Holiday Recommender System

Applied Data Science Capstone

Arnaud Cave

March 2019

Agenda

- Introduction
- Data
- Methodology
- Results
- Discussion
- Conclusion

Introduction

- Big idea: Use artificial intelligence to make customized suggestions of travel destinations
- The technology exists! Recommender systems
- Implementation: Web-based solution for potential clients or as an assistant tool for a travel agent
- Ultimate goal: improve services provided by travel agencies and drive additional revenues

Data

- Wikipedia: list of cities with high number of international visitors
- Foursquare: list of venues for each city
- Focus on cities with at least 50 venues and venue categories with a minimum number of venues

Data

- Transform venue categories into quintiles
- Final dataset is called **VenueTable**
- Top five row are as follows:

	City	arts_entertainment	food	nightlife	parks_outdoors	shops	travel
0	Amman	1	5	4	2	1	1
1	Amsterdam	3	2	5	1	4	1
2	Athens	4	4	5	1	1	1
3	Auckland	5	4	1	2	3	2
4	Barcelona	2	3	4	5	1	3

Methodology

- Preliminary analysis: K-means
- Main contribution: Content-based recommender system
- Input user:

	City	rating
0	Amsterdam	5
1	London	4
2	New York City	2
3	Paris	4
4	Toronto	1

Results

- **K-means**
- Centroid values: average the occurrence variables for each venue category and each cluster

	arts_entertainment	food	nightlife	parks_outdoors	shops	travel
Cluster Labels						
0	3.192308	31.192308	4.192308	2.384615	4.846154	3.192308
1	3.760000	22.440000	4.040000	2.360000	11.600000	4.280000
2	8.500000	20.777778	4.388889	5.111111	4.666667	4.388889

Results

- **Recommender system**
- User profile

arts_entertainment	49
food	36
nightlife	54
parks_outdoors	50
shops	56
travel	28

- Top 10 recommendations

	City	arts_entertainment	food	nightlife	parks_outdoors	shops	travel	Estimated_Ratings
67	Washington D.C.	4	1	4	5	3	4	3.58242
51	Saint Petersburg	5	1	2	5	4	3	3.46886
39	Milan	2	1	4	4	5	4	3.45055
56	Sofia	5	1	5	5	2	1	3.44689
66	Warsaw	5	2	4	3	3	3	3.42491
46	Paris	2	2	4	5	4	1	3.25275
59	Taipei	1	2	4	3	5	4	3.21978
38	Mexico City	5	1	3	3	3	4	3.1978
15	Dublin	4	2	5	1	4	2	3.17949
31	London	5	1	2	5	2	4	3.16117

Discussion

- **K-means**
- Interesting results
- Provide distinct sets of destinations for travelers interested in food (cluster 0), shopping (cluster 1), and arts & entertainment and parks & outdoors (cluster 2).
- Such analysis could be interesting for a travel agent seeking to make travel suggestions to potential clients after asking them their main centers of interest.

Discussion

- **Recommender system**
- Top 10 recommendations align well with the user profile.
- Washington DC is the top recommendation. This destination ranks high with respect to arts_entertainment, nightlife, parks_outdoors ; and shops, which are the four attributes with the highest weights in theuser profile.
- Food venues have a low rank for Washington DC which is consistent with low weight of this attribute in the user profile.
- Travel venues have a high rank for Washington DC despite the low weight of this attribute in my user profile. This is likely due to a positive correlation between the occurrence of travel venues and the occurrence of other venues which are given more weight in theuser profile.

Conclusion

- This is a preliminary work in view of developing a recommender system for holiday destinations.
- Such work could represent significant business value for travel agencies.
- More work would need to be done, especially on finding the best way to characterize a destination in terms of existing venues.