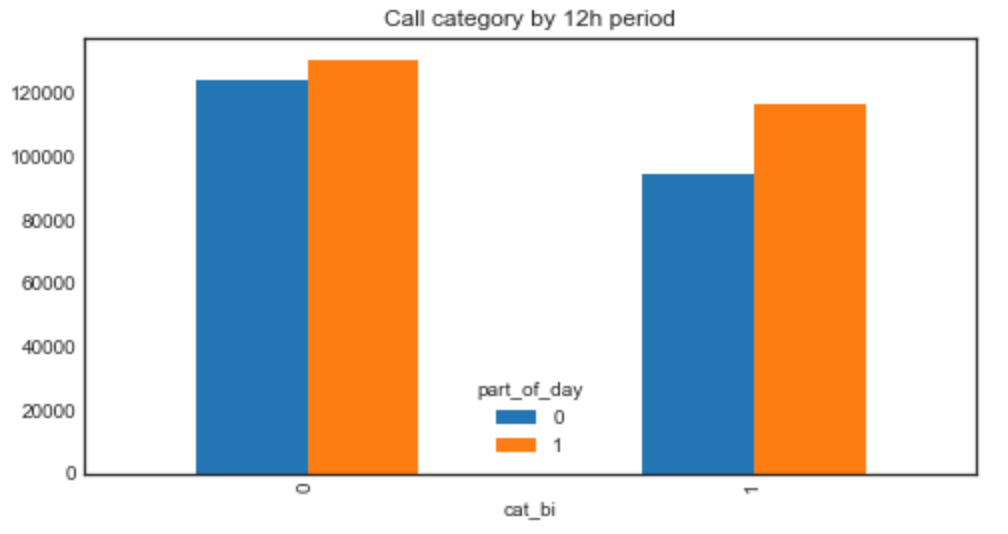
**Capstone project 1 – Apply inferential statistics**

In this project, we analyze 911 calls in order to predict the call category (e.g. prank, fire, assault, etc.). Previous EDA has identified that 911 calls of the category ‘prank/other’ are the highest attested in this data set.

Another important feature in making these predictions with regard to the call category is the time of the day in which the call is made. One way of analyzing this factor is to look at different intervals within the day. Here, we will divide day in two 12 hour intervals. The day interval goes from 6am – 5pm (1) whereas the night interval 6pm – 5am (0). In addition, the call category will also be divide into two groups, prank/other (1) and everything else based (0) on their frequency in the dataset, as show in the graph below.



Thus, we can consider the following hypotheses for testing:

H0: There is no statistically difference between the call category and the time of the day.

H1: There is a statistically difference between call category and time of the day.

First, we check for whether the data is normally distributed or not in order to choose the appropriate statistical test. We find evidence that the data does not come from a normal distribution (p=1.500274390620472e-27). Therefore, a non-parametric test will be used to test the hypotheses. The appropriate test for this problem is the Mann Whitney U test for two groups (‘day’ = 1; ‘night’ = 0).

The result of the Man –Whitney U (z-score = 26022190674.0; p = 8.222773661880775e-155) test provides grounds to reject the null hypothesis that the two categories of 911 calls are not different with regard to their frequency based on the time of the day. In other words, we can conclude that 911 call of certain type are more frequent during the day and less frequent during the night and viceversa.