What makes hiking trails popular?

Predicting trail popularity using key trail characteristics

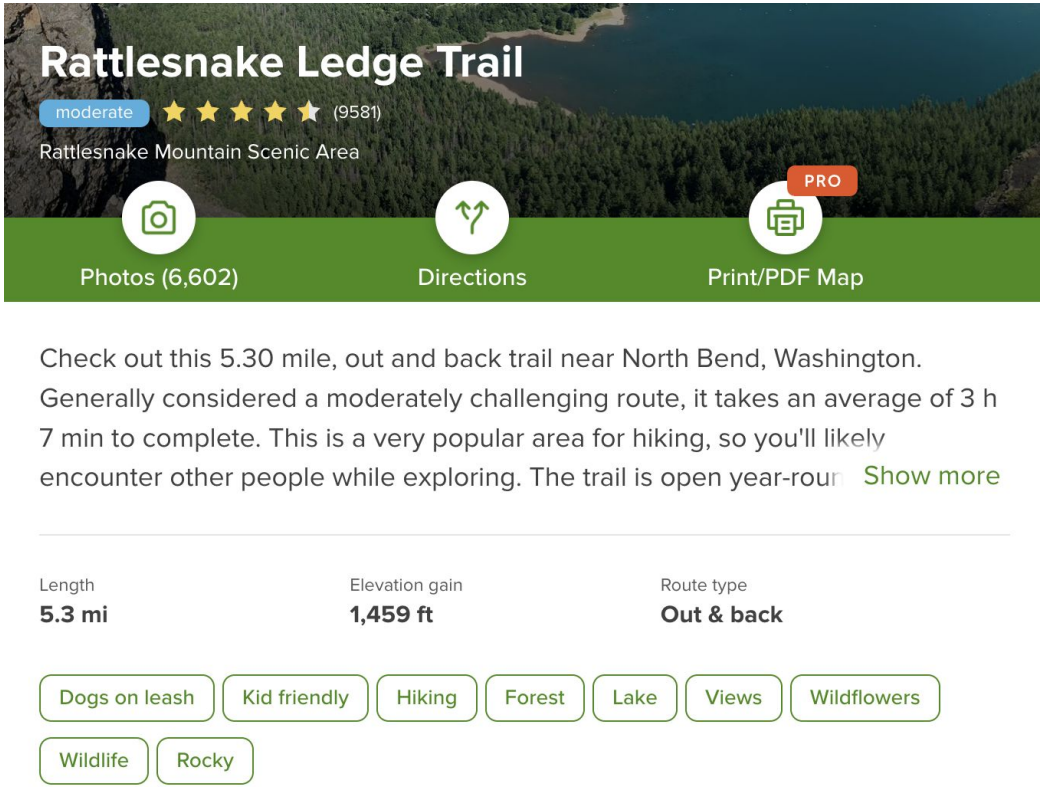
Project Goals

- Create a linear regression model to predict trail popularity
- Better understand what trail characteristics are associated with popular trails
- Use the model to identify undiscovered trails

Data

The **data** for this project includes info on 999 trails in Washington State, scraped from the AllTrails website (https://www.alltrails.com/us/washington)

Target: Number of hikers that have completed the trail



Design

Feature Engineering

- Create dummy variables based off of tags given to the trails
- Log transform target
- Standardize feature values using sklearn's StandardScaler

Model Building & Validation

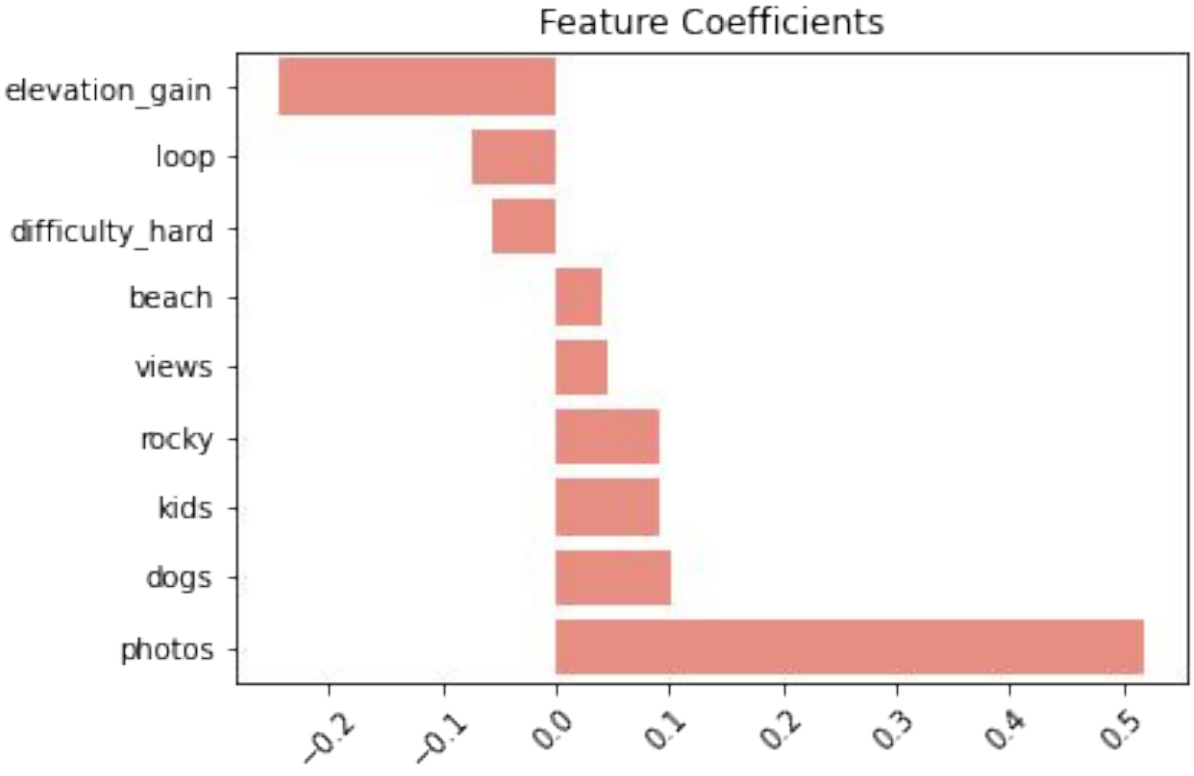
- The dataset was split into 60/20/20 training/validation/test portions. Models were fit using the training data, then tested on the validation data. The 20% test data was used only at the very end to evaluate the performance of the finalized model.
- Elastic net regression ultimately used, as it had the strongest cross-validation performance

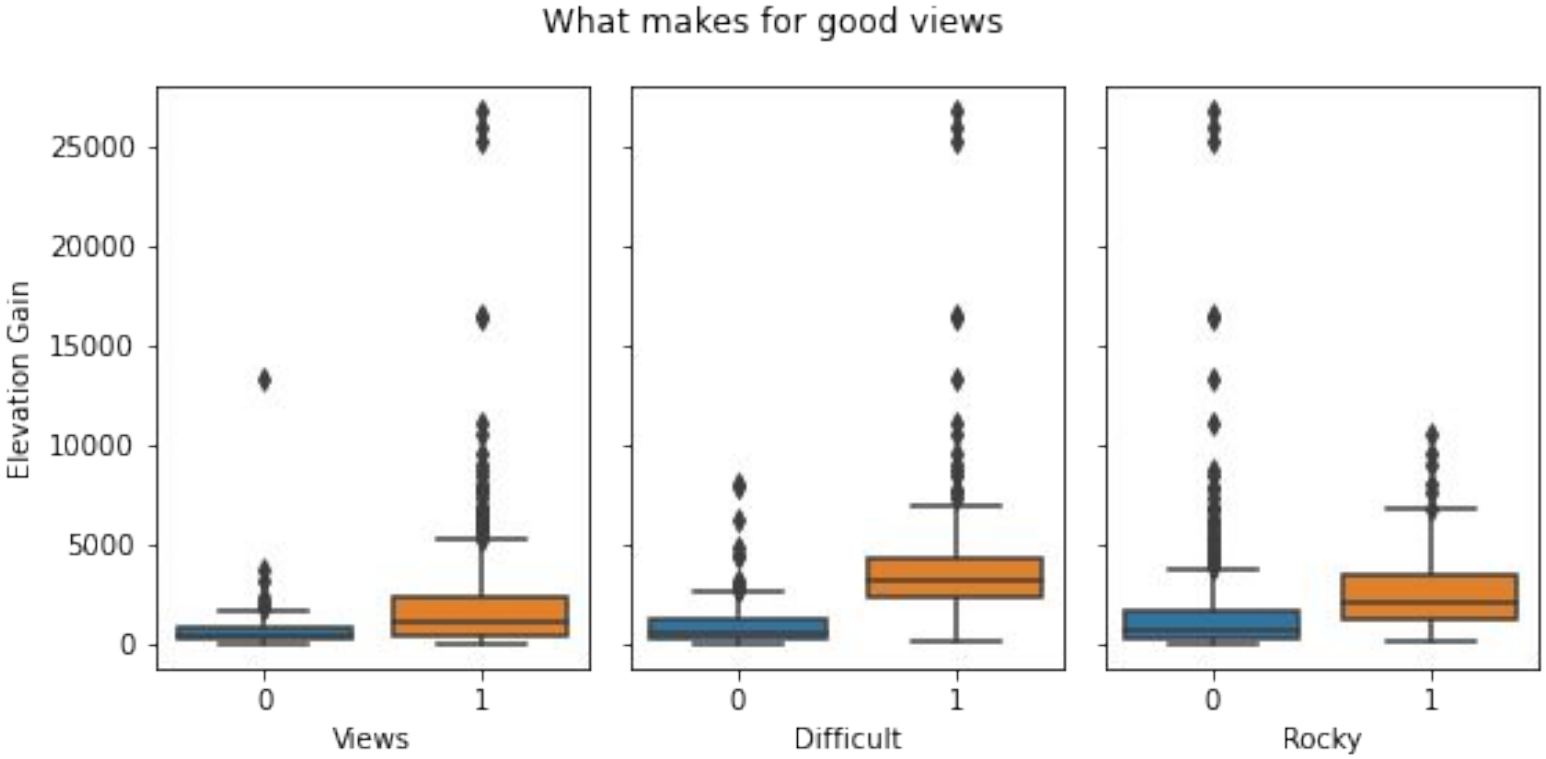
Tools

- BeautifulSoup and Selenium for scraping data from the web
- Numpy and Pandas for data manipulation
- Scikit-learn for modeling
- Matplotlib and Seaborn for plotting

Features

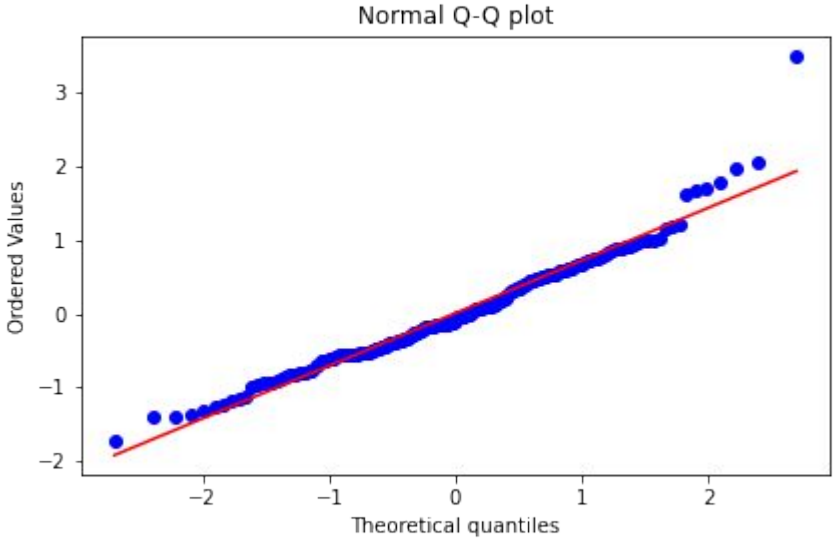
Distance (miles)
Elevation Gain (feet)
Loop
Beach
Views
Rocky
Kid friendly
Dog friendly
Photos (Number of photos)
Fee
Wildlife
Lake
Wildflowers
Moderate Difficulty
Hard Difficulty

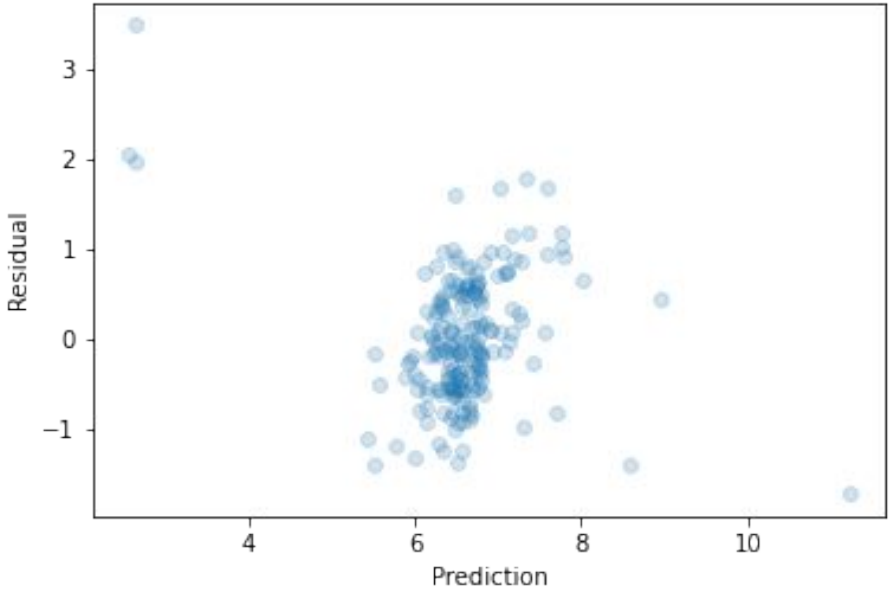



What makes for good views
Elevation Gain
25000
20000
15000
10000
5000
0
0
1
Views
0
1
Difficult
0
1
Rocky

Using the model for prediction

The final model had an adjusted R^2 value of 0.42 when applied to the test data, with a mean absolute error of 928 (hikers)





Trail Name	Number of Hikers	Predicted Number of Hikers
Wright Mountain	1324.0	5347.0
Muir Snowfield Ski Tour	957.0	2191.0
Island Lake Trail	562.0	1488.0
Kalaloch 4th Beach Trail	205.0	721.0
Observation Peak Trail	169.0	664.0
Mirror Lake via PCT Trail	315.0	778.0
Nooksack Falls Trail	338.0	788.0
Twin Sisters Rock	344.0	782.0
Barlow Point Trail	283.0	713.0
Old Sauk River Trail	494.0	917.0

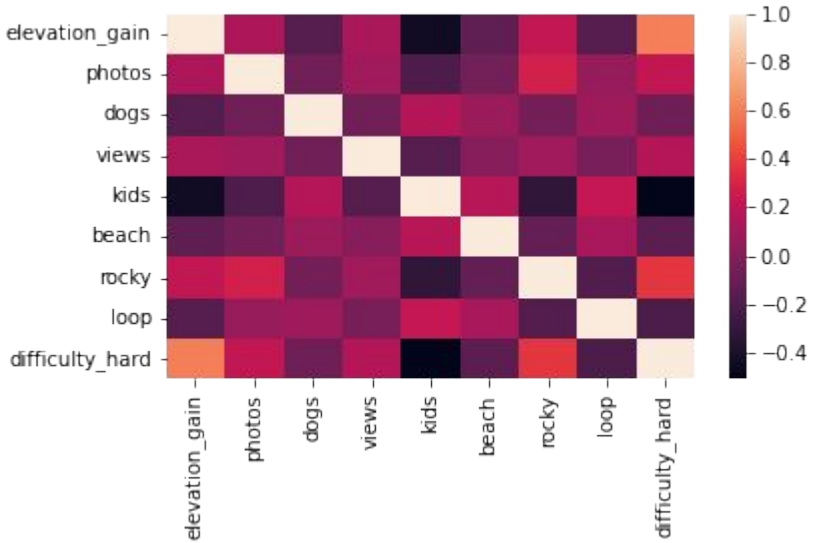
Takeaways & Future Opportunities

- Popular trails have photos (photos provide information & inspiration)
- Making trails kid and dog friendly can help improve the number of users
- Apply the model to data on additional trails to identify undiscovered trails

Appendix

Variable Inflation Factor for Model Features

VIFs were below 5 for all features, suggesting that we would not expect multicollinearity issues with this model



variables	vif
elevation_gain	1.647150
photos	1.137190
dogs	1.052315
views	1.045722
kids	1.518641
beach	1.050342
rocky	1.262083
loop	1.108275
difficulty_hard	1.896036