Analyzing Bear-Human Conflicts in North America: Spatial Trends, Risk Factors, and Strategies for Coexistence

MD Alam Al Saud || 23279118

1 Introduction

Bear attacks in North America are a significant concern for outdoor enthusiasts, wildlife researchers, and policymakers. Understanding the patterns and factors contributing to these incidents is essential for enhancing public safety, protecting wildlife, and fostering coexistence between humans and bears. This report aims to investigate the frequency, geographic distribution, and contextual factors of bear attacks in North America. By analyzing available data, we seek to answer the following key questions:

1. What are the spatial trends in bear attacks across North America?
2. Are there identifiable factors, such as seasonality, age of victim, attack description, associated with a higher risk of bear attacks?
3. What can be learned to mitigate future bear-human conflicts?

2 Data Used

2.1 Data Sources

Data Source 1: Kaggle

- **Metadata URL**: Fatal Bear Attack.
- **Data Type**: *csv*
- **Description**: This dataset contains information on deadly attacks by wild bears in North America. The original data is sourced from Wikipedia, with added latitude and longitude details. The column Bear contains the type of bear (e.g., Brown Bear, Polar Bear, or Black Bear). Cases involv- ing multiple fatalities in one attack are split into one row per deceased individual.
- **Data Structure**:Data on deadly attacks by bears include details on event, date, location and victim demographics. The data contains attacks from the US and Canada.We have these features in our dataset Date(Date of attack),Location(Location and State/Province),Details(Details on the attack),Bear(Type of bear (brown, black or polar)),Latitude(Latitude coordinate),Longitude(Longitude coordinate),Name(Victim's name),Age(Victim's age),Gender(Victim's gender). Here We have categorical data type (Name, location, type of bear , gender,Age) and continuous data type (lat, lang). Also have a column with descriptive text.
- **License**: Data is under an open license, such as CC BY-SA 3.0, which permits free use for research and educational purposes, including AI model training, as long as the license terms are adhered to. Attribution is pro- vided by referencing the relevant sources or license documentation. For more information, visit CC BY-SA 3.0 License.

Data Source 2: Kaggle

- **Metadata URL**: Bear Attacks North America
- **Data Type**: *csv*
- **Description**: This data set shows every recorded killing by a black, brown, or polar bear from 1900-present day in North America.
- **Data Structure**:This data is structured with different types of features. The features are Name(Victim Name),age(Victim age),gender(Victim gender),Date(Attack event Date),Month(Attack event Month),Year(Attack event Year),Type(wild or captive),Location(Event Location),Description(Attack

incident Description),Type of bear(Type of bear Brown or Black), hunter, Grizzly, Hikers. Here We have categorical data type (Name, location, type of bear ,date, month, year, gender,Age) and continuous data type (lat, lang). Also have a column with descriptive text.

- **License**: The ODC-BY 1.0 license explicitly allows the use of data for any purpose, including commercial use, as long as attribution is provided. OpenAI complies with this by ensuring that data sources under such li- censes are acknowledged, either explicitly or by referencing the general license terms when applicable. To fulfill the obligations of the ODC-BY 1.0 license, I ensure proper attribution by referencing open data sources or guiding users to license details when relevant. I provide transparency on the origin of the data and remind users to comply with attribution requirements if they reuse the content. This ensures adherence to the license and supports responsible use of open data.

2.2 Data Preparation:

After loading the csv data we needed to clean all the data sets. Some columns had invalid data type and noisy data. Like age had some extra texts and decimal numbers. That was a crucial part for data validation. And also for place names, we needed to extract the perfect state name from the address.There were some missing values inside the datasets in age and time. Those were filled with interpolation methods. There were some outliers which were also solved. I have normalized the data according to the need for good output and reports. Then merged both clean data into a single data frame. The merge was tricky. Because they had different column names and data types.

3 Analysis

On the merged data, we will run our analysis process. We will try to get some insight from these data that will help us to find the answer to the question we mentioned earlier. We will consider a few factors in the analysis, those are spatial trends, seasonality, victim age and description of the attack.

3.1 Spatial trends

3.1.1 Methodology:

We will try to find spatial trends on our data set that are correlated with bear attack. As we have a state name, the frequency of attack will be considered to determine the trend. So we will check which state has how many Bear attack incidents. After plotting the data in a Heatmap we will set it into a real map.

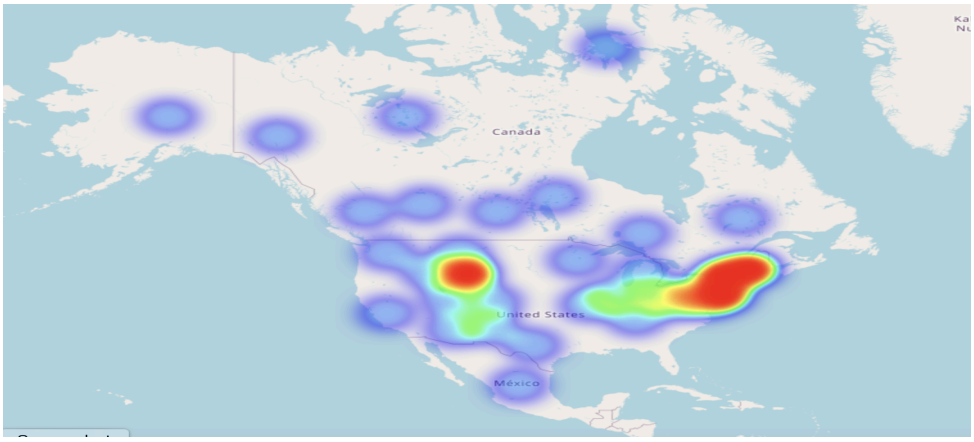

Figure 1: Heatmap of Bear Attack.

3.1.2 Results:

If we see the chart we can realise that Alaska has the highest Bear attack incidents alone. After that Montana, British Columbia, Alberta and Wyoming states have the highest attack. Those states share most of the attack.

3.1.3 Interpretation:

From the result it is clear that Alaska is the most dangerous place for Bear attack incidence. One important thing we can notice is that the next five states are in the same region in Canada and the USA. The all share the Mountain Ranges. So the Mountain range areas are also dangerous places. Besides that the region of New York and Toronto is also a dangerous spot of bear attack.

3.2 Seasonality

3.2.1 Methodology:

In Our data we have a time frame of the incident. Now we will try to find the correlation between seasonality and Bear attack. For doing this we will divide all the data attack incidents into months. So we will take only a month from each data row. We will plot those month data into a bar chart and see whether there is any correlation or not.

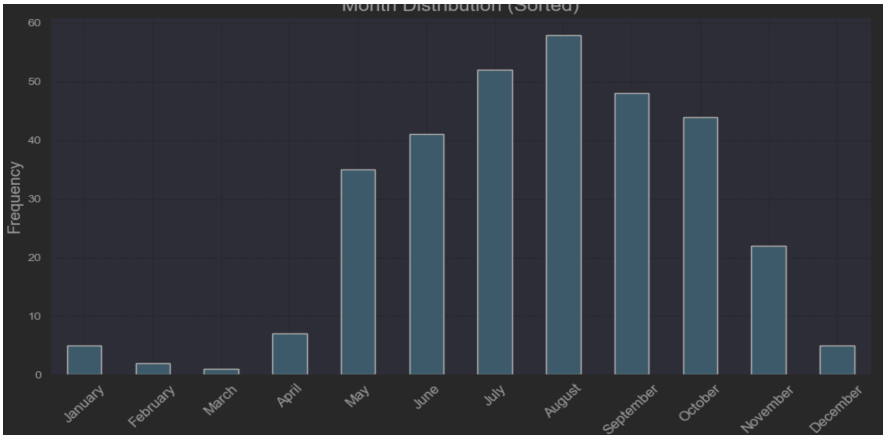

Figure 2: Seasonality of bear attack.

3.2.2 Results:

As we can see in the chart, Bear attacks are significantly higher in summer. Especially July, August and September are on the peak of attack. But in the winter season the number of attacks is very low. January, February and March are the lowest month of attack.

3.2.3 Interpretation:

After observing the result we can say that there is an impactful correlation between seasons and bear attack. Most of the attacks have occurred in the Summer season.

3.3 Victim age and other findings

3.3.1 Methodology:

In that analysis we are going to find out the impact of age in our data set. For sophisticated visualization we have considered the range of age. Here we have taken 5 years of interval. Then we have plotted the data into a Bar chart. Beside That we will have a look inside our description data of every incident and try to get useful insight. We have chosen some words from description to characterize the attack.

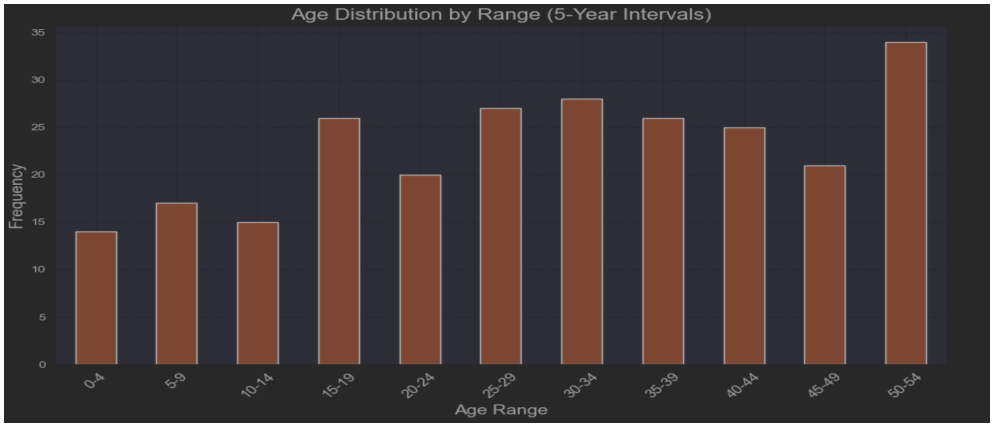

Figure 3: Correlation of bear attack.

3.3.2 Results:

From the chart we can observe that almost every age of people are getting attacked. There is no simple pattern that we describe.

3.3.3 Interpretation:

We saw no significant correlation between the age of the victim and Bear attack. But we can say older people are getting more attacked than younger people. From the description we have got some insight that lots of attack incidents occurred inside the Zoo. Moreover most of the attacks happened while hiking or camping. And most of the time the victim was alone.

4. Conclusion and Limitation

4.1 Conclusion

After all the analysis we got some informative insight. Not all the factors were so impactful. But still all the factors we have checked gave us some information. So far we have checked Spatial trend, Seasonality, Victim Age and attack de- scription. With the analysis we can definitely answer the questions that we wanted to. We have spatial trends of bear attack that has been shown on Heatmap. Seasonality also has great impact on it. We found no correlation between victim age and attack.

Now How can we avoid the bear attack? We can follow those points:

- We have to be more cautious while hiking and camping in dangerous places like Montana, British Columbia, Alberta and Wyoming. And also areas beside New York and Toronto.
- Avoid activity during summer , specially in July, August and September. Those Months are the most Dangerous months of bear attack.
- While going into the Zoo , always stay away from Bear case and never cross a certain line.

4.2 Limitation

After In the data sets we don't have the details of victim activity. With that we could predict the exact reason for the bear attack and could know in detail about what to avoid. Also there is an important thing that needs to be mentioned in data is the medical condition of attacking bears. So we can be more careful about the interaction with bears.