mean or vedge mean help the decision better?
3. Also want to do such analysis with other columns of data available?
4. What combination of data would make a better linear model?

**The problem I am trying to solve:**

 The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze,

Segmentation is generally the first stage in any attempt to analyze or interpret an image automatically. Segmentation bridges the gap between low-level image processing and high-level image processing.

This has application in various areas:

Industrial inspection  
Optical character recognition (OCR)   
Tracking of objects in a sequence of images   
Classification of terrains visible in satellite images.   
Detection and measurement of bone, tissue, etc., in medical images.

I downloaded the data available in the following link for machine learning purpose:

<https://archive.ics.uci.edu/ml/datasets/Image+Segmentation>

brief approach in classifying this data:

Create a numeric field to assign a numeric value to each class, which is currently a character field.

Use the R functions like: summary, cor, plot, lm to understand the relationship between different columns and the Class variable.

Create a model that will classify the data into one of the expected seven classes.

Test the model on the test data.