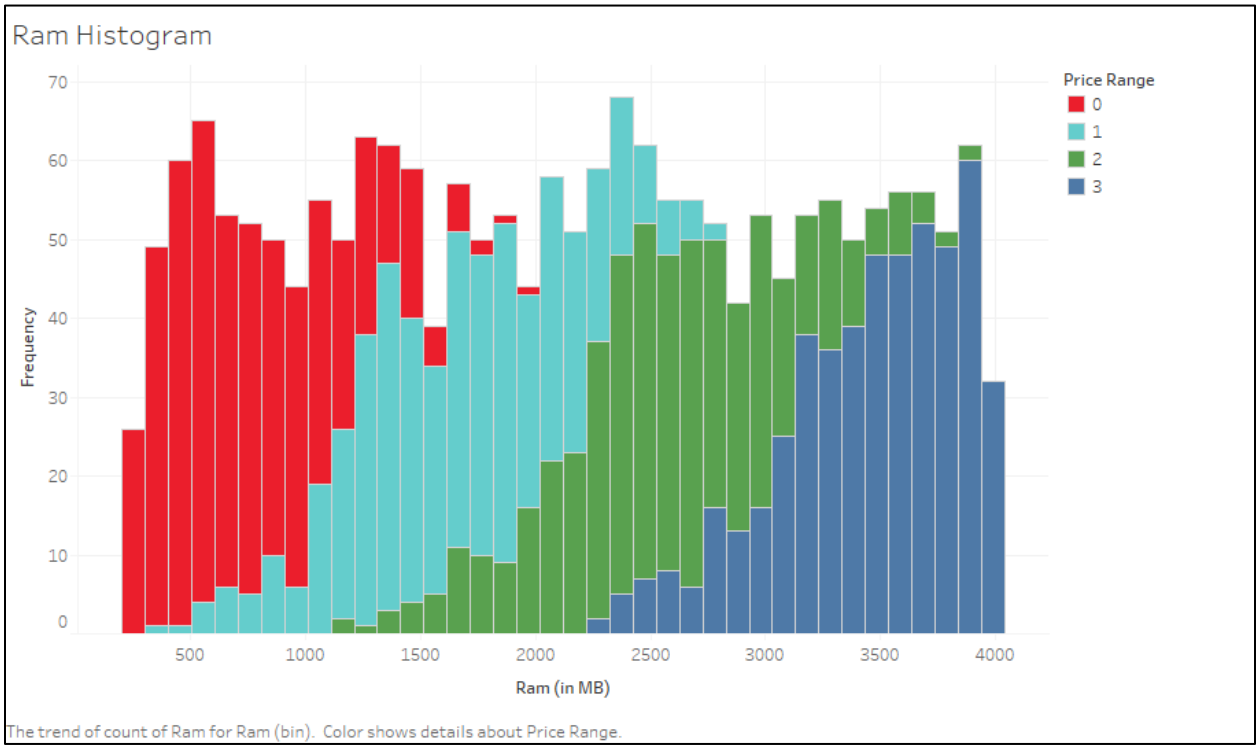
Data Mining Project Visualizations

Team: Niken Shah, Rohit Sanvaliya

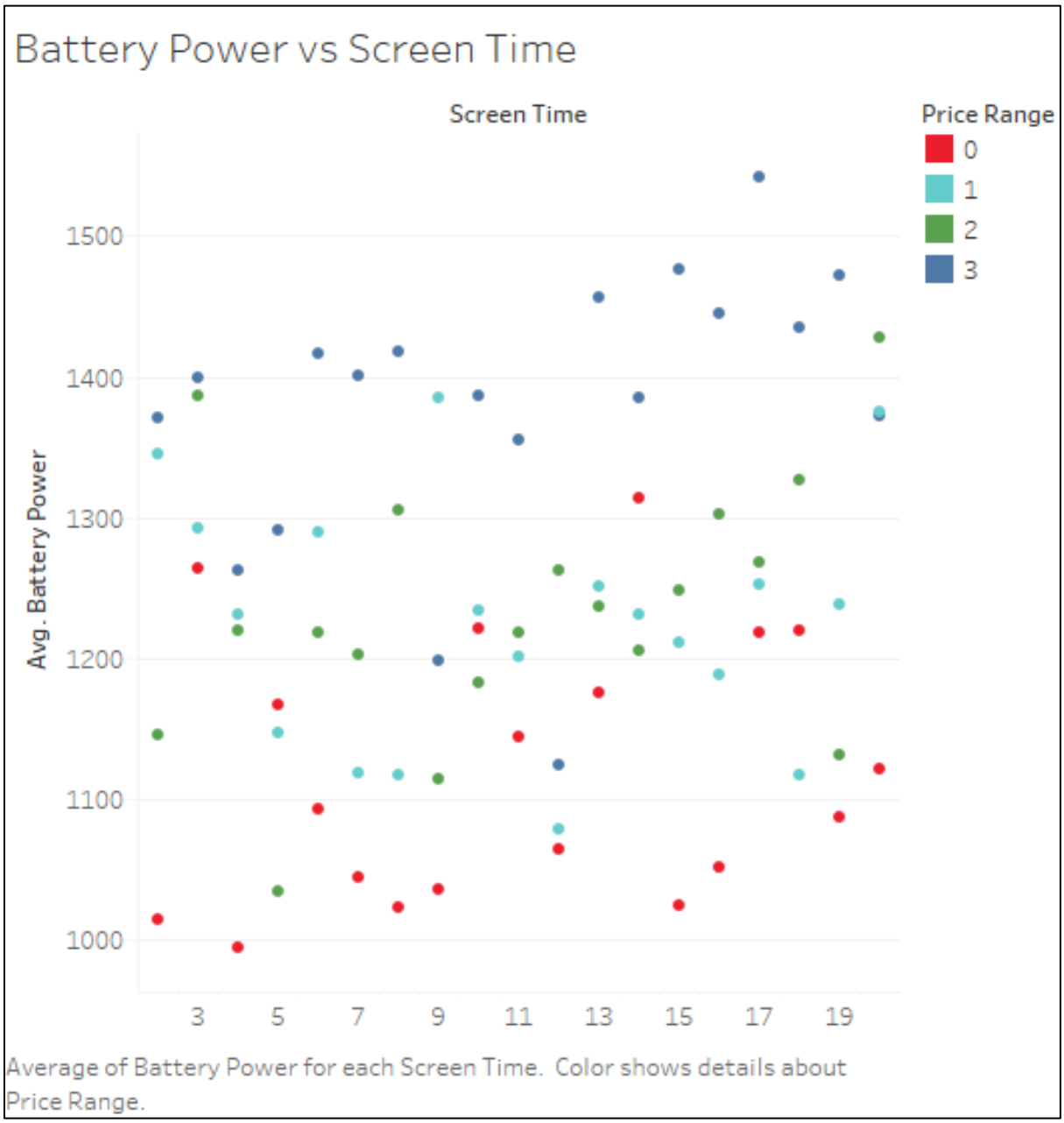
The class label "Price Range" shows 4 category of the mobile phones, based on their price. Here, the value "0" represents lowest priced category of the phones and "3" represents highest priced category of the phones.



The above Histogram represents size of RAM ranging from 0 to 4000 MB. In this plot we have plotted the frequency of the mobile phone based on their price range categories. It can be clearly seen that the cheaper phones have smaller RAM and phones on the expensive side of the spectrum have larger RAM.

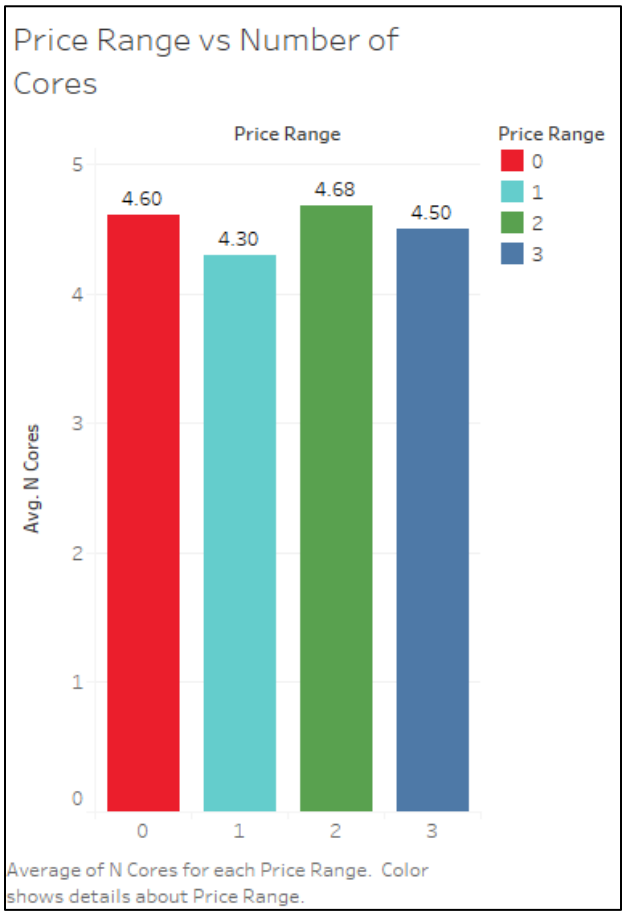
Therefore, from the histogram we can say that the RAM is a good attribute for mobile price classification problem.

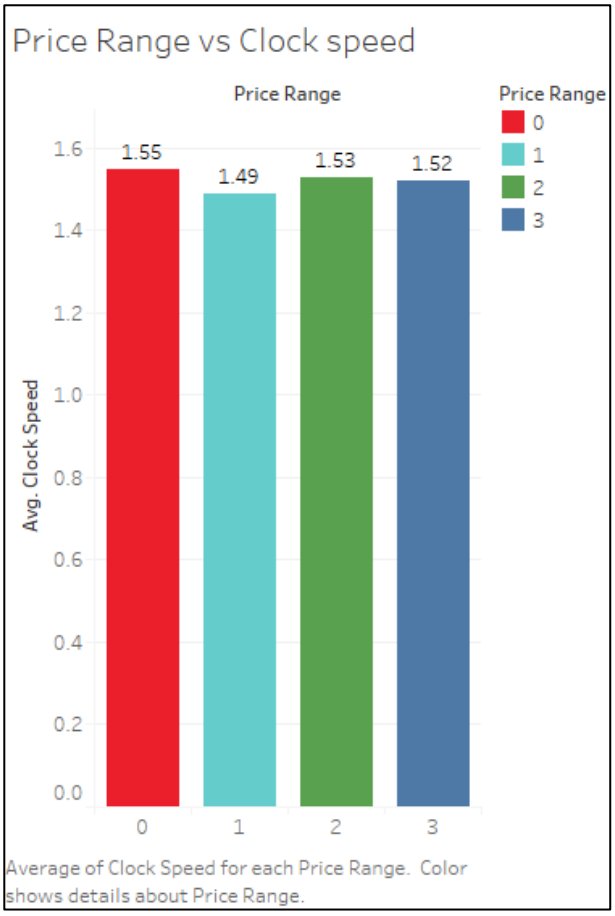
Data Mining Project Visualizations



This scatter plot represents the screen time on X axis and the average battery power on the Y axis. the color is used as a filter to distinguish between the price categories ranging from 0-3. though there is no direct correlation between the average battery power and the screen on time, one thing which is identifiable is that most phones which are in the higher priced categories have more battery power and some have higher screen time as well. The lower price range products have comparatively low average battery power.

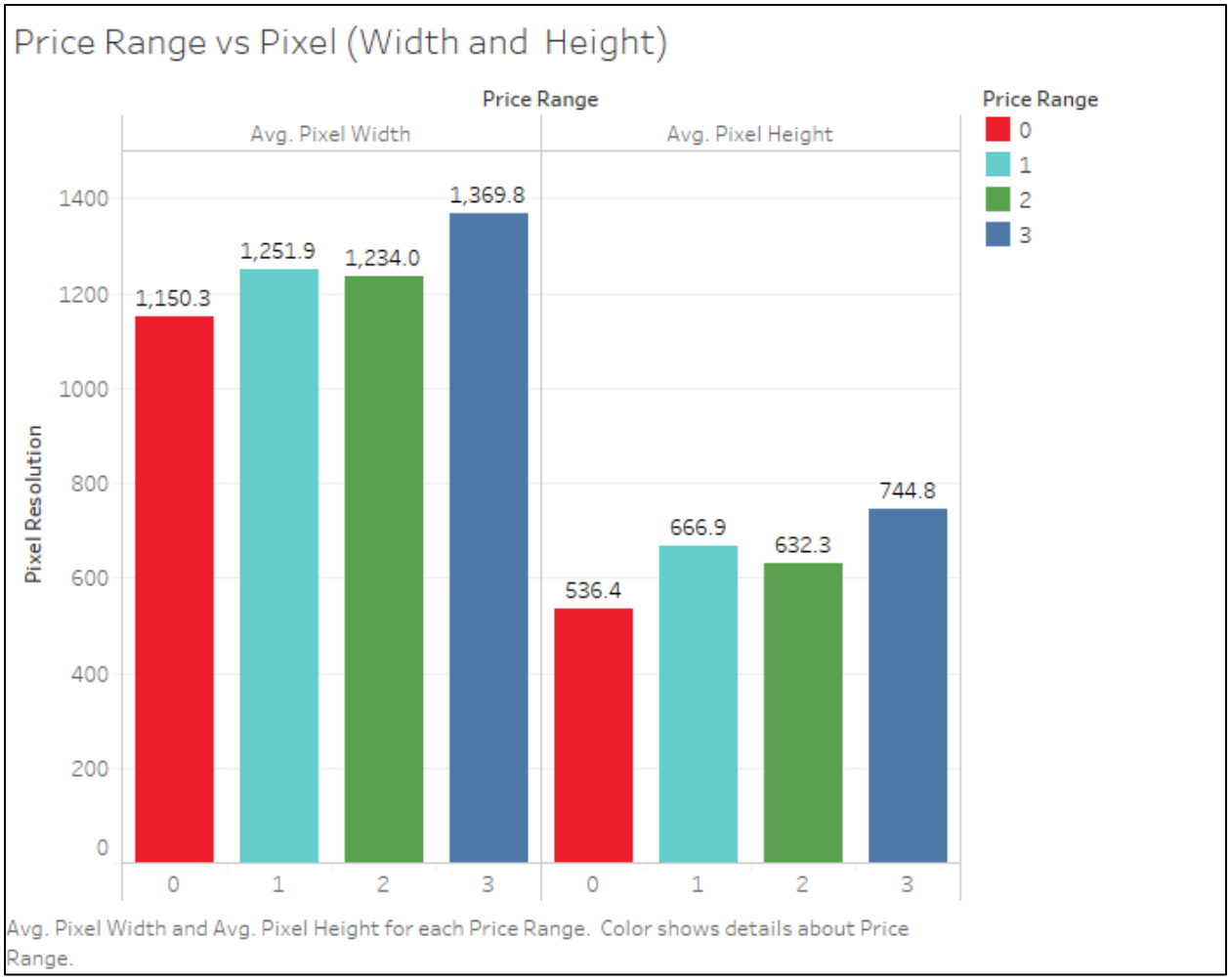
Hence, we can say that the Battery power can be a good attribute for our machine learning model.





For these graphs, we compare the price range with the number of cores and the Clock speed in the mobile phone. The general expectation would be that as the price range category increases from zero to three the number of cores as well as the Clock speed would increase. But in these graphs, we can see that the average number of cores is higher for category 0 and category 2 and lower for 1 and category 3. However, the variation across categories is not much. Similarly, for the Clock speed, the range across categories is from 1.55 to 1.49 which also indicates that the Clock speed across categories is constant. Hence, these attributes would not greatly contribute for the classification in our model.

Data Mining Project Visualizations



These bar graphs you represent the average pixel width and the average pixel height based on their price range categories. Though they seem to be linear i.e. the pixel resolution increases as we go from price range category 0 to price range category 3. Both the pixel width and the pixel height do show an exception when we move from category 1 to category two of the price range, where the pixel resolution for both height and width columns of category one is unexpectedly higher than that of category 2. We can consider these attributes for our project but they do not represent our notion strongly.