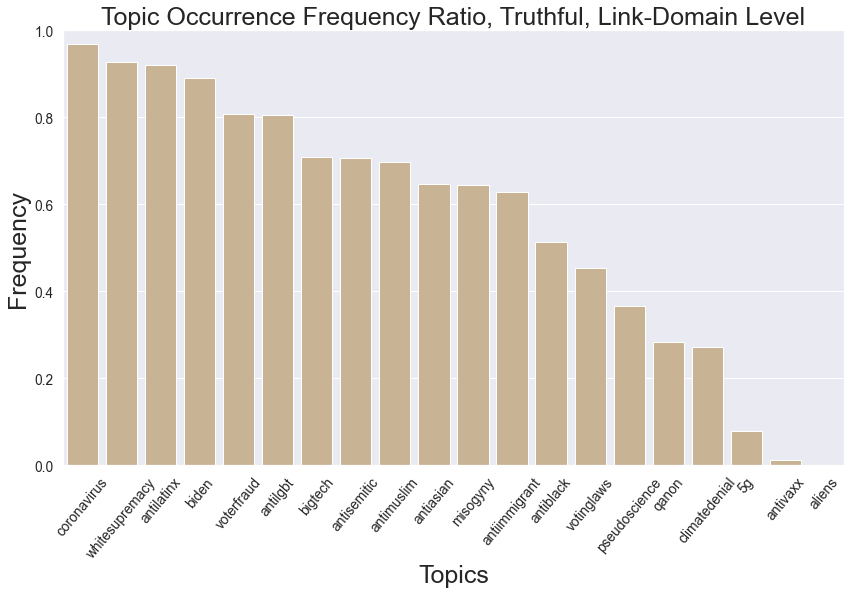
Python Exploration

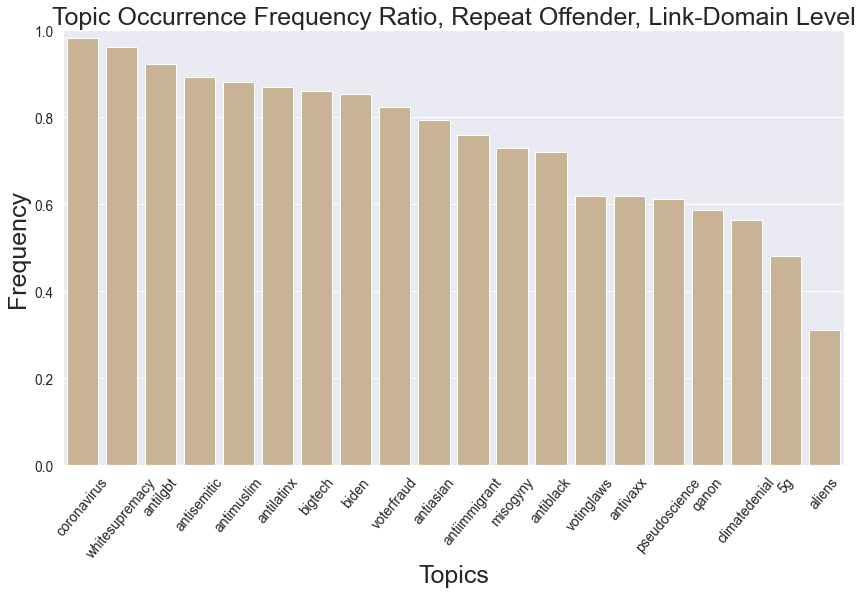
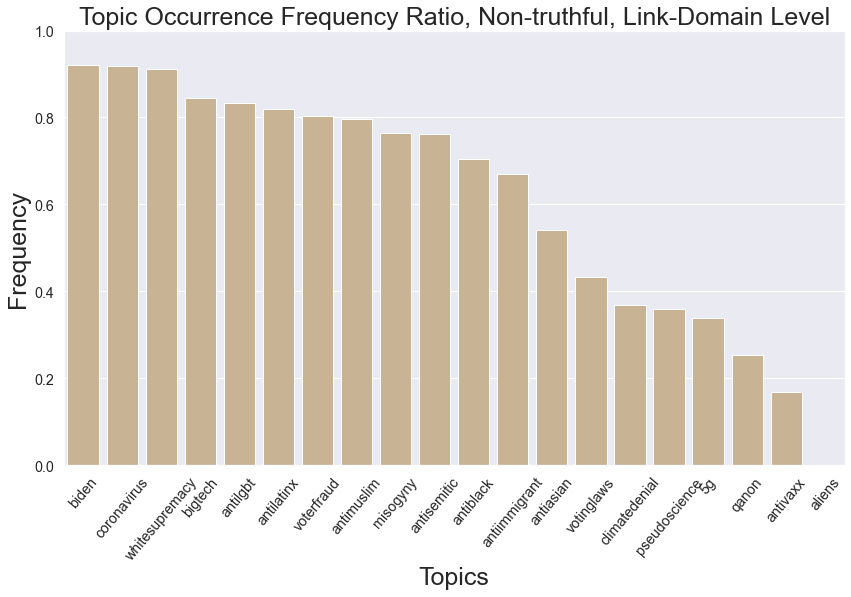
1. Topics Exploration

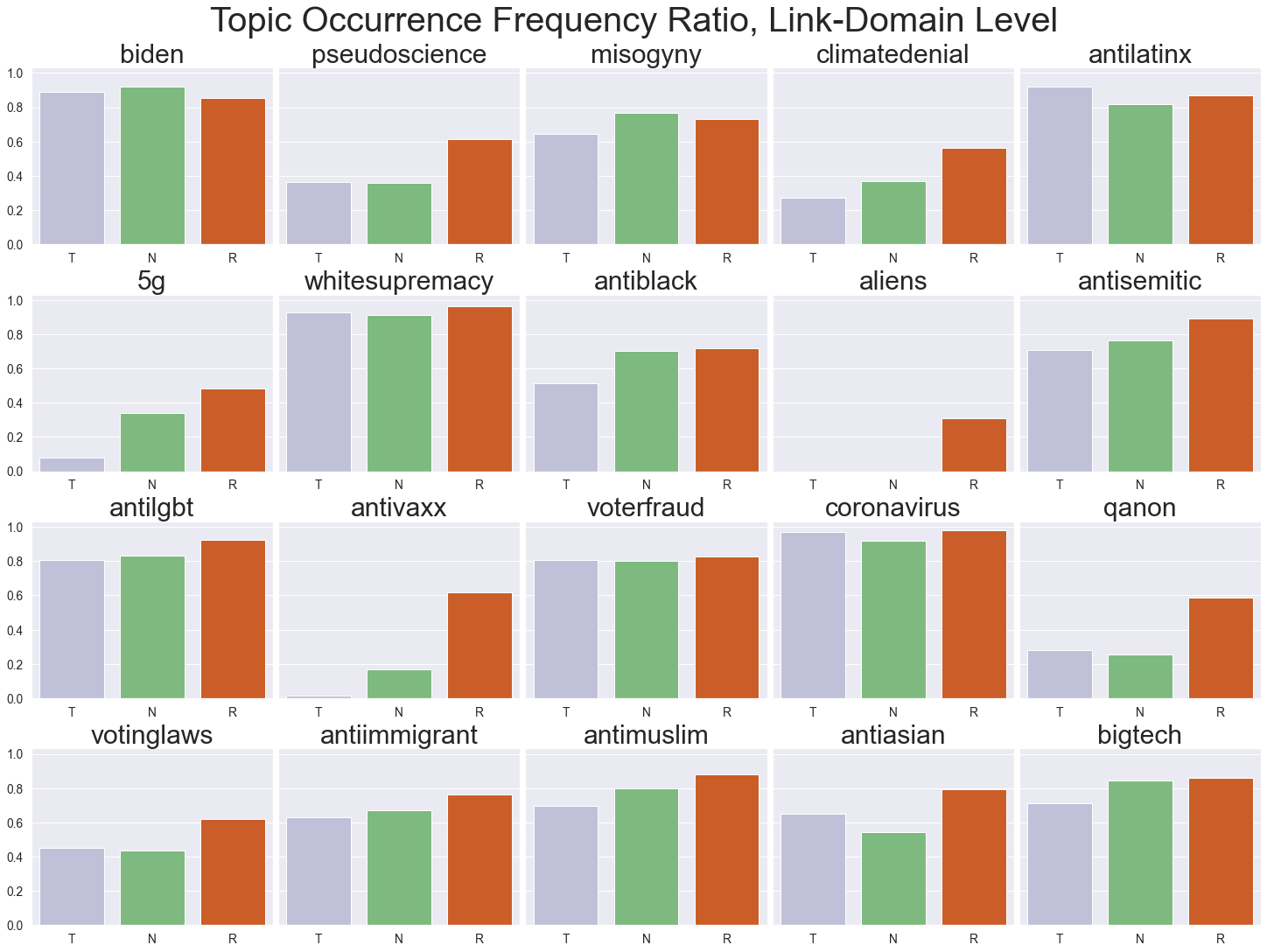
Visualizations in this section forces on explore the topic information provided by GDI in the master dataset. There are two statistics I used in this section:

* **Occurrence Frequency Ratio, Link-Domain Level:** The occurrence frequency of domain topic among all the recorded instances in the TCC master data that have topic information provided by GDI.
  + Formula: (Count of topic occurrence for instance with rating T/N/R) / (Total counts of instance with rating T/N/R)
* **Occurrence Frequency Ratio, Domain Level:** The occurrence frequency of domain topic among all the unique domains that have topic information provided by GDI.
  + Formula: (Count of topic occurrence among all the unique domains with rating T/N/R) / (Total counts of unique domains with rating T/N/R)
  1. Link-Domain Level exploration

The following three plots visualized the Occurrence Frequency Ratio, Link-Domain Level for all 20 topics for Truthful, Non-truthful and Repeat Offender Domains. We can see that the occurrence frequencies of topics high occurrence frequency are similar across all three ratings. Truthful domains’ median topic frequencies are much lower than non-truthful domains and repeat offender. In addition, topics with low occurrence, such as 5g, antivaxx and aliens, are much higher in repeat offender domains than in the other two types of domains.

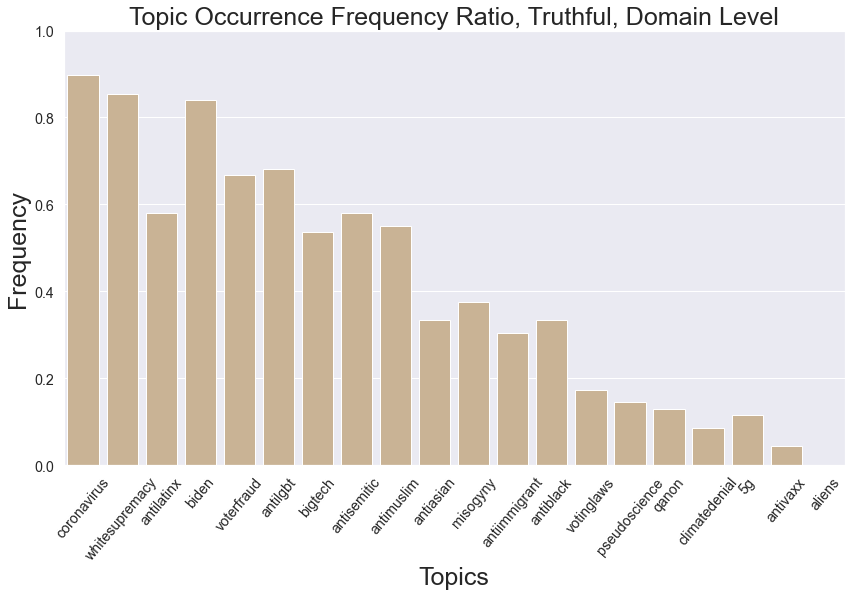


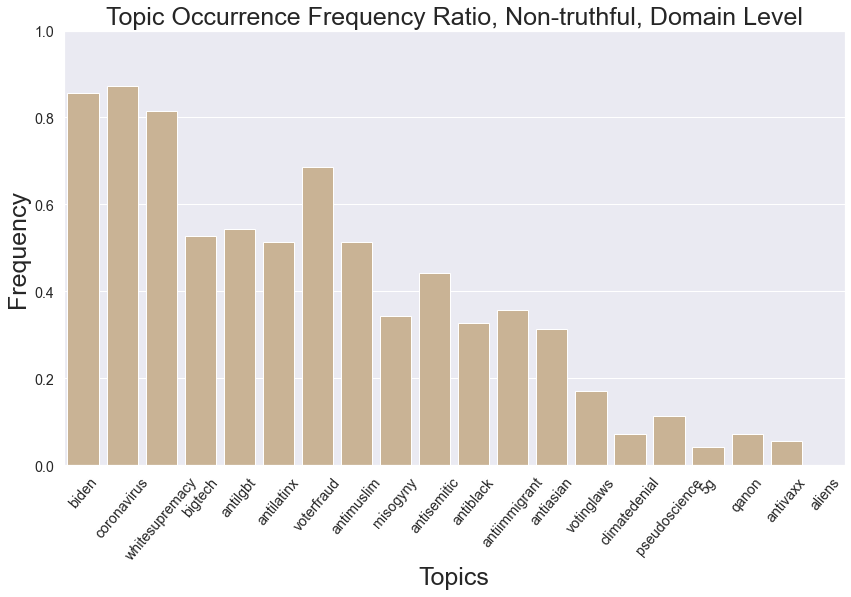


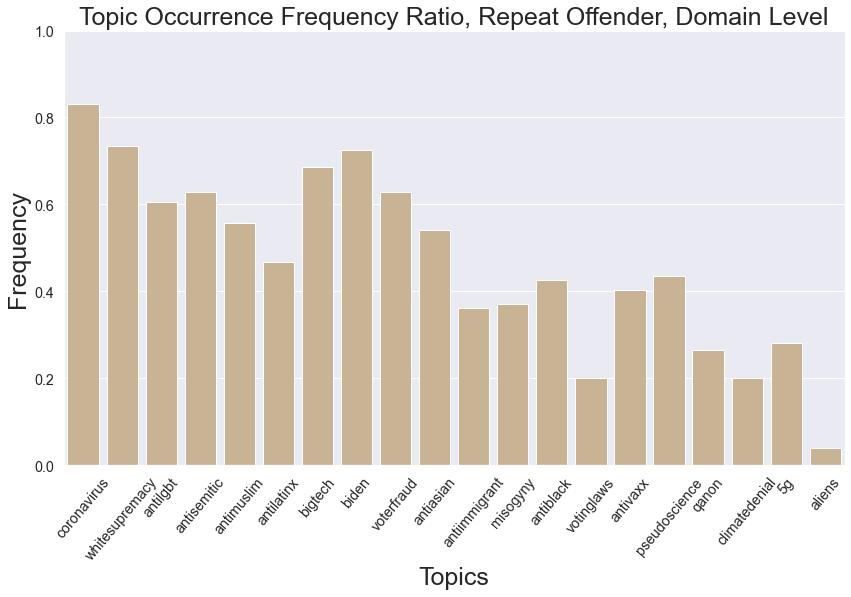
The next plot also visualizes the occurrence frequency of all 20 topics. Each block represents one topic in the 20 topics. We can again see that comparing to Truthful and Non-truthful domains, topics in repeat offender are more evenly occurring. Also, 5g, aliens and antivaxx have a much higher occurrence in Repeat offenders than in the other two types of domains.  


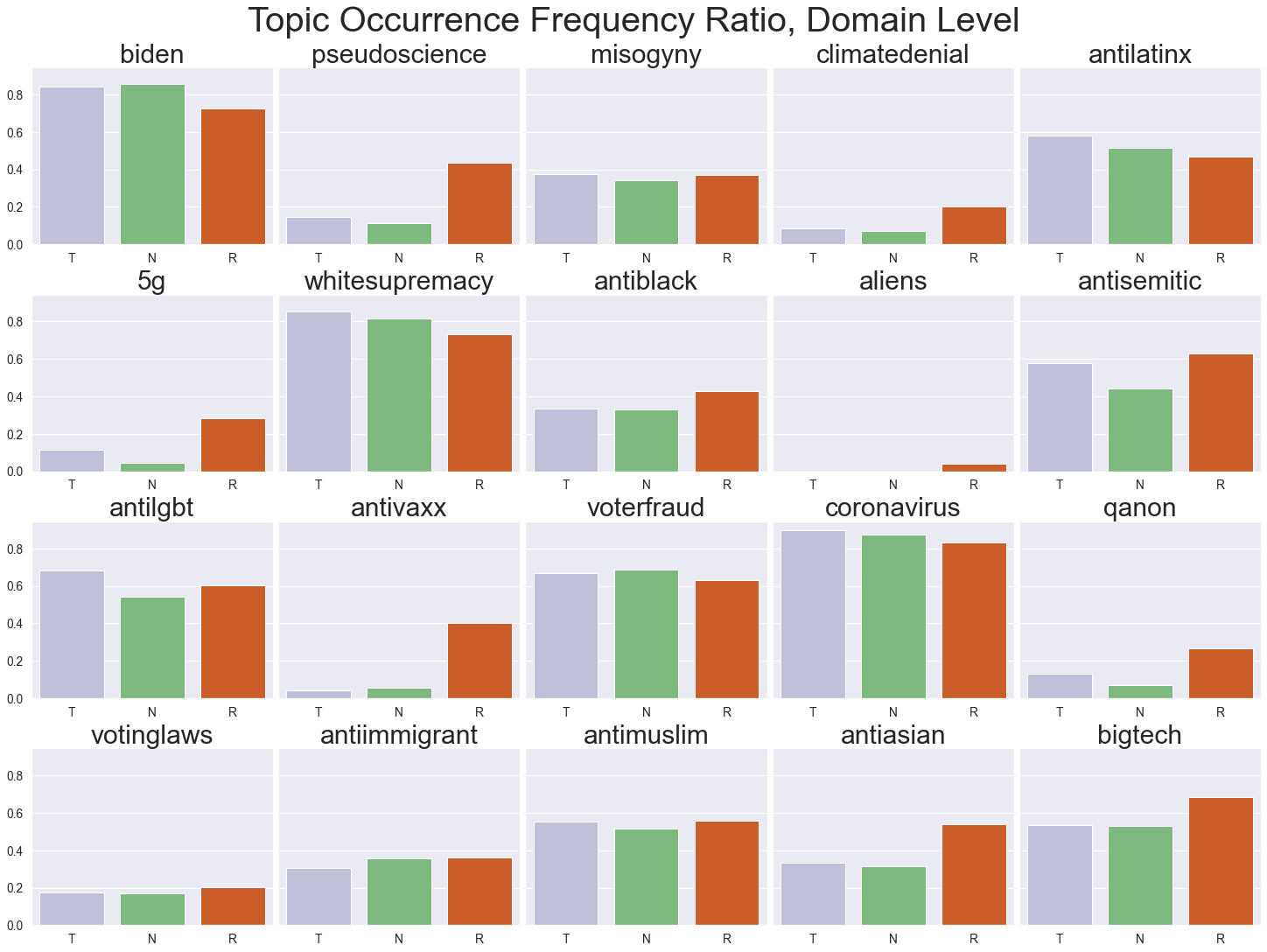
1. Domain level exploration

The following plots visualized the Occurrence Frequency Ratio, Domain Level for all 20 topics for Truthful, Non-truthful and Repeat Offender Domains. The bars in the bar plots are ordered in the same order in the Link-domain level plots. We can see that the domain level frequencies are generally lower than the link-domain level frequency in domains with all three rating. This finding may indicate that news links from domains with these topics are shared more than links from other news site, and topics have a more significant impact on the likelihood of sharing than whether the domain is Truthful, Non-truthful of Repeat offender.

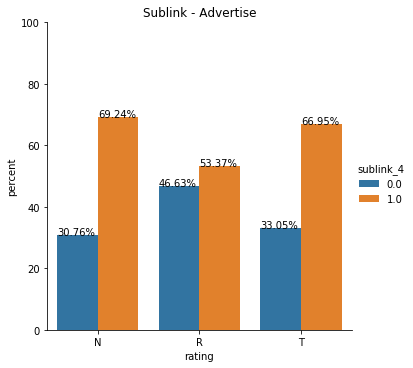
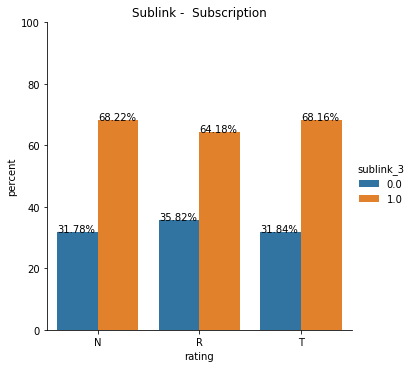
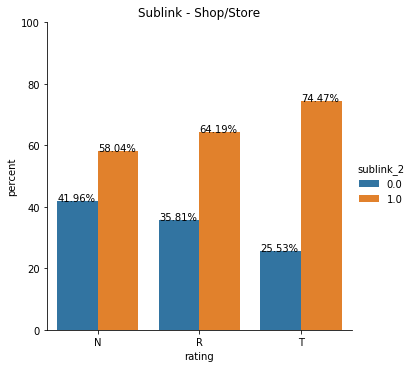
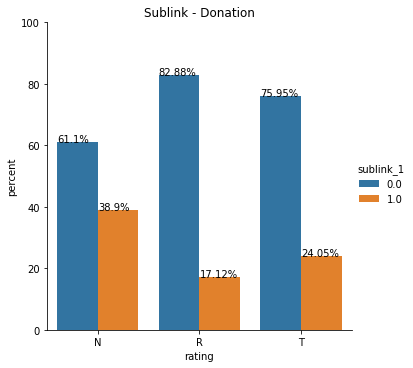


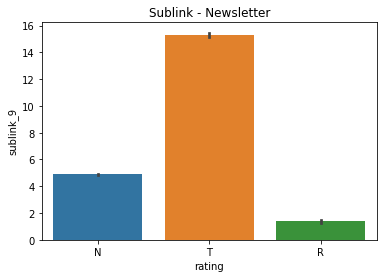
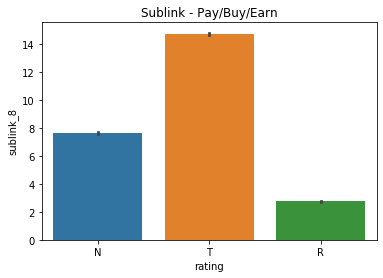
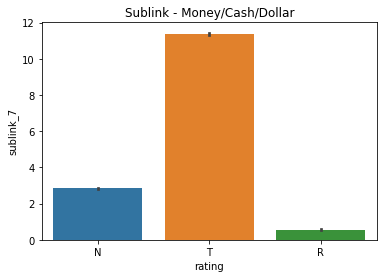
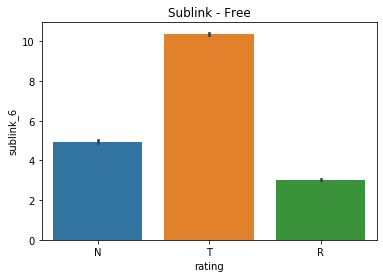
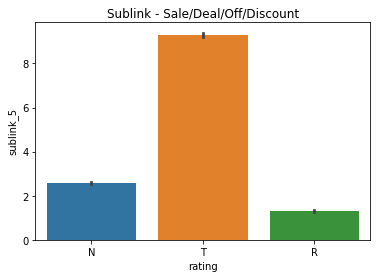


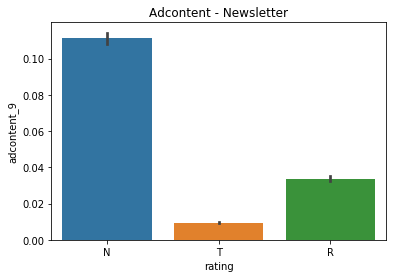
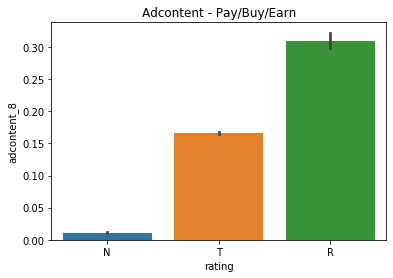
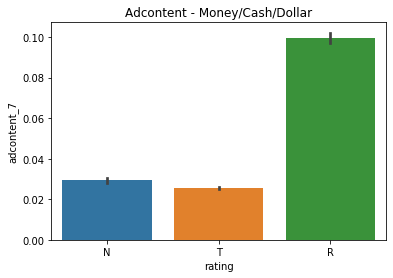
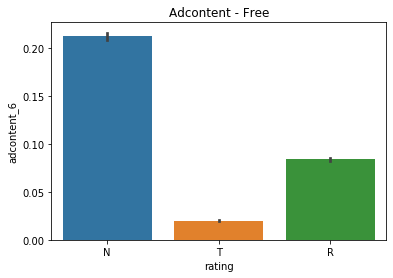
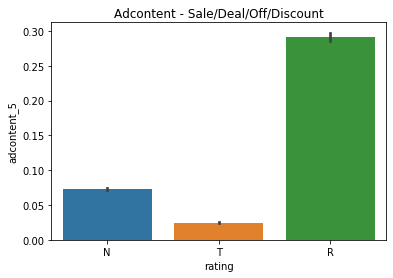
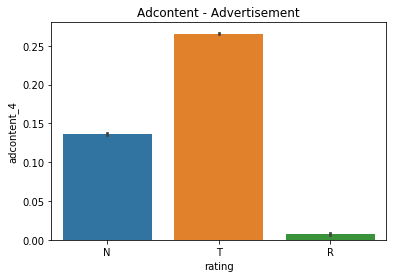
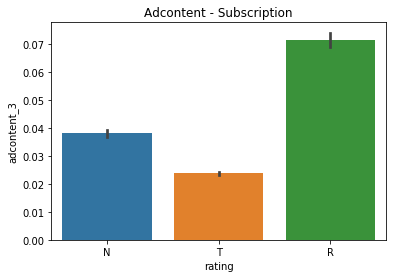
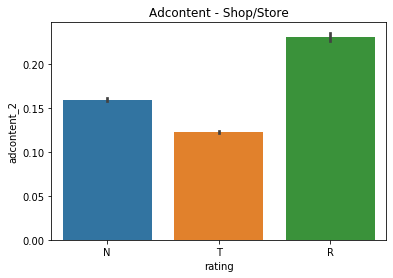
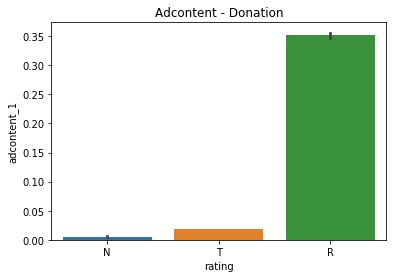




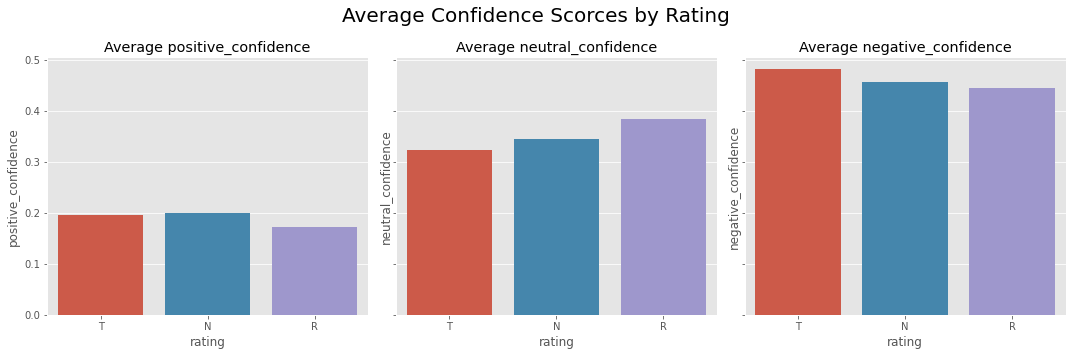
1. Financial-related terms





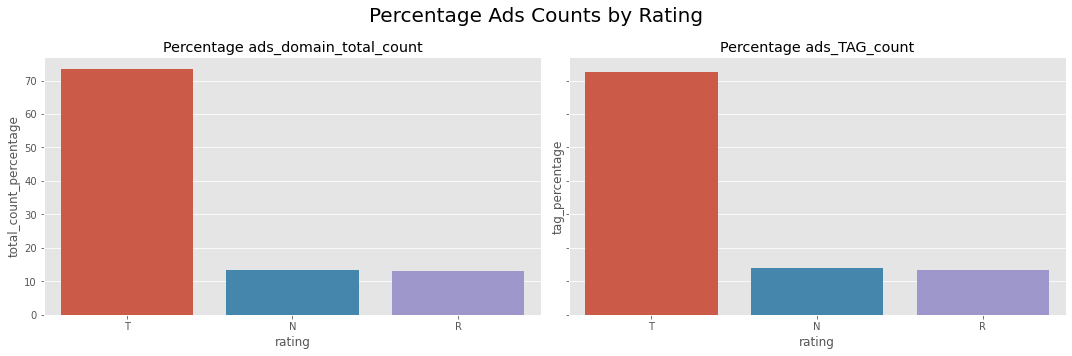


Exploration of Sentiment Confidence Scores



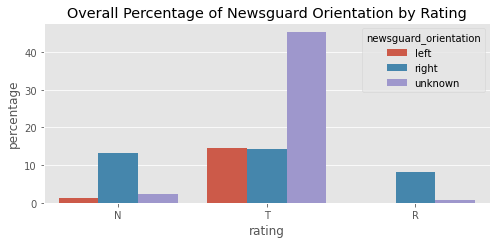
The above plots show the average confidence scores (positive\_confidence, neutral\_confidence and negative\_confidence) for Truthful, Non-truthful and Repeat Offender domains. We can observe in general all three types of domains have lowest scores in positive\_confidence, and the highest in negative\_confidence, and all three types of domains have relatively similar confidence scores in each type of confidence score. Repeat Offer domains have the lowest average positive\_confidence and Non-truthful domains have the highest. Repeat Offer domains have the highest average neutral\_confidence. Truthful domains have the highest negative\_confidence.

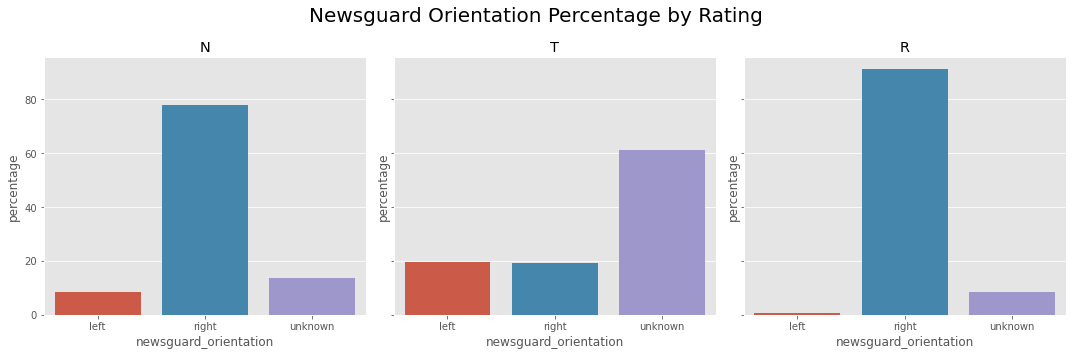
Exploration of Ads Count



The above plots show the percentage of ads\_domain\_total\_count (Total number of digital sellers that placed advertisements on the domain) and ads\_TAG\_count (Total number of digital sellers that placed advertisements on the domain that has a certification authority from the Trustworthy Accountability Group) for Truthful, Non-truthful and Repeat Offender domains. We can observe that for both ads\_domain\_total\_count and ads\_TAG\_count, Truthful domains have the highest percentage of ads counts, indicating more digital sellers place advertisements on Truthful domains.

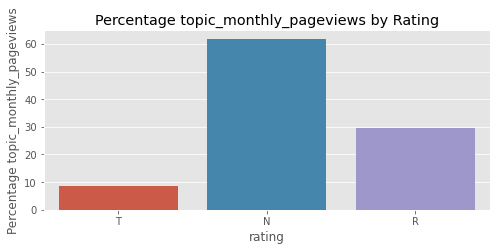
Exploration of Newsguard Orientation





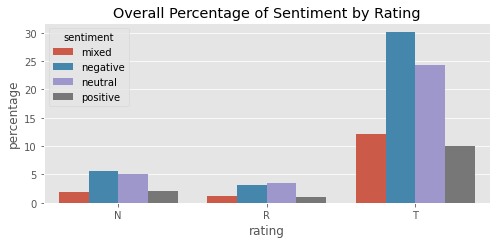
The above plots show the distribution of newsguard\_orientation (The orientation of the domain) by domain ratings (among Truthful, Non-truthful and Repeat Offender domains). The only difference is the first graph shows the overall percentage of distribution (all orientation in all rating groups sum up to 100%) and the second graph breaks overall count and only examines the distribution of newsguard\_orientation among each of the ruthful, Non-truthful and Repeat Offender domain type (each domain type sums to 100%). We can observe first that there are more data for Truthful domains. We can also observe that among Non-truthful domains, most domains have a right orientation. Among Truthful domains, most domains do not have a clear orientation. Among Repeat Offender domains, most domains also have a right orientation.

Exploration of topic\_monthly\_pageviews



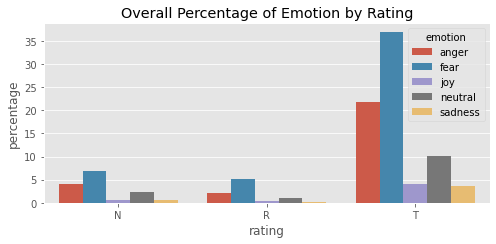
The above plot shows the percentage of topic\_monthly\_previews count for Truthful, Non-truthful and Repeat Offender domains. We can observe that Non-truthful domain have more topic\_monthly\_previews, taking more than 60% of the total views. Repeat Offender domain takes about 30% of the total topic\_monthly\_previews and Truthful domains only take less than 10% of the total topic\_monthly\_previews.

Exploration of Sentiment



The above plot shows the overall percentage of sentiment distribution for Truthful, Non-truthful and Repeat Offender domains (all sentiment percentage for all types of domains sum up to 100%). We can observe that there are more sentiments toward Truthful domains, and for each domain type, “negative” and “neutral” are the top 2 sentiments. People have more negative feelings towards Truthful and Non-truthful domains and tend to stay neutral for Repeat Offender domain.

Exploration of Emotion



The above plot shows the overall percentage of emotion distribution for Truthful, Non-truthful and Repeat Offender domains (all emotion percentage for all types of domains sum up to 100%). We can observe that there are more emotions toward Truthful domains. For each domain type, “fear” is the top emotion and “anger” is the second top emotion.