

Human and Wildlife Coexistence in Canadian National Parks:

Discovering trends in reported incidents and identifying target areas for promoting health and safety of humans and wildlife and mitigating negative incidents in our National Parks

Emma Billard

Student Number: 501114915

Toronto Metropolitan University

CIND820 XJH – Big Data Analytics Project – W2023

Professor Tamer Abdou, PhD

January 23, 2023

Abstract

Time spent in nature is wondrous. Whether you're drawn to witness epic mountainscapes, giant old growth forests, or wildlife in their natural habitat, there is always wonder to be found in the wild. But we must not forget that being able to witness our beautiful natural world is a privilege and a gift. We must care for our natural planet so that it continues to thrive year after year.

In Canada, our National Parks are maintained by the Parks Canada Agency. According to Government of Canada (2022), the Agency's mandate includes acting as guardians of the national parks and protecting our natural places to ensure they remain healthy and whole. Under this mandate, Gummer & Nicholl (2022) indicate that between 2010 – 2021, Parks Canada compiled four datasets of incidents of human-wildlife co-existence for the evaluation of trends to inform Parks Canada policies and to ensure safe visitor experiences while conserving wildlife and integrity of our ecosystems.

For my data analytics project, I will be using these four datasets, which contain 70,000+ records of incidents, responses, animals involved, and human activities related to human-wildlife coexistence. I will combine these datasets using the unique Incident Number's associated with each record. I will not be referring to the thirteen other datasets included in the same Open Record and which contain compiled summaries of the incidents as I will be conducting my own pattern mining and summarizing.

The main problem I seek to address is to identify target areas for promoting health and safety of humans and wildlife and for mitigating negative incidents in our National Parks. I will be using the theme of predictive analysis, specifically pattern mining and causality. The goal is to determine which incidents are the most serious, and find patterns and correlations that will allow me to make predictions on when, where, and why these incidents occur. This information will

allow me to develop recommendations for park visitors and park employees to help ensure the health and safety of both humans and wildlife. I will tackle this by focusing on the following research questions:

1. What patterns can be found in location and time of year for each of the following variables: human activities, animals involved, cause, and type.
2. What incidents are the most concerning (i.e. where there is potential risk for humans or animals)?
3. For incidents found to be the most concerning, what variables are most correlated with their occurrence? Can we accurately or reliably predict future similar incidents will occur near that location or that time of year?

Using the data analysis process discussed in Babaoglu (2018) as an outline, I will go through the following steps and techniques for my analysis. I will use Python and will be incorporating visualizations into various stages of the analysis.

1. Data cleaning and univariate analysis: Looking for special characters and missing values – using hot decking or similar technique to deal with findings. Applying type conversion and character manipulation to ensure consistency and correctness.
2. Exploratory analysis: Use various plots and sub-setting to explore patterns in the data.
3. Multivariate analysis: Using the Kruskal-Wallis Test to determine correlation between variables.
4. Dimensionality reduction: Using recursive feature elimination to help understand which variables are most and least important.
5. Determine experimental design: At this stage, I will decide how to split the data for training and testing and whether considerations are needed for imbalanced data.

6. Modeling: Using random forest decision trees (classification) to predict where, when, and why future “serious” incidents will occur.
7. Evaluation and improvement: calculate evaluation metrics (accuracy, recall, and prediction) based on the confusion matrix of my model(s). Consider how I can improve the model to achieve more desirable metrics.

References

Babaoglu, Ceni. (2018, January 5). *How to Conduct Data Analysis Process Systemically*.

YouTube. <https://www.youtube.com/watch?v=NMffLbFql5k>

Government of Canada. (2022, November 19). *The Parks Canada mandate and charter*.

<https://parks.canada.ca/agence-agency/mandat-mandate>

Gummer, D., & Nicholl, S. (2022, September 15). *Human-wildlife coexistence incidents in selected national parks from 2010 to 2021*. Government of Canada.

<https://open.canada.ca/data/en/dataset/cc5ea139-c628-46dc-ac55-a5b3351b7fdf>