Will Manley

Dr. Cates

COMSC 415

Project 2 Report

**Part 1:**

The training data provided within the dataset contains 505,290 items which are either classified as a star or a quasar.For every item, there is a value for the following color categories: u-g, g-r, r-i, and i-z.

**Describe bayes stuff**

To organize the data, I separated it so that there was a data frame containing every star and another data frame containing every quasar. The following two tables contain the color averages for both separated datasets:

**Star Training Average**

|  |  |
| --- | --- |
| **Color** | **Average** |
| u-g | 1.57 |
| g-r | 1.03 |
| r-i | 0.69 |
| i-z | 0.36 |

**Quasar Training Average**

|  |  |
| --- | --- |
| **Color** | **Average** |
| u-g | 0.44 |
| g-r | 0.19 |
| r-i | 0.12 |
| i-z | 0.06 |

As you can see, it appears that stars, on average, have higher values for every color. To visualize this, I created histograms containing the distribution of each color range. These distributions can be seen in the following four graphs:







