Welcome to SMU Project 3 - Creating an interactive dashboard about US Tornados!

A child looking at a tornado

Description automatically generated

**Project Overview**

Our team has diligently collaborated to delve into the Tornados 1950-2022 dataset from Kaggle.com. We chose this dataset because it has a wide breadth of data and our curiosity about categorizing tornados. We aimed to uncover US tornado patterns through exploratory data engineering, data analysis, and data visualization.

**Exploratory Data Analysis Highlights**

Our analysis commenced with a comprehensive exploration of the dataset. We cleaned the data by dropping off one identified duplicate and 27,926 missing values and creating three new columns. Next, we loaded the transformed data into an SQLite database and had the Flask app listen to the request. The Flask parsed out the input and turned it into an SQL statement: select \* from the database. Once we received the data, we created the visualizations.

As we explored more profoundly, the visualizations focused on the relationships between variables: the number of tornadoes in a US state or region and the month and magnitude scale rating, aiming to tell the data story of the US Tornados over the last seven decades.

**Instructions on how to interact with the project**

Visit htmllink for project

Our landing page consists of xx. The dashboard page displays a bar chart, box plot chart, sunburst chart, and map. It consists of two dropdowns (filter): one established by region and the second one by state. Select any filter for either dropdown and the bar, box plot, and sunburst charts, and the map will update.

**Key Findings**

* Between 1950 and 2022, there were 68,693 tornados.
* The highest concentrations of tornadoes occurred in Texas, Florida, Oklahoma, Mississippi, Iowa, and Louisiana.
* Texas had the highest concentration of tornadoes, with 4,601 tornados, while DC and Alaska had the lowest concentration of tornadoes, with 2 and 1, respectively.
* The magnitude ratings show a strong upward trend where the magnitude is between 0-2 on the rating scale.

**Implications and Recommendations**

Our findings emphasize that tornados occur in all US states, but most occur in the Midwest region. Also, the upward trend in tornado magnitudes was ratings 0-2, meaning less severe tornadoes dominate our data. Consequently, caution is advised of accounting bias – according to the American Meteorological Society, the reporting of tornado occurrence is recorded at the county level, and all tornadoes, regardless of Enhanced Fujita scale (EF) rating, are counted. However, weak tornados with minimum impact/losses may go underreported because of selective reporting. Feel free to explore our analysis and findings within this repository. We welcome collaboration, feedback, and further contributions to enhance our data's story.

**Ethical considerations made in the project**

Several ethical considerations were made to ensure the responsible use of data. First, our team ensured that all data sources used were obtained legally and with proper permissions. Additionally, we tried to minimize any potential bias that could arise from the data analysis. We carefully examined the data for any potential algorithmic bias and took steps to mitigate it. Overall, our team prioritized ethical considerations to ensure the responsible and ethical use of data throughout the project.

**References for the data source**

The National Weather Service Storm Prediction Center

1. Your GitHub repo must include a README.md with an outline of the project including:
   * An overview of the project and its purpose
   * Instructions on how to use and interact with the project
   * Documentation of the database used and why (e.g. benefits of SQL or NoSQL for this project)
   * ETL workflow with diagrams or ERD
   * At least one paragraph summarizing efforts for ethical considerations made in the project
   * References for the data source(s)
   * References for any code used that is not your own
2. The GitHub repo has a README.md that includes the following: (10 points)
   * An overview of the project and its purpose
   * Instructions on how to use and interact with the project
   * At least one paragraph summarizing efforts for ethical considerations made in the project
   * References for the data source(s)
   * References for any code used that is not your own