**Predicting Climbing Success on Himalayan Expeditions**

This project uses historic data of expeditions in the Himalayas from 1905 to 2018 to predict the success of individual climbers reaching the summit.

Go to Report: [PDF](https://github.com/jacqpool/Springboard/blob/master/expedition_success_himalayas/Report_Himalayan_Summit_Success.pdf)   
and presentation: [PDF](https://github.com/jacqpool/Springboard/blob/master/expedition_success_himalayas/Presentation_Himalayan_Summit_Success.pdf)

**Directory Content**

* **Database guide**: The database guide describing the data and how to download it, is here:
  + [Himalayan\_Database\_Guide.pdf](https://github.com/jacqpool/Springboard/blob/master/expedition_success_himalayas/Himalayan_Database_Guide.pdf)
* **Input data**: All the data is stored in this directory. The datasets used are:
  + [peaks.csv](https://github.com/jacqpool/Springboard/blob/master/expedition_success_himalayas/peaks.csv)
  + [expeditions.csv](https://github.com/jacqpool/Springboard/blob/master/expedition_success_himalayas/expeditions.csv)
  + [members.csv](https://github.com/jacqpool/Springboard/blob/master/expedition_success_himalayas/members.csv)
* **Cleaned data**: The data was cleaned and merged, then pre-processed for machine learning.
  + [DF\_Himalayas\_Expeditions.csv](https://github.com/jacqpool/Springboard/blob/master/expedition_success_himalayas/DF_Himalayas_Expeditions.csv)
  + [DF\_Himalayas\_Expeditions\_MLready.csv](https://github.com/jacqpool/Springboard/blob/master/expedition_success_himalayas/DF_Himalayas_Expeditions_MLready.csv)
* **Notebooks**: The notebooks for the data wrangling, exploration and model generating needs running in the following sequence:
  + [Himalayan\_expeditions\_cleaning.ipyn](https://github.com/jacqpool/Springboard/blob/master/expedition_success_himalayas/Himalayan_expeditions_cleaning.ipynb)
  + [Himalayan\_expeditions\_story.ipynb](https://github.com/jacqpool/Springboard/blob/master/expedition_success_himalayas/Himalayan_expeditions_story.ipynb)
  + [Himalayan\_expeditions\_stats.ipynb](https://github.com/jacqpool/Springboard/blob/master/expedition_success_himalayas/Himalayan_expeditions_stats.ipynb)
  + [Himalayan\_expeditions\_ML\_cleaning.ipynb](https://github.com/jacqpool/Springboard/blob/master/expedition_success_himalayas/Himalayan_expeditions_ML_cleaning.ipynb)
  + [Himalayan\_expeditions\_ML.ipynb](https://github.com/jacqpool/Springboard/blob/master/expedition_success_himalayas/Himalayan_expeditions_ML.ipynb)

**Project Proposal**

**What is the goal?**

What makes some members in mountain trekking expeditions successfully reach the summit and others not? Assuming equal ability and determination between climbers, it is interesting to determine what external factors contribute to summiting success. With this insight an analytical model can predict the outcome an individual member’s attempt to summit a specific peak in the Himalayas.

**Who cares?**

According to the website [The Himalayan Database ©](http://himalayandatabase.com/index.html), “The records in the Himalayan Database will be of considerable significance to climbers planning expeditions, to journalists and mountaineering historians needing ready access to historical records, and to medical researchers elucidating patterns of accidents, fatalities, and supplemental oxygen use.”

**What data are you going to use?**

The Himalayan Database is a non-profit organisation that kindly allowed the use of their data for this project. They have continued the work of Elizabeth Hawley, a journalist living in Kathmandu, collecting information on expeditions in the Himalayas from 1905 to 2018. The data is downloaded in CSV format from an application they have developed.

**What is your approach?**

To have workable data the three CSV files, peaks, expeditions and members are cleaned, merged and wrangled. The new dataset DF\_Himalayas\_Expeditions.csv is then used to visualise some features to gain insight. Once it is clear how the data is structured some statistical inferences can be applied to gain further understanding of how some important features relate.

With a clearer grasp of the data it is cleaned again and made ready for machine learning analysis.

**The Results**

The results are explained in the report. There are some significant features identified that contribute to summit success, such as a member’s age, the number of members in the expedition and oxygen use, especially in the climb.

**Found any issues?**

Any feedback or criticism is welcome. Please email me, [jacquespoolman@gmail.com](mailto:jacquespoolman@gmail.com), or find me on [LinkedIn](https://www.linkedin.com/in/jacques-poolman-a895331b).