Website

STRUCTURE

Title screen -> with video

Intro paragraph

Title of research

Source used

* Nuforc report
* Shapeform file from

Methodology

* Explain cleaning process
* Flowcart for methology
* Graphs and figures

Conclusion

Credits

Media containing aliens have influenced ufo sightings?

Pick movies – highest grossing domestically

<https://www.the-numbers.com/box-office-records/worldwide/all-movies/creative-types/science-fiction>

Avatar, avatar 2, star wars

https://www.boxofficemojo.com/release/rl3372254721/

Scape data from website

Clean

Do ufo sighting / population yearly

Do before and after, see if there is an increase

Graph the popularity of said movie

Exclude from years before ->

UFO and light pollution?

UFO and weather balloons?

<https://www.weather.gov/upperair/nws_upper>

<https://www.noaa.gov/NOAA-Communications>

https://www.weather.gov/upperair/nws\_upper#

website

home

* Geographic plots
  + Light pollution
  + Weather balloons
  + Testing sites?
* Media
  + Sci-fi movies
    - Interactive?
  + Google trends

Plan on what to do

Understand how do hypothesis testing, get calculator

Sightings per population, population is taken as yearly population

Hypothesis test the data from movie vs sightings

Make a conclusion

Write up methology

<https://jgscott.github.io/teaching/writeups/write_ups/>

Visualisation

* Simple: just do the graphs and upload them as files to website, carousel style
* Animated graph
* Interactive graph

Clean files

population

<https://www.census.gov/data/tables/time-series/demo/popest/2020s-state-total.html>

<https://www2.census.gov/programs-surveys/popest/datasets/2010-2020/state/totals/>

1. Title and Abstract:

Start with a clear and concise title that describes the essence of your project.

Write a brief abstract summarizing the project's objectives, methods, key findings, and implications.

Title: Comparing the daily gross of the top 10 grossing sci-fi films with extraterrestrial life to see if they have affected the amount of UFO sightings between 2009 – 2022 in the USA?

Abstract:

2. Introduction:

Provide context for your data science project by explaining the problem or question you're addressing.

State your research objectives and hypotheses.

Hypotheses

3. Data Collection:

Explain how you obtained permission or rights to use the data, especially if it's sensitive or proprietary.

UFO Sightings Data:

One of the most important dataset were the UFO sightings. This data was gathered from a cleaned .csv file from Timothy Renner [<https://data.world/timothyrenner/ufo-sightings/workspace/file?filename=nuforc_reports.csv>], which was originally published from the National UFO reporting centre [https://nuforc.org/]

List of Top Grossing Films:

To compile a list of the top-grossing science fiction films featuring extraterrestrial life, IMDb’s website was utilised as it provides information on the box office gross of various films.

<https://www.imdb.com/search/title/?title_type=feature&genres=scifi&sort=boxoffice_gross_us>

Daily Film Gross Data:

To gather daily domestic box office gross data for the top 10 sci-fi films identified above, I designed a Python web scraper to collect the date and daily domestic gross figures from the Box Office Mojo website and save each movie as individual .csv files.

<https://www.boxofficemojo.com/?ref_=bo_nb_rl_mojologo>.

The website does not prohibit automated data retrieval, and no sensitive or proprietary data was accessed.

4. Data Preprocessing:

Detail the steps you took to clean and prepare the data. This may include handling missing values, data transformation, and normalization.

Mention any data exploration techniques you applied to understand the dataset better.

NUFORC cleaning:

Removed the coloumns ‘stats’, ‘posted’, ‘text’ and ‘reported\_link’ as it restates information already given.

Only keeps the USA.

Splits the ‘date\_time’ column into ‘date’, ‘time’ and converts them to datetime.time objects

Gross cleaning

Ensured date was converted using datetime.time object

Data exploration:

Plotted a whole graph that displayed daily ufo sightings in usa and 10 films daily domestic gross against date, gave a brief overview.

Bar graph for ufo sightings, and scatter plot for daily domestic gross

Individual graphs for each movie

Ensured to show ufo sightings 2 months before and after movie release, shows impact before and after movie release.

Found r-value ( correlation coeeficient) of each movie

Did t-test hypothestis testing on each film to see if the movies had an impact on daily ufo sighting

Analysed using a significant level of 5%, right-hand test

Correlation Matrix???

HYPOTHESIS TESTING

Spearman correlation test, measures the strength and direction of a monotonic association between two ranked variables.

5. Feature Engineering:

Explain the process of selecting, creating, or transforming features for your analysis.

Justify your feature selection and engineering choices based on domain knowledge or data exploration.

Data Preparation: Ensure that you have time-series data for both daily movie gross and UFO sightings, with each data point associated with a specific date or time.

Cross-Correlation Analysis: Use a cross-correlation function, which is available in libraries like numpy or scipy, to calculate the cross-correlation coefficient for different time lags.

Hypothesis Testing: Determine if the cross-correlation coefficient at a specific time lag is statistically significant. You can perform hypothesis testing to check if the correlation is different from zero (indicating a significant relationship).

Interpretation: Analyze the results to understand the strength and direction of the correlation and whether there is a time lag between the two variables.

Visualization: Create visualizations, such as cross-correlation plots, to help interpret the results visually.

6. Model Selection and Training:

Describe the machine learning or statistical models you used for your analysis.

Explain how you split the data into training and testing sets for model evaluation.

Specify hyperparameters and any cross-validation techniques employed.

7. Evaluation Metrics:

Discuss the metrics you used to evaluate model performance (e.g., accuracy, precision, recall, F1-score, RMSE).

Explain why these metrics are appropriate for your problem.

8. Results and Discussion:

Present the main findings of your data analysis. Use visualizations, tables, and graphs to illustrate your results.

Interpret the results in the context of your research objectives.

Discuss any limitations or challenges you encountered during the analysis.

9. Conclusion:

Summarize the key takeaways from your analysis.

Revisit your research objectives and state whether you achieved them.

Offer insights into the broader implications of your findings.

10. Recommendations:

Provide actionable recommendations based on your analysis, if applicable.

Suggest areas for further research or improvements.

11. Methodology:

In a separate section, describe the methodology you followed in detail.

Include code snippets, algorithms, and step-by-step instructions to reproduce your analysis.

Mention the software tools and libraries you used.

12. References:

Cite all the sources of data, research papers, and tools you referenced in your report.

13. Appendices:

Include any supplementary material, such as additional charts, code, or data, in the appendices.

14. Acknowledgments:

Thank individuals or organizations that contributed to your project, such as data providers or collaborators.

Ensure your report is well-organized, uses clear and concise language, and includes appropriate visual aids to make the content more accessible. Additionally, consider your audience, whether it's technical experts, stakeholders, or a broader audience, and tailor your report accordingly.

GEO

Years I will be looking at

2008,2013,2018,2022

To find the location of the weather ballon release sites, I used the map provided in the nws website, after verifying by emailing them that they have released the balloons in the same sites since 2008-Present.

[https://www.weather.gov/upperair/nws\_upper#](https://www.weather.gov/upperair/nws_upper)

As I wanted to find the precise co-ordinates of the release sites, I used the weather.gov website to gather the longitude and latitude of each site, along with the

I manually wrote down each stid written on the map and used a web scrapper to get the data bellow.

Weather balloons are released 7 AM and 7 PM EDT and 6 AM and 6 PM EST, according to <https://www.weather.gov/chs/upperair>

Creating the map

<https://medium.com/@kadircolak1999/find-markers-in-google-maps-api-circle-fa8dc89a37d4>

<https://www.google.com/earth/outreach/learn/visualize-your-data-on-a-custom-map-using-google-my-maps/>

[only allows 2000, per layer for a max of 10000]  
<https://pfisterer.dev/posts/cache-map-clusters/>

<https://plotly.com/python/maps/>

<https://deck.gl/docs/api-reference/layers>

<https://stackoverflow.com/questions/32565950/loading-100-200k-markers-on-google-map>

<https://kepler.gl/>

I used this website when gathering data, as it showed years when station was active

<https://explore.synopticdata.com/KSLE/metadata>

<https://kepler.gl/demo/map?mapUrl=https://dl.dropboxusercontent.com/scl/fi/2bngyuk63fnog0qpmffbj/sightings-co-ordinates.json?rlkey=qv6e02h1yk4agdvnwk79r156o&dl=0>

<https://kepler.gl/demo/map?mapUrl=https://dl.dropboxusercontent.com/scl/fi/zocxa9rw53lzd2jbf2lb5/jsds-Copy.json?rlkey=4qn8k07ogpn6f01ym5of3g481&dl=0>

track weather balloons

https://sondehub.org/#!mt=Mapnik&mz=9&qm=3h&mc=30.59064,-92.06818&f=23068783

|  |  |  |
| --- | --- | --- |
| Weather Ballon | Data | Justification |
| Stid | Id of station based on map | More easily identifies individual stations  (primary key) |
| Longitude |  | Plotting geographically |
| Latitude |  |  |
| Report start | Status of Station | Ignore stations is not within date |
| Report end | Status of station |  |
| Timezone |  | Different timezones have different time of days where ballon is more visible |
| State |  | Alligns with state sighting |
| Percentage active |  |  |

|  |  |  |
| --- | --- | --- |
| nuforc | Data | Justification |
| date | Date of sighting | Compare with location of weather balloons |
| City\_Longitude |  | Plotting geographically |
| City\_latitude |  |  |
| State |  | Alligns with state sighting |
| time |  |  |

Stations

KOAK, KNKX