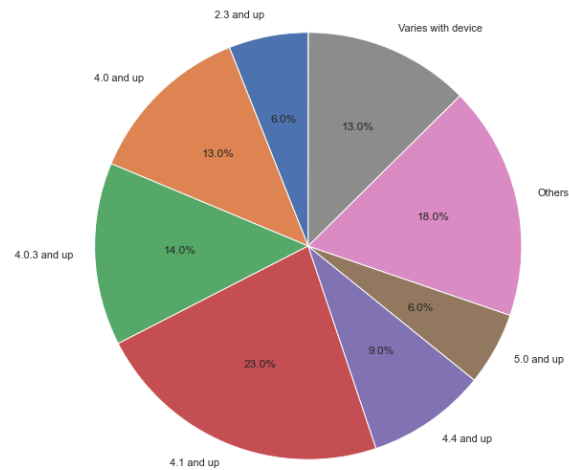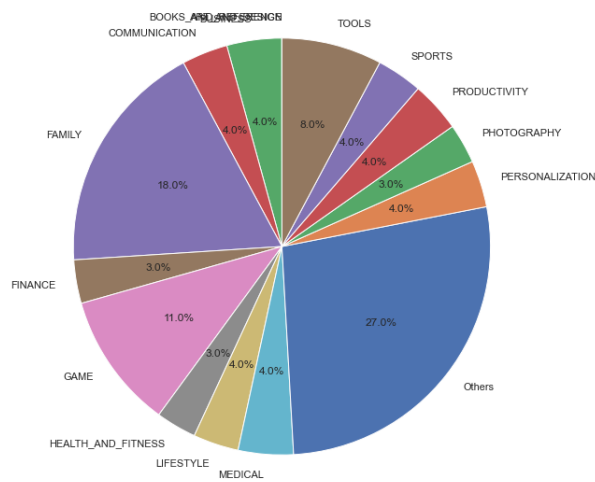김나경(2018-15185)

데이터과학

2020 년 10 월 5 일

Homework03

Step 4 Figure

In the figure above, we can see that 63% of the apps require 4.0 and up android version. In other words, people with outdated android version cannot use more than 63% of apps.

Step 5 Figure

18% of apps were under Family category, which took the largest pie except for Others, and 11% of apps were under Game category while others were distributed similarly. Others took 27% of the pie, which consists of categories under 3%. Here we can imply that there are various categories of apps and except for two major categories, Family and Game, apps are similarly distributed across all the other categories.

Step 6 Figures
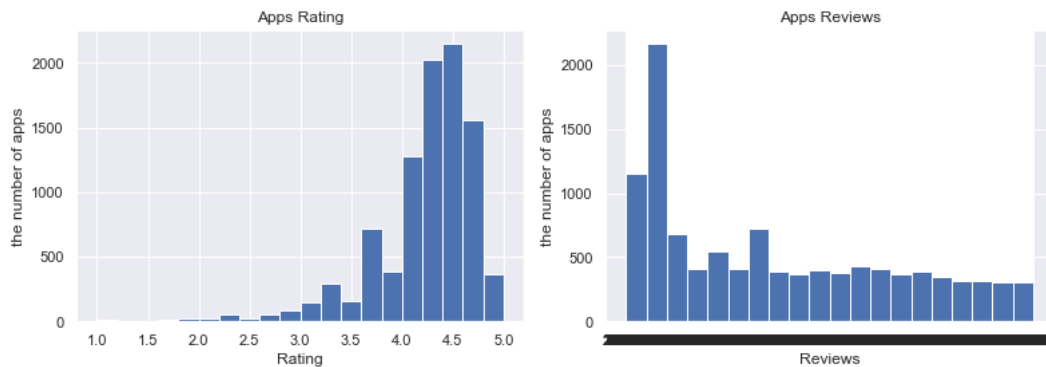
Step 6 figures, which are histograms, show the rating distribution as well as reviews distribution. With the first figure, which is a histogram related to rating, it is obvious that most of the apps are distributed among rating range 4 to 5 while there is no app with 1 rating. Here I imply that the users tend to rate highly and they might think around 4 point is an enough rating for average apps.

The next figure, a histogram with reviews, show that a lot of apps distributed in a low - ranged category. But as I looked at the data, there were a lot of reviews with only a few words, such as 'Good' while some are really long. It's necessary to consider the length of reviews because some of the applications nowadays require reviews to the users and offer them benefits inside apps.

Step 7 Figure
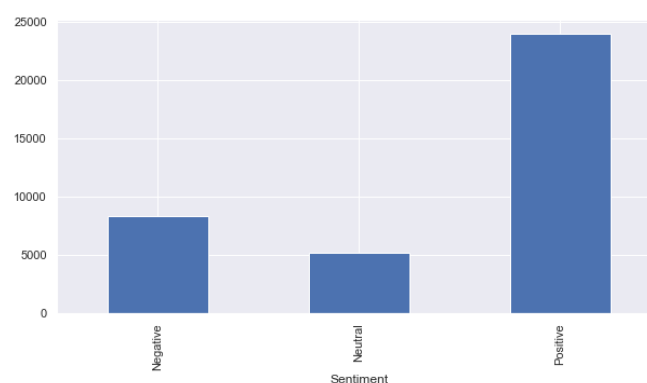
Unlike the histogram upward, we can see that there are much more apps with negative sentiment reviews than expected, which makes the histogram less reliable and less accurate. As I referred to before, we can anticipate that there might be some other reasons on rating the apps with high points(4 to 5) while users are not really satisfied with them. But still, we can see that more than half of the users have positive sentiments on the apps they are using.

Step 9 Figure
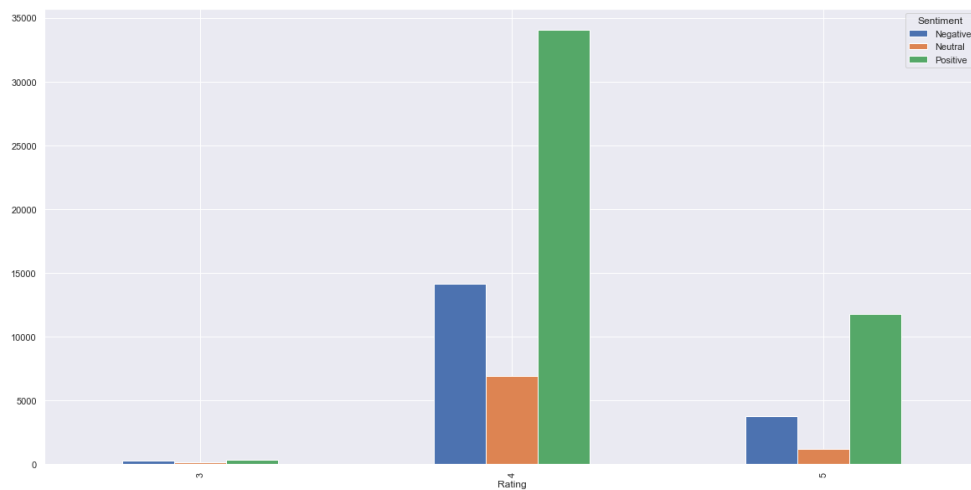
I think the last figure is what we do really have to focus on. This bar graph is drawn based on rounded rating and shows that there are a lot of people who have negative or neutral sentiments on the apps they are using while they rated them highly. It is reasonable that there are more reviews with positive sentiments on highly rated apps, so I won't consider it. But this figure shows that only considering ratings of apps is not reliable enough and implies that there must be some other reasons on rating. Based on my personal experiences, I saw some users rate an app highly and write bad reviews because they thought the developers read reviews with high ratings first.