Analysing Fatalities in the World's Eight-thousanders

FIT5147 – Data Exploration Project

Table of Contents

1.0 Introduction

Summiting the world's tallest peaks, the eight-thousanders with peaks over 8,000 meters, is a coveted feat in extreme mountaineering. However, this urge comes at a significant cost. In the past few years, more people have died in these mountains than before (Redakton and Reuters, 2023). This project delves into the tragic reality of climber deaths on eight-thousanders over these years.

A multifaceted motivation drives this research. As an avid climber myself, with aspirations for mountaineering, understanding the fatality patterns is crucial for safety and informed decision-making for this sport. Furthermore, this project aims to benefit a wider audience by revealing these insights. Climbers, adventure leaders, and policymakers will acquire valuable insights for enhancing security protocols and guiding ethical behaviour. The goal is to empower the outdoor recreation community to navigate the inherent risks of high-altitude mountaineering with a greater sense of awareness.

This project tackles two primary questions:

1. **What patterns have emerged in fatalities on the world's highest mountains over the past five years?** We will explore these patterns from various perspectives, including geographic location, climber nationality, and the specific timeframe within climbing seasons.
2. **What environmental factors contribute to these fatalities and why?** We will investigate the climatic conditions under which these tragedies most frequently occur.

Combining data from multiple sources, including dedicated climbing accident datasets, detailed mountain information, and comprehensive climate data, I attempt to paint a clearer picture of the risks associated with scaling the world's highest peaks. This project will utilise Python for data manipulation, and R and Tableau for analysis, and visualisation.

2.0 Data Wrangling

This section will outline the data description, sources, and the key steps taken in Python for cleaning and transforming the data.

2.1 Loading and combining datasets:

1. Utilising the **Mountain Climbing Accidents Dataset**, a CSV file with 1700+ rows and 5 columns, detailing fatalities on eight-thousanders, including mountain names, climber details (name, nationality), and fatality specifics (date, cause).
2. A supplementary dataset, the **Mountain Location dataset,** also in CSV format, furnishing geological characteristics like height (in meters and feet) and location of various mountains, was incorporated. Employing a left join, this data was amalgamated with the primary dataset to enrich it further.
3. To pinpoint precise geographic coordinates for each mountain and nationality, the **ArcGIS API** in Python was deployed. Leveraging the geocode() function, the geographic coordinates were imported, crucial for visualising the data on a map and for subsequent weather data extraction (*Find Location Coordinates—ArcGIS Online Help | Documentation*, n.d.).
4. To uncover climatic conditions during fatalities and identify patterns, historical weather data (1940 – 2023) for each mountain was sourced. This entailed utilising the reliable **Open-Meteo API** (Historical Weather API | Open-Meteo.com, n.d.). Daily temperature and precipitation data were enriched for each mountain's exact coordinates, augmenting dataset depth.

The primary dataset is left-joined with each subsequent dataset according to the join format shown in the image below:

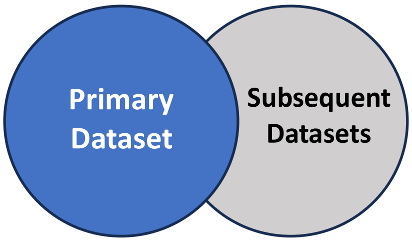

Image 1: Left join visual representation

2.2 Cleaning the final data:

The below steps were executed to clean the data and improve its usability:

- **Identifying Data Structure**: The dataset's structure was assessed, examining column names and data types to ensure compatibility with analysis tools.
- **Improving Readability**: Column names were refined for clarity using the rename function from the 'dplyr' package, enhancing the dataset's understandability and ease of analysis.
- **Ensuring Proper Data Formats**: Correct data formats are vital for accurate analysis. For example, date columns were converted from text to date format using the as.Date function to facilitate chronological analysis.

2.3 Data Transformations:

1. **Temperature Category**: Categorises temperatures into four groups: 'Extremely Cold', 'Cold', 'Warm', and 'Hot' (below 0 °C, 0 °C to 17 °C, 17 °C to 24 °C, above 24 °C respectively), providing a clear understanding of temperature ranges.
2. **Precipitation Category**: Divides precipitation levels into four categories: 'No Precipitation', 'Slight Precipitation', 'Moderate Precipitation', and 'High Precipitation' (0 mm, 0-0.8 mm, 0.8-2.9 mm, above 2.9 mm), offering insight into the intensity of precipitation.

A literature review on 'Climatic Thresholds' was used to determine the thresholds for categorisation (Asseng et al., 2021).

3. **Weather Condition**: Combines temperature and precipitation categories into a single descriptor, facilitating a comprehensive overview of weather conditions at each mountain location.
4. The '**cause of death' column** was condensed from 176 unique causes of death into 9 broader categories. By grouping similar causes, such as 'fall,' 'fell,' and 'rope,' into 'Fall', the dataset was better modelled for analysis. This involved handling missing values by adding them to a tag 'Unknown', converting text to lowercase, and streamlining the original lengthy descriptions into more manageable groups.

The final dataset has the dimensions (1700 × 14), and the following columns:

Columns	**The purpose of adding**
Mountain Name	These are columns from the primary dataset and provide information about the fatalities, including the cause, date, mountain range and the deceased individual's name and nationality.
Accident Date	
Climber's Name	
Climber's Nationality	
Cause of Death	
Height in Meters	The altitude of the mountains will aid in understanding the technical difficulty level of climbs.
Height in Feet	
Location (Country)	To identify hotspots of eight-thousanders.
Mountain Coordinates	Map the location of the mountains to identify accurate weather conditions, related risks, and optimal climbing seasons.
Country Coordinates	The longitude and latitude of the country that the deceased person belongs to allow us to visualise where people are travelling from to climb these mountains.
Temperature	Insights into weather conditions during time and location of accidents.
Precipitation	
Temperature Category	Indicates the severity of the weather during the accident.

Precipitation Category	Indicates the severity of rainfall during the accident.

A secondary dataset, comprising over **400K rows and 6 columns** containing daily weather information, has been utilized to examine climatic patterns across these eight-thousanders.

3.0 Data Checking

Going further in this report, all functions are done in R.

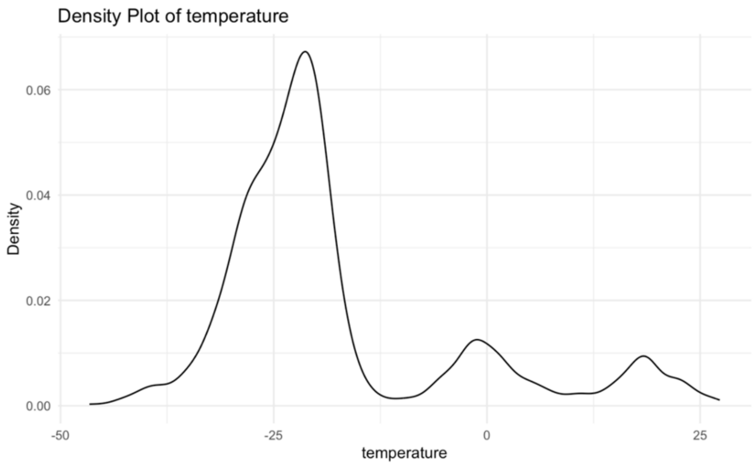

Image 2: Sample density plot of the 'temperature' column for outlier detection

- **Data Distribution and Outlier Detection**: The ggplot2 package was employed to analyse the distribution of all numerical columns within the dataset. It was noticed that none of the columns followed a normal distribution; instead, they displayed skewness. Outliers were detected *(Image 3)* in the temperature and precipitation data. However, these outliers are not considered erroneous within the project's context since extreme climates are expected at different altitudes. They offer valuable insights that enrich our understanding of accidents in mountainous regions.

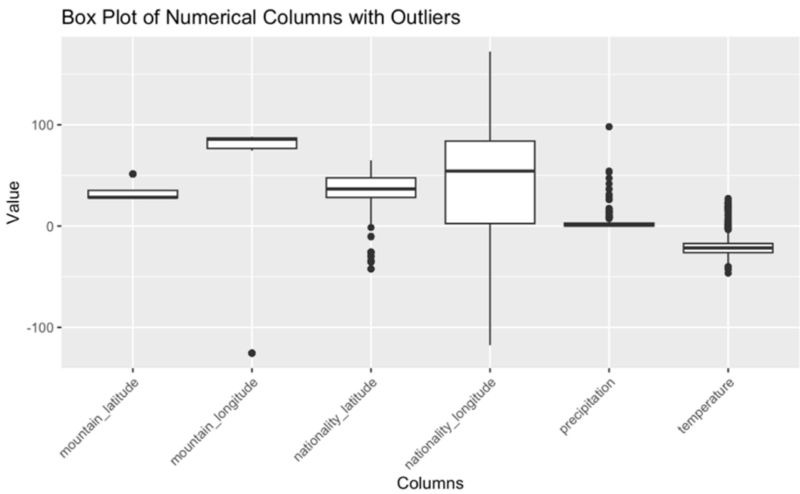

Image 3: Box plot for numerical columns portraying existing outliers.

- **Duplicate data**: The duplicate values were examined across all columns of the dataset using the duplicate() function in R. The analysis revealed no duplicate rows, indicating the dataset's integrity.
- **Data Consistency and logic**: The consistency and logical coherence of relevant columns, such as recorded temperature and precipitation values, align with known climate patterns for the mountainous region, climber nationality, and cause of death, to ensure logical consistency and accuracy in the dataset. This was a manual check done in Excel keeping the context of the project in mind.
- **Null Values**: The vismiss() function was utilised to detect null values in each column of the dataset *(Image 4)*. Given the data does not conform to a normal distribution, imputing missing values using mean or other methods was considered unsuitable. Consequently, the rows with null values were retained to maintain the dataset's integrity, as removing them could potentially compromise valuable information. It is important to note that the null values were found in data that was subsequently added and not in the primary dataset. Hence, specific null values were handled in a column-wise manner by grouping them into **'Others'.**

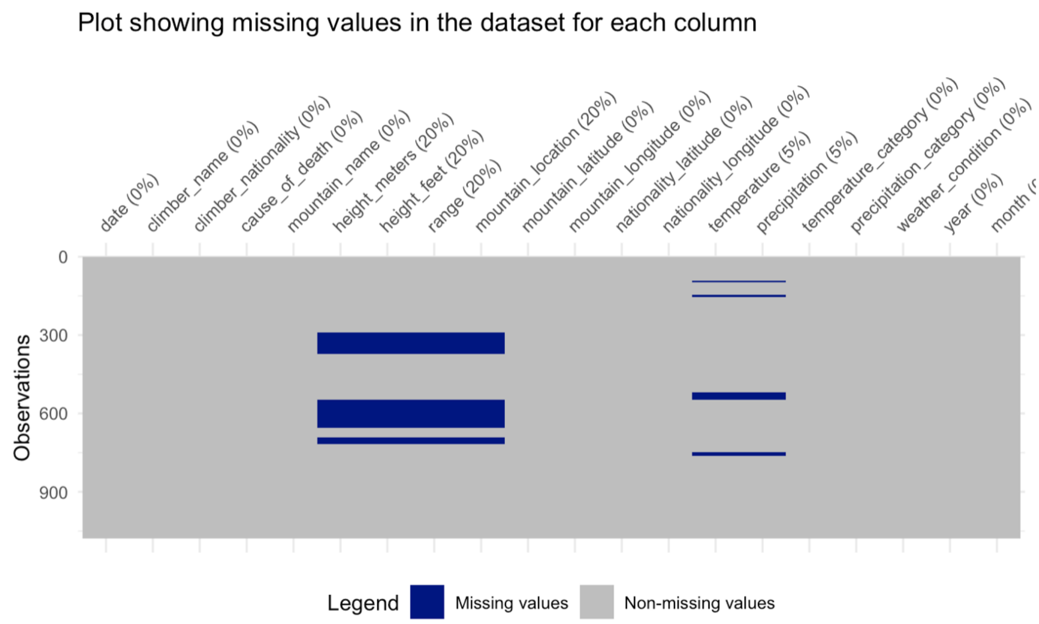

Image 4: Plot shoeing Missing Values in Each Column of the Dataset

4.0 Data Exploration

In this section, we will analyze fatality patterns on the world's highest mountains over the past years, employing a multifaceted approach. Using R and packages such as 'ggplot2' and 'dplyr', we will visually map geographic distributions, analyze temporal trends, and scrutinise demographic data. Through meticulous visualisations and statistical analyses, our exploration aims to uncover a comprehensive understanding of fatality occurrences while respecting the sensitivity of the subject.

4.1 Fatality Patterns Across Mountain Ranges and Time:

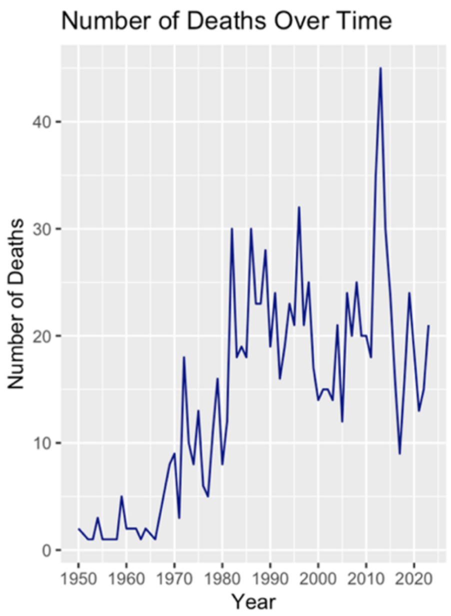

Image 5: Line Plot showing the Number of Accidents by Year.

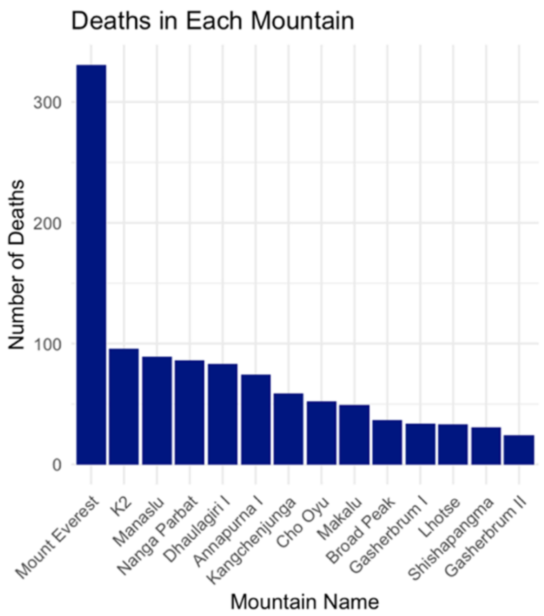

Image 6: Bar Plot displaying the Number of Accidents in Each Mountain

The adjacent **caret plot** provides a comprehensive overview of accidents occurring across various mountain ranges recently. In tandem, the line plot *(Image 5)* depicts an uptick in accident frequency post-1980, suggesting multiple plausible contributing factors, including heightened climbing endeavours, climatic fluctuations, or evolving expedition methodologies. Notably, the year 2012 witnessed a surge in fatalities, attributed to a significant climatic event—a sudden Tsunami at Mount Everest, claiming 11 lives in a single day (Krakauer, 2017).

Furthermore, the subsequent bar chart *(Image 6)* accentuates the nature of Mount Everest, with the highest fatality count, followed by K2 and Manaslu. This revelation underscores the need for delving into the motivations behind the popularity of these peaks for climbing and the distinctive variables—such as elevation, weather dynamics, technical intricacies, and rescue capabilities—that shape each expedition.

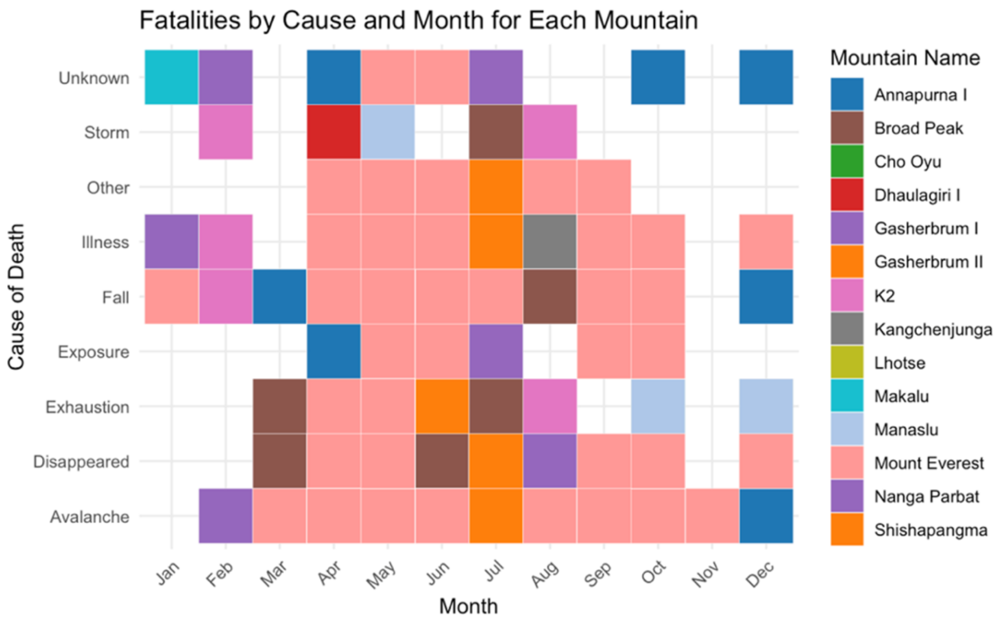

Image 7: A Tile map of Fatalities by Cause and Month for Each Mountain

On this journey of exploring fatality patterns, it is imperative to comprehend the cause of death among climbers. Climbers are primarily hiking from **April to July**, likely to be the warmer seasons *(Image 7).* The tiled heat map shows that **avalanches, falls, and illnesses (x-axis)** emerge as the primary causes of fatalities, in that order.

Research from 2023 on causes of avalanches highlights that warm temperatures and high precipitation in the lower mountain zones have a high probability of avalanches, particularly during the **April-May season** (Acharya et al., 2023). Climber Falls, notably on **Mount Everest** (indicated by pink colour), is often associated with symptoms of high altitude during the descent, exacerbated by factors like fatigue and reduced visibility (Firth et al., 2008). On the other hand, Illness and exposure issues result from decreased barometric pressure and oxygen levels, with insufficient **acclimatisation** among climbers, mainly those under 40 years, contributing to 64% of fatalities (Windsor et al., 2009).

This emphasizes the importance of considering **acclimatisation** and **timely medical assistance**, especially in the Asian mountain ranges like Mount Everest and K2 during their peak climbing seasons from April to June.

4.1 Patterns in Fatality Cause and Geological Factors:

The above-mentioned insight that the number of fatalities varies by mountain range and month suggests that the causes can be attributed to the sheer number of people attempting to ascend these mountains or the various obstacles. To investigate this, we examine the **impact of altitude on the difficulty of climbing**. The below heat map *(Image 8)* shows that most fatalities occur at higher altitudes, in this case, at **29K feet**. It is interesting to note that **falls are the main cause of death at the highest altitudes**, despite avalanches being the primary cause overall. Avalanches are the main concern at 26.6k feet and over 28k feet. This visualisation confirms our insight above where avalanches occur at lower altitudes with higher precipitation.

Illness along with fall is a key concern at **mid-range altitudes (between 27k and 28k feet)**, contributing to 70% of total mortality at that altitude. Illness continues to be a concern at higher altitudes, making it a reason for outdoor communities to prioritise and be cautious about. These insights stoke curiosity about the effects of weather at different altitudes on mortality, which will be explored further.

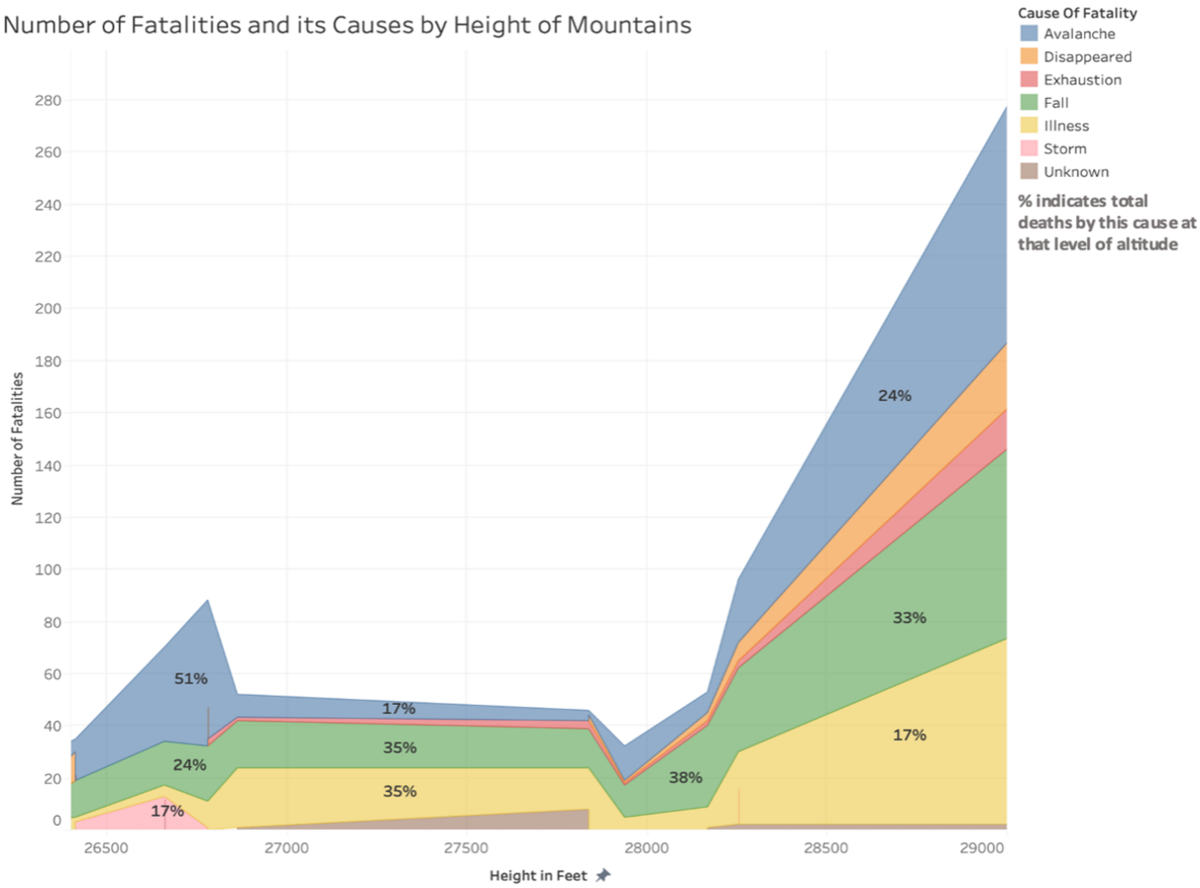

Image 8: An Area Plot showing the Number of Fatalities and Cause of Fatalities at Different Altitudes

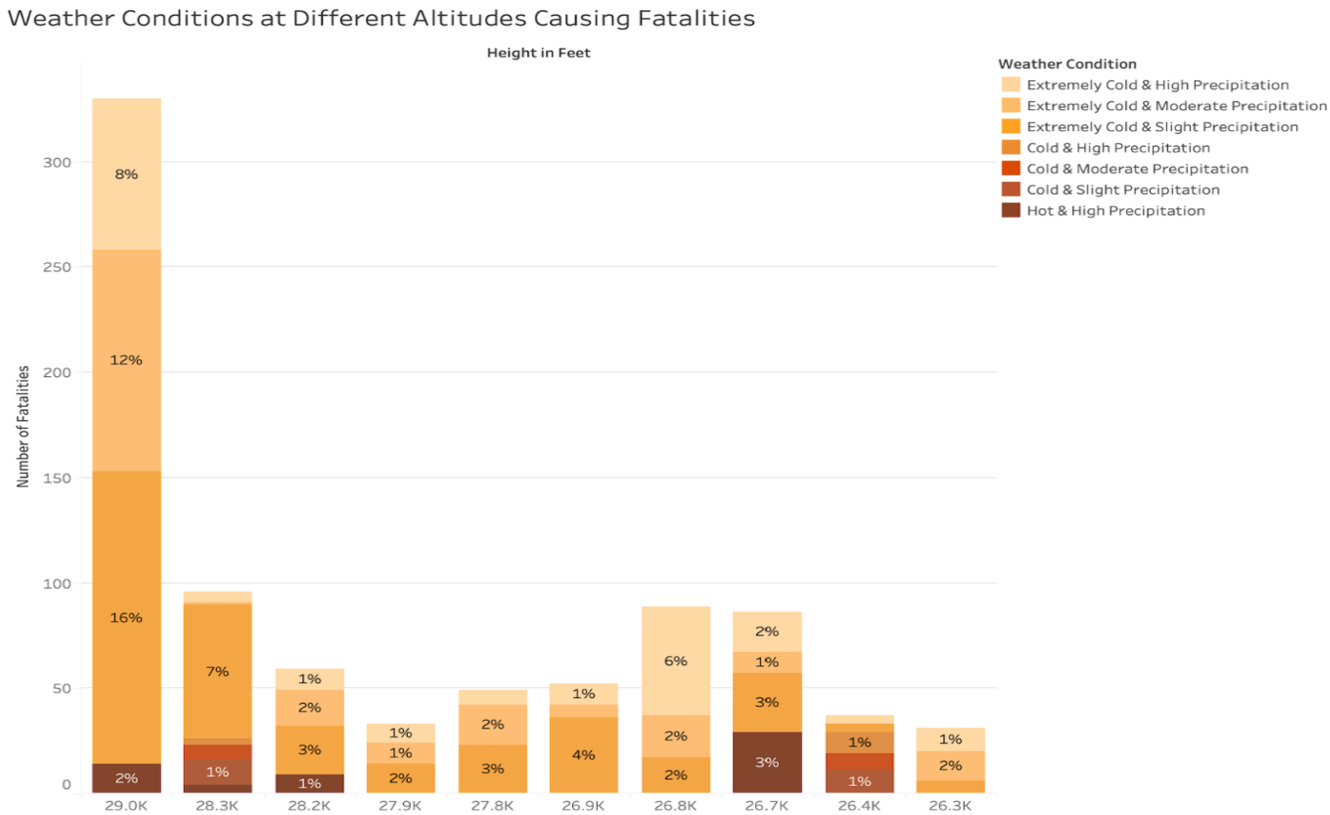

Image 9: A Stacked Bar Chart showing the Number of Fatalities by Altitude and Weather Conditions.

Climate conditions like extreme cold, precipitation, and wind pressures have a large impact on climbers, making the climb more difficult. Weather fluctuations due to altitude and season shifts result in alterations in barometric pressure and oxygen levels, making survival difficult (Szymczak et al., 2021). The stacked bar graph *(image 9)* shows a **positive correlation** between altitude and the number of fatalities, which was previously observed. Upon further examination of the challenges posed by weather conditions (which range in hue from an extremely cold climate to a darker-coloured warmer climate), most fatalities occur at **high altitudes with extremely cold temperatures and slight precipitation** (16% of fatalities in 29k feet). This climatic condition is the most fatal, even as altitudes drop. **Extreme cold temperatures along with moderate precipitation** also account for a substantial amount of climber deaths, specifically in altitudes between 26k-27k feet.

In general, based on the established thresholds, **temperatures below 0 °C** and precipitation **above 2.9 mm** constitute the primary cause of mortality at **altitudes above 28 feet**, whereas the same temperatures with precipitation between **0.8-2.9 mm** pose significant concerns at **lower altitudes**. It can also be noted that most fatalities have occurred at freezing temperatures, but with varying amounts of precipitation. This suggests that **precipitation fluctuations play a more important role**.

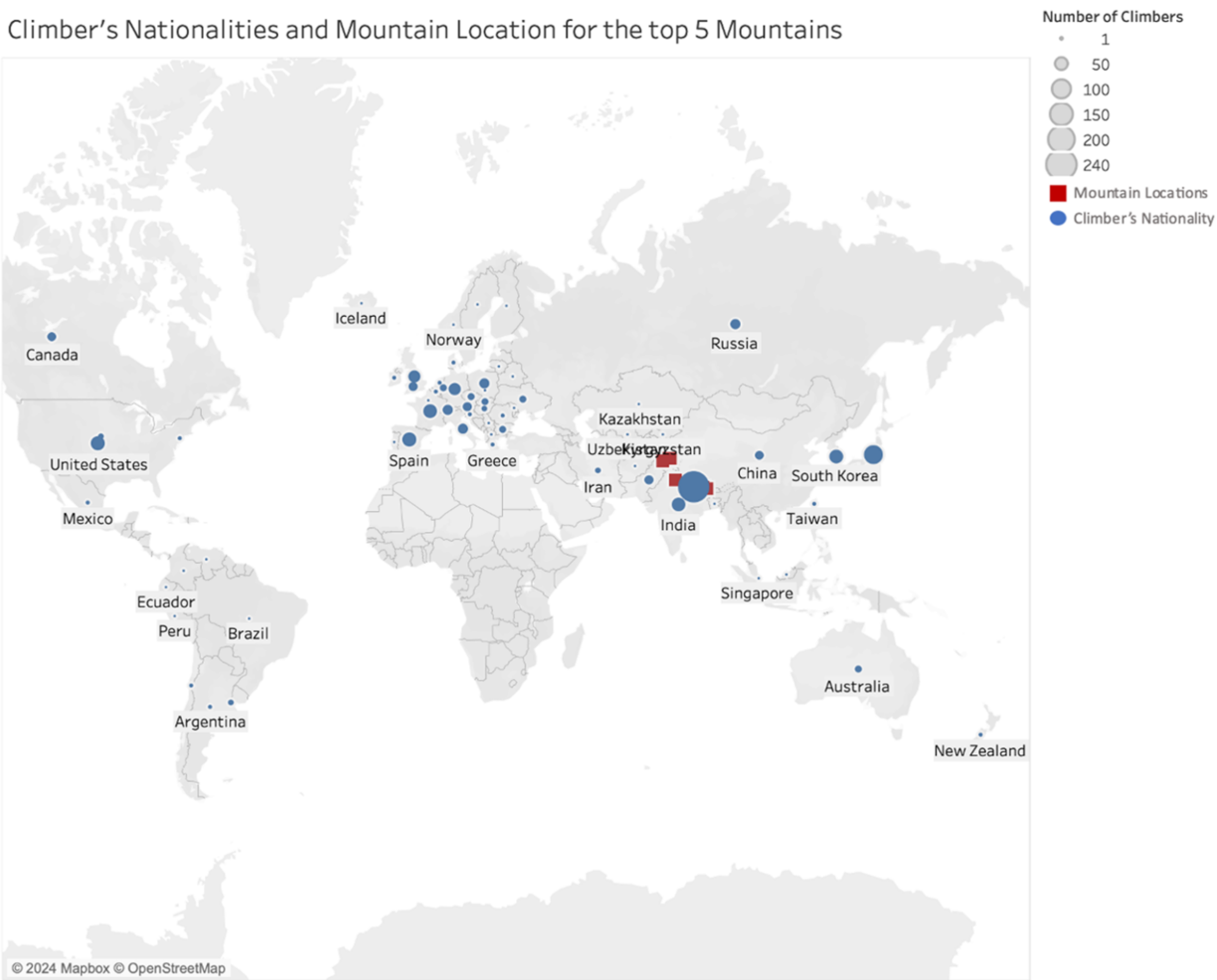

Image 10: A Map showing the Nationality of Climbers Travelling to the Eight-thousanders.

To delve deeper into understanding where people are travelling from to summit these peaks, our analysis focuses on the top five mountains with the highest number of fatalities: Mount Everest, K2, Manaslu, Nanga Parbat, and Dhaulagiri. A **map** was used to analyze where climbers were coming from and the impact of geographic locations on climber motivations. The blue markers indicate locations people are travelling from and the red markers are the locations of mountains.

Examining the map (Image 10), we note a diverse distribution of nationalities represented, with a notable concentration of climbers from **India and Nepal** (30%), indicated by the largest circle markers. This could be attributed to their geographical proximity to these formidable peaks. Additionally, a significant presence of **Japanese nationals** is evident (10%), closely followed by climbers from **South Korea** (5%), **Spain** (5%), and the **United States** (4%).

Understanding these nationality patterns underscores the need for **international collaboration and cooperation in mountain safety and rescue efforts**. Higher levels of safety can be ensured by increasing training and resources for climbing communities in these countries.

4.2 Clustering Analysis of Fatality Patterns:

This study employs **clustering**, an unsupervised learning technique, to group information based on numerical and categorical characteristics. The objective of this analysis is to understand the common patterns among fatalities and to use these segments to identify **high-risk conditions for mountaineering**. For instance, clusters of fatalities in specific mountain ranges can indicate hazardous climate for the climb and assist in route planning.

The analysis was initiated by **pre-processing the data**, including converting categorical values to numerical format and scaling (standardising) numerical features. For clustering, multiple clustering algorithms were explored, including **hierarchical clustering and partitioned clustering**. The initial algorithm was unsuitable for large datasets such as this because it used a **bottom-up approach**. Moreover, managing outliers and generating homogeneous clusters was not a benefit of this algorithm. On the contrary, K-means clustering employs **partitioning based on centroids** and generates a comprehensive set of clusters for features including temperature, precipitation, altitude, mountain ranges, and the causes of fatalities.

In **k-means clustering**, the number of clusters (K) is a crucial parameter for making meaningful cluster assignments. However, determining the optimal values of k requires a balance between model complexity and cluster interpretability.

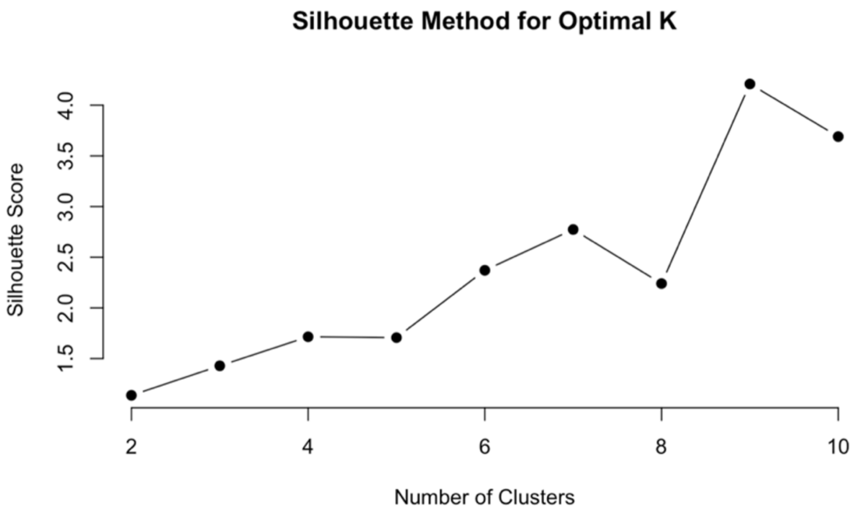

Image 11: A Line Plot for the Silhouette Coefficients against the Number of Clusters.

The **Silhouette method** assists in determining this optimal value by calculating a coefficient for each data point, and how close each data point is to its cluster compared to other clusters. The coefficient ranges from -1 to 1, where a high value indicates that the data point is well clustered. Based on the line graph (Image 11), we can notice that **9 clusters are optimal** for this study.

The algorithm was then fit with 9 as the clustering parameter to partition the data. *Image 12* shows the clusters that were formed.

cluster <int>	weather_mode <chr>	nationality_mode <chr>	cause_of_death_mode <chr>	min_temperature <dbl>	max_temperature <dbl>
1	Extremely Cold & Slight Precipitation	Pakistan	Fall	-20.8	-0.3
2	Cold & Slight Precipitation	Poland	Fall	-12.5	12.6
3	Extremely Cold & Slight Precipitation	Japan	Fall	-46.6	-13.7
4	Extremely Cold & Slight Precipitation	Nepal	Fall	-41.8	-18.6
5	Extremely Cold & High Precipitation	Nepal	Avalanche	-37.3	-17.9
6	Extremely Cold & Slight Precipitation	Nepal	Fall	-34.2	-14.8
7	Extremely Cold & Slight Precipitation	Nepal	Illness	-36.0	-19.3
8	Extremely Cold & High Precipitation	Nepal	Avalanche	-35.9	-16.5
9	Extremely Cold & High Precipitation	Germany	Avalanche	-27.7	-13.9

Image 12: A Table with Cluster Profiles

The clusters formed were valuable to the study, but **suboptimal** due to multivariate data complexity, outliers valuable to the domain, imbalanced data distribution, and the inherent complexities of mountaineering incidents. Despite encountering challenges, the formed clusters offer valuable insights into diverse weather conditions, altitudes, and causes of death in mountaineering incidents. Notably, intriguing patterns emerge, such as Nepal nationals may appear across multiple clusters, but it's essential to highlight that their fatalities are influenced by **extremely cold temperatures ranging from -42 °C to -15 °C (cluster 4 – 8)**, whereas fatalities for other nationals occur at lower temperatures. Additionally, observing *Image 13*, mountain ranges like K2 and Broad Peak (cluster 7) experience **relatively warmer climates** compared to others, despite their geographical distance. Surprisingly, **Poland nationals** constitute the highest number of fatalities at these peaks, even though the peaks are not geographically proximate to the country.

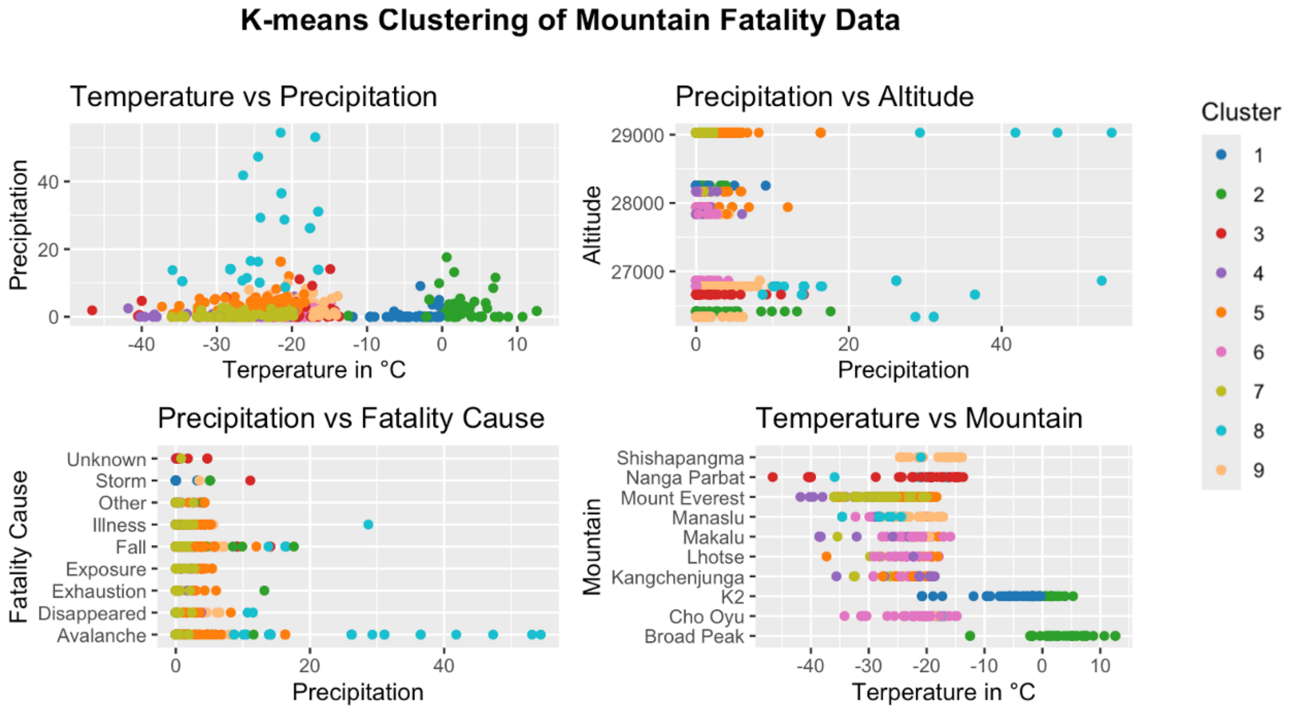

Image 13: A Grid of Scatter Plot of Features Key for Risk-Mapping

Overall, the clustering analysis provides insights into the complex interplay between essential factors causing fatalities in mountains. These clusters can prove valuable for providing safety information and better planning of expeditions. Further investigation into factors like **climbers' features, and safety protocols** in each of these countries can help address customise fatality concerns.

5.0 Conclusion

In conclusion, this study sheds light on the concerning rise in fatalities on the world's highest peaks. By analysing 14 mountain ranges, it pinpoints **Everest, Manaslu, K2, and Dhaulagiri** as particular danger zones, shedding light on the pressing need for enhanced safety measures in these areas. The research unveils a complex interplay of factors contributing to these deaths such as mountain ranges, weather conditions, altitude, and need for climber fitness. **Avalanches, falls, and altitude sickness** emerges as leading threats, with their prominence shifting across altitude ranges.

A crucial takeaway lies in the **impact of climate change**. While **extreme cold climates** at high altitudes can certainly lead to illness, it's not the sole culprit. The combination of **thin air (caused by high altitude) and low pressure** creates a harsh environment for survival. Conversely, lower altitudes may seem safer, but **warm temperatures and increased precipitation** are causing more frequent avalanches, defying traditional expectations. This underscores the importance of meticulous expedition planning that considers not just the mountain itself, but also seasonal variations in **precipitation** mainly.

Furthermore, the study highlights a trend of fatalities among climbers from **Nepal, Japan, and South Korea**. This suggests a need for targeted safety protocols in these regions, delivered in native languages. The cluster profiles from the analysis hold promise for **optimising safety strategies** based on factors like mountain range, season, and altitude. Thus, the research paves the way for customised safety plans based on specific mountain ranges, seasons, and altitudes.

Looking ahead, further research can refine these insights. Analysing climber profiles by gender and age can illuminate fitness levels, while injury data can pinpoint particularly hazardous situations. This comprehensive understanding will be instrumental in establishing best practices for safer mountaineering expeditions.

6.0 Reflection

This project was a rewarding journey that allowed me to delve into a topic I'm truly passionate about, fostering creativity and a real-world connection. I embraced a user-centric perspective, shaping the report with their needs in mind. A significant takeaway was my improved approach to data visualisation. Previously, I prioritised the message over the data types involved in the data. This project highlighted the crucial role data types play in selecting the most impactful visualisations. Furthermore, I honed my attention to detail, focusing on aspects like colour choice, legends, and titles.

Looking ahead for this study, I would prioritise in-depth research on multivariate data analysis techniques and explore the potential of newer algorithms for a better fit with my data. Perhaps incorporating an additional data source as mentioned above could enrich the analysis as well.

7.0 References

Acharya, A., Steiner, J. F., Walizada, K. M., Zakir, Z. H., Ali, S., Caiserman, A., & Watanabe, T. (2023). *Review article: Snow and ice avalanches in high mountain Asia – scientific, local and indigenous knowledge*. https://doi.org/10.5194/nhess-2022-287

Asseng, S., Spänkuch, D., Hernandez-Ochoa, I. M., & Laporta, J. (2021). The upper temperature thresholds of life. *The Lancet Planetary Health*, *5*(6), e378–e385. https://doi.org/10.1016/s2542-5196(21)00079-6

Cronin, C. (1991). Sensation seeking among mountain climbers. *Personality and Individual Differences*, *12*(6), 653–654. https://doi.org/10.1016/0191-8869(91)90264-c

Find location coordinates—ArcGIS Online Help | Documentation. (n.d.). Doc.arcgis.com. Retrieved April 17, 2024, from https://doc.arcgis.com/en/arcgis-online/get-started/find-location-coordinates-mv.htm

Firth, P. G., Zheng, H., Windsor, J. S., Sutherland, A. I., Imray, C. H., Moore, G. W. K., Semple, J. L., Roach, R. C., & Salisbury, R. A. (2008). Mortality on Mount Everest, 1921-2006: descriptive study. *BMJ*, *337*(dec11 1), a2654–a2654. https://doi.org/10.1136/bmj.a2654

Historical Weather API | Open-Meteo.com. (n.d.). Open-Meteo.com. https://open-meteo.com/en/docs/historical-weather-api

Krakauer, J. (2017). *Death and Anger on Everest*. The New Yorker. https://www.newyorker.com/news/news-desk/death-and-anger-on-everest

Redaktion. (2023, December 16). *Deadliest season on Everest*. Lacrux Klettermagazin. https://www.lacrux.com/en/alpinism/deadliest-season-on-everest/

Reuters. (2023, October 8). Two die in avalanches on Shishapangma mountain in Tibet. *The Guardian*. https://www.theguardian.com/world/2023/oct/08/two-die-in-avalanches-on-shishapangma-mountain-in-tibet

Szymczak, R. K., Marosz, M., Grzywacz, T., Sawicka, M., & Naczyk, M. (2021). Death Zone Weather Extremes Mountaineers Have Experienced in Successful Ascents. *Frontiers in Physiology*, *12*. https://doi.org/10.3389/fphys.2021.696335

Windsor, J. S., Firth, P., Michael PW Grocott, & Montgomery, H. (2009, July). *Mountain mortality: A review of deaths that occur during recreational activities in the mountains*. ResearchGate; BMJ Publishing Group. https://www.researchgate.net/publication/26293826_Mountain_mortality_A_review_of_deaths_that_occur_during_recreational_activities_in_the_mountains