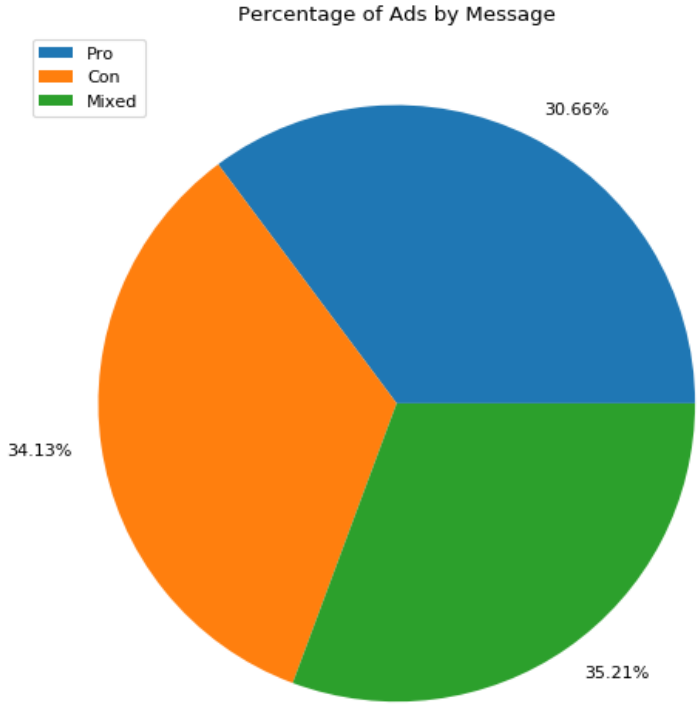
2016 Political Ad Message Analysis

What type of Ads to people think Americans want to see? In this Analysis we will explore a data set from the US 2016 national and statewide elections including 300,000 Ads aired and descriptive elements for each Ad including: location, program_type(news or not news) sponsor_type (who paid for the Ad), and AirTime (the time of day the Ad was shown). We will focus on the overall "Message" (Pro, Con, or Mixed) of the Ads. Let's take a look at the column names to understand the data available to analyze.

What is the distribution of Ad's 'Message' throughout our entire data set?



From the graph above the messages appear to be evenly distributed throughout our data.

Now we will begin grouping our Ad's by their other descriptive elements. Below you can see the count of Ads by program type and message:

message	con	mixed	pro
program_type			
news	84591	84944	72388
not news	43517	47207	42703

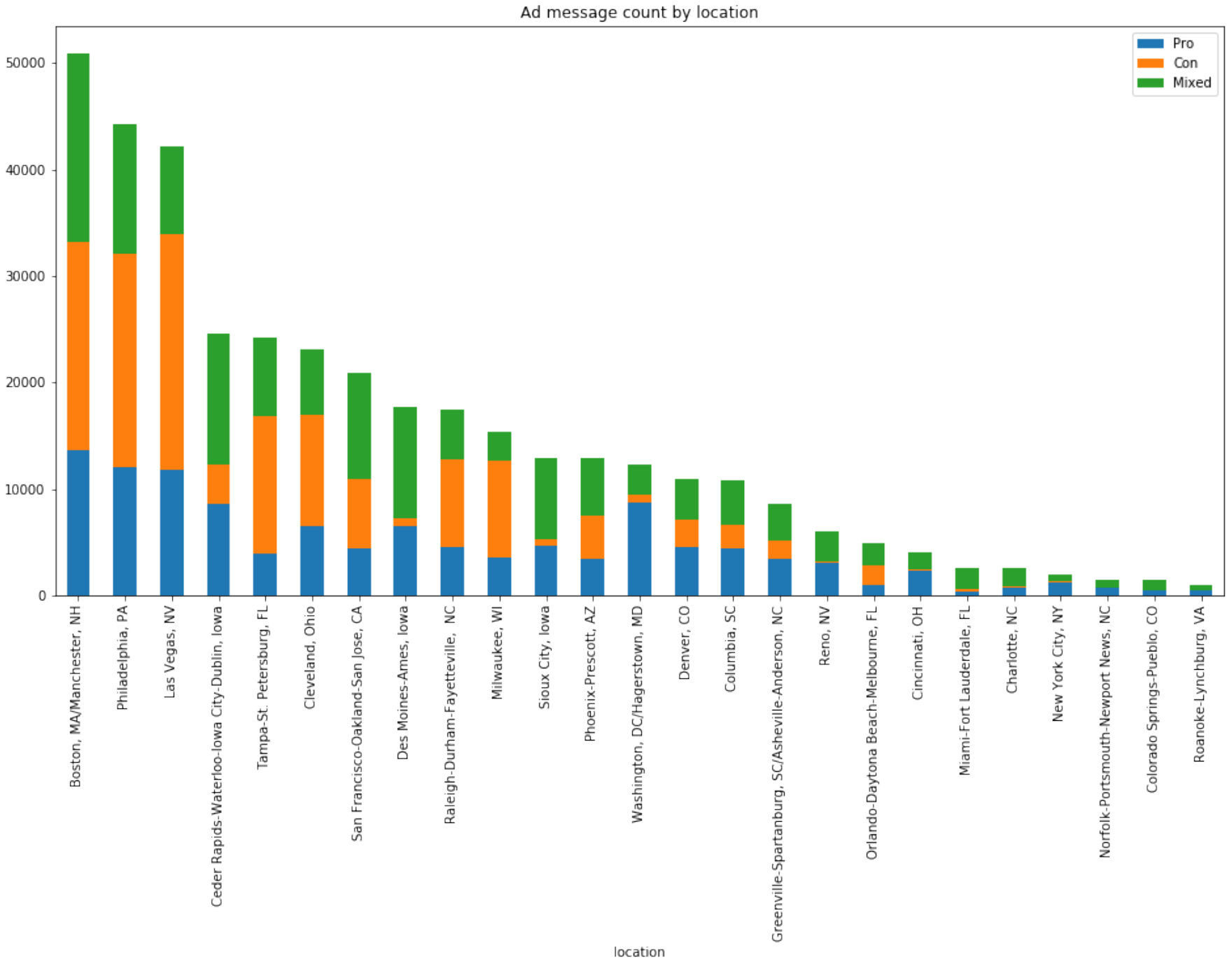
We can apply the chi squared test to take a look at the relationship between program_type and message. Our null hypothesis: there is no difference among the number of ads with con, mixed and pro message and between program type (news vs not news) and hypothesis A: there is a difference. The p-value returned from the chi squared test is 8.08 x 10^ -58 and is less than alpha = 0.05. We conclude that the null hypothesis is rejected.

Will this hold true for other elements of our Ads? What factors are drive those differences? Now we will aggregate our Ad start_time data into our AirTime field. Start times from 4am-12am will be morning; 12am to 9pm will be afternoon, and 9pm to 4am will be night. Below you can see the count of Ads, by AirTime and message.

message	con	mixed	pro
AirTime			
Afternoon	36972	36999	33649
Evening	52131	55089	47400
Morning	39005	40063	34042

We apply the chi squared test with a similar null hypothesis: there is no difference among the number of ads with con, mixed and pro message by airtime (Afternoon, Evening, Morning) and hypothesis A: there is a difference. Since the p-value (4.73 x 10^ 14) is less than alpha = 0.05, we conclude that the null hypothesis is again rejected.

What is the average number of Ads shown in a location? Does the location of the Ad or the volume of Ads shown have an effect on the message of the Ad? Now we will look into the message of Ads shown by location:


Ad message count by location
Pro
Con
Mixed
50000
40000
30000
20000
10000
0
Boston, MA/Manchester, NH
Philadelphia, PA
Las Vegas, NV
Ceder Rapids-Waterloo-Iowa City-Dublin, Iowa
Tampa-St. Petersburg, FL
Cleveland, Ohio
San Francisco-Oakland-San Jose, CA
Des Moines-Ames, Iowa
Raleigh-Durham-Fayetteville, NC
Milwaukee, WI
Sioux City, Iowa
Phoenix-Prescott, AZ
Washington, DC/Hagerstown, MD
Denver, CO
Columbia, SC
Greenville-Spartanburg, SC/Asheville-Anderson, NC
Reno, NV
Orlando-Daytona Beach-Melbourne, FL
Cincinnati, OH
Miami-Fort Lauderdale, FL
Charlotte, NC
New York City, NY
Norfolk-Portsmouth-Newport News, NC
Colorado Springs-Pueblo, CO
Roanoke-Lynchburg, VA
location

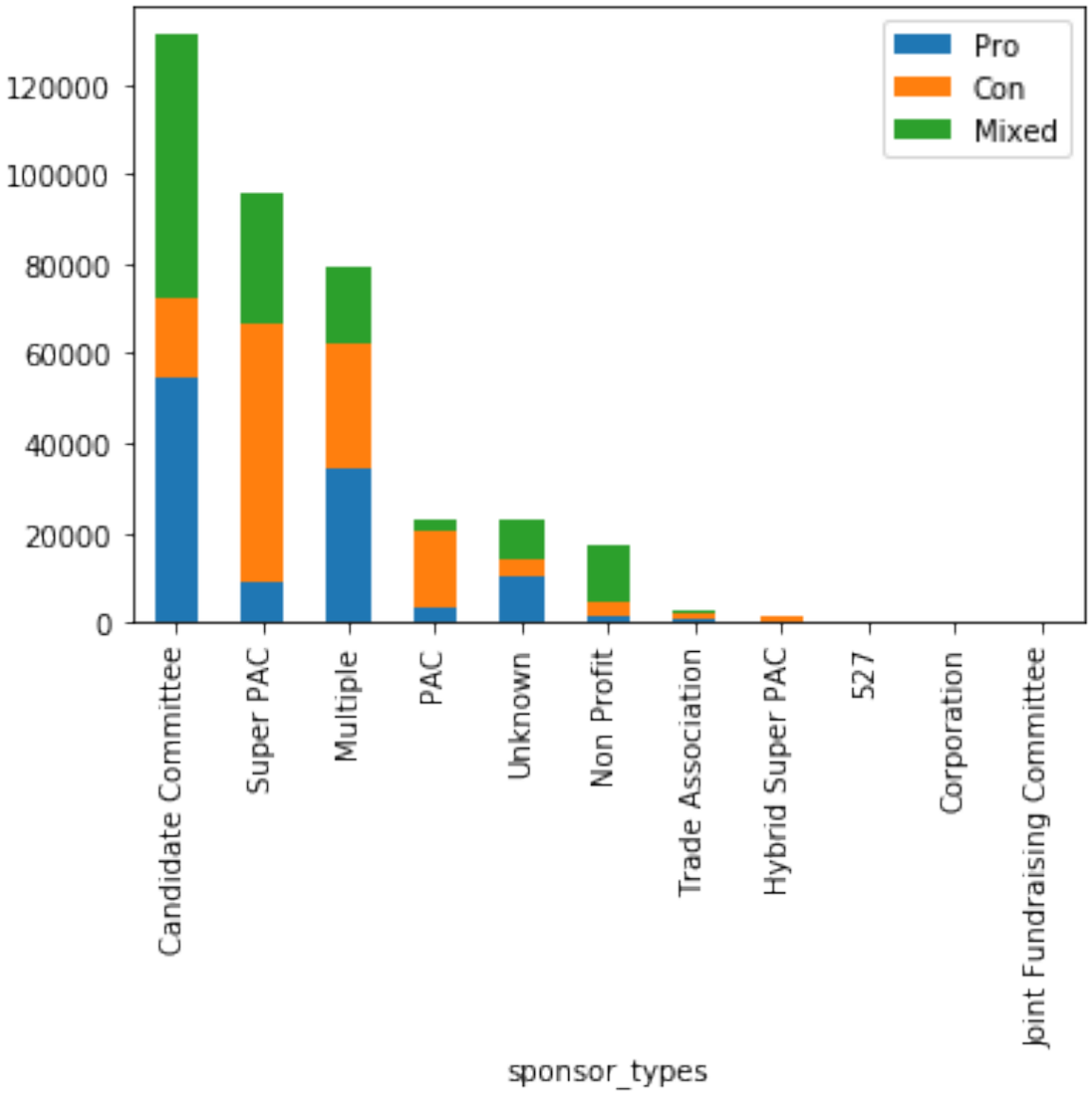
The chart above shows us a few things. First the bottom four locations where fewer Ads are shown do not have any 'Con' Ads. If people have a limited amount of resources to display Ads, do they focus on Pro and Mixed Messages. As more Ads appear by location a larger percentage are allocated to Con Ads. This could indicate that after a campaign is satisfied that it has sufficiently described the positives of its campaign they begin to allocate resources to attacking their opponent.

One exception is Ads from cities in Iowa. Ad Campaigns in Iowa appear to be very positive compared to the rest of our Locations. Are Iowan's culturally less inclined to respond well to Con Ads? It could be due to their proximity to more rural cultures. Looking at other small cities we have: Denver, Columbia, Greenville-Spartansburg/Ashevill-Anderson, Raleigh/Durham, and Milwaukee all also have lower occurrence of Con Ads. It seems appears that smaller and more rural cities are not targeted with as many Con Ads as a whole. This supports our hypothesis from our observations in Iowa. There are some exceptions; let’s examine Milwaukee and Raleigh-Durham. Why do they have such high occurrences of Con Ads relative to other small cities? It is possible that competition in the campaign together created the negativity; Once one side begins using Con Ads does this escalate the usage by their opponent creating a Con feedback loop?

Let's take a look at the more populous cities: Boston, Philadelphia, Las Vegas, San Francisco-Oakland/San Jose, New York, Washington DC. Conversely to smaller cities, larger more populous cities appear to have larger uses of Negative Ads. The exceptions are New York and DC which have more Pro Ads than Con, however, these cities have significantly fewer total Ads shown than the rest. In our larger cities with high volumes of Ads shown the majority have A significantly higher occurrence of Con ads. Could this be opposite to the rural trend we saw earlier in our smaller cities? Campaign Advertisers may view people in big cities as more receptive to Con Ads than those in rural areas. This reinforces our hypothesis that culture could determine the message Advertisers use.

We have looked at location and have found some potential trends with our Advertisers. Now let’s look at the different types of Advertisers(sponsors) to see if how message is distributed among them:

Ad message by sponsor Type



We can see significant differences here in the Message based on the sponsor type. Candidate Committee and Non Profit funded Ads are predominantly Pro and Mixed while Super PAC and PAC funded Ads. If a Candidate is paying for his own Ads it would make sense that they would want to keep a Pro Message to preserve the image of running a clean campaign. Non Profits have to consider their donors, a Pro Message is less likely to offend their donors. On the other hand, PACs and Super PACs are vehicles that corporations and donors use to funnel large amounts of money to a specific candidate or cause. They are not run by the Candidate and thus are less connected to their Campaign. This could explain why they are inclined to sponsor more Con Ads. They do not have an image to protect, only a message to convey or a goal to achieve i.e electing a candidate.

While we started with an even distribution of Ads across all message types. We looked at the type of show (News or Not News) and AirTime, statistically, to see if we should expect a difference in message type based on those characteristics. Our analysis showed us that yes there is a difference based on the characteristics. We then dove into the details of the data and we discovered that depending on the Location of the Campaign or the people paying for it we discovered that the distribution of message type can vary greatly. We discovered a potential trend in Advertising targeting based on location size and culture. Additionally we looked into the types of Advertisers responsible for paying for these Ads and addressed the motivations that would lead them to advertise with a certain message type.

To explore further I would want a larger data source covering more Ads in more campaigns to see if the trends we discovered hold or more trends emerge across the country. We would need this larger data set to provide recommendations based on this data and would need to research which candidates won each race and by what margin to determine if there was any correlation between winning a race and the characteristics of Ads shown in our data. Additionally I would want to dig into the strategy, goal, and motivation of these Advertisers(Campaigns).