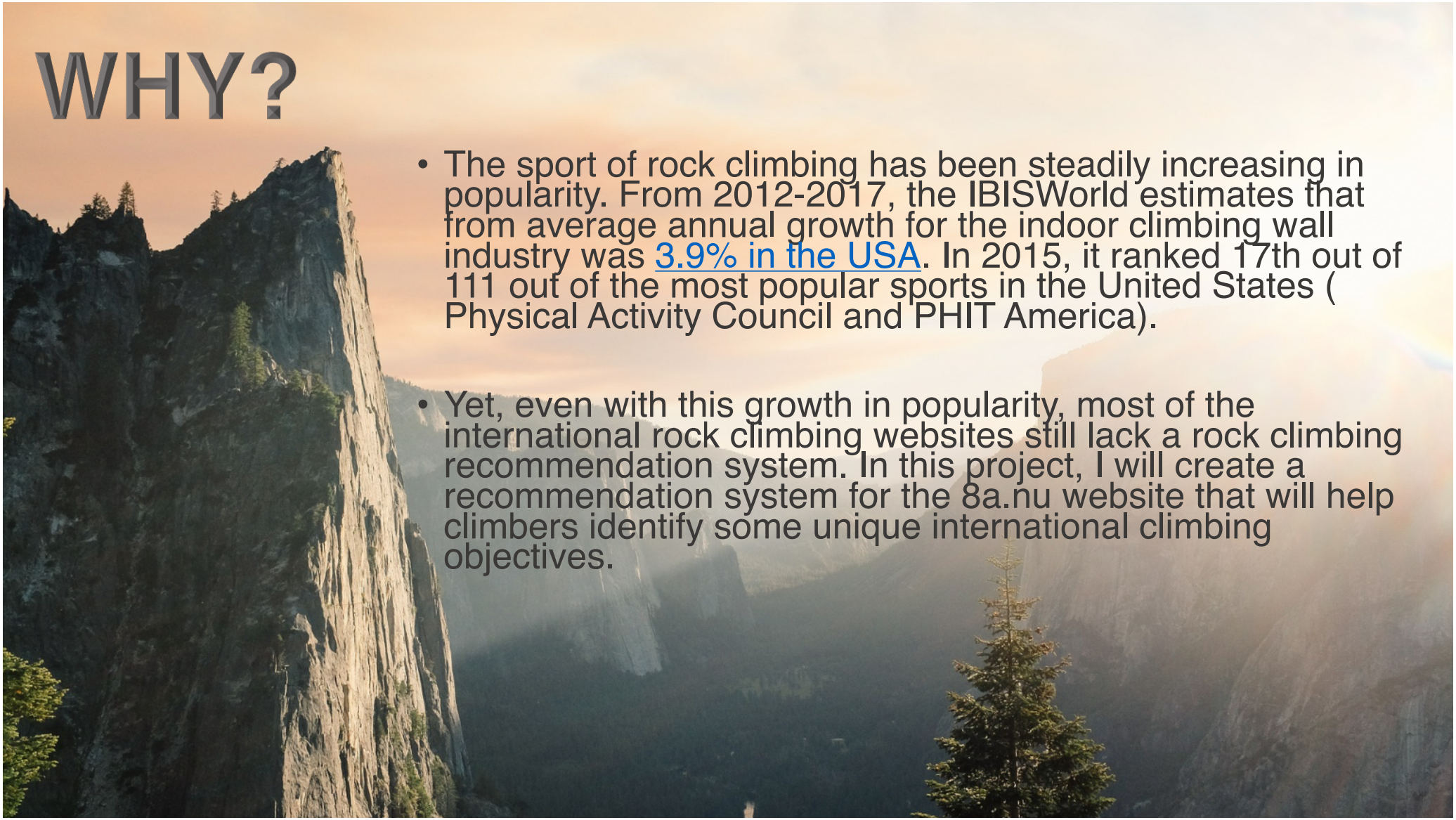

WHERE IN THE WORLD
SHOULD
YOU CLIMB NEXT?
INTERNATIONAL ROCK CLIMBING CLIMBING
RECOMMENDATION SYSTEM
An exercise in user-based collaborative filtering
Kristen N. Colley

WHY?



- The sport of rock climbing has been steadily increasing in popularity. From 2012-2017, the IBISWorld estimates that from average annual growth for the indoor climbing wall industry was 3.9% in the USA. In 2015, it ranked 17th out of 111 out of the most popular sports in the United States (Physical Activity Council and PHIT America).

- Yet, even with this growth in popularity, most of the international rock climbing websites still lack a rock climbing recommendation system. In this project, I will create a recommendation system for the 8a.nu website that will help climbers identify some unique international climbing objectives.

DATA

8a.nu is one the world's largest international rock climbing websites. With over 4 million entries of climbs and ratings, this Kaggle webscraping project is a sufficient size to develop a good predictor model. To view the 8a.nu website, the original Kaggle four SQLite tables created by David Cohen, or the import report using the Kaggle API click on the links below:

- 8a.nu website
- Kaggle Dataset
- Data Import Report

METHOD

There are three main types of recommenders used in practice:

1. **Content-based filter:** Recommending future items to the user that have similar innate features with previously "liked" items.
2. **Collaborative-based filter:** Recommending products based on a similar user that has already rated the product.
3. **Hybrid Method:** Leverages both content & collaborative based filtering.

WINNER: User-based collaborative-filtering

	item_1	item_2	item_3	item_4
user_1				★★
user_2	★			
user_3		★★		
user_4	★★★			★★★

DATA CLEANING

[Full Data Cleaning Report]

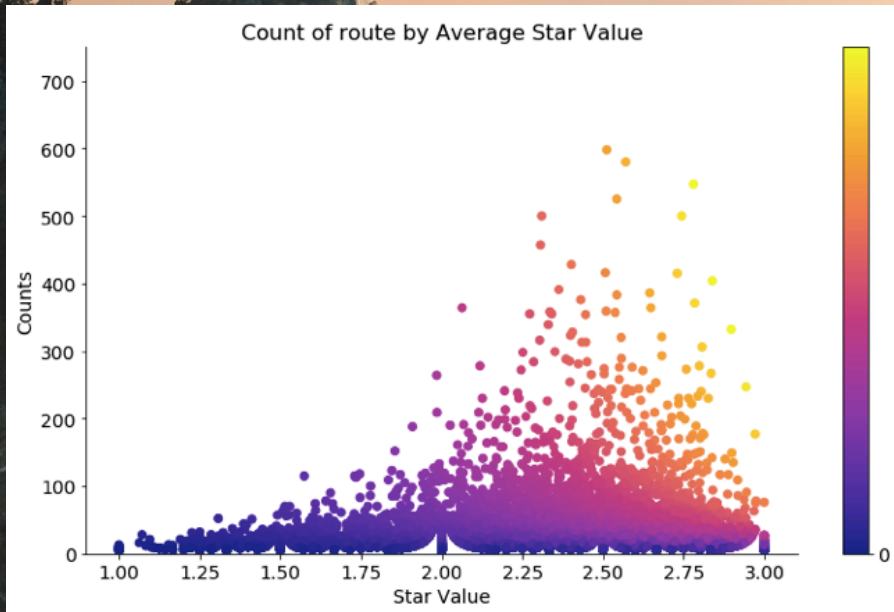
Larger Issues Encountered:

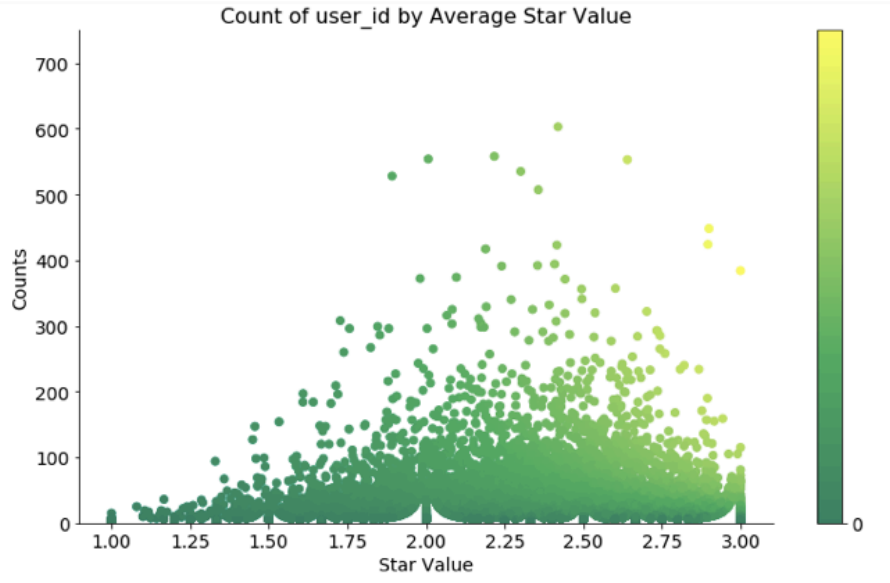
- **Problem 1:**This dataset is all user-entered information and subject to invalid information
  - ***Solution:*** *groupby three reference columns and impute the mode of the column to increase the accuracy of the dataset*
- **Problem 2**: Being this is an international rock climbing website, the names of the rock climbing routes were differing based on if the user enters accent marks or not.
  - ***Solution****: normalize all names to the ascii standards.*
- **Problem 3:** Spelling issues with the route names. For example: if there was a route named "red rocks canyon" it could be spelled "red rock", "red rocks", "red canyon" etc.
  - ***Solution****: I tried phonetic spelling algorithms (soundex & double metaphone). However, my final solution was to create an accurate filter for route names. The logic being that if up to x number of users all entered that *exact same* route name, the chances were good that it was an actual route spelled correctly.*

A	B	C
USER_ID	ITEM_ID	RATING
1	100	***
2	101	**
3	102	*

EDA

[Full EDA Report]





ALGORITHMS

Machine Learning Report

I tested four different filtered datasets on the 11 surprise library recommendation algorithms:

WINNER:
SVD++ Algorithm

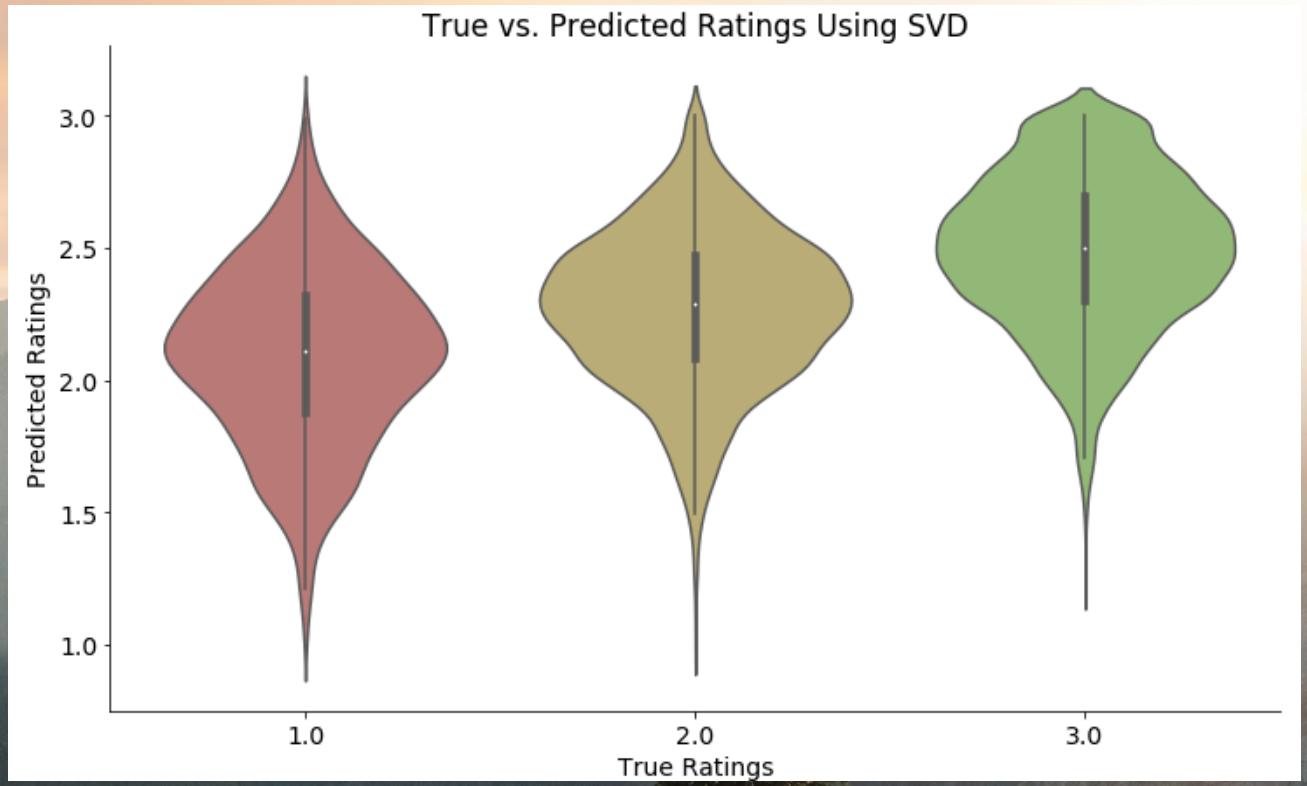
$$\hat{r}_{ui} = \mu + b_u + b_i + q_i^T \left(p_u + |I_u|^{-\frac{1}{2}} \sum_{j \in I_u} y_j \right)$$

This algorithm is an improved version of the SVD algorithm that Simon Funk popularized in the million dollar Netflix competition that also takes into account implicit ratings (yj).

Algorithm	test_rmse	fit_time	test_time
SVDpp	0.656549	16.115670	0.648576
BaselineOnly	0.660056	0.066147	0.135528
SVD	0.661268	2.013415	0.109702
KNNBaseline	0.697995	0.582456	0.640773
CoClustering	0.705715	1.005888	0.139992
KNNWithMeans	0.707114	0.610985	0.545816
KNNWithZScore	0.707959	0.649346	0.553792
NMF	0.730429	2.461001	0.122456
SlopeOne	0.733207	0.774625	0.367432
KNNBasic	0.743560	0.553092	0.482563
NormalPredictor	0.939817	0.043034	0.129717

ACCURACY

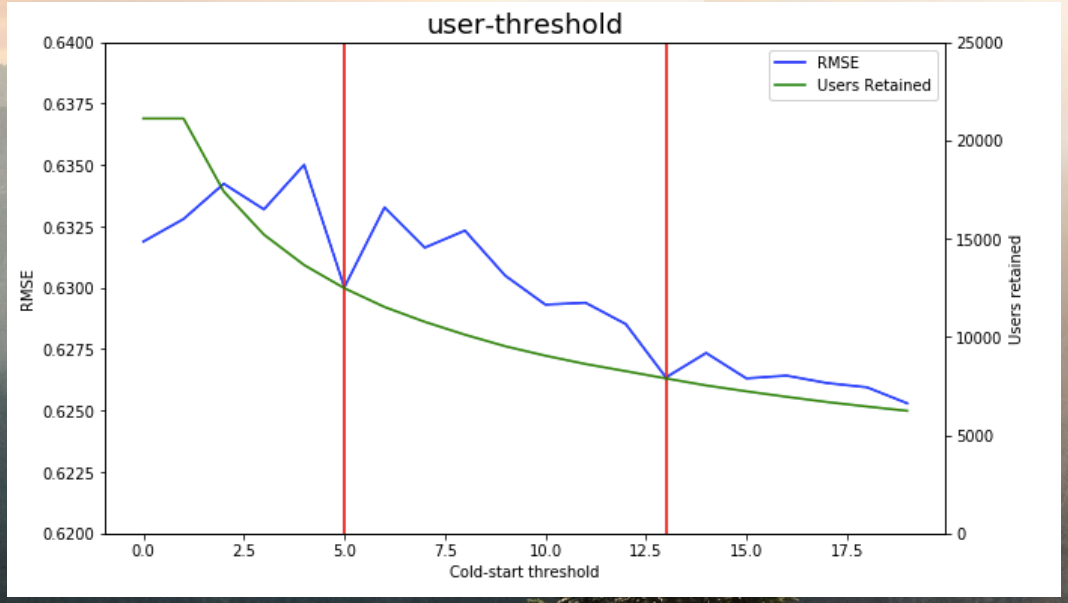
Machine Learning Report



Cold-Start Threshold

Machine Learning Report

- **What is it?** When only using collaborative based filtering there is a problem to consider: *what to recommend to new users with very little or no prior data?*
- Remember, to fix the spelling issue we already set the route cold-start threshold by choosing the dataset that filtered out any route occurring less than 6 times.



PREDICTIONS

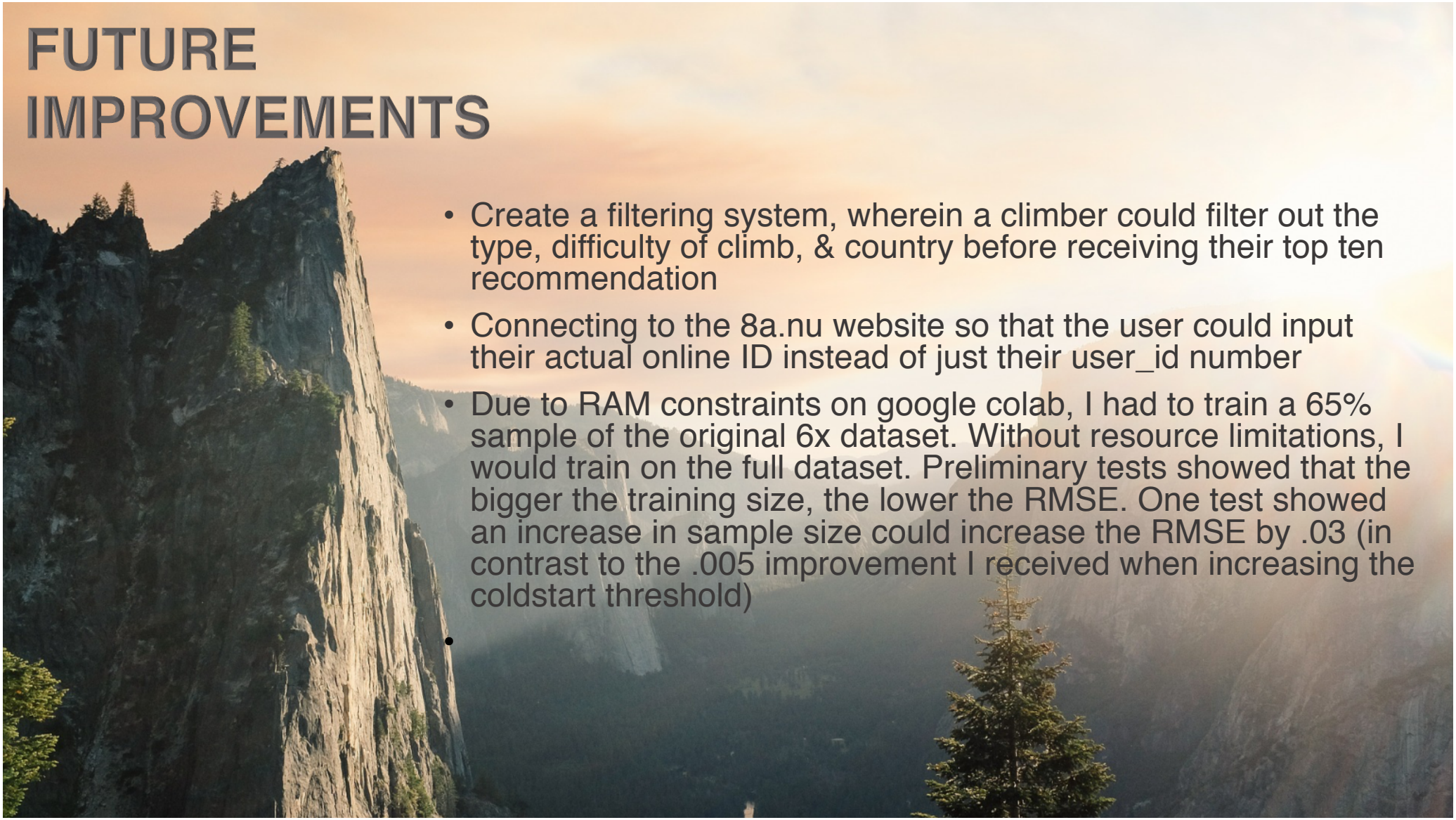
Predictions Notebook

Where in the world should you climb next?
Top Ten Rock Climbing Recommendations for User #12:

	route	user_rate_predict	avg_rating	num_users_rate	climb_type	usa_routes	usa_boulders	crag	country
1	Abstrakt	2.965450796247701	2.931	29	Rope Climb	5.13b	V11	Hylteberget	SWE
2	Hegar	2.809462278617374	2.778	27	Rope Climb	5.10d	V4/V5	Nissedal	NOR
3	Thaiboxing	2.767406704232296	2.811	37	Rope Climb	5.12b	V8	Jarlsberget	NOR
4	Vestpillaren	2.7656782513690805	2.919	86	Rope Climb	5.10c	V4	Lofoten	NOR
5	Quimera	2.745291868794412	2.900	20	Rope Climb	5.11a	V5	Pego	ESP
6	Trampoline	2.7450755717280315	2.000	1	Bouldering	5.12a	V7	Smuggs	USA
7	Magnetfinger	2.7443795704718594	2.939	49	Rope Climb	5.13a	V10	Pfalz	DEU
8	Villskudd	2.7339687745935484	2.820	128	Rope Climb	5.10a	V3	Bohuslan	SWE
9	Prismaster	2.731965630823839	2.649	114	Rope Climb	5.10a	V3	Bohuslan	SWE
10	Kachoong	2.724216995120893	2.850	40	Rope Climb	5.10d	V4/V5	Arapiles	AUS

In the final predictions notebook, the user can enter their user_id number and receive a list of top ten routes recommended to them

FUTURE IMPROVEMENTS



- Create a filtering system, wherein a climber could filter out the type, difficulty of climb, & country before receiving their top ten recommendation
- Connecting to the 8a.nu website so that the user could input their actual online ID instead of just their user_id number
- Due to RAM constraints on google colab, I had to train a 65% sample of the original 6x dataset. Without resource limitations, I would train on the full dataset. Preliminary tests showed that the bigger the training size, the lower the RMSE. One test showed an increase in sample size could increase the RMSE by .03 (in contrast to the .005 improvement I received when increasing the coldstart threshold)
-


TIME TO GET OUTSIDE!
Kristen N. Colley