In [ ]:
```
#The data that we are currently investigating and exploring is FCC.csv and IRS.CSV.
# We wanted to get a better idea of what this dataset contained. After exploring it we
# saw that it mainly gave the total number of different types of unwanted calls for time
# zipcode distribution from fcc datasets. This exploration helps us to know if there is correla
# between the number of unwanted calls based on geolocation from fcc datasets and income range
# irs datasets.The goal is to comeup which kind of machine learning model can give us better
# prediction fot the number of unwanted calls can citizens receive based on their income range
```

In [ ]:
```
#***EDA Scope ***
#The scope of this data doesnt directly show whether or not the citizens in a negative or posi
# from the fcc data that we collected, but we tried to see some features can correlate from irs
# from this we can see or understand what kind of distribution of unwanted calls in our region
# , as well as certain states, happening frequent unwanted calls than other states. This allows
# #to ask more questions and try to see if their are relationships between certain locations o
# unwanted calls and certain location have range of income conditions occuring.
```

In [ ]:
```
#**** ML ****

#We currently do not have a complete Machine Learning analysis so far,
# but we do have some ideas that we want to implement because we have a lot of features within
# The current Machine learning analysis that we are working on is to find a relationship
# between if citizens get to the same levels of unwanted calls, for example
# what is the density location for unwanted calls happening as the first fcc datasets containi
# zip, location, state. Since we need to know if there is unwanted calls happening with
# in the range of citizens range income. we train two features which can retrieve from
# fcc datasets and irs datasets.
# Finally, we decide logistic regression to classify if unwanted calls occuring in certain ran
# income.
#to use modeling such kind logistic regression help us estimate a probability
#of falling into a certain level of income range. inorder to get accuracy in
#our prediction we need two features
#the first features from fcc datasets to based on the location of unwanted
#calls received and on the second feature from the irs we retrieve the
#range of income, i believe since both features are independent we need
#to check whether it is correlate or not by using scattered plot
#after that we can apply logistic regression model to predict how many
#number unwanted calls can receive with in the range of income
```