Travel Attraction Recommendation System

Phillip Ojo

Business Case

Travel trends amongst millennials have been increasing over the years. Different travel trends like Weekend Getaways, Romantic Getaways, Insta-Tourism, and Business-Leisure Travel. The goal is to build and intelligent travel recommendation system that will provide the user details on the most popular attractions based on the type of travel trend you fall under.

Research/Data

Research was performed on multiple travel blogs to identify the various locations in the United states that travelers used in these different trends. Data was then pulled from Tripadvisor website and made into a csv file.

- Attraction
- Rank
- Rating
- Address
- Label
- Overview

Model and Technologies Used

- **Selenium Web Driver was used to web scrape Tripadvisor website.**

- **NLTK was used to clean and tokenize the text for content based modeling.**

- **FuzzyWuzzy was used to help the model receive inputs.**

Collaborative Approach

Model #1 Recommendations

```
Recommendations for Garden and Glass:
1: Paul and Lulu Hilliard University Art Museum, with distance of 0.5491651826662838
2: Moxie Falls, with distance of 0.5491651826662838
3: Fountains of Bellagio, with distance of 0.54426728481235
4: Wednesday Wines on 3rd St, with distance of 0.42264973081037416
5: ZooWorld Zoological Conservatory, with distance of 0.42264973081037416
6: Mendocino Coast Model Railroad & Historical Society, with distance of 0.42264973081037416
7: Otis Johnson Park, with distance of 0.42264973081037416
8: Pedego Electric Bikes, with distance of 0.42264973081037416
9: The Metropolitan Museum of Art, with distance of 0.42264973081037416
10: Sarah P. Duke Gardens, with distance of 0.42264973081037416
11: Museum of Life + Science, with distance of 0.42264973081037416
12: Little Rock Central High School National Historic Site, with distance of 0.4226497308103741
13: Village of Baytowne Wharf, with distance of 0.42264973081037416
14: Summer Waves Water Park, with distance of 0.42264973081037416
15: El Paso Holocaust Museum and Study Center, with distance of 0.42264973081037416
16: Mission San Buenaventura, with distance of 0.42264973081037416
17: The Ark Thrift Shop, with distance of 0.42264973081037416
18: San Buenaventura State Beach, with distance of 0.42264973081037416
19: River Market District, with distance of 0.42264973081037416
20: Cattail Cove State Park, with distance of 0.42264973081037416
```

Content based Approach

Model #2 Recommendations

```
Recommendations for Garden and Glass:
1: Eyrie Vineyards, with distance of 0.871435130693355
2: The Museum of Flight, with distance of 0.871435130693355
3: Mission San Buenaventura, with distance of 0.8694417580332265
4: Gracianna Winery, with distance of 0.8666337326576838
5: Wild Goose Casino, with distance of 0.8666337326576838
6: The Getty Center, with distance of 0.8651600275073517
7: Elevation Sports, with distance of 0.8598000725534437
8: The Kitchen Shoppe Cooking School, with distance of 0.8598000725534437
9: U.S. Space and Rocket Center, with distance of 0.8578661890962598
10: Mendo Insider Tours, with distance of 0.8492443277111182
11: Universal Studios Florida, with distance of 0.8492443277111182
12: Silver Queen Gondola, with distance of 0.8492443277111182
13: Von Braun Astronomical Society, with distance of 0.8492443277111182
14: Minnehaha Falls, with distance of 0.8492443277111182
15: Dark Moon Craft Beer, with distance of 0.8477137740317069
16: Antique Adventures LLC, with distance of 0.8468320837865033
17: Driftwood Beach, with distance of 0.8358780120275564
18: Carlisle Events, with distance of 0.8358780120275564
19: Wild Horse Renewable Energy Center, with distance of 0.8139031579203059
20: Windansea Beach, with distance of 0.8093074821508816
```

Conclusions

The content based approached recommendation score was determined but the description of the different attractions. The collaborative approach was determined using the "user-label" rating of each attraction. Notice we have a higher scores vs when using the collaborative approach. Based on the content, has the model has captured that theses are similar attractions as percent match.

Since we were iterating through a list of cities the webdriver did perform the web scraping task quite slowly. Hopefully next time we can use a REST api to be able to call the data from the database much more quickly.

Tf-idf Vectorizer versus the Countvectorizer had an impact on how the model performed in the end. Both suffered by the curse of dimensionality but they help to eliminate the cold start problem most recommendation systems run into. Countvectorizer performed better since it wasn't so detailed.

Future Work

- Implement a hybrid recommender system and this hybrid recommender can offer both popular and less-know content to users.

- There are several other website that can contain more relevant travel information on the items and hence can be be used for better predictions.

- Model to also predict flight prices and "best" times to purchase a flight.

- Eliminate under 18 recommendations and emphasize on suggesting kid friendly places.

Thank You