*The dataset Call\_Center.csv contains information for close to 33000 calls to a call center. Information included in this file includes: a timestamp, call-centers city, channel, city, customer name, reason, response time, sentiment, state, call duration in minutes, csat (customer satisfaction) score.*

I analyzed the data with my primary question being: What makes the customers happy?

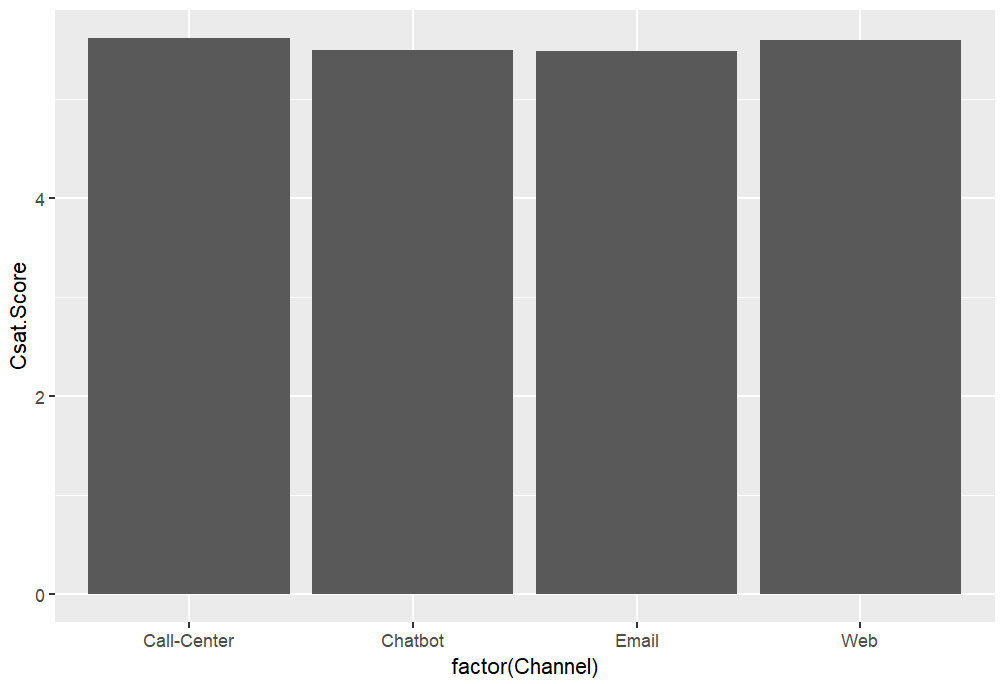
There were two scores of customer satisfaction in the data: Sentiment and CSat (customer satisfaction) score and I used both in my analyses. The CSat Score was a continuous variable, which would usually be better for analysis, but there were many N/A values in that column which meant that analyses using that score could only utilize some of the data. I therefore sometimes looked at Sentiment, sometimes at CSat Score and sometimes at both.

* *Analysis 1: Is the type of customer service channel significant to customer satisfaction?*

I first looked at the correlation between the channel and the CSat score, which meant that I had to use modified dataset, dropping the rows with an N/A value for CSat score.

The first step was to make a linear regression model with Channel as the predictor and CSat Score as the response variable. The adjusted R Squared of that model was 0.0003844, which is very low. This indicates that the model is not good, or in other words, the channel is not significantly correlated with the CSat Score.

I also looked at a bar graph of mean CSat Score for each channel, and it was consistent with the previous conclusion, with the means appearing very similar.



Just to double-check my conclusion that the channel is not significant to customer satisfaction, I looked at the correlation between Channel and Sentiment, which would not require dropping rows. For that, I used the chi-squared test, which compares categorical variables. The resulting p-value was 0.3756, far higher that .05, which again indicated that the customer service channel is not significantly correlated with customer satisfaction.

* *Analysis 2: Are some call centers better than others i.e. is the call center city significant in the sentiment?*

I used similar steps in this analysis as I had in the previous one. I first made a linear regression model between Call Center City and CSat Score. The adjusted R Squared value was -2.501e-05, which is extremely low, even lower than in the previous analysis. The chi-squared test between sentiment and Call Center City bore this out, with a p-value of 0.4075, which is a lot higher than the .05 which would indicate significance.

The conclusion of this analysis was that the Call Center City did not significantly affect customer satisfaction.

* *Analysis 3: What is the most significant factor for customer satisfaction?*

The previous two analyses showed that both channel and call center city were not significantly correlated with customer satisfaction. Still looking for the answer to what makes the customers happy, I ran a model containing all the factors that may affect customer satisfaction and looked for the factors that were statistically significant. The resulting model had a very low adjusted R squared, which means that it was a bad model. Only one of the factors was significant, one of the Call Center Cities.

The results were ultimately inconclusive, with it appearing that none of the various factors are significantly correlated with customer satisfaction.