MTB Trail Recommender

Ryan McDonald – Data Scientist

RYAN MCDATA .COM

Agenda

- Problem Statement
- Methodologies / EDA
- Recommenders
- Streamlit WebApp
- Conclusions & Next Steps

Problem Statement

Choosing which trail to ride can be daunting!

This project developed two tools for riders:

- Content-based trail recommender system for Arizona and Utah MTB trails.
- Collaborative user-based binary recommender system for finding similar trail users

"Which trail should I ride?"

"That trail was great! What's next"?

"Am I going to hurt myself on this trail?"

"Is this trail going to offer the training I need?"

Methodologies

Data Retrieval with OctoParse API

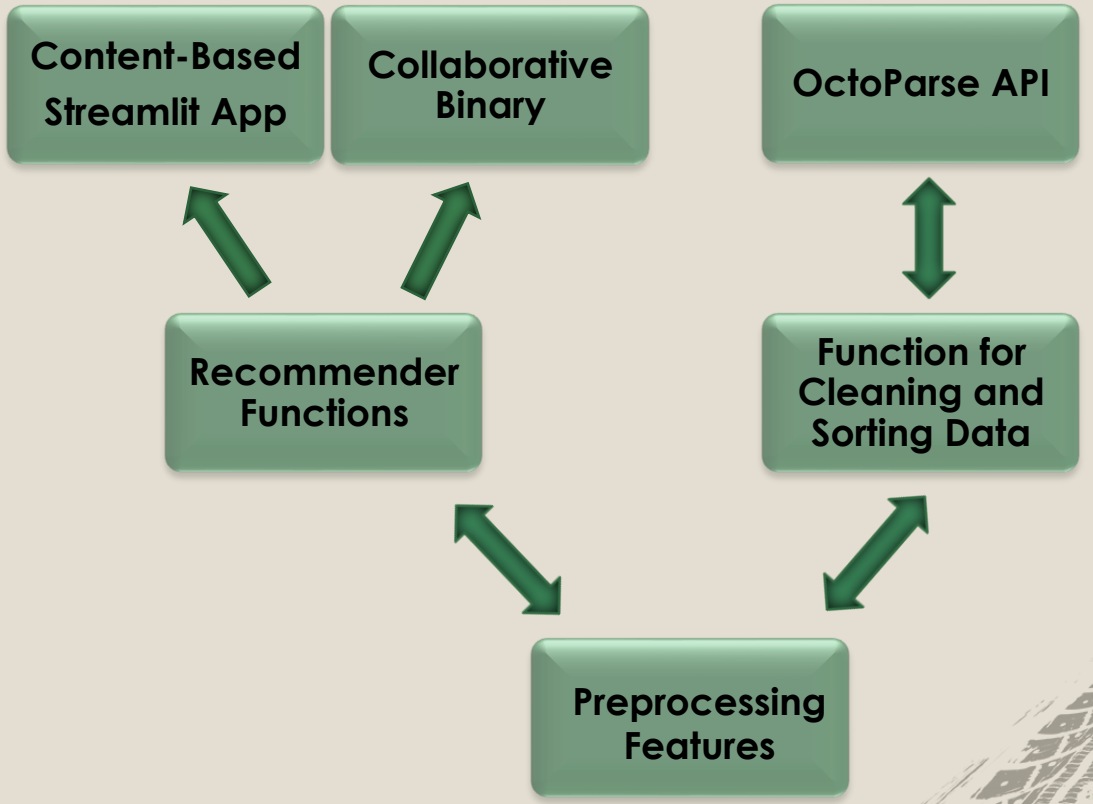
- Pulled all relevant trail stats information for ~2000 trail in AZ and Utah and ~12,000 usernames and ratings

Scalable functions for data cleaning, sorting, and saving to various CSVs

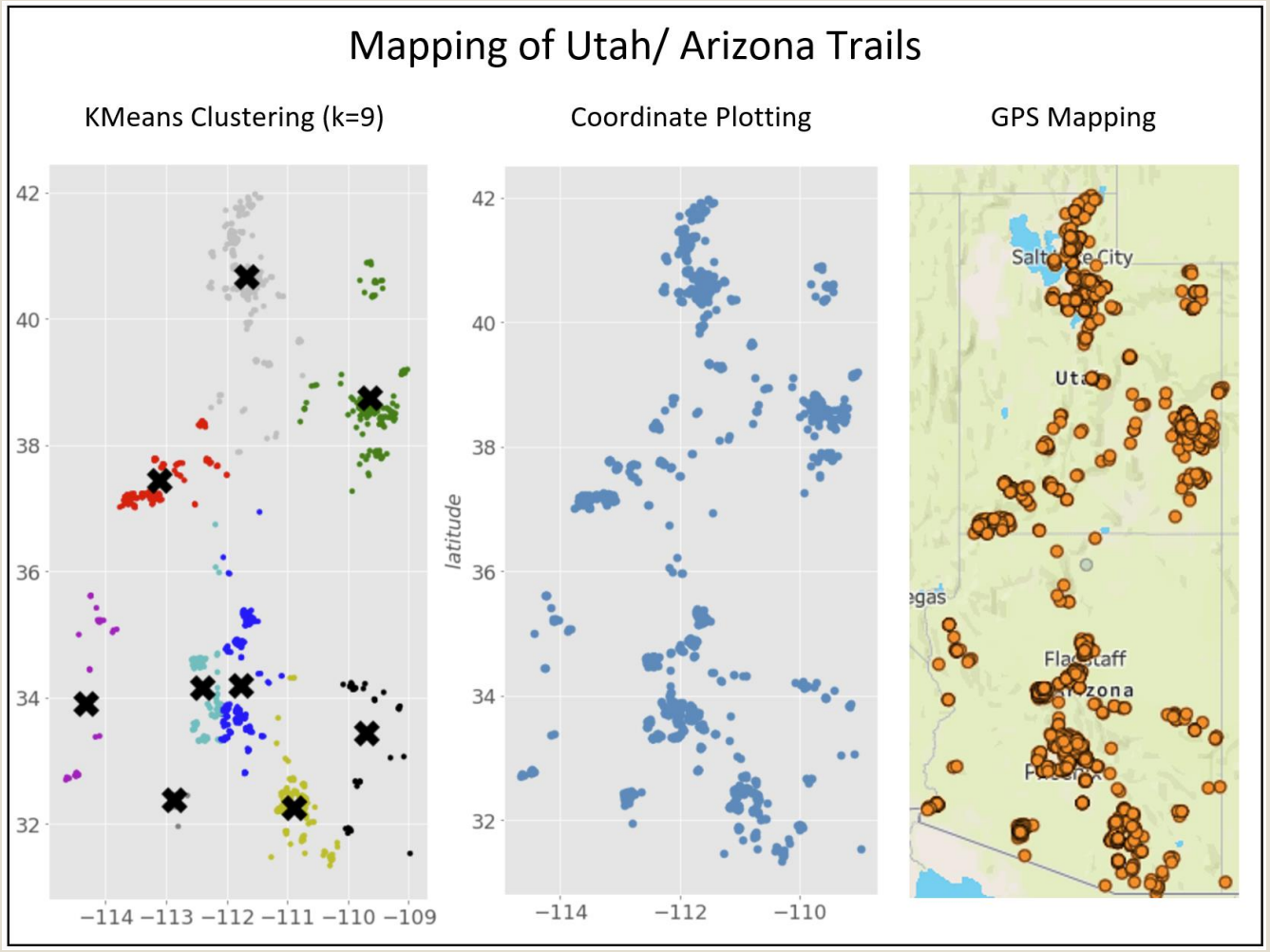
- Unscaled data for trail dashboard
- Scaled, OHE data for content recommenders
- Split user ratings for collaborative recommenders

Streamlit WebApp deployed to Heroku

- Trail recommender live!



EDA / Feature Development



Nulls and datatype discrepancies corrected in cleaning functions.

- KNN Imputation for most missing data
- Scaled, OHE data for content recommenders
- Split user ratings for collaborative recommenders

Interactive features exploration available on Streamlit

- Mountain Bike Trail Filter allows users to
  view all stats for trails

Recommenders

Content – based trail recommender

- 929 trails for AZ, 916 for UT
- 24 trail features used
- Capable of taking in trails from other states once data is gathered
- Deployed to Streamlit

User – based collaborative recommender (binary)

- 5,192 users in AZ, 7346 in UT
- Binary ratings (1 = trail checked in)
- Every trail had users represented

	John Connolly
User Name	**Cos. Similarity**
David Darling	**0.374372**
Nick Wilder	**0.350500**
Atfarley Farley	**0.300567**
Matt Castelli	**0.270135**
Tim Mccabe	**0.257325**

To Streamlit !

https://rcm-mtb-trail.herokuapp.com/

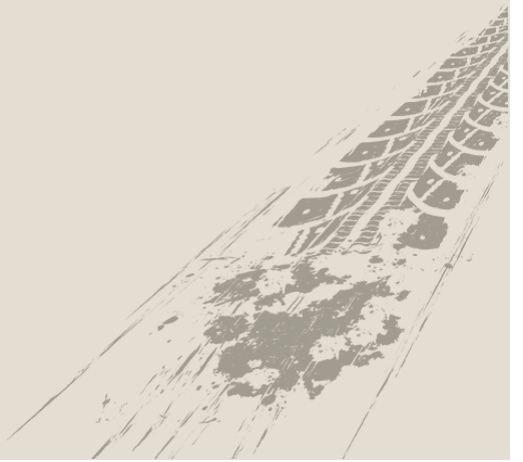
Next Steps / Future Development

More Trails! More Features!

- 13 days are needed to pull the remainder of US trails (40,000)
- Location- based search function
  - All states can be on one recommender
  - Trail systems near state lines can be together
- Users from the entire country

More search criteria!

- NLP on trail descriptions for users to search more specific trail stats (ex. ‘ledge’, ‘jump lines’, ‘drops’)
- Trail areas can be chosen based on KMeans clustering results



Conclusions

- Choosing which trail to ride shouldn't be a guess. With the recommenders developed in this project and deployed onto the production webapp, users now have the confidence needed to try new trails and experience great solo-riding in new areas!

- The tools utilized in this project can very easily be carried over into other recommenders for hiking, climbing, or any other outdoor activity. As more data is collected, options are expansive, but the recommender provided is a great start!