Investigating Trends in UFO sightings and US elections

Shelby Bearrows, Sam Cramer, Gurhar Khalsa, Kara Wolley

Description

- Investigate trends in UFO sightings in the United States
- Better understand patterns of sightings, occurrences and sighting location
- Identify trends in UFO shape to better understand advances in alien technology or at least how human perception of aliens have changed over time.
- Predict where and when UFO sightings occur based on the outcomes of US presidential and general elections.

Questions

- Are UFO sightings increasing over time?
- How has the UFO shape changed over time?
- If UFO sightings do increase over time, how is that correlated with the turn out rate for elections?
- Do states that vote Republican have significantly more UFO sightings?
- How does term number effect the number of UFO sightings? Is this also dependent on the political party?

Prior Work

- Some of the most commonly explored topics on UFO sightings are frequency of sightings by location, word cloud analysis on the description attribute, shape of UFOs by time of year, sightings by shape and season, and shape frequencies, just to name a few avenues of exploration. Perhaps the most explored question is whether sightings have increased over time. Work done by others seem to indicate that they have.
- Among the correlational questions explored are chi-square analysis revealed a relationship between sightings and region(www.kaggle.com/adhok93/eda-and-chi-squared-test).
- Geospatial analysis of UFO sightings has also been explored previously using heatmaps (www.kaggle.com/abigaillarion/ufo-reports-in-united-states).

Datasets

- Main Dataset:
  - https://www.kaggle.com/NUFORC/ufo-sightings
  - Available as csv files on personal computers of team members as well as on our GitHub page.
- Supplemental Datasets
  - https://www.kaggle.com/tunguz/us-elections-dataset
  - Available as csv files on personal computers of team members and on our GitHub page.
- The merged dataset (“merged_ufo_elect_data.csv”) can be found on our GitHub

Tools

- Python
- Tableau
- Matlab

Proposed Work

- Cleaning
  - The UFO dataset is especially 'dirty’. Reformatting of dates, state name and removal of random symbols (&, !, #) from the comments attribute is necessary. Cleaning will be necessary prior to merging with the US Elections dataset. The US Elections dataset appears to be much cleaner.
- Preprocessing
  - Correct for missing values. Some values may remain null, but many are predictable, such as if the 'county' value for a row of data is blank, but we know the state is 'TX,' it's safe to input the USA for the country value.
  - Identify erroneous data and outliers.
- Integration
  - Merge the UFO dataset with the US elections dataset. This will drop some UFO sightings that occurred outside of the US states.

Evaluation

- Histograms to identify classification of UFO shape or visualizations by election terms (bins of years)
- Heatmaps to visualize geographical location of UFO sightings to support hypotheses.
- Apiori to quantify minimal support
- Correlation measures including Pearson’s and lift.
- Student t test for the comparison of means