Mountaineering: Be lucky first, prepared second*

Preperation and Timing are more important than Gender Identity, Age, Amount of expedition members, and height of peak when mountaineering in the Himalayas.

Daniel Xu

December 4, 2024

This study analyzes over a century of Himalayan mountaineering expeditions (1905–2019) to identify key factors influencing success. Using logistic regression, we find that preparation and timing—such as the use of supplemental oxygen and climbing during optimal seasons—are more important predictors of success than individual traits like age, gender, or the height of the peak. These findings reveal that success in high-altitude climbing depends more on strategic choices than personal attributes. This insight underscores the value of informed planning and environmental awareness for achieving mountaineering goals.

1 Introduction

Mountaineering has been growing as an activity since the 90s, the views and sense of challenge have brought people around the globe to summit more and more mountains every year. It is no surprise then, that when combined with a mountain range containing the highest above sea level peak in the world the Himalayas have had a rich documented history of mountaineering attempts. Mountaineering is a dangerous hobby however, with unpredictable weather and temperatures so high up it is natural that not all attempts to summit mountain peaks succeed here. This paper will find what factors have affected the climbers of the past to either conquer or be conquered.

This paper uses data sourced from Cookson (n.d.), based on The Himalayan Database (Salisbury, n.d.) about expeditions from 1905-2019 to analyze which factors help or deter climbers

*Code and data are available at: https://github.com/possibleburger2/.

from succeeding. We use a binomial logistic regression model to determine the probability of success based on a variety of demographic, enviromental, and organizational factors.

The estimand is the relationship between said factors and the success of a mountaineering expedition. By analyzing the data set, we can identify which factors contribute to a success or failure and by how much.

My analysis shows that Mountaineering is positively impacted by oxygen used, younger age, going in the fall and spring, and being a man while is is negatively impacted by the height of the mountain and the amount of people not hired in the expedition.

These findings practically inform climbers about the impact of good preparation and timing over personal factors like age and sex, showing that regardless of who you are, what you do is the biggest factor in succeeding in high-altitude mountaineering.

The paper is further organized into four sections. Section 2 discusses how the dataset used for the analysis was obtained and pre-processed. I will explain the variables of interest in the dataset used for the analysis. Section 2.3 describes the model being used for the analysis. Section 2.5 then highlights and discusses the trends and associations found during the analysis. Lastly, Section 3 talks about some interesting trends found in Section 2.5 in depth, linking it to the real world and also highlights the weaknesses and future of my analysis.

2 Data

The datasets were cleaned and analysed using the statistical programming software `R` (Team 2023) along with the help of `tidyverse` (CiteTidyverse, n.d.), `knitr` (CiteKnitr, n.d.), `ggplot2` (CiteGgplot, n.d.), `here` (CiteHere, n.d.), `dplyr` (CiteDplyr, n.d.), `rstanarm` (CiteRstanarm, n.d.), `arrow` (CiteArrow, n.d.) and `kableExtra` (CiteKableExtra, n.d.).

2.1 Overview

The raw datasets were obtained from Cookson (n.d.), who sourced them from The Himalayan Database (Salisbury, n.d.).

The Himalayan Database is a compilation of records for all expeditions that have climbed in the Nepal Himalaya. The database is based on the expedition archives of Elizabeth Hawley, a long-time journalist based in Kathmandu, and it is supplemented by information gathered from books, alpine journals and correspondence with Himalayan climbers.

The original database currently covers all expeditions from 1905 through Spring-Summer 2023 to the most significant mountaineering peaks in Nepal. Also included are expeditions to both sides of border peaks such as Everest, Cho Oyu, Makalu and Kangchenjunga as well as to some smaller border peaks. Data on expeditions to trekking peaks are included for early attempts, first ascents and major accidents. The updates to this database are published bi-annually.

My dataset, derived from Cookson's, contains the entries from 1905 through Spring 2019.

A person becomes an entry in my analysis dataset if, between 1905 and Spring 2019, they attempted to climb any one of the many Himalayan peaks in Nepal.

The Peaks data includes several variables. The "peak_id" is a unique identifier for each peak. "peak_name" refers to the common name of the peak, such as "Mount Everest." "peak_alternative_name" is the alternative name of the peak, like "Sagarmatha" for Mount Everest in Nepal. "height_metres" is the height of the peak in metres, and "climbing_status" indicates whether the peak has been climbed or not. The "first_ascent_year" is the year when the peak was first successfully climbed, and "first_ascent_country" shows the country (or countries) of the climbers who made the first ascent. The "first_ascent_expedition_id" is a unique identifier for the expedition that achieved the first ascent.

The Expeditions data contains variables such as "expedition_id," which is a unique identifier for each expedition. "peak_id" links to the unique identifier for the peak being climbed, and "peak_name" refers to the common name of the peak. "year" is the year the expedition took place, and "season" indicates the season when the expedition occurred (such as Spring or Summer). The "basecamp_date" is the date the expedition arrived at basecamp, and "highpoint_date" is the date the expedition reached the highest point or summit of the peak. "termination_date" shows when the expedition ended, and "termination_reason" describes why the expedition was called off. "highpoint_metres" is the highest elevation reached by the expedition, while "members" refers to the number of people in the expedition, typically the number of foreigners in Nepal or non-hired members in China. "member_deaths" is the number of members who died, and "hired_staff" is the number of hired staff who went above basecamp. "hired_staff_deaths" shows the number of hired staff who died, and "oxygen_used" indicates whether any member of the expedition used oxygen. "trekking_agency" is the name of the agency organizing the expedition.

The Members data includes variables such as "expedition_id," which is a unique identifier for the expedition, linking to the Expeditions data. "member_id" is a unique identifier for each person, though it may differ across expeditions. "peak_id" links to the unique identifier for the peak being climbed, and "peak_name" is the common name of the peak. "year" is the year of the expedition, and "season" refers to the season of the expedition. "sex" is the sex of the person, and "age" is the person's age during the expedition based on available data (such as summit date, date of death, or basecamp arrival). "citizenship" is the person's nationality, and "expedition_role" is their role in the expedition. "hired" indicates whether the person was hired by the expedition, and "highpoint_metres" is the highest point reached by the person. "success" shows whether the person succeeded in summiting the main peak or a sub-peak, depending on the expedition's goal. "solo" indicates whether the person attempted a solo ascent, and "oxygen_used" shows whether the person used oxygen. "died" indicates whether the person died during the expedition, while "death_cause" shows the primary cause of death. "death_height_metres" is the height at which the person died, and "injured" indicates whether the person was injured. "injury_type" refers to the main cause of injury, and "injury_height_metres" is the height at which the injury occurred.

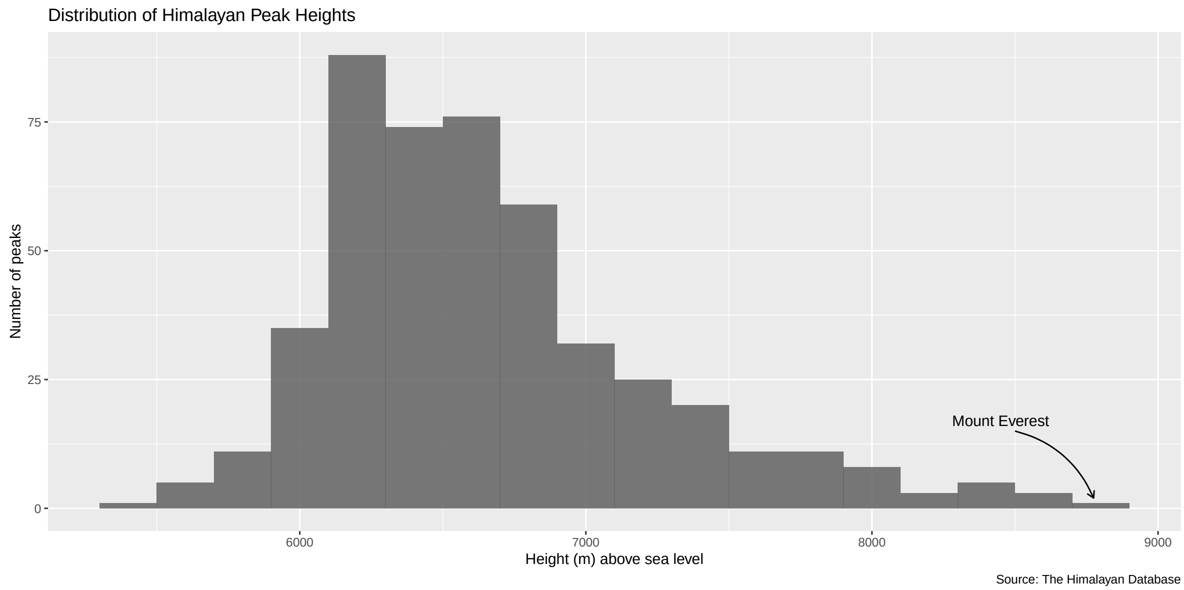

Figure 1: Distribution of Himialayan peaks, mountains recorded are centered around 6200 M above sea level..

Figure 1 shows most peaks sit around 6000 meters above sea level, with a right skewed distribution and a maximum of almost 9000 meters and a minimum of 5250 meters.

The Himalayan mountain range spans a large amount of distance, which means that the graphed distribution does not represent the distribution of ALL peaks in the Himalayas. Instead this distribution represents the distribution of peaks commonly attempted/climbed.The Himalayas starting at X M starting above 2000m sea level combined with the lowest height required to be considered a peak, the distribution starts at 5000m above sea level.

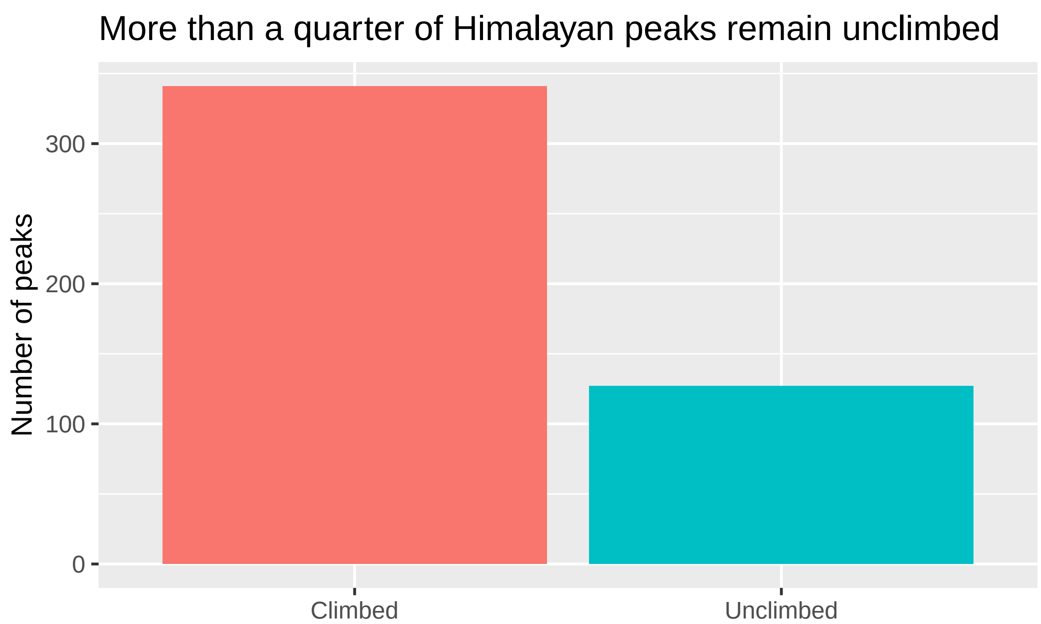

Figure 2: Amount of unclimbed Himalayan peaks

This graph shows that despite over 100 years of climbing history, not all attempted peaks have been conquered. Why do so many climbs fail however? The following graph shows the most common reasons.

Table 1: Common reasons why attempts fail

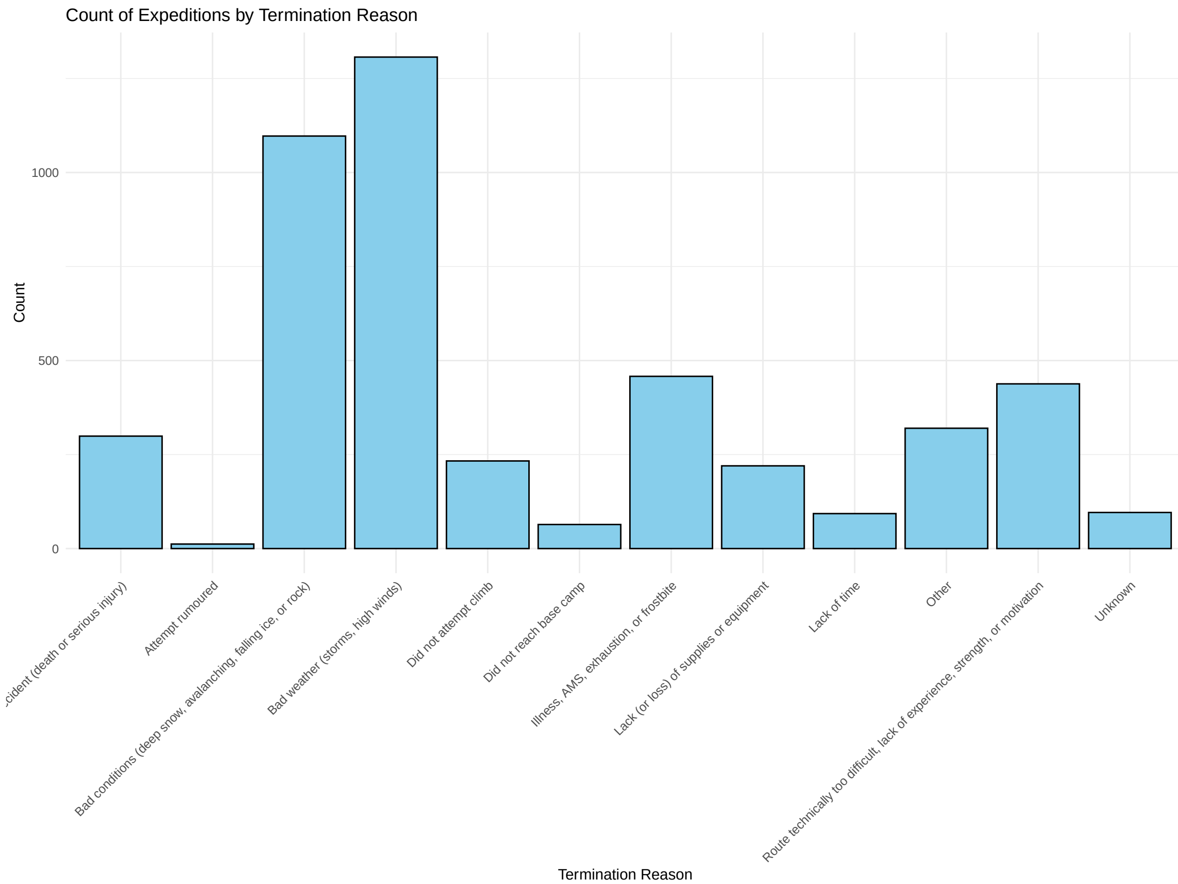

Table 1 shows that most attempts fail due to partially random events like weather, or random events such Avalanches.

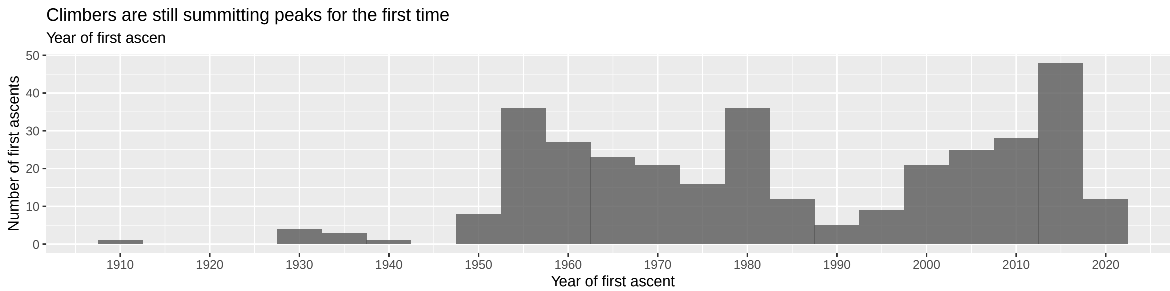

Figure 3: Distribution of First ascents across years

Figure 3 shows that while there was an intial surge then drop off of first ascents, there was again a steady increase in first ascents actually peaking in 2019, and the 2020 data was inconclusive since not all of the year was recorded in the data. This shows that first ascents are currently trending upwards

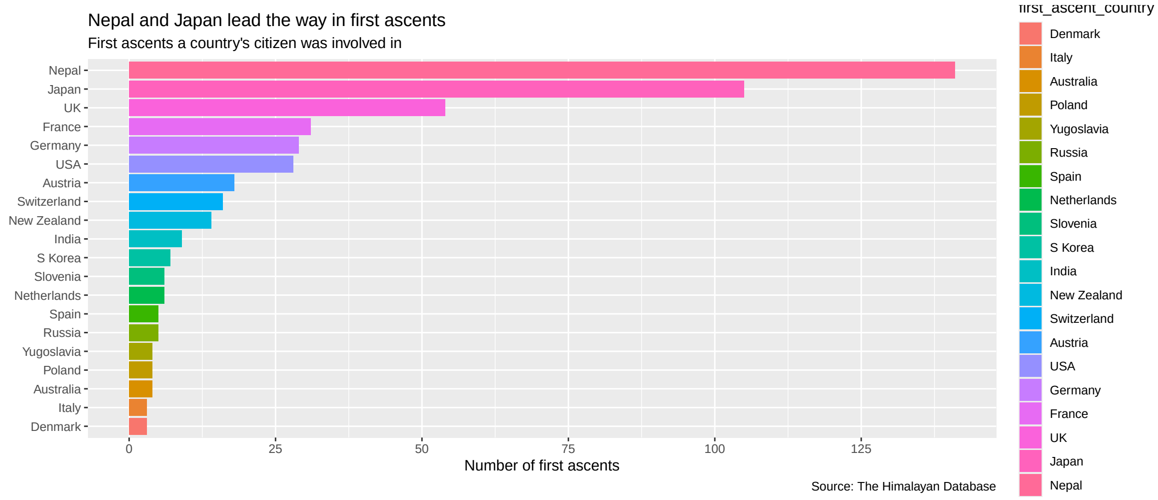

Figure 4: Distribution of first ascents by country of citizenship

Figure 4 Expectedly most first ascents are from the country that the Himalayas reside in, followed by Japan, UK, France, Germany, then the USA..

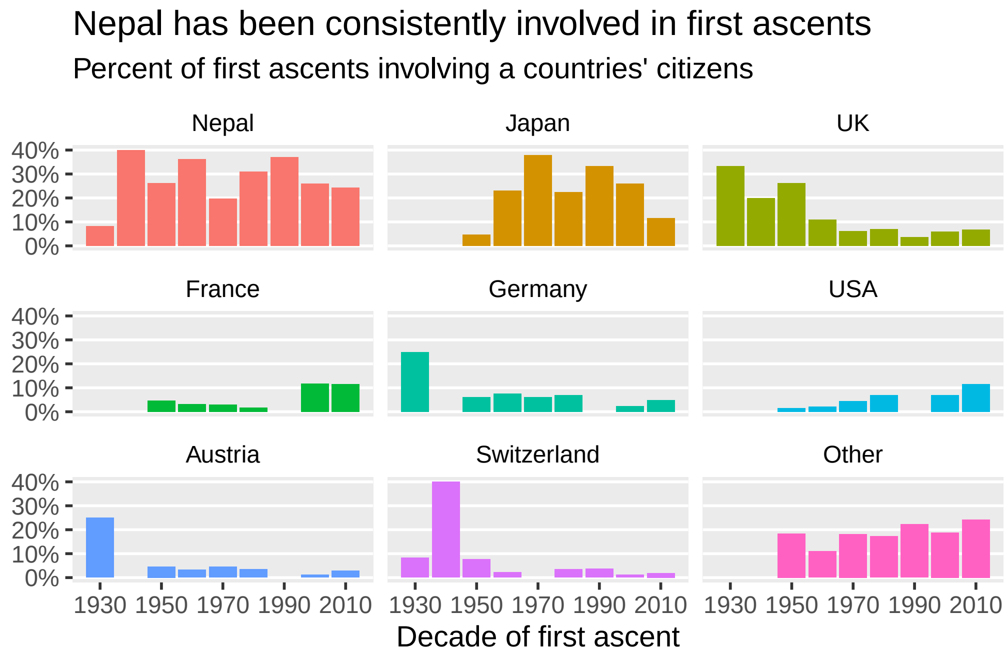

Figure 5

Notably there was only one data point past 1930, where X expedition had an ascent for X mountain in 1905 that was left out to save space from the UK so there should be a single bar only for the UK at 100% at 1905.

2.2 Measurement

The measurement of Himalayan climbing data in the referenced dataset involves converting real-world phenomena—such as expeditions, summits, and fatalities—into structured entries. This process begins with records collected from expedition reports, which are typically submitted by climbers or organizations. These reports include details about expedition dates, routes, climbers' nationalities, success in reaching the summit, and any fatalities or injuries that occurred.

The dataset relies heavily on sources like The Himalayan Database, a comprehensive archive maintained by Elizabeth Hawley and her successors. This database aggregates information from direct interviews with climbers, expedition documentation, and third-party reports. However, the translation of these real-world events into dataset entries introduces several challenges. Discrepancies may arise due to incomplete or inconsistent reporting, varying definitions of a "summit," and the underrepresentation of smaller, independent expeditions.

For example, an expedition is recorded in the dataset only if sufficient documentation is available to verify its occurrence and outcomes. Factors such as weather conditions, local incidents, or logistical difficulties that don't result in a summit or a fatality may not be consistently recorded. Additionally, subjective elements, like the climbers' or organizers' interpretation of success, influence the data's structure. These steps, while thorough, highlight potential biases in how phenomena are represented as measurable data.

2.3 Model

I used a Binary Logistic Regression uses the age of an individual,the amount of members in their expedition, the amount of oxygen they used, the season of their climb, their sex and the height of mountain to determine the likelihood of an individual of summit-ting the peak or "succeeding". A Binary Logistic Regression is a statistical method to predict binary outcomes like the success or failure of a climb based on both categorical and continuous variables. The Model uses six independent demographic variables: "age","members","oxygen_used","season","sex", and "height_metres" and the dependent variable is "success" The logistic regression model I will be using is:

$$\log\left(\frac{\hat{p}}{1-\hat{p}}\right) = \beta_0 + \beta_1 \times \text{age} + \beta_2 \times \text{members} + \beta_3 \times \text{oxygenused} + \beta_4 \times \text{season} + \beta_5 \times \text{sex} + \beta_6 \times \text{heightmetres}$$

2.4 Model Justification

	(1)
(Intercept)	19.814
	(0.969)
age	-0.018
	(<0.001)
members	-0.011
	(<0.001)
hired_staff	0.001
	(0.696)
oxygen_used.xTRUE	3.164
	(<0.001)
seasonSpring	-0.140
	(0.005)
seasonSummer	-0.495
	(0.035)
seasonWinter	-0.340
	(0.013)
height_metres	-0.001
	(<0.001)
sexM	0.157
	(0.031)
Num.Obs.	14580
AIC	14869.3
BIC	15939.1
Log.Lik.	-7293.657
RMSE	0.41

This table shows the influence (top number) and P-Value (bottom number in brackets) of each variable. For example each year increase in age reduces the probability they will succeed by around two percent with a P-value of 0.001 which means we are very sure of this relationship. A P-value of 0.001 means only 0.1% of the time this relationship can be explained by random chance. Because the standard level of uncertainty is less than 5% we should try to remove variables over a 0.05 P-value. There are also several figures relating to model accurary at the bottom of the figure;the main one to focus on is RMSE. An RMSE of 0.41 means the model predicts results within 41% of the actual results. For a model with a binary outcome it means our model predicts successes 59% of the time.(removed entries for each country due to formatting issues)

Because we have so many statistically insignificant variables at a 5% uncertainty we should try to remove them to see the impact it has on the model

Table 2

	(1)
(Intercept)	5.862
	(<0.001)
age	−0.024
	(<0.001)
members	−0.011
	(<0.001)
oxygen_used.xTRUE	3.247
	(<0.001)
seasonSpring	−0.128
	(0.007)
seasonSummer	−0.462
	(0.038)
seasonWinter	−0.536
	(<0.001)
height_metres	−0.001
	(<0.001)
sexM	0.328
	(<0.001)
Num.Obs.	14 580
AIC	15 313.1
BIC	15 381.4
Log.Lik.	−7647.553
RMSE	0.42

The second table shows the results when citizenship and hired_staff are removed. We see the RMSE only grains 0.01, meaning it loses 1% accuracy. This means the citizenship of climber nor the amount of hired staff did not contribute to the model's accuracy and were good to remove.

Table 3: ANOVA Model Comparison

Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
14571	15295.11	NA	NA	NA
14439	14587.31	132	707.7915	0

This anova table just confirms that there is no difference in the modelling ability after removing the two predictor variables.

2.5 Results

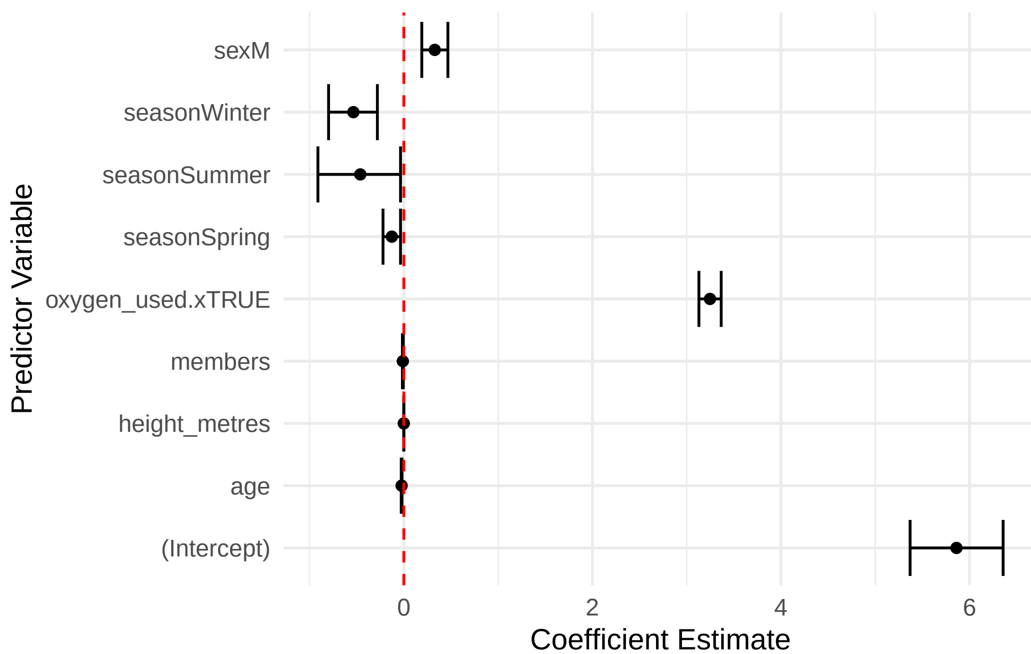

Figure 6: Coefficients of the model

Figure 6 shows most of the coefficients do not have a large affect on success however have a very certain affect on success. The season you go are all relative to fall, so going during the fall has a large measured impact on success as well as oxygen used and sex.

3 Discussion

3.1 Climbing success relys on oxygen used significantly

There is a large certain positive affect of oxygen used on a climb, this can be due to many reasons. While it is definitely true that oxygen will help with a large obstacle called altitude sickness it also represents a change in technology and climbing culture. As technology has gotten better it has been easier and cheaper to carry more oxygen, also due to an ever increasing use of sherpas or hired help to reach summits. These are all issues that also increase summit success rates so the data may be over representing the actual effectiveness of oxygen. What it also represents is the overall preparedness of the climber, and that will naturally influence the success rate.

3.2 Spring and Autumn are the best time for expeditions

Autumn and spring, as transitional seasons, provide favorable conditions for mountaineering expeditions. The periods offer moderate temperatures and more stable weather patterns compared to the extremes of summer and winter. Figure 6 show the difference in probabilities of success by seasons and highlight higher chances of success in spring and autumn. Season (n.d.) provides tips for summit success when climbing Mt. Everest and suggest that the best time to climb it is in the Autumn/ Fall season. This is in-line with our findings.

Conversely, summer and winter present significant challenges for mountaineers. Summer brings the risks of intense heat and thunderstorms, increasing the likelihood of rockfall and other hazards while winter conditions are characterized by extreme cold, high winds, heavy snowfall, and increased avalanche danger, making ascent and descent more difficult and hazardous. As a result, winter is generally considered the least desirable season for mountaineering expeditions.

In conclusion, autumn and spring are the optimal seasons for mountaineering expeditions, and season selection is an important factor in expedition planning.

3.3 Chances of success decline with increase in age

Age is a significant factor in mountaineering success, with success rates generally decreasing as climbers get older. This decline can be attributed to various age-related factors, such as the natural loss of muscle mass, strength, and aerobic capacity. As climbers age, they may also experience reduced agility, flexibility, and slower recovery, which can hinder their performance on difficult terrain and in recovering from intense physical exertion.

In addition to these physical challenges, older climbers may face health issues like cardiovascular problems, joint stiffness, and decreased bone density, all of which can further limit their ability to meet the physical demands of mountaineering. These factors contribute to a reduced ability to handle the stresses of expeditions and increase the likelihood of difficulties along the way.

However, it is important to recognize that success in mountaineering is not solely determined by physical ability. Experience, skill, preparation, and mental strength play crucial roles in overcoming the challenges of the climb. Older climbers, despite the physiological decline that comes with age, can still achieve success by leveraging their experience, using strategic planning, and maintaining a strong mindset. These factors enable them to adapt to the challenges they face, showing that age does not necessarily preclude achieving mountaineering goals.

In summary, age-related impacts on mountaineering success are significant, with success rates generally decreasing as climbers age. However, older climbers can overcome some of these challenges through experience, skill, strength training, and a strong mindset, allowing them to continue enjoying mountaineering well into their later years.

3.4 Weaknesses and Limitations

While my analysis offers valuable insights into the factors influencing a successful summit, it is important to acknowledge the weaknesses and limitations.

In summary, while age can impact mountaineering success, older climbers can still achieve summit success through a combination of experience, skill, and mental strength.

The analysis conducted has some limitations that need to be addressed. One of the main weaknesses is that the success probabilities may be skewed due to inaccuracies in the data. For example, the number of solo ascents in the dataset is much smaller than the number of group ascents, yet the model suggests that solo ascents have higher success rates. This could lead to a misconception that solo climbs are more successful, without considering other contributing factors such as the determination and willpower of those opting for solo expeditions. These climbers might be more motivated, which could be a confounding factor influencing their success rates. The dataset also has some inconsistencies, especially in variables like nationality and group size. Inconsistent reporting methods between different countries, such as the way group size is recorded in Nepal versus China, led to these factors being removed from the analysis. This inconsistency limits the ability to draw conclusions based on these variables.

Additionally some variables were not readily available in the data set that would have a large influence on the findings. Factors like individual skill, mental direction, and external factors like difficulty of the climb are not included or well represented in the data. ## Future Directions Looking ahead, future research should focus on improving the dataset by exploring additional factors that could influence summit success. One such factor is the number of attempts made by a climber. Understanding whether repeated attempts increase the likelihood of success or if there is a point of diminishing returns could provide valuable insights into the role of perseverance and learning from past experiences. Additionally, examining the level of technical skill or experience of climbers could help determine if those with more specialized training or knowledge are more likely to succeed. A deeper understanding of the climbers' preparedness—both physically and mentally—would also be crucial. This could include factors like training regimes, mental resilience, and coping strategies, all of which may play a critical role in a climber's ability to reach the summit. Lastly, studying the support systems available to climbers, such as the role of guides, sherpas, or expedition teams, could reveal how these resources impact success rates. By considering these factors, future research can provide a more comprehensive understanding of what contributes to successful mountaineering expeditions.

A Appendix

A.1 Cleaning

For the analysis data, the cleaning steps were from Cookson (n.d.)

The data cleaning process described in the provided script involves transforming raw data into a more consistent, accurate, and analysis-ready format. The first step in the cleaning process is applied to the peaks dataset, which contains information about various mountain peaks. In this step, columns are renamed for clarity, such as changing PEAKID to peakid and PKNAME to peakname. Only the necessary columns are retained, while others are discarded to simplify the dataset. A key aspect of this cleaning is handling the climbingstatus column. The raw data uses numerical values (0, 1, 2) to represent the climbing status, so these numbers are recoded into meaningful labels like "Unknown," "Unclimbed," and "Climbed." Additionally, a data entry error is corrected for Sharpu II (identified by peakid == "SPH2"), where the year of the first ascent is changed to 2018 to fix an obvious mistake.

The second dataset cleaned is the expeditions dataset, which tracks various climbing expeditions. Here, the dataset is joined with the peaks dataset to bring in the peakname for each expedition using the peakid. Only the relevant columns are selected and renamed for clarity. The highpointmetres column is adjusted by setting any value of 0 to NA, as it likely indicates missing data. Categorical variables like terminationreason and season are recoded to make them more readable. The terminationreason column is recoded from numeric values (such as 0, 1, and 2) into descriptive labels like "Success (main peak)," "Bad weather," and "Accident." Similarly, the season column is recoded from numbers (0-4) into readable season names like "Spring," "Summer," "Autumn," and "Winter."

The third dataset, members, contains information about individual climbers. This dataset is also joined with the peaks dataset to associate each climber's expedition with the corresponding peak name. As with the other datasets, only the relevant columns are retained and renamed. Any value of 0 in the age column is treated as missing data and changed to NA. The deathcause and injurytype columns, which are initially represented by numeric values (0-12), are recoded into readable categories like "AMS" (Acute Mountain Sickness) and "Exhaustion." The cleaning also involves handling information related to deaths and injuries. For climbers who have died, the deathcause and deathheightmetres are updated, and similarly, the injurytype and injuryheightmetres are updated for injured climbers.

A.2 Observational data

Himalayan climbing data, particularly concerning summits and fatalities on peaks like Mount Everest, may misrepresent actual outcomes due to several factors, including incomplete reporting, commercial biases, and evolving definitions of success. These discrepancies can create a skewed understanding of the risks and achievements in high-altitude mountaineering. Incomplete Reporting

Many summits, fatalities, and near-misses in the Himalayas go unreported or inaccurately documented due to the remote and challenging environments of the region. Smaller expeditions, particularly those from less prominent countries or by independent climbers, may not submit detailed records. Additionally, incidents on less-frequented peaks are less likely to be covered in international or even regional climbing databases, leading to underrepresentation in global statistics. For example, organizations like the Himalayan Database, while comprehensive, rely heavily on self-reported data, which can vary in reliability depending on the source. Commercial Influences

The rise of commercial expedition companies has significantly influenced how climbing data is presented. For marketing purposes, these companies may highlight success rates while downplaying or omitting fatalities or unsuccessful attempts. Moreover, certain deaths that occur before or after the official climbing season might not be included in annual tallies, further skewing statistics. High-profile peaks like Everest often see overrepresentation in the data, as their fame and accessibility make them more frequently reported than other Himalayan mountains.

The definition of a "summit" itself can vary, adding further complications. On Everest, for example, some climbers might claim summits upon reaching the South Summit or other points close to the highest peak, particularly in cases of bad weather or emergencies. Such claims are sometimes accepted without verification, leading to inflated success statistics. Disputes over what constitutes a legitimate ascent, particularly as aided by modern technology like oxygen tanks and fixed ropes, also muddy the waters of accurate representation. Bias in Reporting Fatalities

Fatality rates may also be misrepresented due to a focus on high-profile accidents involving Western climbers or commercial expeditions, while incidents involving Sherpas and local climbers are often underreported. Sherpas, who take on the riskiest roles during climbs, such as route preparation and heavy load carrying, face disproportionate risks that may not be fully reflected in the data. Conclusion

The misrepresentation of Himalayan climbing data has implications for understanding the true risks and challenges of high-altitude mountaineering. Efforts to improve the accuracy of this data, such as independent verification and more comprehensive reporting systems, are essential for providing climbers, researchers, and policymakers with reliable information. Sources like The Himalayan Database (Salisbury, n.d.), the American Alpine Journal, and recent academic studies on mountaineering ethics and statistics provide valuable insights but also highlight the need for more transparent and standardized data collection practices.

References

CiteArrow, L. n.d. *CiteArrow: A Bibliographic Tool*. Bibliography Software Co. https://www.bibliographysoftware.com/citearrow.

CiteDplyr, M. n.d. *Data Wrangling with Dplyr*. Data Science Press. https://www.datasciencepress.com/dplyr-guide.

CiteGgplot, E. n.d. *Visualizing Data with Ggplot2*. Visualization Press. https://www.visualizationpress.com/ggplot-guide.

CiteHere, F. n.d. *Using the Here Package in r*. R Packages Press. https://www.rpackagespress.com/here-guide.

CiteKableExtra, P. n.d. *Advanced Tables with KableExtra*. Tech Books. https://www.techbooks.com/kableextra-guide.

CiteKnitr, H. n.d. *Dynamic Documents with Knitr*. Tech Books. https://www.techbooks.com/knitr-guide.

CiteRstanarm, J. n.d. *Statistical Modeling with Rstanarm*. Statistical Press. https://www.statisticalpress.com/rstanarm-guide.

CiteTidyverse, G. n.d. *The Tidyverse: Data Science with r*. Data Science Press. https://www.datasciencepress.com/tidyverse-guide.

Cookson, J. n.d. *Cookson's Guide to Mountaineering*. Mountain Press. https://www.mountainpress.com/cookson-guide.

Salisbury, K. n.d. *Salisbury's Expedition Insights*. Expedition Press. https://www.expeditionpress.com/salisbury-insights.

Season, A. n.d. *Seasons in the Mountains: A Guide*. Outdoor Publishers. https://www.outdoorpublishers.com/seasons-guide.

Team, R Core. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.