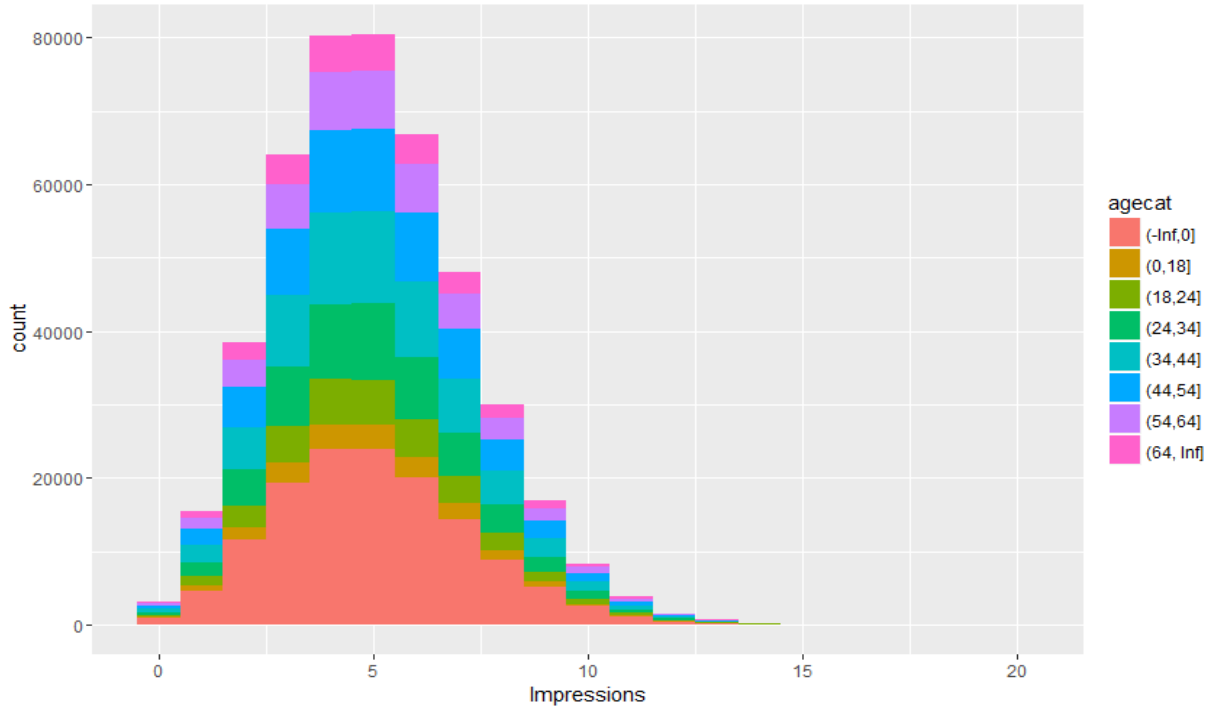
Problem 2: Simple EDA

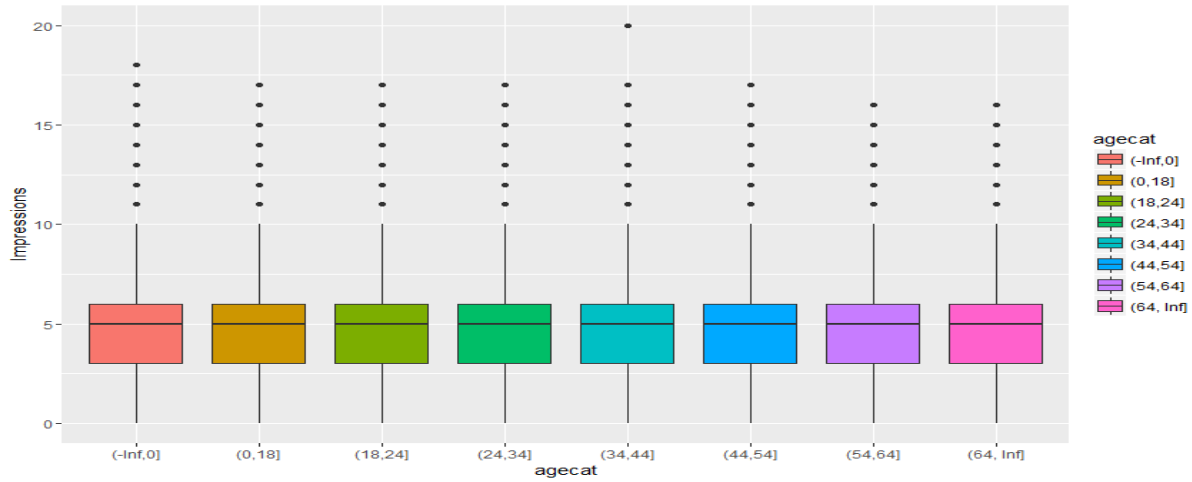
2a) Analysis was performed for one day of the 31 day data provided by the New York Times.

The plots and observations are shown below:

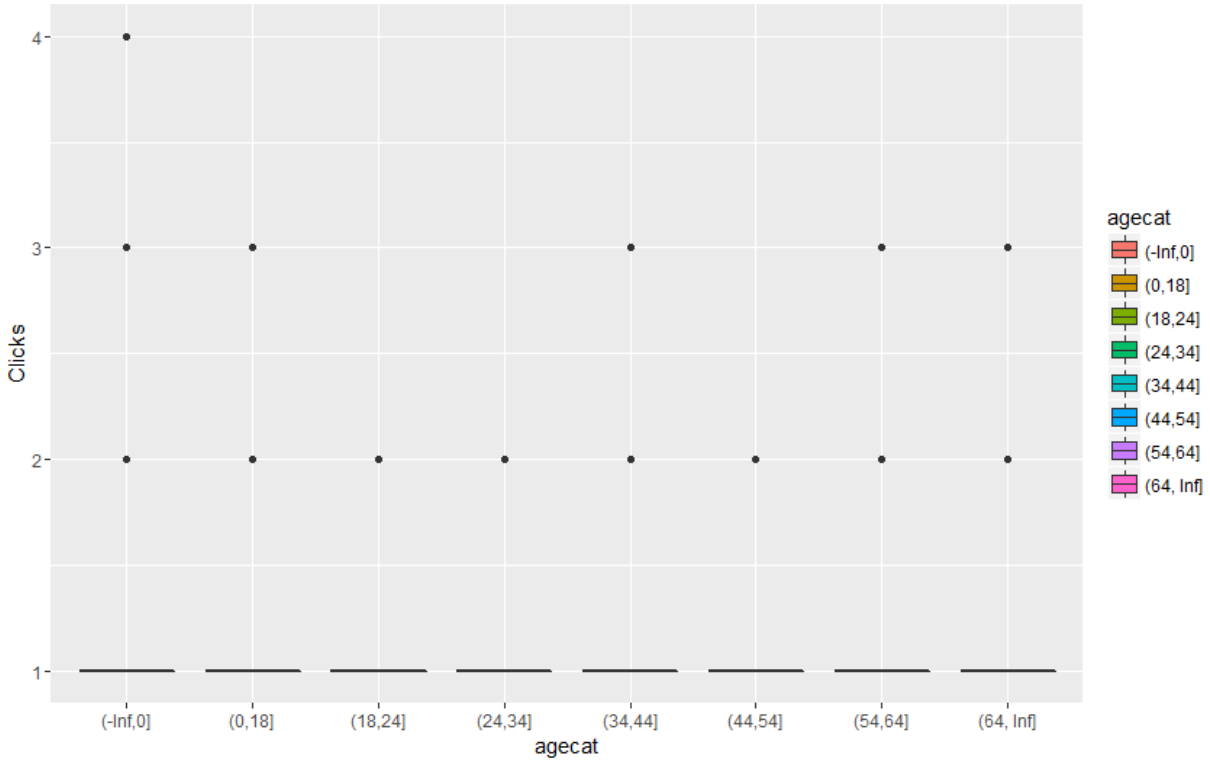
1) Impressions vs agecat



The above graph shows a histogram of the number of impression made by people in each age category. This helps us observe how many people have watched the advertisements and based on the age group, we can make the advertisements more appealing to the majority. We can make the same observations for boxplot below.



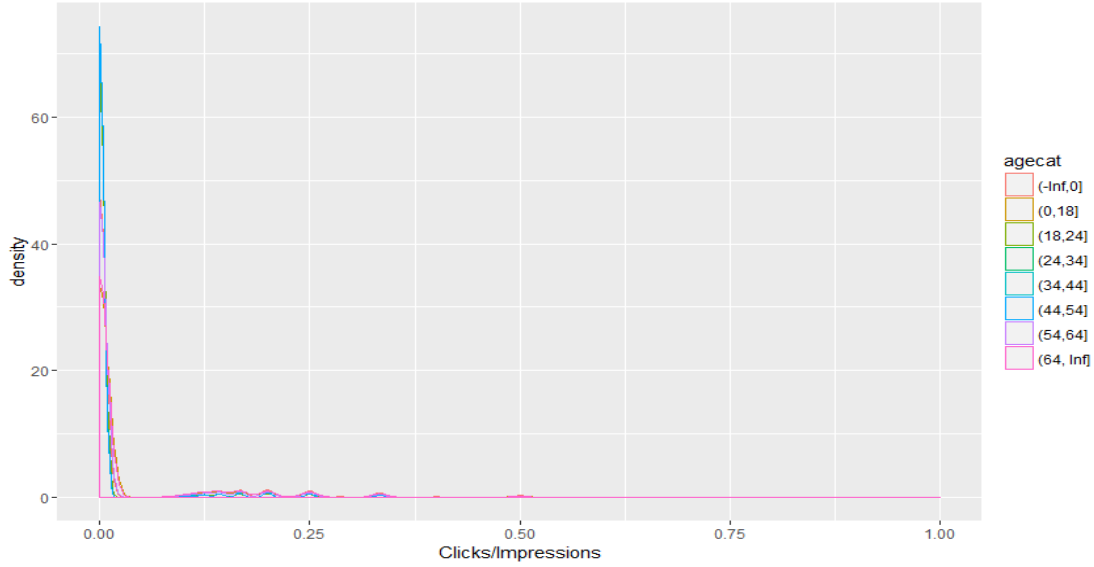
2) No. of clicks vs age category



We can see that, the number of clicks for every age group is very minimal on this day. For this, using the previous analysis, we can make the ads more appealing so that the number of clicks increases.

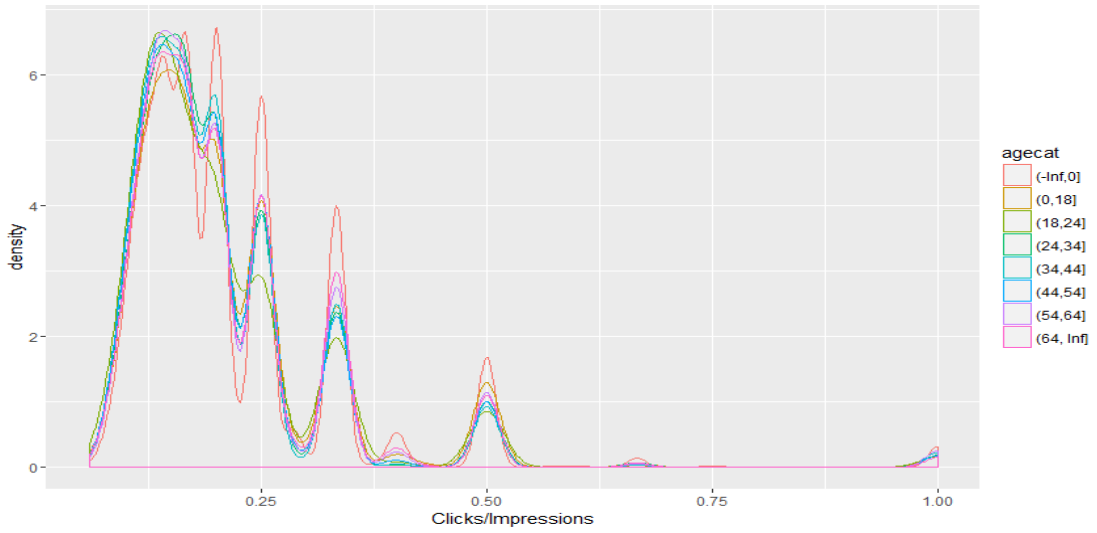
3) Clicks/impressions vs age category

In this graph, we can see that the ratio of clicks to impressions where the impressions are greater than zero, the density is less, which shows that the number of impressions are way larger than the number of clicks per ad.

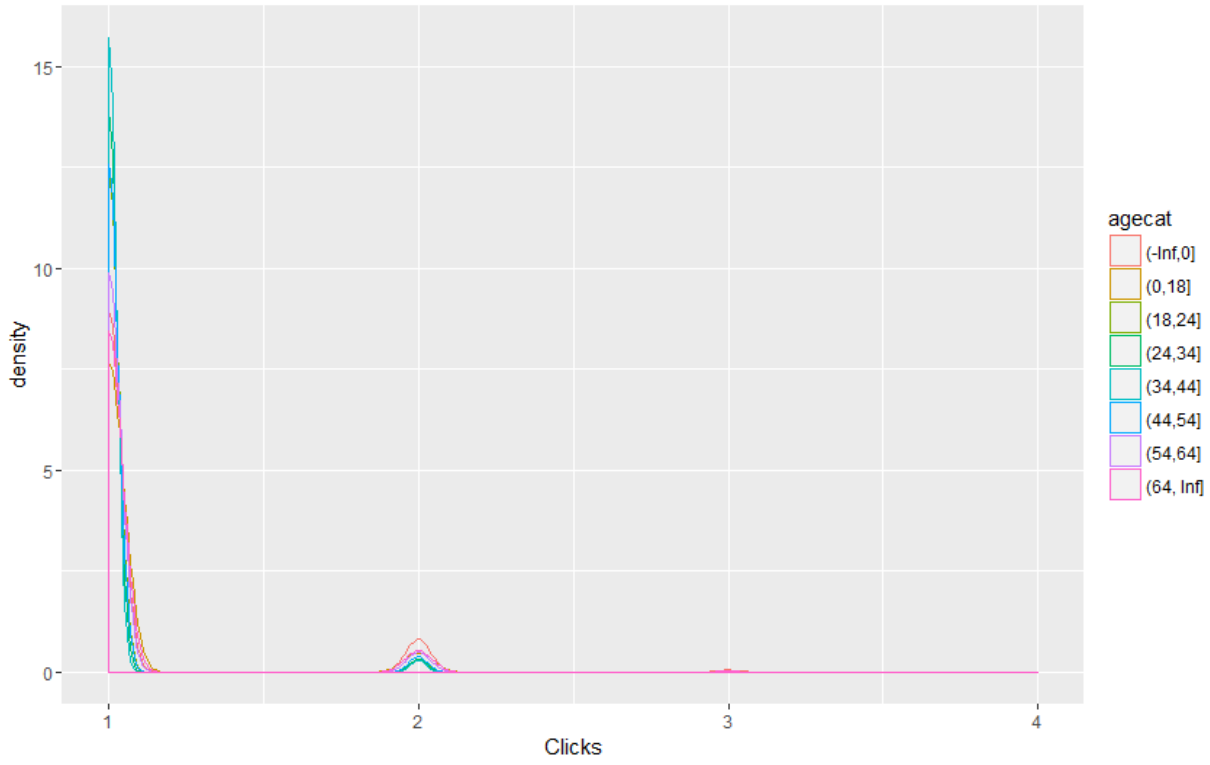


4) Clicks/Impressions vs age category:

In this graph we can see that the density is more since we have considered only those which have clicks greater than zero.



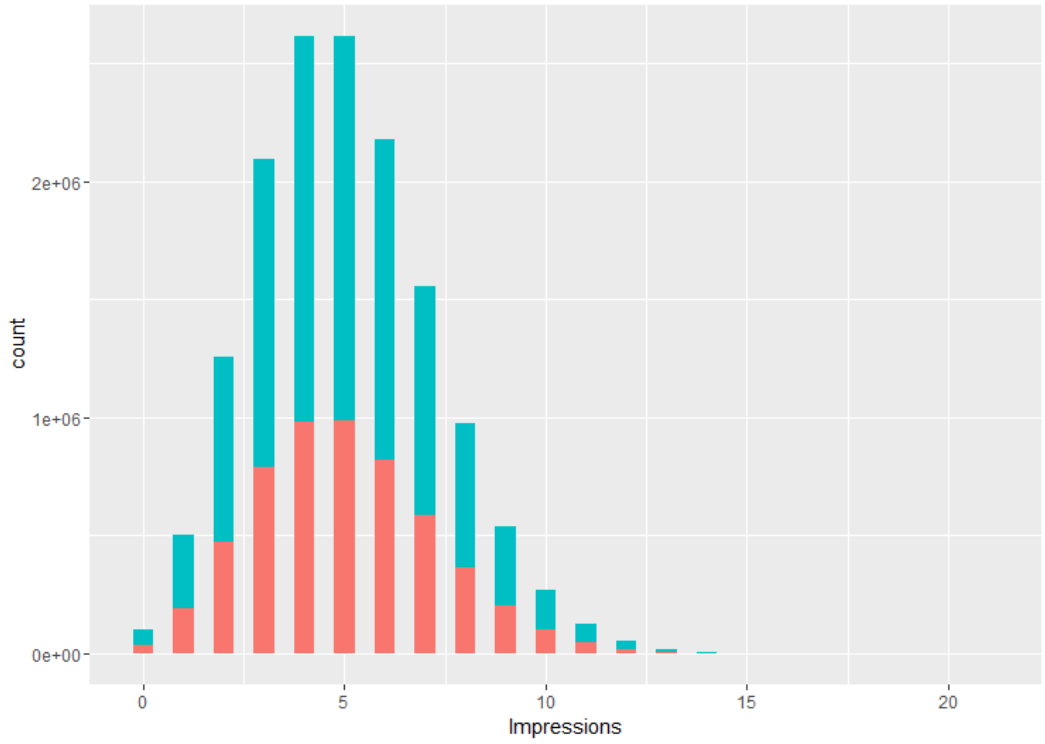
5) Clicks vs age category

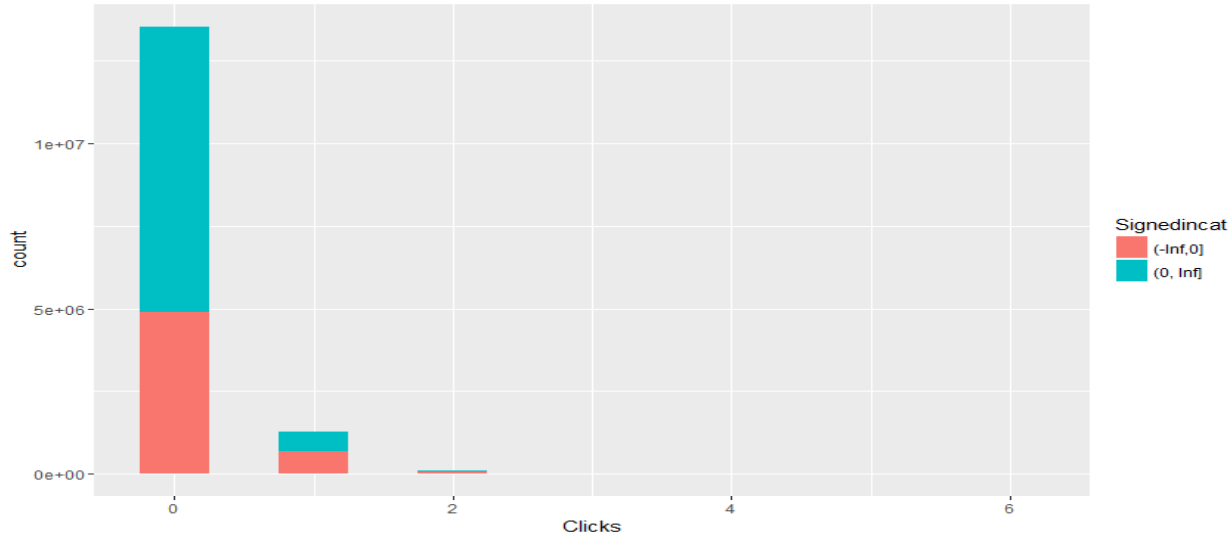


In this graph, like before, we can see that the number of clicks are very less for every age category.

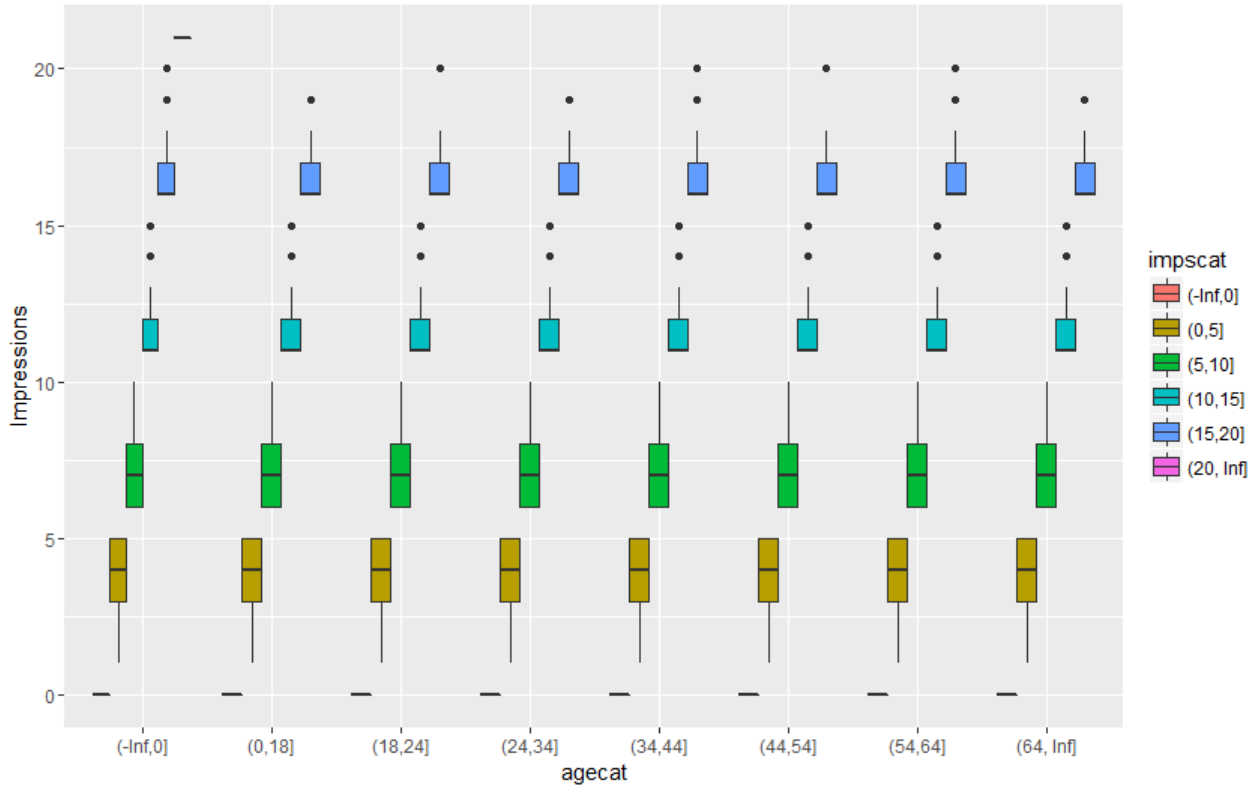
2b) The analysis was extended to include all the 31 days and some more additional observations were made as shown below:

The number of impressions vs the number of people signed in is plotted below and is seen that the number of people who are signed in are viewing the ads the most.

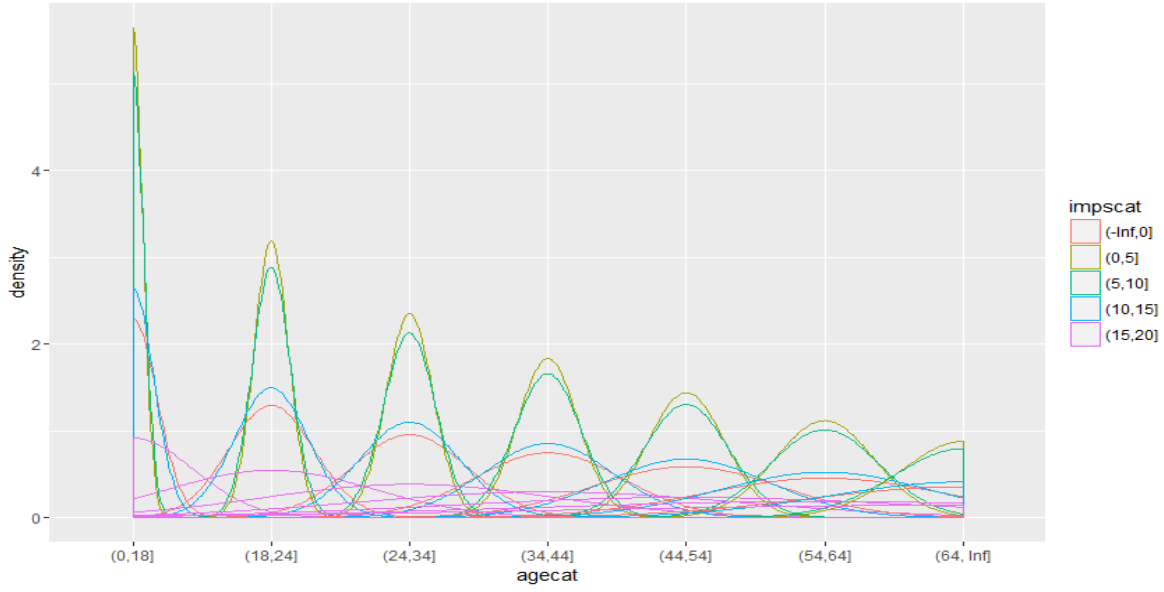




The number of clicks vs the signed in people shows that the people who are signed in have the most number of clicks.

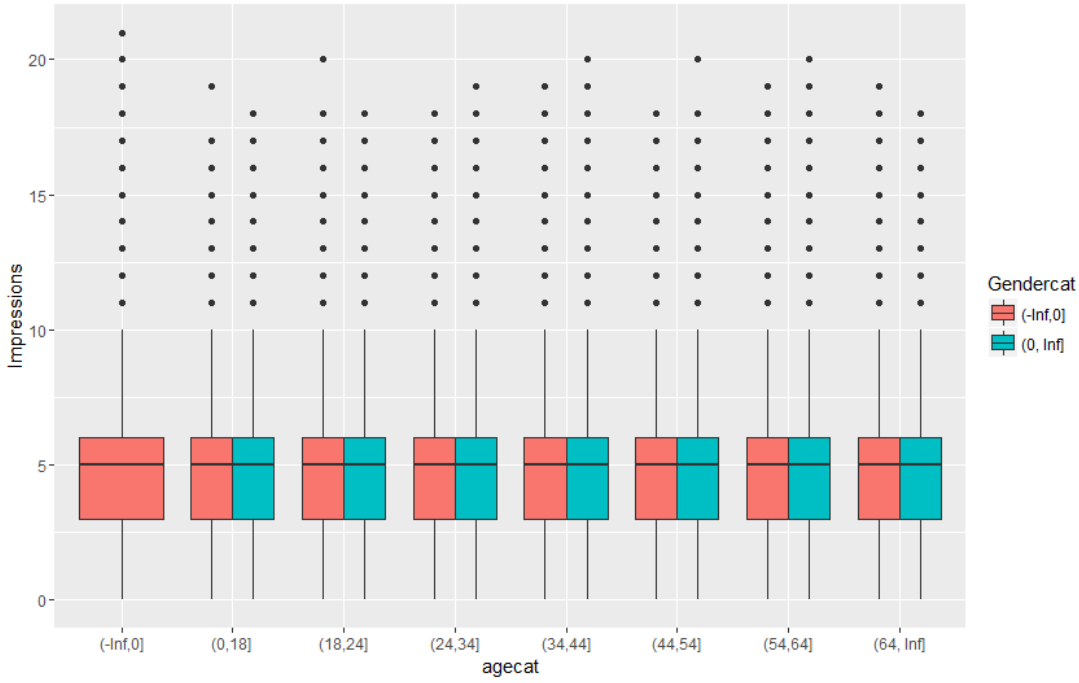


The above boxplot shows the number of impressions with the age category. The people of the majority age group should be targeted with marketing strategies appealing to them.

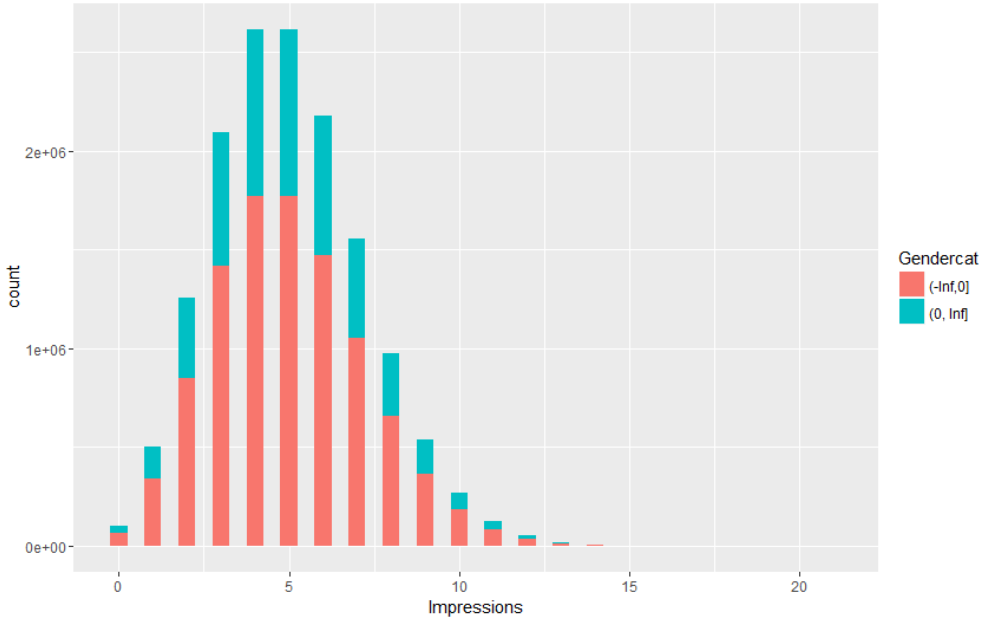


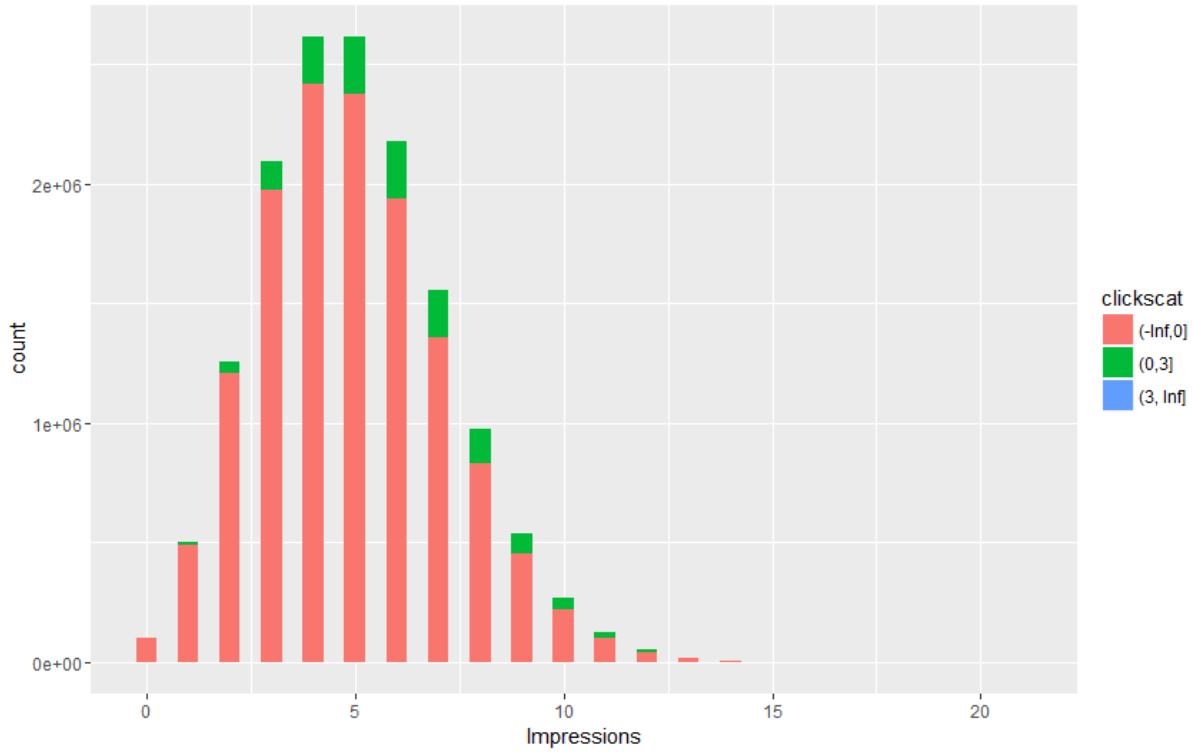
The density of the above graph shows us that a comparatively large amount of people in different age categories have an average number of impressions of the advertisement.

The graph below shows us that the number of impressions made by the males and females in the age categories descibed is alsmost the same. They all hav the same mean and almost same median.

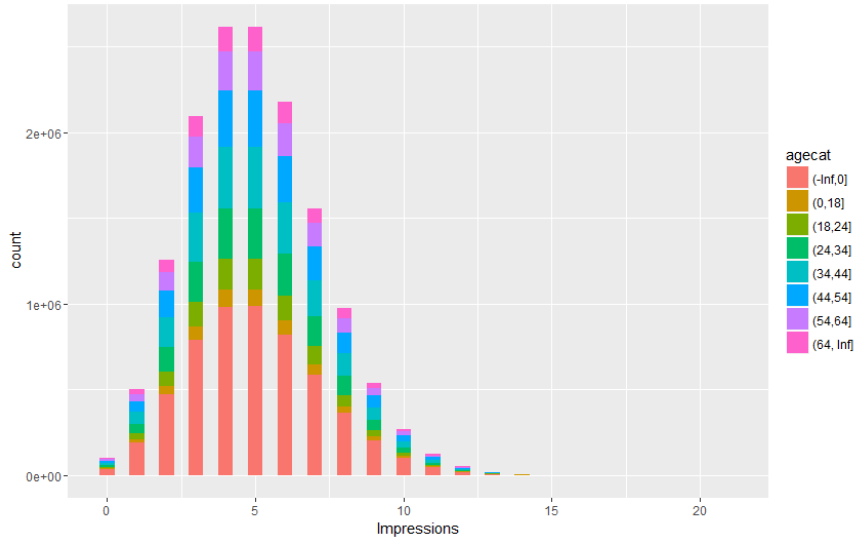


This graph shows us that the number of impressions for the females is more than males.





The first graph shows us that even though a lot of people viewed the ad, they did not click on it.



This graph shows us the distribution of the impressions among the age categories.