

Analyzing Election Ads In Select Markets From 2016 Elections

W200 - Project 2, Nov 2022

Jonathan Phan, Leanna Chragchian, Noor-UI-Ain-Ali, Aska Zaman

Github Repository : UC-Berkeley-I-School/Project2 Ali Phan -Chragchian Zaman

Datasets: All Aired Ads and All Unique Ads Oct 2015 - Dec 2016

INTRODUCTION

The 2016 elections were a contentious period of U.S. history. The Political TV Ad Archive serves as a database that collects information on political ads, such as the amount of times an ad has been fact checked. This exploratory data analysis project focuses on these political ads to uncover insights about the who, what, and when of these ads.

DATA

The Political TV Ad Archive, a project of the Internet Archive, archives political ads and tracks ad airings in select markets in the 2016 elections. The data set provides details about airings of ads on TV, giving information about when and where they aired. The metadata included are network, `airing_id`, market, program, `program_type`, `start_time`, `end_time`, `archive_id`, `embed_url`, sponsors, `sponsor_types`, race, cycle, subjects, candidates, type, and message. Documentation of each variable is available at the archive.

The primary dataset used in this report is an inner joined dataset between all political ads and all aired political ads. Each observation is a single airing.

There are 375350 observations with 18 *relevant* variables. We define the variables primarily used in our analysis here:

- **wp_idenfifer**: Uniquely assigned id per ad
- **network**: Which network aired the ad
- **location**: The market where the ad was aired
- **candidates**: The candidate(s) featured in the ad
- **subject**: Topics the ad touched upon
- **message**: Noted ad message about a candidate. (supportive = “pro”, critical = “con”)
- **candidates**: Candidate the ad supports
- **start_time**: Time and day the ad aired
- **end_time**: Time and day the ad ended
- **air_count**: The number of times an ad aired

RESEARCH QUESTIONS

1. Where were the ads aired?
 - Did a particular candidate air more in some areas than others?
2. Who aired the most ads?
 - Did candidates air ads against each other often? If so, what are the trends?
 - Which candidate got fact checked the most? Were their “pro” or “con” ads fact checked more for each of these candidates?
3. When were the ads aired?

- Who aired the most during certain time periods?
- How long were the ads for during specific hours of the day?

INITIAL DATA EXPLORATION AND CLEANSING DATA ANALYSIS

The dataset downloaded from the Political TV Ad Archive has 375350 rows and 19 columns from Oct. 2015 to Dec. 2016. Each observation in the data frame is a single political ad. All these ads have all the factors described earlier.

We will use this dataset to gain insights into which candidates have a higher amount of ads that mention them when aired. Although certain individuals are mentioned more in ads, this does not always mean it is positive for that candidate. Further in the report we will see which candidates are shown more often in positive or negative sentiments.

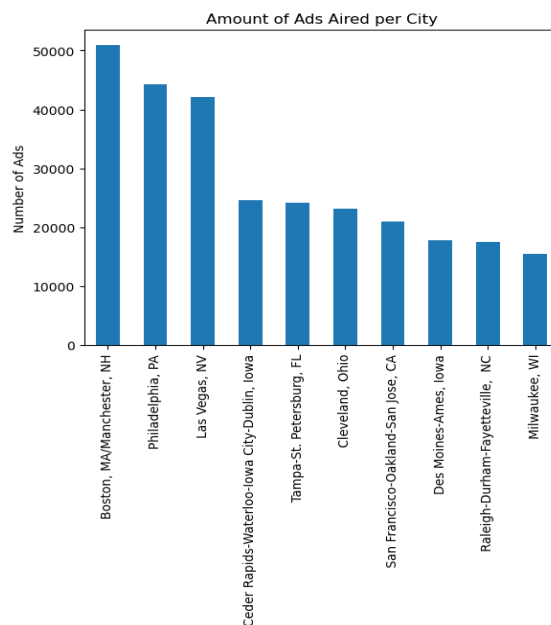
We cleaned our dataset by ensuring our time variables were in the correct datetime format. This included casting our “start_time” and “end_time” variables to a datetime format. We also noticed that there was one typo under the “sponsors” column. More specifically, there were two values very similar to each other “Donald J. Trump for President” and “Donald J Trump for President.” We corrected this by changing one to the other.

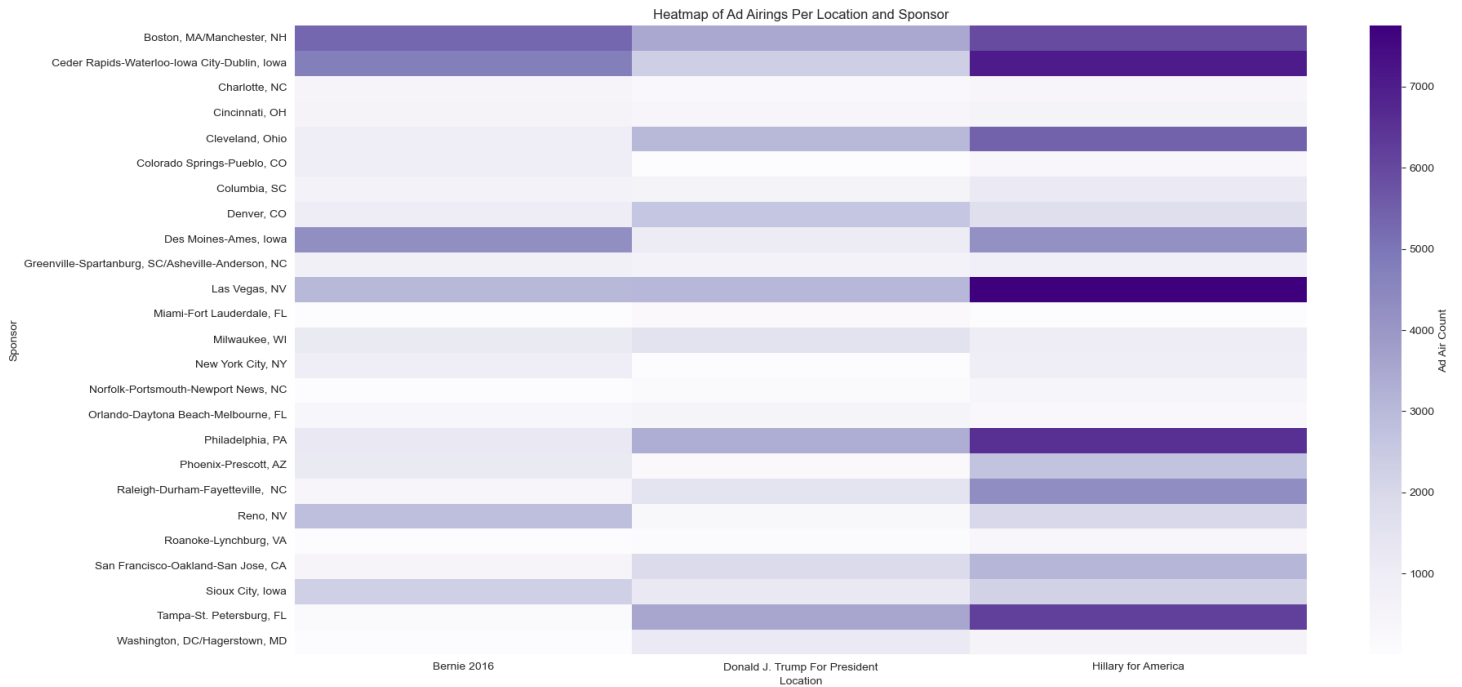
	sponsors
Donald J Trump For President	22677
Donald J. Trump For President	10493

Finally, to ensure we are able to analyze candidates with the most amount of ads we made sure to take the top three candidates that had the most ads aired about them. In many instances we specifically analyze Donald Trump and Hillary Clinton as they were the front runners in the 2016 election.

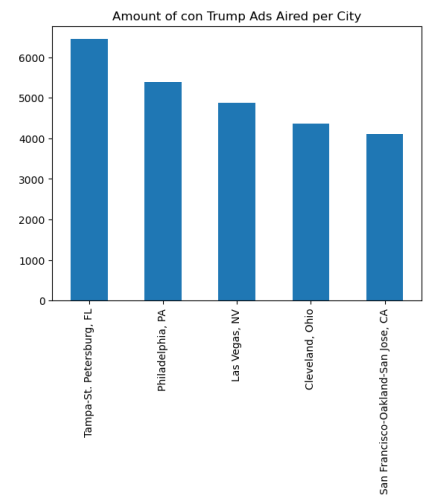
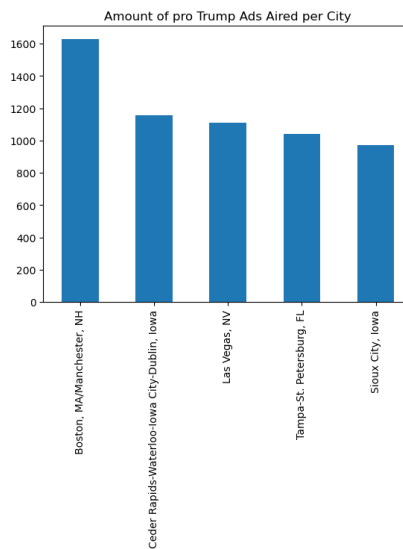
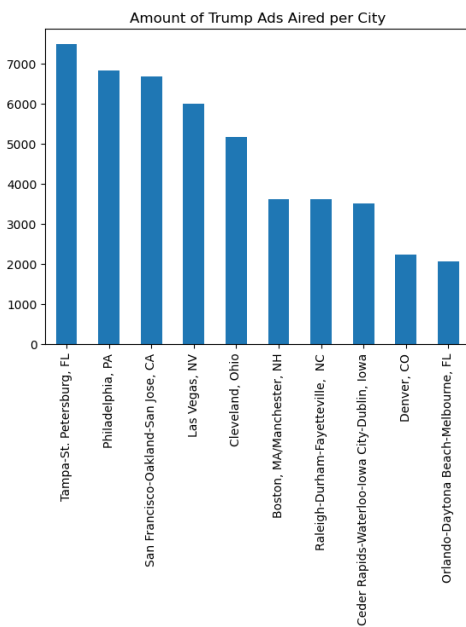
WHERE?

We start with a quick look at the top cities within the U.S in terms of the pure number of political ads aired. The top ten cities are shown to the left and it allows for us to get a high-level view at which cities are most commonly targeted with political ads. An assumption prior to this analysis can be made that metro regions with the largest populations would have a higher amount of ads. However here we can clearly see that political ads tend to be aired the most in metro areas within swing states. New York, Chicago, and the L.A metro areas do not appear in the top ten. We can see the Boston Metro area had the highest number of ads with 50920 recorded followed by Philadelphia and Las Vegas. After the first three cities we can see a steep drop off in the number of ads however the rest of the cities continue to be in states that would be considered crucial to election results such as Iowa.

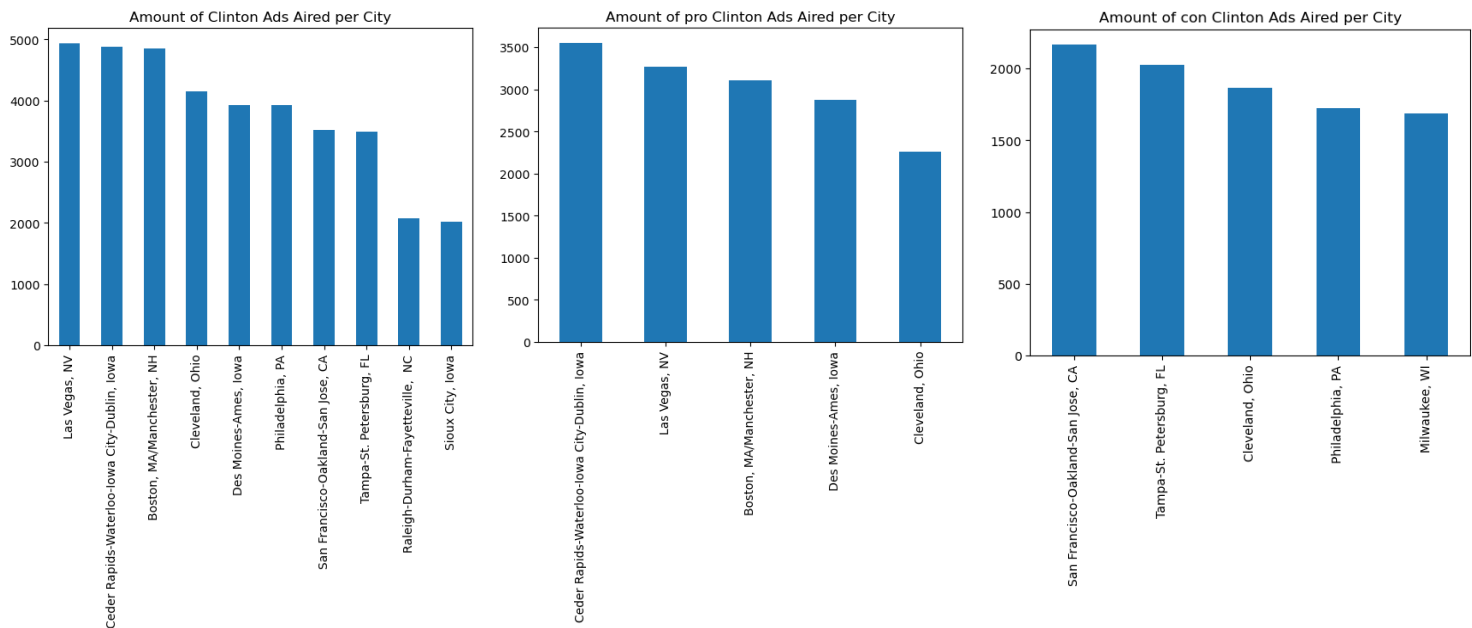




We decided to break the ad airings down further by specifying the candidate in different cities. From this, we can see that the three most popular campaign sponsors include Hillary Clinton, Bernie Sanders, as well as Donald Trump. From the heatmap above we can see all the different cities included in the dataset as well as the respective density in terms of the amount of ads that were run about that candidate. Again we can see that the most popular cities such as Las Vegas, Philadelphia and Cedar Rapids have the darkest cells.

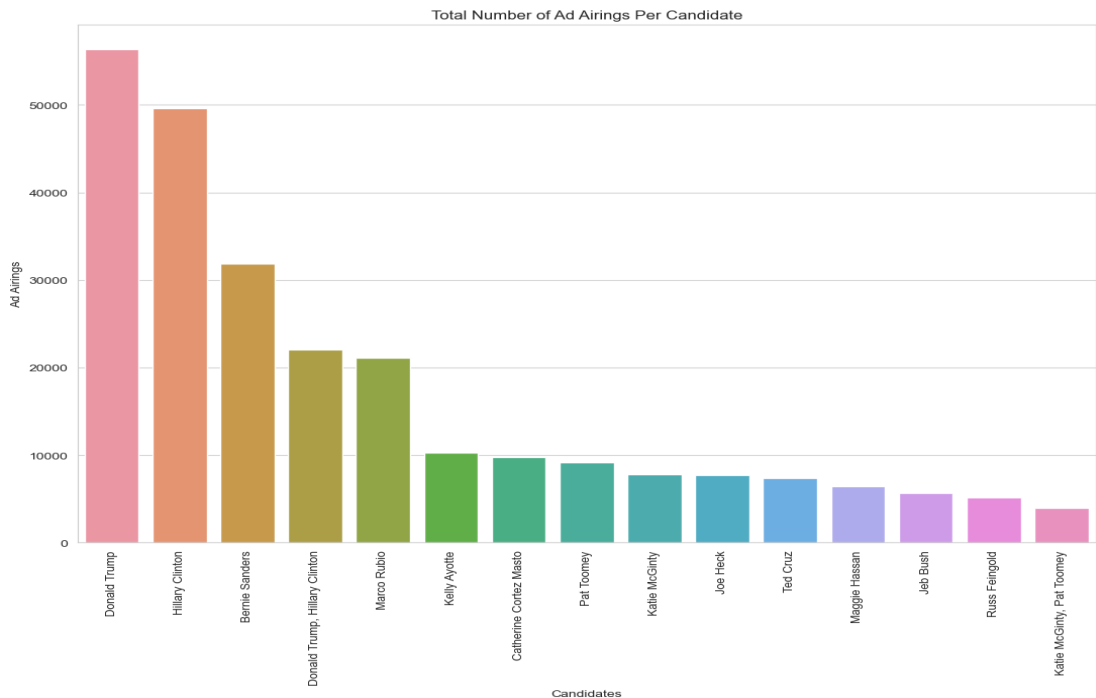


Going further into looking at the front runner candidates, we can see that Tampa, Philadelphia, and the SF Bay area appear to have the most amount of ads regarding Trump being aired. This count does include both pro and con ads for the candidate. Further, the ads can be separated into whether the airing had a pro or con sentiment regarding the candidate the ad was about. It appears that the Tampa, Florida area had the highest number of ads relaying ads that were con Trump.



Further expanding on the front runners, Hillary Clinton had the highest number of ads that aired about them in Las Vegas as well as in Cedar Rapids Iowa. When we take a look at which areas had the most con ads aired about Clinton, the Bay area came in first. This may be due to the assumption for it to have a higher percentage of people who may vote blue which may be a reason to have targeted ads in the San Francisco Area.

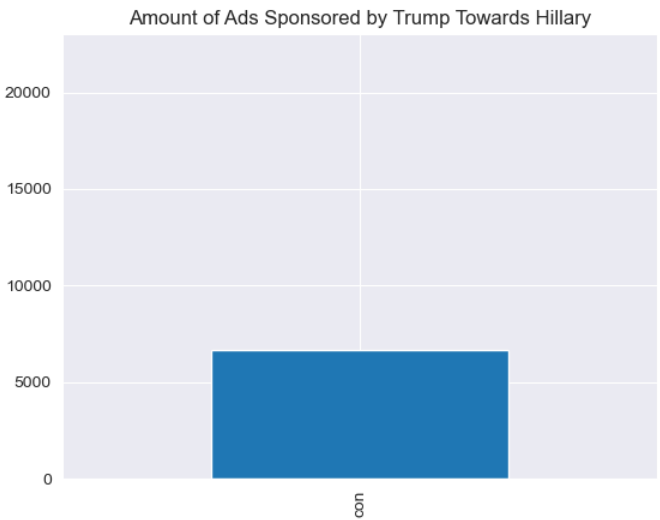
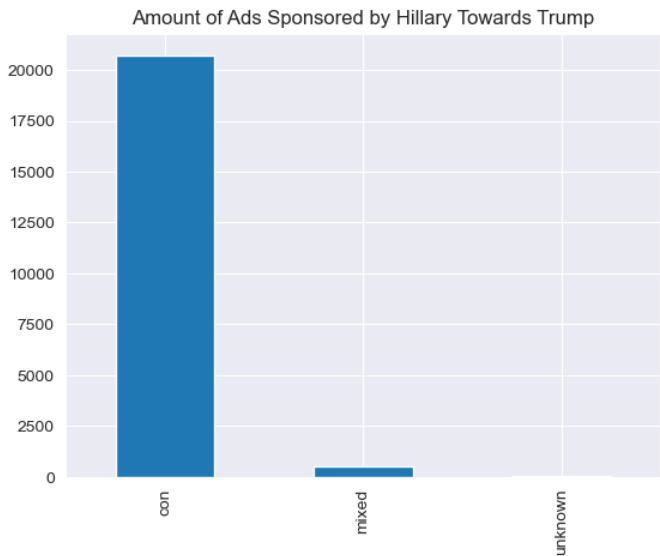
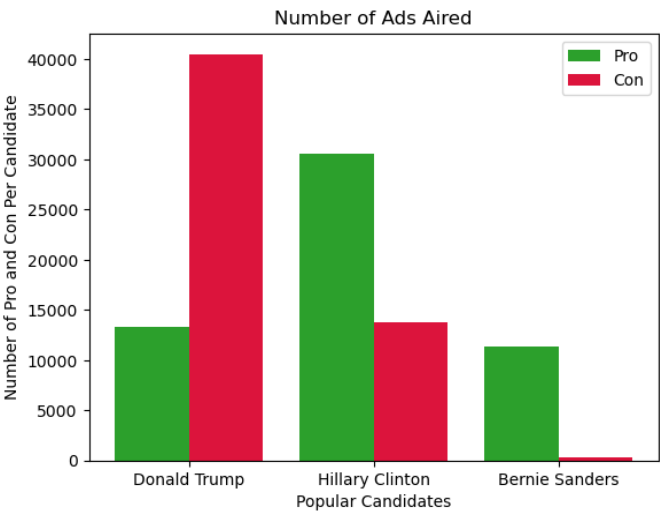
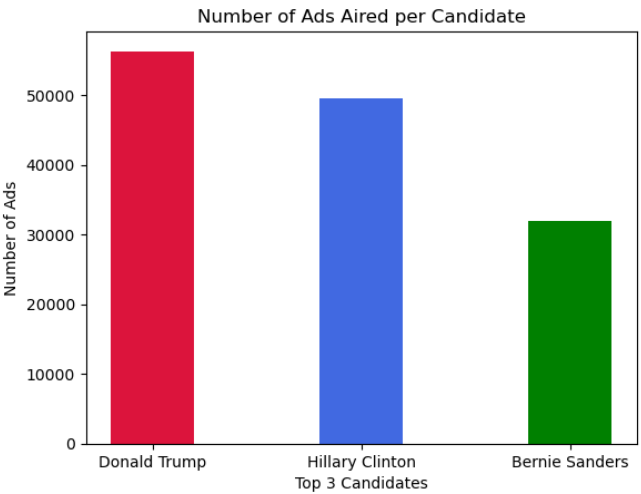
WHO?



Creating a barplot of the top 15 candidates featured in ads tells us that Donald Trump had the most ads aired about him. We were curious to see if this was always the case and further explored this concept by analyzing it between different election periods.

We can see that the top three candidates who appear to be mentioned the most in political ads are Donald Trump with 56,297 ads, Hillary Clinton with 49,568 and Bernie Sanders with 31,886 in that order and this can be aligned with our expectations relating to who was successful in the primaries and other metrics.

Among the three most commonly aired candidates, Bernie Sanders has the highest proportion of pro vs con ads with 11,354 pro ads and only 263 con ads. On the other hand Trump had the highest proportion of con vs pro ads with 13,368 pro and 40,493 con ads. This highlights the high proportion of ads that would target Trump and show their campaign in a negative or “con” position. It is interesting to see the much higher proportion of negative ads that aired when we see the other two popular candidates had much more pro ads aired about them. This can reflect the massive amounts of media attention both positive and negative that a front-runner candidate deals with when campaigning.



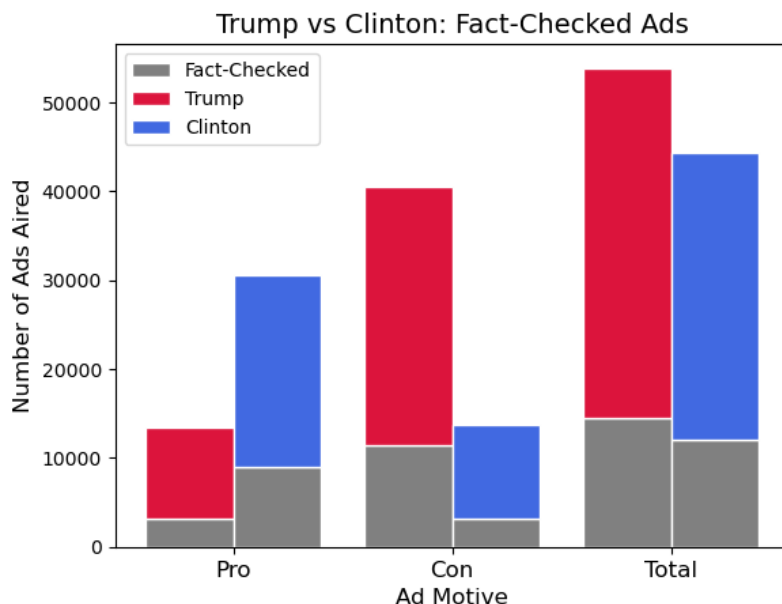
Interesting to see:

- Hillary owned sponsor “Hillary for America” aired many ads that had a negative view on Trump
- Trump owned sponsor “Donald J. Trump For President” aired many negative ads about Hillary, but not as much as she did of him

Ads Featuring Both Trump and Hillary and The Messages Per Sponsor			
candidates	message	sponsors	sponsors
Donald Trump, Hillary Clinton	mixed	Donald J. Trump For President	11086
		Hillary for America	8653
	pro	Hillary for America	3

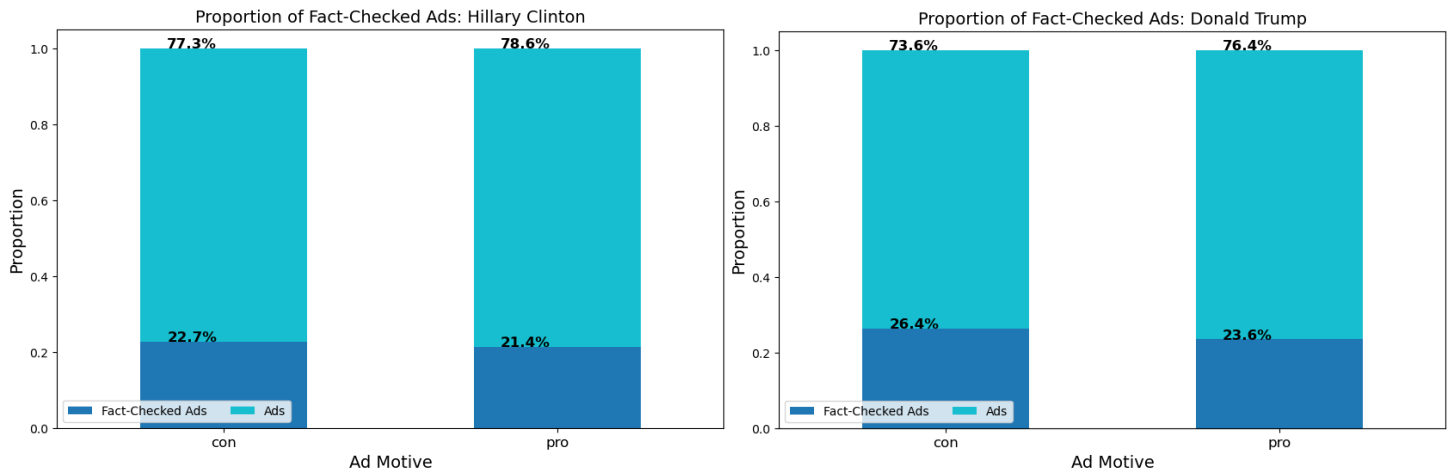
- Aired ads that included both Trump and Hillary were proportionally sponsored more by Trump’s own sponsor

Which candidate got fact checked the most? Were their “pro” or “con” ads fact checked more for each of these candidates?



	Candidates	Fact-Checked Ads
0	Donald Trump	56297
1	Hillary Clinton	49568

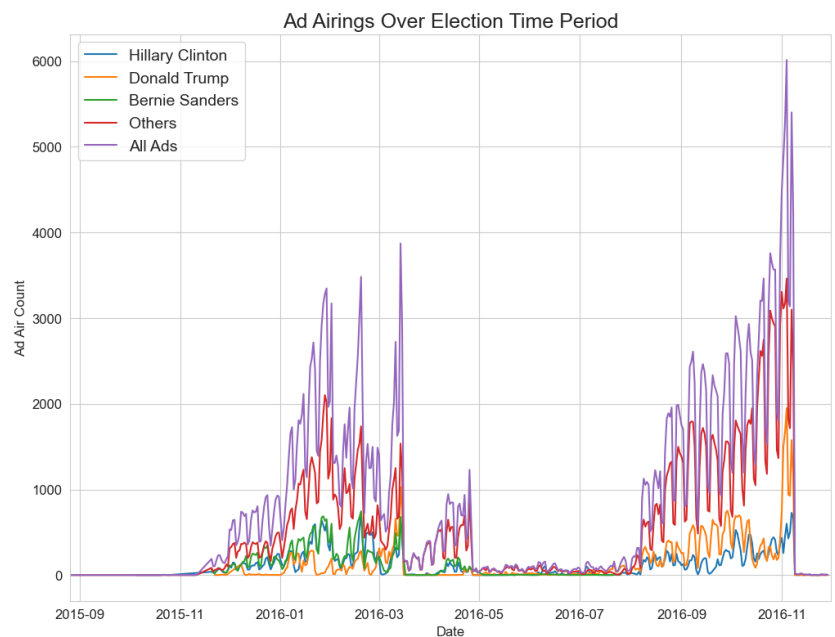
The bar graph above shows the number of fact-checked ads among all aired ads for the top two candidates, Hilary Clinton and Donald Trump, grouped by the message of the ad. The data specifically references the number of ads that were fact checked at least once. While Donald Trump advertisements received a greater number of fact checks in total, it is alarming to see the number of ads that circulated the country with possible factual discrepancies. It is also interesting to note that anti-Trump ads received more fact checks than anti-Clinton ads, and pro-Clinton ads received more fact checks than pro-Trump ads. This could possibly lead to an assumption of some sort of bias in the fact-checking process.

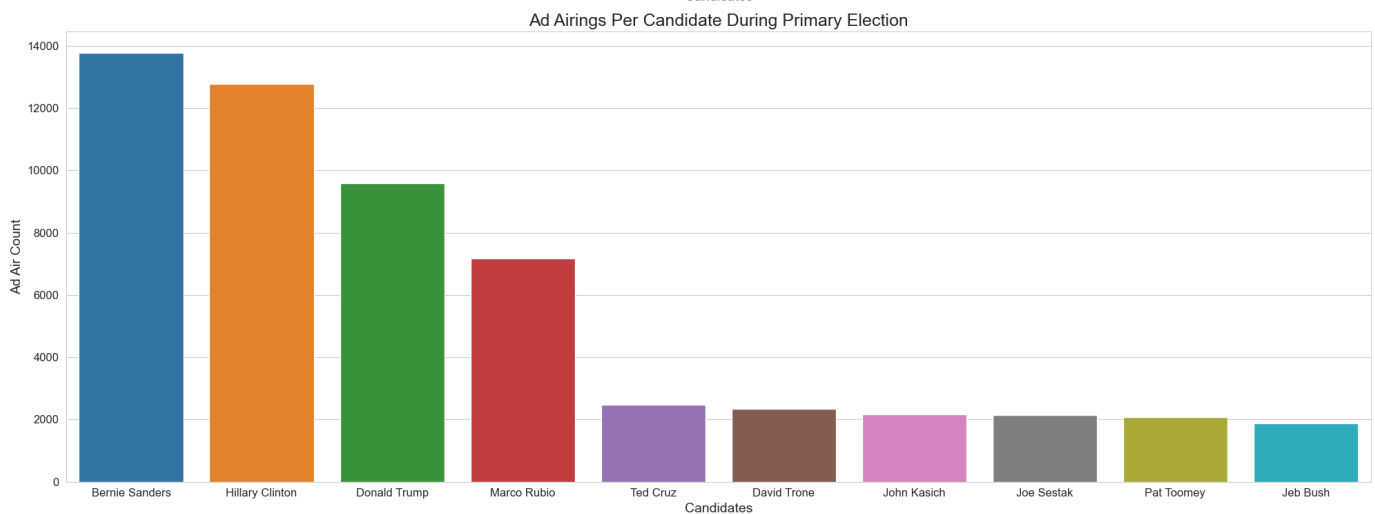
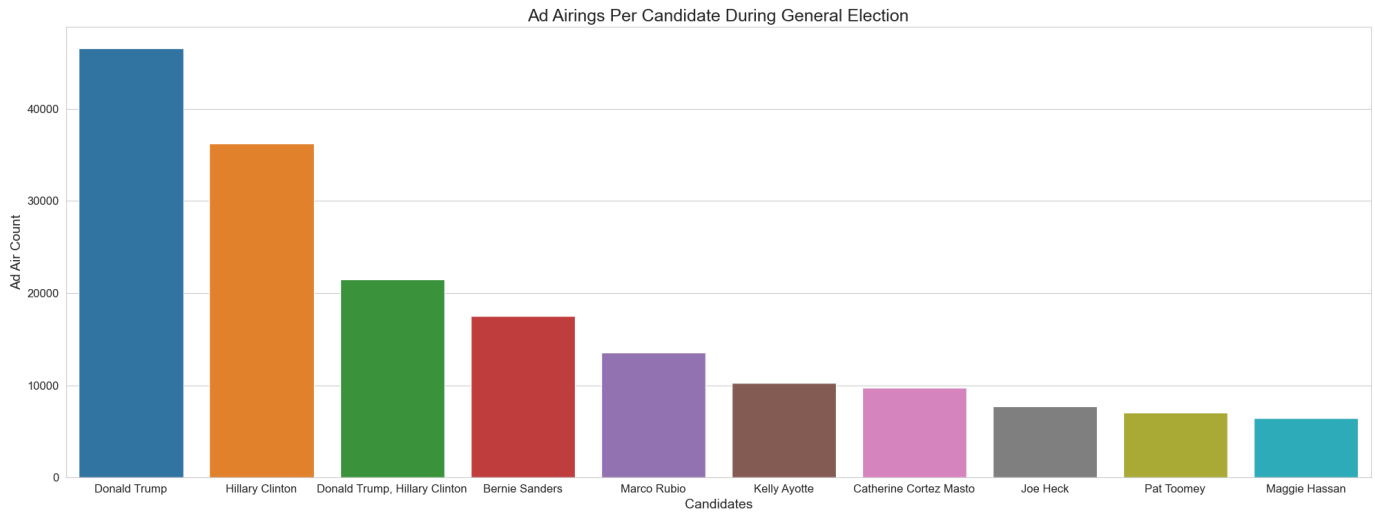


Digging a bit deeper, here we look at the proportion of fact-checked ads as opposed to the sheer number. This representation allows us to view fact-checked vs non-fact-checked ads on a comparable scale. These proportional bar plots show that 26.4% of anti-Trump ads received fact-checks compared to 22.7% of anti-Clinton ads. While this confirms what we saw in the initial “Trump vs Clinton: Fact-Checked Ads” bar plot, it reveals a more accurate magnitude. On the other hand, 21.4% of pro-Clinton ads received fact-checks compared to 23.6% of pro-Trump ads. This is interesting because it shows that proportionally, Trump received more fact-checks on both types of ads. If we were to only look at the first bar chart alone, we would be easily tempted to disregard the proportional aspect simply because it appears obvious that pro-Clinton ads received a significantly greater number of fact-checks. Thus, the idea of a pro-Trump bias during the fact-checking process would likely be a false assumption because the more ads there are, the more opportunities for fact-checking.

WHEN?

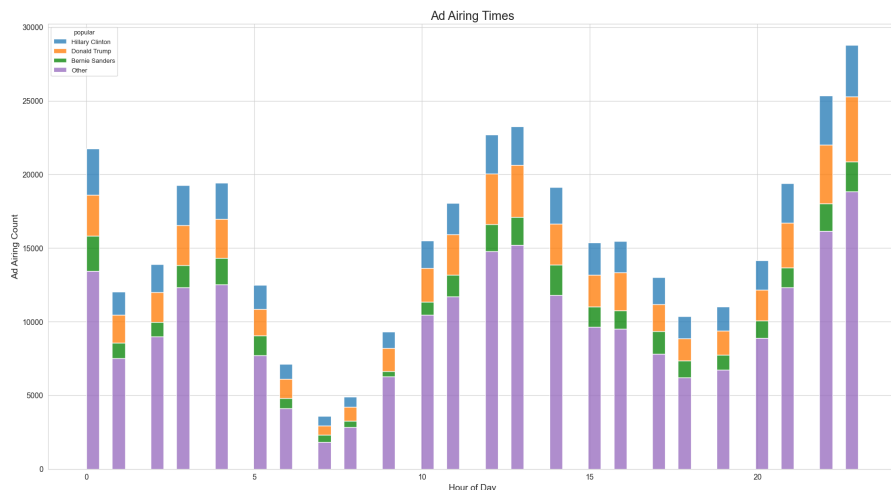
The political ads aired increased near the end of the primary and general election periods. Bernie Sanders (green) accounted for a large proportion of total ads aired in the primary election period. His ads were close to non-existent in the general election period. This is likely due to Sanders not doing well in the primary election and being unable to progress to the general election. Donald Trump and Hillary Clinton are the two prominent candidates in political ads aired in the general election period. Trump has a higher count of ads aired compared to Clinton in the general election. This is likely due to Trump’s sponsors and affiliated organizations spending more on TV ads.





The candidates with the highest political ads aired during the primary elections are Bernie Sanders, Hillary Clinton, and Donald Trump. Bernie Sanders has the highest count of aired ads in the primary elections. Donald Trump has the highest count of aired ads in the general elections. During the general elections the number of ads aired for Bernie Sanders decreased by a large amount.

The count of ads aired is high during the morning hours around 10am-12pm and during the night hours around 8pm-12am. This is likely due to ads being aired when their target audiences are able to watch TV (times of the day when they are not working or busy). It is interesting to note that ads that feature Donald Trump seem to fluctuate more in accordance to peak hours, relative to all other candidates.



CONCLUSION

Through our analysis of the 2016 election campaign political ads, we were able to gain some valuable insights. When looking at ads by location, the cities with the most ad aired appeared to be in metropolitan areas within influential swing states. These include Philadelphia, Las Vegas and Cedar Rapids all of which are not among the largest of U.S. cities as New York and Los Angeles are not on this list. These insights allow us to see how geographic locations influence political ad airings.

Our analysis also uncovered that while Trump was mentioned in the most number of ads and won the election, he had the highest proportion of critical (con) to supportive (pro) ads. This could indicate that his competitors and other active groups were aware of his potential influence. It also reflects the amount of criticism the winning president received. While digging into the number of fact checked ads, it is extremely alarming to see the sheer amount of non-fact-checked ads that could possibly circulate false information throughout the country.

A key insight from analyzing the “when” section of these ads is that there are generally more ad airings during the general elections than in the primary elections. Ads are also aired most during late hours of the day when people are preparing for bed around 10 pm to midnight.

Appendix

	Ad_Length
mean	32.490627
min	14.000000
max	300.000000

