Meteorite Data Impact Application

A Full-stack Data Application Project

Project Team Members

Adriana Khul, Thomas Gartley, Valentina Sanhueza, Victor Rincon

1. Introduction

The purpose of this project was to create a full-stack interactive data application that visualizes and analyzes meteorite impact data from around the world and within a range of years (between 0yrs and 2013) provided by a user. The dataset we used offered valuable insights into meteorite landings, including variables such as the location, mass, fall date, and classification of meteorites. We offered multiple visualizations for 'at a glance' data comparison, including maps with embedded data. These features assist in extracting meaningful patterns and trends from complex datasets. Through the use of interactive features like filters, maps, and charts, users can manipulate the data according to their interests and needs. Interactive dashboards enhance user experience by allowing them to explore data on their terms, fostering deeper insights and more informed decision-making.

2. Dataset

Our dataset was acquired from Kaggle.com, where it had already gone through several cleaning iterations. The original dataset is from the Meteoritical Society website. Their dataset was cleaned by NASA and posted as a dataset in Kaggle.com (NASA). That dataset was further processed by Jay Panderson and posted again in Kaggle.com

Some of the data had null values for latitude and longitude or the year discovered, so these were dropped from the dataset. Additionally, we dropped the name, id, and name_type columns as they were not relevant to the analysis we performed. The final piece of data cleaning that was required was changing the datatype on the Year column from a float to an integer.

3. Research Questions

We had 3 main questions we hoped to answer:

1. Is there a pattern to the location of meteorite impact?
2. How many meteorite impacts occur within a specific time frame?
3. Which meteor class encounters Earth most frequently?

4. 'Meteorite Impact' Application

A. Overview

'Earth Impact Data App' webapp begins by loading the Home page (index.html) with the website description, purpose of the site, and links to other website sections (Dashboard, Maps, Works cited, About us). The interactive dashboard allows users to identify the number of meteorite landings in any given year and see the corresponding meteorite classifications. The visualizations of the sunburst chart help expand on the percentage of meteorite classification in a given year within the initial search range, and the bar graph indicates the top fifteen classes of meteorites that had fallen in the given time frame. The map dashboard has a filter by start and end year to search the dataset. On the map dashboard, users can toggle between a heat map and markers to indicate where meteorites have landed throughout the world.

The main app file (app.py) contains the 9 app routes used to render the website and retrieve data for the visualizations. The first 5 routes render the template HTML files, and the remaining 4 routes manage the API queries. There are 3 for the visualizations on the dashboard page, and 1 for the map. Each API route calls a function from the sql_helper.py file and utilizes the SQLAlchemy library to call data from our SQLite database (meteorites.sqlite). All our queries accept two variables (min and max year) from user inputs.

The first API query retrieves the data for the table on our dashboard. It groups meteorites by their type and year and computes the average and total mass of meteorites for that year. We aggregated the data to improve the loading time on our dashboard and created our table.

```python
def query_data(self, min_year, max_year):
    """Query meteorite data filtered by year."""
    query = text("""
        SELECT
            Year as year,
            rec_class as class,
            COUNT(*) as count,
            SUM(mass) as sum_mass,
            AVG(mass) as avg_mass
        FROM meteorites
        WHERE Year >= :min_year AND Year <= :max_year
        GROUP BY class, Year
        ORDER BY Year ASC;
    """)
    return self._execute_query(query, {"min_year": min_year, "max_year": max_year})
```

For the filters, we defined textboxes and buttons in the HTML files. We allow the user to input a minimum and maximum year and then use the "Enter" button to refresh the data. This re-runs the initialization function that pulls the min and max year values by their IDs and then makes a call to one or more of our API routes depending on the page.

B. Challenges

The most challenging part of the webpage to build was the sunburst chart. The data formatting requirements of the sunburst chart required us to combine multiple queries into a single DataFrame and then jsonify it. One DataFrame contained a year, classification, and the count for that classification and year, and the second DataFrame had an empty column, year, and the count for that year. Combining the two gave us a DataFrame that had a column for the parent (either an empty string “” or the year), an ID (either the classification or the year), and then the count for that ID.

```
query1 = text("""
    WITH TopYears AS (
        SELECT Year
        FROM meteorites
        WHERE Year >= :min_year AND Year <= :max_year
        GROUP BY Year
        ORDER BY COUNT(*) DESC
        LIMIT 15
    ),
    TopClasses AS (
        SELECT rec_class_group, year, sum(count) as count
        FROM (
            SELECT CASE WHEN COUNT(*) < 100 THEN 'Misc.' Else rec_class END as rec_class_group, Year, COUNT(*) AS
            FROM meteorites
            WHERE Year IN (SELECT Year FROM TopYears)
            GROUP BY rec_class, Year
        )
        GROUP BY rec_class_group, Year
    )
    SELECT rec_class_group AS label, count, Year AS parent, rec_class_group || "_" || Year as id
    FROM TopClasses
    ORDER BY Year, count DESC;
""")
```

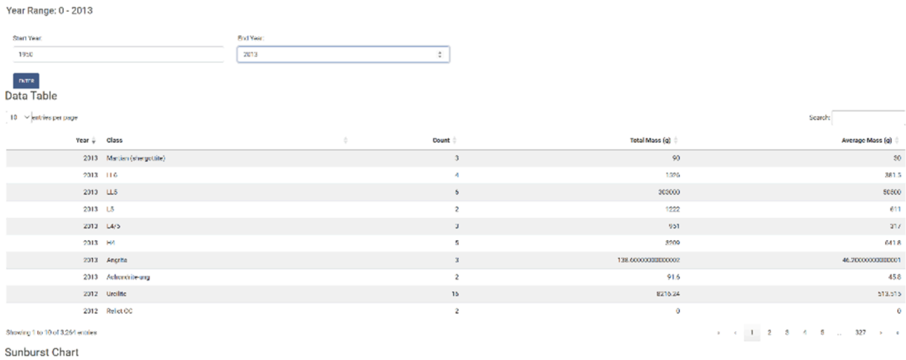
```
query2 = text("""
    WITH TopYears AS (
        SELECT Year
        FROM meteorites
        WHERE Year >= :min_year AND Year <= :max_year
        GROUP BY Year
        ORDER BY COUNT(*) DESC
        LIMIT 15
    )
    SELECT Year AS label, COUNT(*) AS count, '' AS parent, Year as id
    FROM meteorites
    WHERE Year IN (SELECT Year FROM TopYears)
    GROUP BY Year
    ORDER BY count DESC;
""")
df2 = self._execute_query(query2, {"min_year": min_year,"max_year": max_year})
```

The amount of data loaded during the initial pre-loaded query significantly impacted the webpage performance time. Limiting the classifications quantity in the sunburst section resolved this issue, but limited the data by too much. This was resolved by grouping all classifications with less than 100 meteors into a Misc category.. After verifying proper operation of all components locally and deploying the website, the sunburst chart did not render completely. This was resolved by changing the query concatenation function to the || operator.

5. Visualizations

A. Data Table

Data set with a filter start and ending year from 0-2013. Allows users to view and sort all the categories from the dataset. With this table, users can enter their desired date range and sum the count column to find the answer to our second research question.



Year Range: 0 - 2013

Start Year: 1950

End Year: 2013

ENTER

Data Table

10 entries per page

Search:

Year	Class	Count	Total Mass (g)	Average Mass (g)
2013	Martian (shergottite)	3	90	30
2013	LL6	4	1326	381.5
2013	LL5	6	303000	50500
2013	L5	2	1222	611
2013	L4/5	3	951	317
2013	H4	5	3209	641.8
2013	Angrite	3	138.60000000000002	46.200000000000001
2013	Achondrite-ung	2	91.6	45.8
2012	Ureilite	16	8216.24	513.515
2012	Relict OC	2	0	0

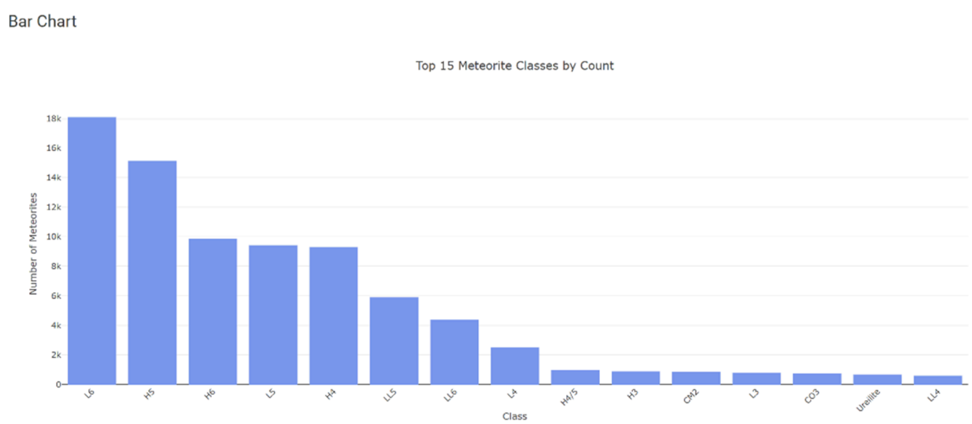
Showing 1 to 10 of 3,264 entries

1 2 3 4 5 … 327

Sunburst Chart

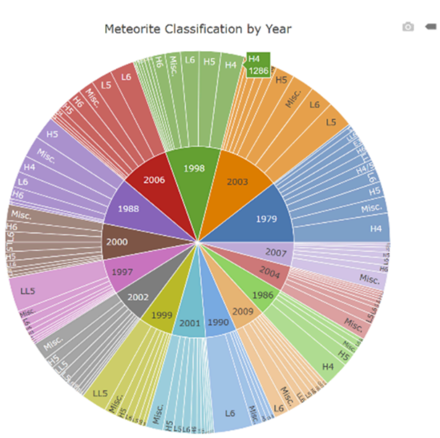
B. Bar Graph

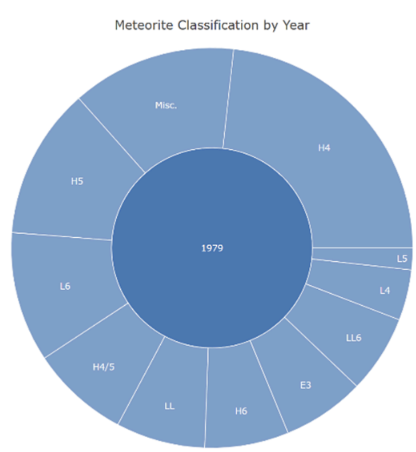
The bar graph shows the top 15 classes of meteorites that fell within the filtered time frame. This is particularly useful in answering our third research question. When setting the start year to 1950 and the end year to 2013 in the dashboard filter, the bar graph will show that the L6 meteorite was the highest occurring meteorite class in that timeframe. For further insight into how this classification varies by year, the dashboard features a sunburst chart. Users can click on individual years, causing the chart to expand into a donut chart that offers an additional, detailed visual breakdown of meteorite classifications.



C. Sunburst Chart

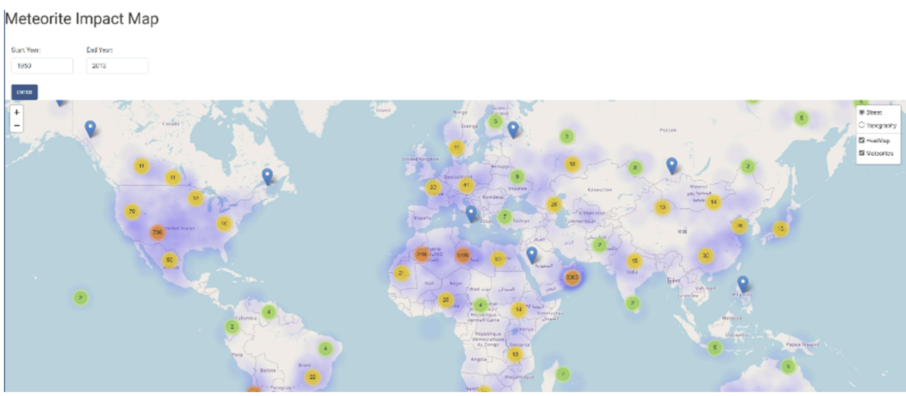
A sunburst chart that shows the top 15 years with the most meteorites from the given time frame and a breakdown of the meteorite compositions.





D. Map

An interactive map where users can pan and zoom worldwide and visualize meteorite impact sites. It also has a heat map to quickly determine the areas with the highest impact count within the user query. To answer our first research question, we can use the map dashboard to help understand the dataset. This can be shown by both the grouped/individual meteorite markers and heat map concentrations.



6. Bias and Data Limitations

Size: It is difficult to visualize the size of the datasets due to the amount of data.

Time frame: The dataset is not current. It includes data up to 2013 but has not been updated with modern findings.

Data Cleaning Bias: In the process of data cleaning, removing columns such as fall, name_type, and unnamed and removing rows with NaN values may have excluded valid data. This impacts the overall results of the dataset.

Human-Reported biases: Human-reported biases occur due to the population size of where meteorites strike. Specifically, in areas with a higher population, there are more people to come across a meteorite. Vice versa, a less populated area where meteorites have landed will be underreported.

Geographic Biases: Geographic biases occur due to the elevation and vegetation differences in meteorite landing areas. For example, if a meteorite lands in an open area, such as a desert, it is more visible than a meteorite landing in the rainforest.

7. Conclusion

Key findings from our analysis fall into two main categories: geographical distribution and meteorite classifications. We observed that meteorite reports are more prevalent in the eastern hemisphere, with significant hotspots in Antarctica and the Middle East. These trends may reflect underlying biases. For example, the flat terrain in both Antarctica (glacial surfaces) and the Middle East (desert landscapes) makes meteorite sightings easier to report. Conversely, regions like Canada, which recorded fewer than 30 meteorite events from 1950 to 2013 despite its vast area, might suffer from underreporting due to lower population densities and more rural landscapes.

Building on these geographic insights, we also examined the classifications of meteorites that have impacted Earth. Utilizing our interactive dashboard, we can explore meteorite classes both by year with a sunburst chart and overall frequency with a bar graph. This analysis revealed that the L6 meteorite is the most commonly encountered class between 1950 and 2013.

In future renditions of our dashboard, we plan to incorporate additional visualizations and more contextual information to enhance the overall user experience. One improvement is to introduce a bubble chart that visualizes the average mass of meteorites, enabling users to quickly gauge the scale of each landing event. Additionally, we intend to reintroduce the meteorite name category from the original dataset, which will offer more detailed insights and identification of individual meteorites. Also, we aim to provide more educational content about meteorite composition and explore the origins of different classifications. This information can be shown through informational panels on our dashboard. Finally, we would like to enhance our website appearance. After seeing the works of previous groups, we have ideas to include animations, such as a meteor shower, to make the homepage more exciting to a new user.

8. Works Cited

Meteoritical Society Meteoritical Bulletin: Search the Database
https://www.lpi.usra.edu/meteor/metbull.php/

NASA Kaggle Dataset Meteorite Landings
https://www.kaggle.com/datasets/nasa/meteorite-landings

Jay Panderson Kaggle Dataset cleaned_meteorite_data
https://www.kaggle.com/datasets/jaypanderson/cleaned-meteorite-data/data