5000 Years of Solar Eclipses Trends Visualization

CSE 564 FINAL PROJECT PROPOSAL

Group 32
Apoorva Kashi (115696552)
Chahat Kalsi (115825394)

Background

The "5000 Years Solar Eclipse Enhanced Catalog" dataset provides a comprehensive overview of eclipse events spanning over 5000 years. It includes detailed celestial and geographical data on each eclipse, making it a valuable resource for temporal and spatial analysis. This proposal outlines a project aimed at creating insightful visualizations of the dataset to explore the frequency and characteristics of eclipse events globally and across decades.

Objectives/Problem Statement

The project aims to achieve the following objectives:

1. **Understand the Global Distribution of Eclipses**: Identify which countries have experienced the most eclipses and visualize this distribution geographically.
2. **Analyze Eclipse Characteristics**: Explore relationships between various eclipse characteristics such as magnitude, type, and geographic coordinates.
3. **Trend Analysis Over Decades**: Investigate how eclipse attributes such as magnitude have evolved over the decades.
4. **Interactive Visualization for Enhanced Usability**: Develop interactive visualizations that allow users to explore data across different dimensions and extract tailored insights.

Data Description

Dataset used:
https://www.kaggle.com/datasets/s34n180/nasa-5000-years-solar-eclipse-event-catalog

The dataset contains 11,898 records of eclipse events with 47 attributes including date, time, type, magnitude, geographic coordinates, and more. Data types include integers, floats, and strings, with some missing values particularly in columns like 'Path Width (km)'.

Key Attributes

- **Catalog Number:** Unique identifier for each eclipse event.
- **Calendar Dates**: Precise dates of each eclipse.
- **Eclipse Time**s: Exact timing of the eclipses.
- **Lunation Number**: A consecutive number given to every new moon.
- **Saros Number**: Used to identify the series to which the eclipse belongs.
- **Gamma Value:** Indicates the distance of the moon's shadow from the earth's center during maximum eclipse.
- **Eclipse Type**: Classification of the eclipse based on its nature and geometry.
- **Eclipse Magnitude**: Ratio of the Sun's obscured diameter by the Moon to the Sun's apparent diameter.
- **Geographical Coordinates**: Coordinates where the eclipse is visible.
- **Sun Altitude and Azimuth**: Position of the sun during the eclipse.
- **Path Width and Central Duration**: Specifics for total and annular eclipses.
- Each record is organized chronologically by decade, making it easier to analyze trends over time.

Data Preparation

Data Preparation

- **Geocoding**: Convert latitude and longitude to country names using the GeoNames or Nominatim API to facilitate country-wise aggregation.
- **Data Cleaning**: Handle missing values and ensure all datetime and categorical variables are correctly formatted.
- **Feature Engineering**: Calculate decade from the year of the eclipse for trend analysis.

Visualizations/Approach

GeoMap (Country vs Frequency of Eclipses)

A choropleth map showing the frequency of eclipses per country, highlighting areas with the highest eclipse occurrences.

Insights:

- Geographical Trends: Determine which countries or regions have historically witnessed the most eclipses. This can offer insights into geographic and celestial dynamics that favor certain locations over others.

Hexbin Map (Various Variables)

Utilize hexbin plots for showcasing the density and relationship of eclipses concerning two key continuous variables (e.g., Latitude vs. Longitude, Eclipse Magnitude vs. Moon Distance).

Insights:

- Density Clusters: Identify where eclipses are most densely clustered geographically and by physical characteristics.
- Variable Relationships: Understand how variables such as eclipse magnitude correlate with the moon's distance from the earth, which might affect the eclipse's visibility and type.
- Spatial Distribution: Mapping variables like latitude and longitude in hexbins can reveal concentrations of eclipses, indicating celestial paths or zones with high eclipse activity.

Pie Chart (Types of Eclipses)

Provide a proportional representation of each type of eclipse (Total, Annular, Hybrid, Partial) within the dataset, showing how common each type is relative to the others.

Insights:

- Eclipse Variety: Highlight the most common types of eclipses and their rarity, which can be crucial for educational and research purposes.
- Predictive Understanding: Help in predicting the likelihood of a specific type of eclipse occurring, based on historical data.
- Celestial Mechanics Insight: Offer insights into the mechanics of the solar system, as certain types of eclipses occur under specific celestial alignments.

Bar Plot and Violin Plot (Magnitude vs Frequency)

A bar plot overlaid with a violin plot to show the frequency and distribution of eclipse magnitudes, providing insights into the most common eclipse sizes.

Insights:

- Magnitude Distribution: Understand which eclipse magnitudes are most common and how they vary in frequency.
- Visibility and Impact: Correlate eclipse magnitude with its potential visibility and impact on the earth, as larger magnitudes mean a more substantial obscuration of the sun.
- Statistical Overview: The violin plot overlay will provide a deeper statistical insight into the distribution, showing peaks and tails that indicate variability in magnitude sizes.

Line Plot (Eclipse Magnitude vs Decade)

A line plot to trace the change in average eclipse magnitude over decades, revealing long-term trends. The goal is to figure out if the solar eclipse magnitude varies by decade. This could help us gain insights into the sun, moon and earth's celestial movements.

Insights:

- Trend Analysis: Identify whether the magnitudes of eclipses are increasing, decreasing, or remain stable over time, which can help infer changes in celestial mechanics or earth's orientation.
- Celestial Predictions: Insights from this analysis could be used to predict future eclipse characteristics and their potential visual and climatic impacts.
- Historical Celestial Dynamics: Understand if there are any long-term cycles in eclipse magnitudes that correlate with other astronomical cycles like solar cycles or moon phases.

Expected Outcomes

The visualizations will allow us to:

- Identify hotspots for eclipse occurrences and relate eclipse frequency with geographic locations.
- Understand the distribution of eclipse types and magnitudes, providing insights into their commonality and variability.
- Discover trends in eclipse characteristics over time, particularly looking at changes in magnitude across decades.

Conclusion

This project will leverage advanced data visualization tools to interpret the rich dataset of solar eclipses. By mapping the global frequency of eclipses, analyzing their physical characteristics, and observing trends over time, we aim to provide a comprehensive visual exploration of solar eclipses as documented over the past 5000 years. These insights could be beneficial for educational purposes, research, and enhancing public knowledge on celestial events.
