TOURIST GUIDE

By: Amgad Mohamed

INTRODUCTION

- Some people travel to visit historical places and museums, spend time at famous bars/nightclubs, or for 'Food Tourism'
- In this project a list of large cities from all over the world is considered
- Cities will be grouped according to the type of tourism they offer
- This recommender system can be of interest to tourists willing to get suggestions for cities to visit which satisfy their touristic interests.
- This can also be of interest to people willing to invest in any of these categories of tourism (historical places/museums, bars\nightclubs and food)
- The Foursquare API can be used to obtain the list of venues for each of the cities

DATA SOURCES

- The list of different cities in the world is obtained from the 'World City Database' from the site https://simplemaps.com/data/world-cities
- It is built using authoritative sources such as the NGIA, US Geological Survey, US Census Bureau, and NASA.
- It was last refreshed in April of 2019. The database contains around 13 thousand entries.
- The file containing the list of cities and countries with their corresponding longitudes and latitudes is in CSV format with no missing data in these columns.
- This can be imported and then the Foursquare API can be used to get the venues for each city.

DATA CLEANING

- The column used to get the names of the cities from the dataset is 'city_ascii' and not the 'city' column .Therefore, the columns to keep are 'city_ascii', 'country', 'lat' and 'lng'. Then, change 'city_ascii; column name into 'city'.
- Duplicate city names are removed
- Cities that 0 hotels or equivalent are removed from the dataset
- To decide if a given venue can be considered as an accomodation for tourists, the category of the venue will be checked if it belongs to the following list of words (hotel, motel, hostel, auberge, inn, lodge, tavern, guesthouse, B and B, resort, camp, room, appartment, mansion) obtained from https://www.merriam-webster.com/thesaurus/hotel https://relatedwords.org/relatedto/hotel
- For each of the cities, venues that belong to the three categories will be counted, while other venues will not be included.

DATA CLEANING

- For a venue to belong to the 'food' category, it's category needs to have one of the following words (restaurant, bistro, pizza, chicken, beef, seafood, ice cream, sushi, barbeque, noodle, steak, diner, bbq, wings, burger, buffet, grill, grills, grilled, steakhouse, fish, tacos, pasta).
- And for a venue to belong to the 'bars\nightclubs' category, it's category needs to have one of the following words (bar, nightclub, liquor, brewery, pub, disco, discotheque, wine, dance, casino, beer, cocktail, cabaret, brasserie, lounge).
- Finally, for a venue to belong to the 'historical places/museums', it's category needs to have one of the following words (museum, historical, history, monument, site, historic, monuments, gallery, palace, hall, library, archeological, castle, chateau, fortress, fountain)

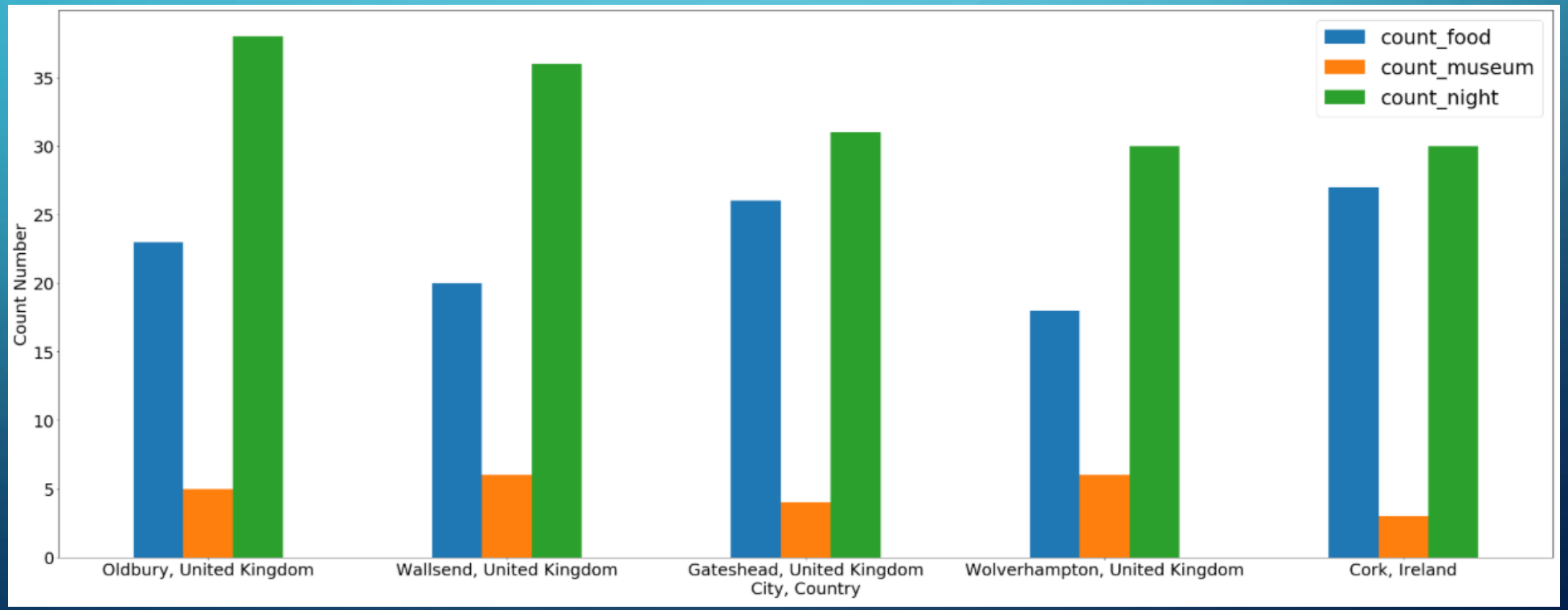
METHODOLOGY (EXPLORING DATA)

- **Cities to get the ones with best nightlife**

	city	count_accom	count_food	count_museum	count_night	country	lat	lng	city, country
0	Oldbury	2	23	5	38	United Kingdom	52.5000	-2.0167	Oldbury, United Kingdom
1	Wallsend	4	20	6	36	United Kingdom	54.9914	-1.5597	Wallsend, United Kingdom
2	Gateshead	3	26	4	31	United Kingdom	54.9450	-1.6175	Gateshead, United Kingdom
3	Wolverhampton	2	18	6	30	United Kingdom	52.5833	-2.1333	Wolverhampton, United Kingdom
4	Cork	2	27	3	30	Ireland	51.8986	-8.4958	Cork, Ireland
5	Galway	9	24	0	29	Ireland	53.2724	-9.0488	Galway, Ireland
6	Katerini	6	31	0	28	Greece	40.2723	22.5025	Katerini, Greece
7	Bradford	2	28	4	28	United Kingdom	53.8000	-1.7500	Bradford, United Kingdom
8	Walsall	4	13	4	28	United Kingdom	52.6000	-2.0000	Walsall, United Kingdom
9	Birmingham	2	27	4	28	United Kingdom	52.4750	-1.9200	Birmingham, United Kingdom

METHODOLOGY (EXPLORING DATA)

- Top 5 cities **with best nightlife**



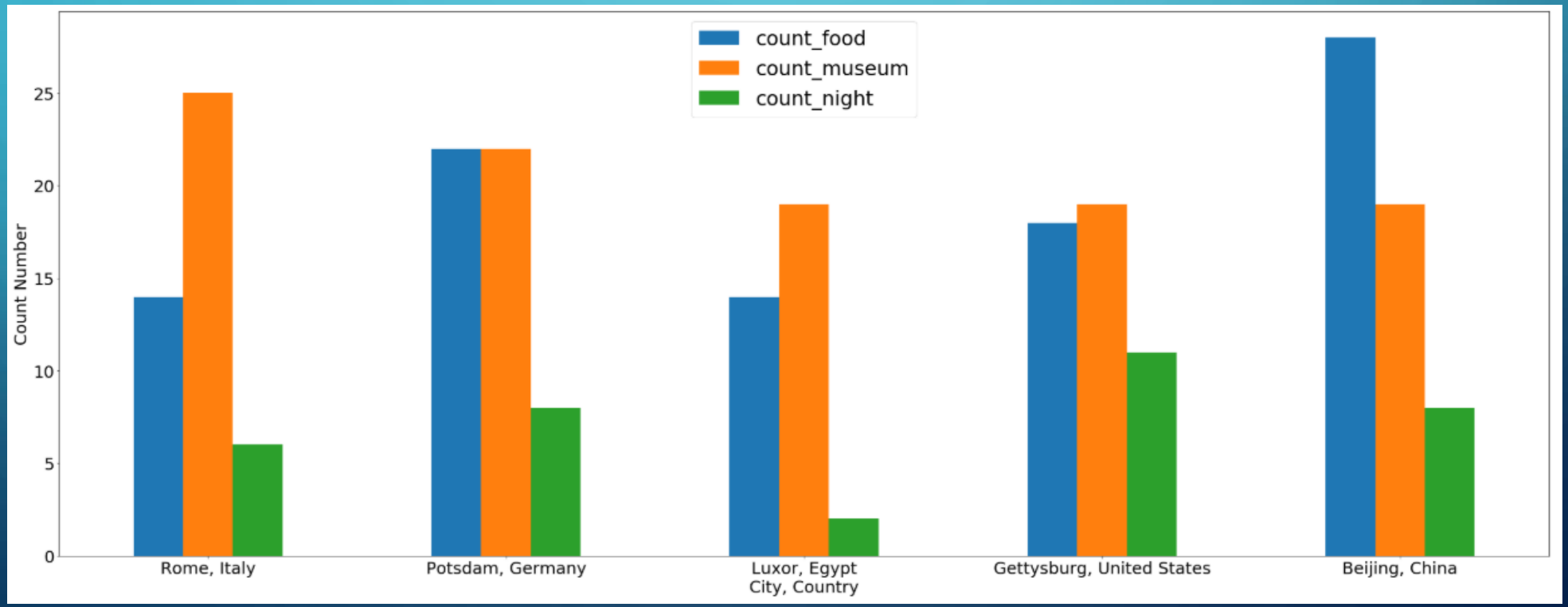
METHODOLOGY (EXPLORING DATA)

- **Cities to get the ones with the most historical places**

	city	count_accom	count_food	count_museum	count_night	country	lat	lng	city, country
0	Rome	1	14	25	6	Italy	41.8960	12.4833	Rome, Italy
1	Potsdam	3	22	22	8	Germany	52.4004	13.0700	Potsdam, Germany
2	Luxor	10	14	19	2	Egypt	25.7000	32.6500	Luxor, Egypt
3	Gettysburg	7	18	19	11	United States	39.8304	-77.2339	Gettysburg, United States
4	Beijing	17	28	19	8	China	39.9289	116.3883	Beijing, China
5	Washington	6	23	18	8	United States	38.9047	-77.0163	Washington, United States
6	Ankara	1	14	18	13	Turkey	39.9272	32.8644	Ankara, Turkey
7	Diyarbakir	6	22	18	4	Turkey	37.9204	40.2300	Diyarbakir, Turkey
8	St. Petersburg	8	2	17	9	Russia	59.9390	30.3160	St. Petersburg, Russia
9	Al Quds	15	19	17	7	West Bank	31.7764	35.2269	Al Quds, West Bank

METHODOLOGY (EXPLORING DATA)

- Top 5 cities **with the most historical places**



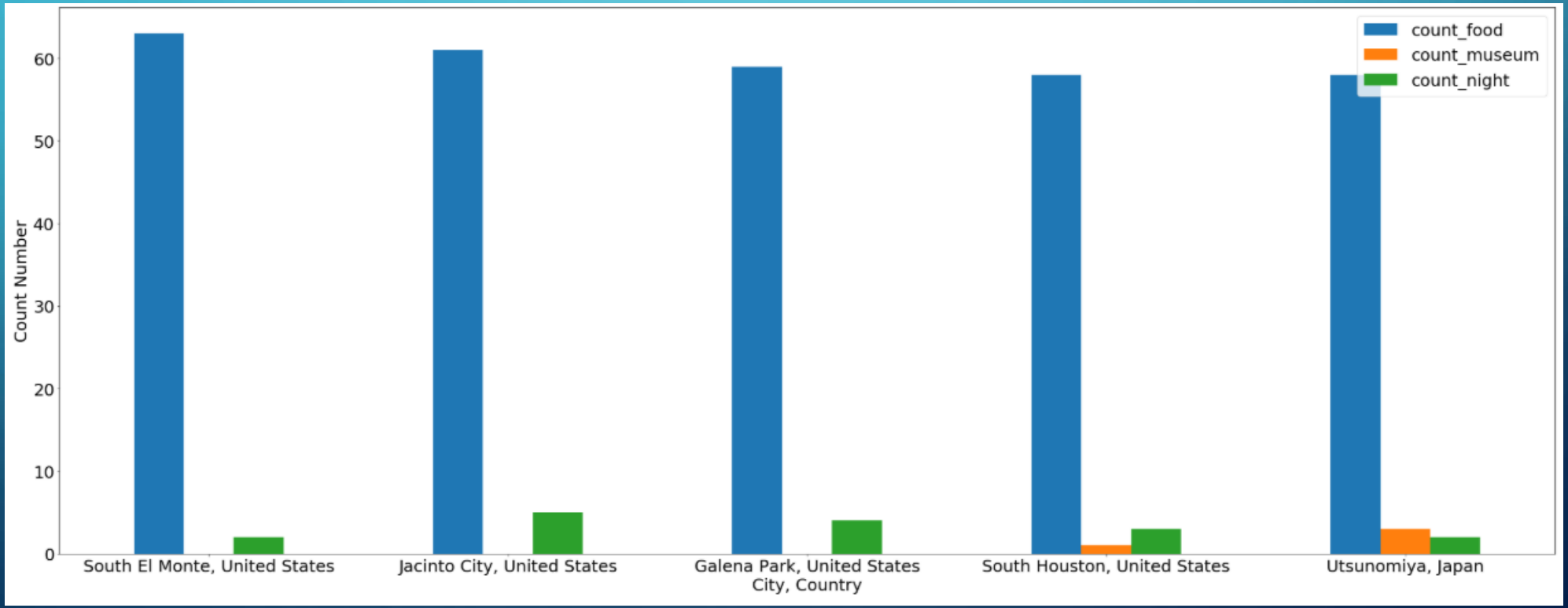
METHODOLOGY (EXPLORING DATA)

- **Cities to get the ones with best food tourism**

	city	count_accom	count_food	count_museum	count_night	country	lat	lng	city, country	trending_food
0	South El Monte	1	63	0	2	United States	34.0493	-118.0484	South El Monte, United States	0
1	Jacinto City	1	61	0	5	United States	29.7663	-95.2410	Jacinto City, United States	0
2	Galena Park	1	59	0	4	United States	29.7452	-95.2333	Galena Park, United States	0
3	South Houston	1	58	1	3	United States	29.6611	-95.2285	South Houston, United States	0
4	Utsunomiya	2	58	3	2	Japan	36.5500	139.8700	Utsunomiya, Japan	0
5	Chon Buri	1	57	0	2	Thailand	13.4004	101.0000	Chon Buri, Thailand	0
6	Fukuoka	7	56	3	6	Japan	33.5950	130.4100	Fukuoka, Japan	0
7	Addison	1	56	1	3	United States	32.9587	-96.8356	Addison, United States	0
8	Marbella	4	55	1	7	Spain	36.5166	-4.8833	Marbella, Spain	0
9	Kangar	4	55	1	1	Malaysia	6.4330	100.1900	Kangar, Malaysia	0

METHODOLOGY (EXPLORING DATA)

- Top 5 cities **with best food tourism**



METHODOLOGY (NORMALIZE DATA)

	city	count_accom	count_food	count_museum	count_night	country	lat	lng	city, country	ratio_food	ratio_museum	ratio_night
0	South El Monte	1	63	0	2	United States	34.0493	-118.0484	South El Monte, United States	0.969231	0.000000	0.030769
1	Jacinto City	1	61	0	5	United States	29.7663	-95.2410	Jacinto City, United States	0.924242	0.000000	0.075758
2	Galena Park	1	59	0	4	United States	29.7452	-95.2333	Galena Park, United States	0.936508	0.000000	0.063492
3	South Houston	1	58	1	3	United States	29.6611	-95.2285	South Houston, United States	0.935484	0.016129	0.048387
4	Utsunomiya	2	58	3	2	Japan	36.5500	139.8700	Utsunomiya, Japan	0.920635	0.047619	0.031746

METHODOLOGY (USE K-MEANS TO GET 4 CLUSTERS)

	count_accom	count_food	count_museum	count_night	lat	lng	city, country	ratio_food	ratio_museum	ratio_night	Cluster Labels
0	3	50	1	5	35.4480	-94.3529	Van Buren, United States	0.892857	0.017857	0.089286	3
1	5	36	7	11	35.8869	14.4025	Imdina, Malta	0.666667	0.129630	0.203704	1
2	4	36	6	9	4.5964	-74.0833	Bogota, Colombia	0.705882	0.117647	0.176471	1
3	6	49	1	9	20.5504	-97.4700	Poza Rica de Hidalgo, Mexico	0.830508	0.016949	0.152542	3
4	2	5	2	2	48.3318	40.2518	Kamensk Shakhtinskiy, Russia	0.555556	0.222222	0.222222	1
5	3	37	5	16	37.6897	-97.3442	Wichita, United States	0.637931	0.086207	0.275862	1
6	2	40	0	2	34.9514	-89.9787	Southaven, United States	0.952381	0.000000	0.047619	3
7	5	37	1	5	26.0293	-80.1678	Hollywood, United States	0.860465	0.023256	0.116279	3
8	7	37	14	7	37.5663	126.9997	Seoul, Korea, South	0.637931	0.241379	0.120690	1
9	1	14	0	2	43.0282	-74.9928	Herkimer, United States	0.875000	0.000000	0.125000	3
10	2	1	0	0	42.8071	22.3247	Crna Trava, Serbia	1.000000	0.000000	0.000000	3
11	4	38	0	6	42.3765	-122.9109	Central Point, United States	0.863636	0.000000	0.136364	3
12	18	38	0	5	42.5361	1.5828	Encamp, Andorra	0.883721	0.000000	0.116279	3
13	1	37	0	8	41.7980	-87.9569	Clarendon Hills, United States	0.822222	0.000000	0.177778	3
14	4	5	0	2	54.8612	-6.2763	Ballymena, United Kingdom	0.714286	0.000000	0.285714	1
15	3	21	0	4	-25.3800	-51.4800	Guarapuava, Brazil	0.840000	0.000000	0.160000	3
16	1	38	1	12	41.6000	-87.6905	Markham, United States	0.745098	0.019608	0.235294	1
17	5	28	2	3	18.4504	-97.3800	Tehuacan, Mexico	0.848485	0.060606	0.090909	3
18	2	8	1	4	46.4367	15.9536	Dornava, Slovenia	0.615385	0.076923	0.307692	1
19	1	35	3	10	41.9663	-87.8057	Harwood Heights, United States	0.729167	0.062500	0.208333	1

There are 204 cities that belong to cluster 1

RESULTS (CLUSTER 1)

	count_food	lat	lng	city, country	ratio_food	ratio_museum	ratio_night	Cluster Labels
66	10	50.7004	-3.5300	Exeter, United Kingdom	0.333333	0.066667	0.600000	0
125	0	-5.9897	39.2519	Mahonda, Tanzania	0.000000	0.000000	1.000000	0
178	0	-0.5196	37.4500	Embu, Kenya	0.000000	0.000000	1.000000	0
249	3	0.0204	37.0600	Nanyuki, Kenya	0.428571	0.000000	0.571429	0
375	3	46.5103	15.0806	Slovenj Gradec, Slovenia	0.375000	0.000000	0.625000	0
383	0	57.7647	36.6900	Bezhetsk, Russia	0.000000	0.000000	1.000000	0
500	6	54.8800	-2.9300	Carlisle, United Kingdom	0.315789	0.105263	0.578947	0
515	1	2.7800	32.2800	Gulu, Uganda	0.333333	0.000000	0.666667	0
521	0	58.4503	-130.0333	Dease Lake, Canada	0.000000	0.000000	1.000000	0
569	0	-20.0809	146.2587	Charters Towers, Australia	0.000000	0.000000	1.000000	0

There are 204 cities that belong to cluster 1

There are 204 cities that belong to cluster 1

RESULTS (CLUSTER 2)

	count_food	lat	lng	city, country	ratio_food	ratio_museum	ratio_night	Cluster Labels
1	36	35.8869	14.4025	Imdina, Malta	0.666667	0.129630	0.203704	1
2	36	4.5964	-74.0833	Bogota, Colombia	0.705882	0.117647	0.176471	1
4	5	48.3318	40.2518	Kamensk Shakhtinskiy, Russia	0.555556	0.222222	0.222222	1
5	37	37.6897	-97.3442	Wichita, United States	0.637931	0.086207	0.275862	1
8	37	37.5663	126.9997	Seoul, Korea, South	0.637931	0.241379	0.120690	1
14	5	54.8612	-6.2763	Ballymena, United Kingdom	0.714286	0.000000	0.285714	1
16	38	41.6000	-87.6905	Markham, United States	0.745098	0.019608	0.235294	1
18	8	46.4367	15.9536	Dornava, Slovenia	0.615385	0.076923	0.307692	1
19	35	41.9663	-87.8057	Harwood Heights, United States	0.729167	0.062500	0.208333	1
22	37	41.9193	-88.3110	Saint Charles, United States	0.804348	0.000000	0.195652	1

There are 2835 cities that belong to cluster 2

There are 204 cities that belong to cluster 1

RESULTS (CLUSTER 3)

	count_food	lat	lng	city, country	ratio_food	ratio_museum	ratio_night	Cluster Labels
29	1	57.3364	23.1235	Mersrags, Latvia	0.333333	0.666667	0.000000	2
49	1	28.6800	115.8800	Nanchang, China	0.500000	0.500000	0.000000	2
65	0	39.4763	75.9699	Kashgar, China	0.000000	1.000000	0.000000	2
80	12	40.8182	-74.0022	Fairview, United States	0.428571	0.428571	0.142857	2
151	14	39.9272	32.8644	Ankara, Turkey	0.311111	0.400000	0.288889	2
200	0	-28.8780	28.0560	Hlotse, Lesotho	0.000000	1.000000	0.000000	2
215	1	55.9077	21.8456	Plunge, Lithuania	0.333333	0.333333	0.333333	2
272	1	28.0304	73.3299	Bikaner, India	0.333333	0.666667	0.000000	2
285	1	43.9443	116.0443	Xilinhot, China	0.333333	0.666667	0.000000	2
338	14	25.7000	32.6500	Luxor, Egypt	0.400000	0.542857	0.057143	2

There are 183 cities that belong to cluster 3

There are 204 cities that belong to cluster 1

RESULTS (CLUSTER 4)

	count_food	lat	lng	city, country	ratio_food	ratio_museum	ratio_night	Cluster Labels
0	50	35.4480	-94.3529	Van Buren, United States	0.892857	0.017857	0.089286	3
3	49	20.5504	-97.4700	Poza Rica de Hidalgo, Mexico	0.830508	0.016949	0.152542	3
6	40	34.9514	-89.9787	Southaven, United States	0.952381	0.000000	0.047619	3
7	37	26.0293	-80.1678	Hollywood, United States	0.860465	0.023256	0.116279	3
9	14	43.0282	-74.9928	Herkimer, United States	0.875000	0.000000	0.125000	3
10	1	42.8071	22.3247	Crna Trava, Serbia	1.000000	0.000000	0.000000	3
11	38	42.3765	-122.9109	Central Point, United States	0.863636	0.000000	0.136364	3
12	38	42.5361	1.5828	Encamp, Andorra	0.883721	0.000000	0.116279	3
13	37	41.7980	-87.9569	Clarendon Hills, United States	0.822222	0.000000	0.177778	3
15	21	-25.3800	-51.4800	Guarapuava, Brazil	0.840000	0.000000	0.160000	3

There are 3432 cities that belong to cluster 4

RESULTS

- The first cluster includes mainly the cities that have higher ratio of venues that can be categorized as nightlife. As for the third cluster is dominated by the cities that have higher ratio of museums and historical places. While the fourth cluster is dominated by cities that have higher ratio of restaurants.
- The second cluster contains cities that have almost equal ratios of nightlife venues, museums and historical places, while the ratio of restaurants are dominating.
- The number of cities that belong to the clusters 2 and 3 are more than those that belong to clusters 1 and 4 and that is probably due to the fact that cities with higher ratio of restaurants are usually more than cities that have dominating ratios of nightlife and historical sites venues.

CONCLUSION & FUTURE PERSPECTIVES

- Using the results of this project, people can have a better idea of what to expect when visiting certain cities, or can target specific cities when seeking a certain type of tourism (historical places/museums, bars/nightclubs and food) or investment. This helps makes things easier for strangers visiting different cities.
- More work can be done to include more cities and to filter the venues to get only the trending ones or those with high number of tips.