**Abstract**

This project provides useful insights into the critical issue of animal vehicle collisions, a phenomenon posing significant risks to both human safety and wildlife conservation. The core of the study is the development of an innovative machine learning mode, employing logistic regression techniques, to predict the likelihood of animal vehicle collision taking geographical factors into account. The report offers a comprehensive approach to understanding and addressing the factors that contribute to these hazardous incidents.

In addition to the logistic regression model, the project includes an array of graphical representation to identify the underlying causes of these incidents. The graphical representations not only elucidate patterns and trends in animal-vehicle collisions, but also highlight the critical areas where these incidents are prevalent. These visuals can serve as a crucial component in understanding the dynamics of animal-vehicle collisions facilitating better-informed decision-making for wildlife protection and traffic safety measurements. The project is a multi-faceted approach to mitigate animal-vehicle collisions blending machine learning techniques with practical, data-driven insights.

**Introduction**

The escalating issue of animal-vehicle collisions (AVCs) presents a dual threat, imperiling both human lives and wildlife populations. AVCs not only lead to significant economic losses but also contribute to the decline of various animal species. Recognizing the gravity of this issue, this project is anchored around the development and deployment of a machine learning model to predict the likelihood of such collisions. By focusing on geographic coordinates (longitude and latitude), the model provides a novel approach to identify potential collision hotspots, thereby enabling preemptive measures to safeguard both human and animal lives.

**Objectives**

The primary objectives of this project are.

* To harness the predictive power of logistic regression in machine learning to analyze and forecast AVC risks based on geographical data.
* To identify the hotspots for these incidents, where animal vehicle collisions are prevalent.
* To identify the patterns and trends involved in Animal Vehicle Collisions.
* To suggest preventive measures, which would ensure traffic and animal safety.

This approach stands out by integrating specific location data, thereby offering a targeted and efficient way to identify areas with a high risk of AVCs. The model's ability to process and analyze large datasets allows for a comprehensive understanding of the patterns and trends in AVC occurrences, which is crucial for developing effective mitigation strategies.

In addition to the predictive model, this project incorporates a series of data visualizations and graphical analyses, which play a pivotal role in interpreting the complex data involved in AVC incidents. These visual tools not only aid in understanding the distribution and frequency of AVCs across different regions but also assist in communicating these insights effectively to stakeholders, including wildlife conservationists, urban planners, and policymakers. This aspect of the project is vital in translating data-driven findings into actionable insights and recommendations.

Finally, the broader aim of this initiative is to contribute to the fields of wildlife conservation, public safety, and data science by providing a scalable and replicable model that can be adapted to different geographic contexts. The project underscores the potential of machine learning in addressing real-world problems and sets the stage for future innovations in the intersection of technology and environmental conservation. The goal is to establish a framework that not only predicts but also helps in preventing AVCs, aligning with broader efforts to promote biodiversity and road safety.

**What is the problem?**

The problem at the heart of this project is the increasing incidence of animal-vehicle collisions (AVCs), a phenomenon that poses significant risks both to human life and wildlife. AVCs are not merely isolated incidents; they reflect a growing challenge in the interface between expanding human habitats and the natural territories of various wildlife species. These collisions often result in severe consequences, including loss of human life, serious injuries, and substantial property damage.

For wildlife, AVCs contribute to the decline of animal populations, particularly affecting species that are already vulnerable or endangered. The problem is exacerbated by factors such as habitat fragmentation, increasing vehicular traffic, and often the lack of awareness or predictive measures to prevent such incidents. The frequency and distribution of AVCs are influenced by various geographical and environmental factors, making it a complex problem to address.

**Why this is a project related to this class**

This project “Mitigation of Animal Road Accidents” is intrinsically related to the field of data analytics as it fundamentally relies on the collection, analysis, and interpretation of large datasets to address the issue of animal-vehicle collisions (AVCs). By employing logistic regression, the project applies sophisticated analytical techniques to discern patterns, trends, and correlations within the data. This approach exemplifies the core principles of data analytics: transforming raw data into meaningful insights. The project not only utilizes historical data to predict future occurrences but also integrates various data points such as geographical locations (longitude and latitude) to enhance the accuracy and relevance of its predictions.

In essence, this project is a practical application of data analytics, leveraging the power of machine learning to solve a real-world problem, demonstrating how data-driven strategies can be employed effectively in the realm of public safety and wildlife conservation.

**Statement of the problem**

The statement of the problem addressed in this project is the significant and growing challenge of animal-vehicle collisions (AVCs), which present serious threats to human safety and wildlife populations. These incidents, often occurring due to the intersection of wildlife habitats with roadways, lead to substantial human casualties, wildlife fatalities, and economic losses. The complexity of predicting AVCs, influenced by environmental and geographic factors, necessitates a data-driven approach to identify risk areas and implement preventive measures.

This project aims to tackle this issue by developing a machine learning model using logistic regression, focused on analyzing geographic data to predict and mitigate the risks associated with AVCs.

**Methodology**

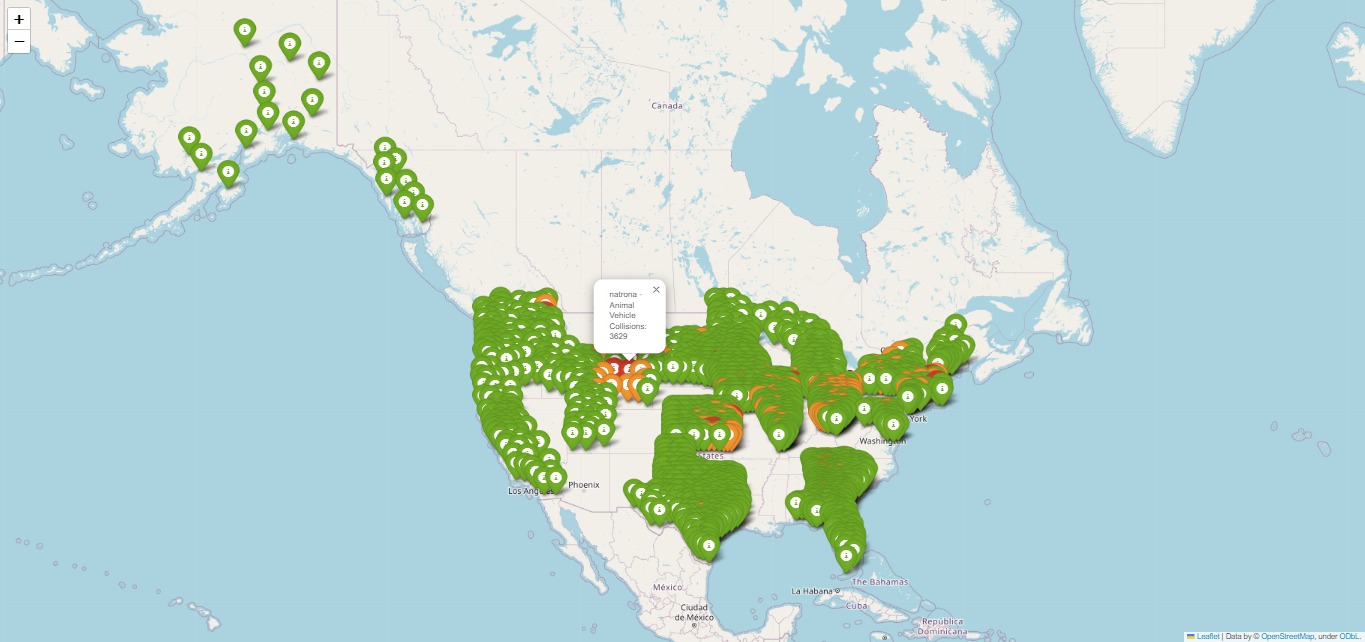
The methodology of this project revolves around the use of logistic regression in machine learning to predict the likelihood of animal-vehicle collisions (AVCs) based on geographical data. Initially, the project involved the collection and preparation of a comprehensive dataset, focusing on the longitude and latitude coordinates of known AVC incidents. The preprocessing of this data included cleaning, normalizing, and structuring the data in a format suitable for analysis. Key steps in this phase included handling missing values, categorizing data based on various parameters such as time and geographic location, and transforming categorical data into a format amenable to machine learning algorithms.

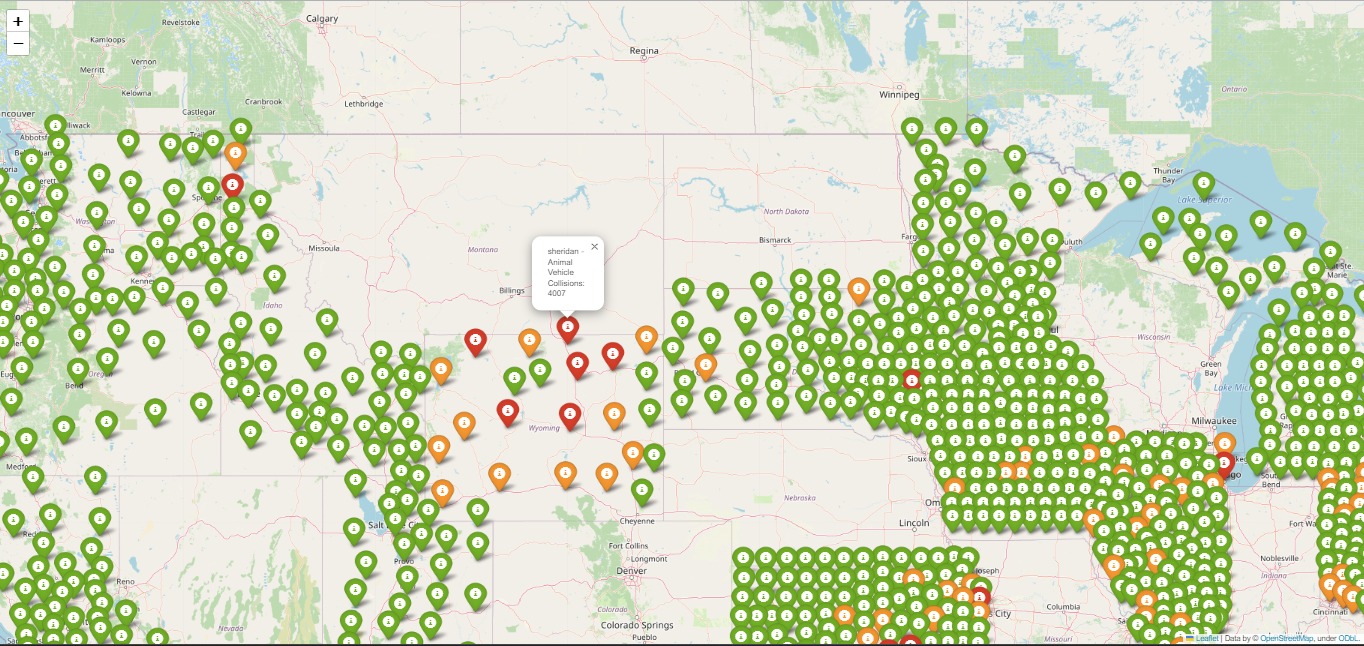
Following data preparation, logistic regression was employed to develop a predictive model. Logistic regression, a powerful statistical method for binary classification, was chosen for its efficacy in handling binary outcomes, in this case, the occurrence or non-occurrence of an AVC. The model was trained using a subset of the dataset, with careful partitioning into training and testing sets to validate the model's accuracy and robustness. To enhance the model's predictive capability, various features such as time of day, month, and specific geographic coordinates were integrated, allowing for a nuanced analysis of collision risks. The final phase of the methodology involved evaluating the model's performance using standard metrics like accuracy and classification reports. The integration of machine learning with geographic data analysis in this project showcases an innovative approach to tackling the complex issue of AVCs.

**Experiments and Results**

1. **Folium County Map**: There were several issues that we encountered with the datasets, the most prevalent was the Null values in geographical data i.e., longitudes and latitudes. We handled these missing values by using the longitudes and latitudes of the county where were the nearest geographical co-ordinates from the collisions, this helped us in removing the inconsistencies of the dataset. By doing this we were able to identify the hotspot counties in various states across the US where these accidents were most prevalent.

We created a folium map of the US which shows the counties on the map along with the number of collisions which have occurred there. In this way hotspots were identified for animal vehicle collisions.

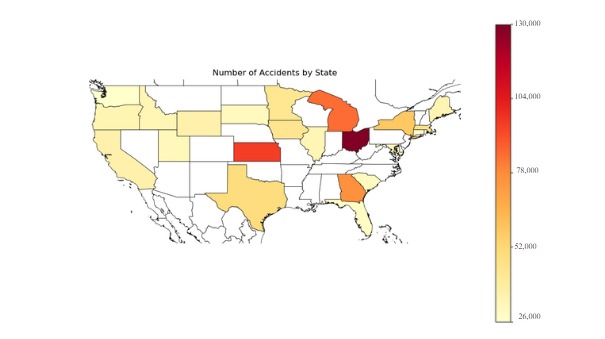




The colors represent the intensity of the accidents in these counties, where red indicates high number of collisions (3500+), while yellow is for mild number of collisions (1800+) and green represents number of collisions below 1800 in the given time frame of 1994-2021.

1. State Heat Map

This map allows for a quick understanding of collison distribution, helping to identify high-risk areas or states that require more attention in terms of safety measures and traffic management. The color gradient provides an immediate overview of the relative animal-vehicle collision frequency across the states, aiding in informed decision-making and resource allocation for safety improvements.

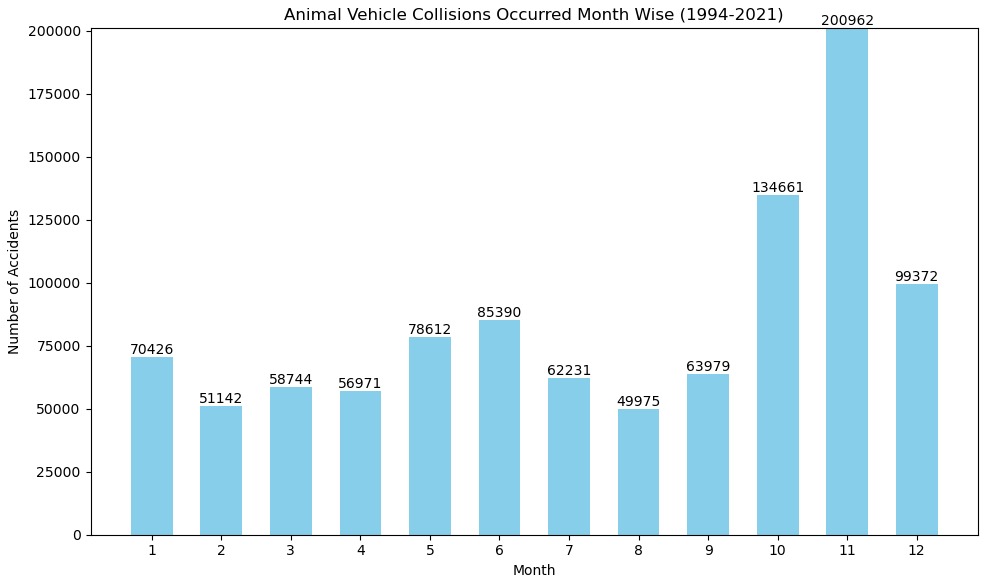


The following bar chart gives the highest five numbers of collision for each state, we can use this to understand the underlying causes in these states and their relation with population road conditions and weather patterns.

A graph of blue rectangular bars

Description automatically generated with medium confidence

1. The graph presents data on animal-vehicle collisions on a month-by-month basis spanning from 1994 to 2021. From a glance, there is a discernible pattern where the number of accidents involving animals increases significantly during the winter months. The months of November (Month 11) and December (Month 12) show the highest numbers of accidents, with November peaking at 200,962 incidents. The graph demonstrates a clear seasonal trend, with the least number of accidents occurring in the months of April (Month 4) through September (Month 9), indicating a higher frequency of collisions in the colder months.



This pattern could be attributed to several factors related to winter conditions. Shorter daylight hours in winter mean that drivers are more likely to be on the road during dusk or dawn when animals are most active, increasing the likelihood of collisions. Additionally, winter driving conditions can reduce visibility and vehicle handling, making it more difficult for drivers to avoid animals on the road. Animals may also be more likely to enter roads during winter months in search of food, particularly in areas where natural food sources are scarce due to snow cover. The data shown in the graph is a clear indicator that animal-vehicle collisions are a significant concern during the winter season and suggest that extra caution and preventive measures may be needed during this time.

1. The bar graph reinforces this distribution, showing a significantly higher number of incidents at night compared to other times of the day, with evening and morning following in frequency.

A graph showing the number of vehicles falling

Description automatically generated

The pie chart and bar graph presented illustrate the distribution of animal-vehicle collisions throughout different times of the day from 1994 to 2021. According to the pie chart, the majority of collisions occur at night (34.7%) and evening (31.3%), with morning collisions accounting for 27.0%, and afternoon collisions being the least frequent at 6.9%.

A pie chart with numbers and text

Description automatically generated

The reasons behind this distribution are multifaceted. Nighttime driving is associated with reduced visibility, making it more challenging for drivers to spot and react to animals on the road. Furthermore, many animals are nocturnal or crepuscular, meaning they are more active during twilight and night hours, increasing the likelihood of crossing roads when visibility for drivers is compromised. In the evening, reduced light conditions and the potential increase in traffic as people return home from work may contribute to a higher rate of collisions. Morning incidents are also relatively high, potentially due to animals returning from nocturnal activities. The afternoon sees the fewest incidents, likely due to better visibility and a decrease in animal activity during these hours. These patterns underscore the need for heightened awareness and caution when driving during times of limited light and high animal activity.