In [647]:

```
output:
```

I created a barplot that displays the proportion of unique ham and spam emails. To my surprise, there is only a small 10% difference between them, with around 80% unique ham emails, and 70% unique ham emails. Originally, I was thinking that there would be distinctive spam email content that I could use as my feature to distinguish spam emails, but this disproves my belief.