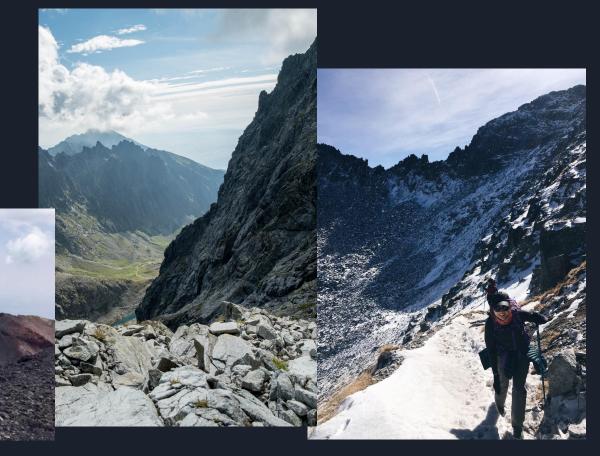# SummitPost Mountain Recommender

Angela Huang

Background

- Objective:
  Create recommender based on type of adventure

Data & Methodology

- SummitPost.org
    - Climber's logs (67,500) for top mountains (860)
        - Annapurna
        - Olympus Mons
- Pre-processing
    - Lowercased, removed line breaks, numbers & punctuation
    - Bi-grams & tri-grams
    - Extended stop word list
    - Lemmatized
- MongoDB
- Topic Modeling
    - LSA, NMF, LDA
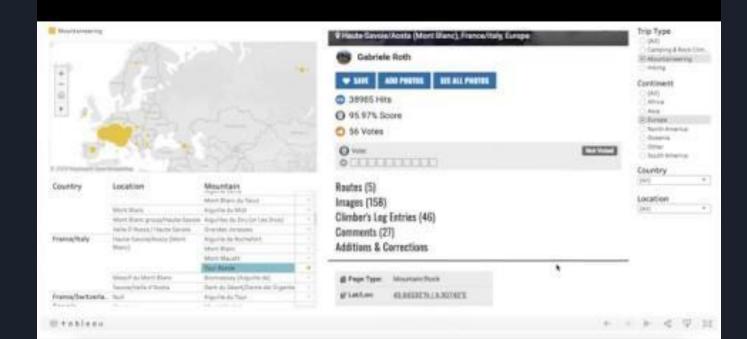
Results

Latent Dirichlet Allocation Topics

- **Hiking**: drive, road, walk, highpoint, tram, mile, hiking
- **Camping & Rock Climbing**: lake, camp, traverse, scramble, class, pitch, exposure, approach
- **Mountaineering**: glacier, hut, condition, ice, normal_route, crevasse, refuge, guide
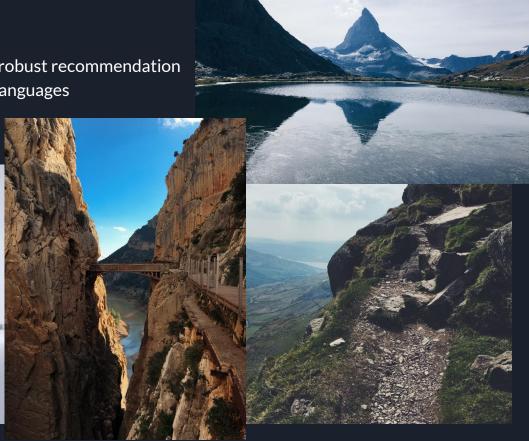
Tableau Dashboard Demo

Future Work

- Add lat/long to Tableau
- Secondary data source for more robust recommendation
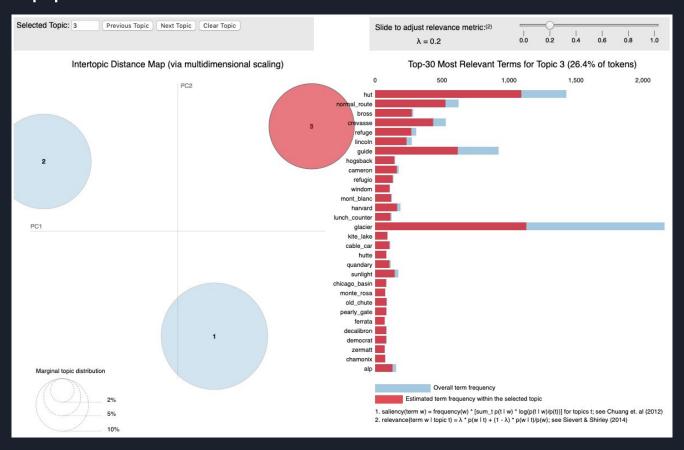- Spelling corrections & translate languages
- More mountains!

Thank You!

Appendix - LDA Visualization

Appendix - Tools Used

- AWS
- Pandas
- Numpy
- BeautifulSoup
- WordCloud
- NLTK
- spaCy
- Gensim
- pyLDAvis
- Tableau
- MongoDB