

CEBD-1260 Final Presentation

KIRSTEN ZERNASK

Dataset Description

- More than 1500 museums scraped from TripAdvisor, with general museum information, including:
 - Museum name, address, latitude and longitude, rank, number of reviews, average rating, and number of times featured by Trip Advisor
- Additional data per museum describing traveler and review information, including:
 - Detailed ratings and traveler types, museum categories, and tag clouds

<https://www.kaggle.com/annecool37/museum-data>

Problem Definition

- Cluster museums based on traveler type and tags in order to recommend types of museums to visit

Solution

- kmeans cluster analysis (k=8)
- Selection of 16 attributes (tags + traveler type)
- Useful tags selected based on cluster analysis with 100 top tags
- Dimensionality reduction by grouping similar tags together
 - 'art lovers' + 'works of art' etc. -> 'art'

Results

cluster	0	1	2	3	4	5	6	7
special exhibit	-0.31	0.19	-0.16	-0.02	1.46	-0.26	-0.12	0.44
interactive	-0.09	0.03	-0.26	1.33	-0.41	-0.11	-0.38	-0.14
family	-0.31	0.98	-0.36	1.89	-0.5	-0.46	-0.45	-0.24
audio guide	-0.07	-0.15	0.2	-0.31	-0.02	-0.24	0.02	0.82
art	-0.29	-0.3	0.1	-0.32	0.94	-0.29	-0.02	0.02
display	-0.04	-0.01	0.02	-0.27	0.52	0	-0.6	0.55
rainy day	0	0.7	-0.08	0.4	-0.07	-0.29	-0.4	-0.17
natural	-0.23	4.35	-0.23	-0.23	-0.23	-0.23	-0.23	0.28
eye opening	-0.03	-0.15	0.05	-0.15	-0.15	0.12	0.02	0.61
learn	0.25	0.07	-0.09	0.2	-0.2	0.04	-0.2	-0.2
history	2.44	-0.19	-0.41	-0.36	-0.41	0.24	-0.27	-0.41
permanent collection	-0.28	-0.21	-0.34	-0.34	2.97	-0.31	0.16	0.03
TopCouples	0.7	-0.66	0.7	-1.12	0.62	-1.44	-1.44	-1.44
TopFamilies	-0.65	0.74	-0.65	1.22	-0.57	1.55	-0.65	-0.65
TopFriends	-0.16	-0.16	-0.16	-0.16	-0.16	-0.16	6.23	-0.16
TopSolo	-0.08	-0.08	-0.08	-0.08	-0.08	-0.08	-0.08	13.27
Count	166	79	704	224	147	224	40	9
	history	natural history	art	science	art	specialty	specialty	specialty
	landmarks		history	children's		history	art	
	historic sites		landmarks	natural history		military		

Discussion

- In general, the analysis creates clusters of family-oriented or non-family-oriented museums, broken down by several key attributes
 - Family-oriented: tends to be interactive and educational (e.g. science, natural history, or children's museums)
 - Non-family-oriented: art collections, historical artifacts, audio guide / tours, popular sights / landmarks
- Further opportunities would be to increase the number of clusters and inputs (including location) for a more granular analysis, and to recommend specific museums in addition to the types

Data App

<https://museum-recommender.herokuapp.com/>