Data Memo

2024-01-22

Dataset Overview

My dataset includes information about rock climbing routes from around the world. This includes categories such as rating, grade, location, and quality votes. I aquired the data from Kaggle, but the original dataset is from REI's Mountain Project. The data consists of 116,700 observations, but I might exclude observations that have a small amount of rating votes for the sake of accuracy. There are 8 predictors. I will be working with both numerical and categorical variables. There are about 15,200 observations with missing data for the lengths of the climb, but I will most likely exclude these observations. There are also some columns that contain information about the individual routes, but I will exclude these since they are not useful.

Link: https://www.kaggle.com/datasets/pdegner/mountain-project-rotues-and-forums?select=mp_routes.csv

Source: REI Mountain Project

Research Questions

I am interested in predicting the ratings of routes based on their characteristics. I want to find out if harder climbs will be rated higher. My response variable is avg stars, which is the average user rating of the route. To answer my question, I will be using regression since the rating is numerical. I think that the most useful predictor will be rating, but I think that pitches and length will also prove useful. This model will be inferential since I want to test the relationship between the quality of the route and the other predictor variables.

Project Timeline

Week 3: Load data and introduction

Week 4: Exploratory data analysis

Weeks 5-6: Fit models

Week 7: Model selection

Weeks 8-9: Write-up and edits

Week 10: Conclusion and finalize project