Where to go on vacation

PREDICTING BEST DESTINATIONS BASED ON USER RATINGS OF PAST VACATION DESTINATIONS

Presentation content:

- Introduction, background and problem
- Data acquisition
- Cities and features selection
- Methodology
- Analysis
- Results, discussion and conclusion

Introduction, background & problem:

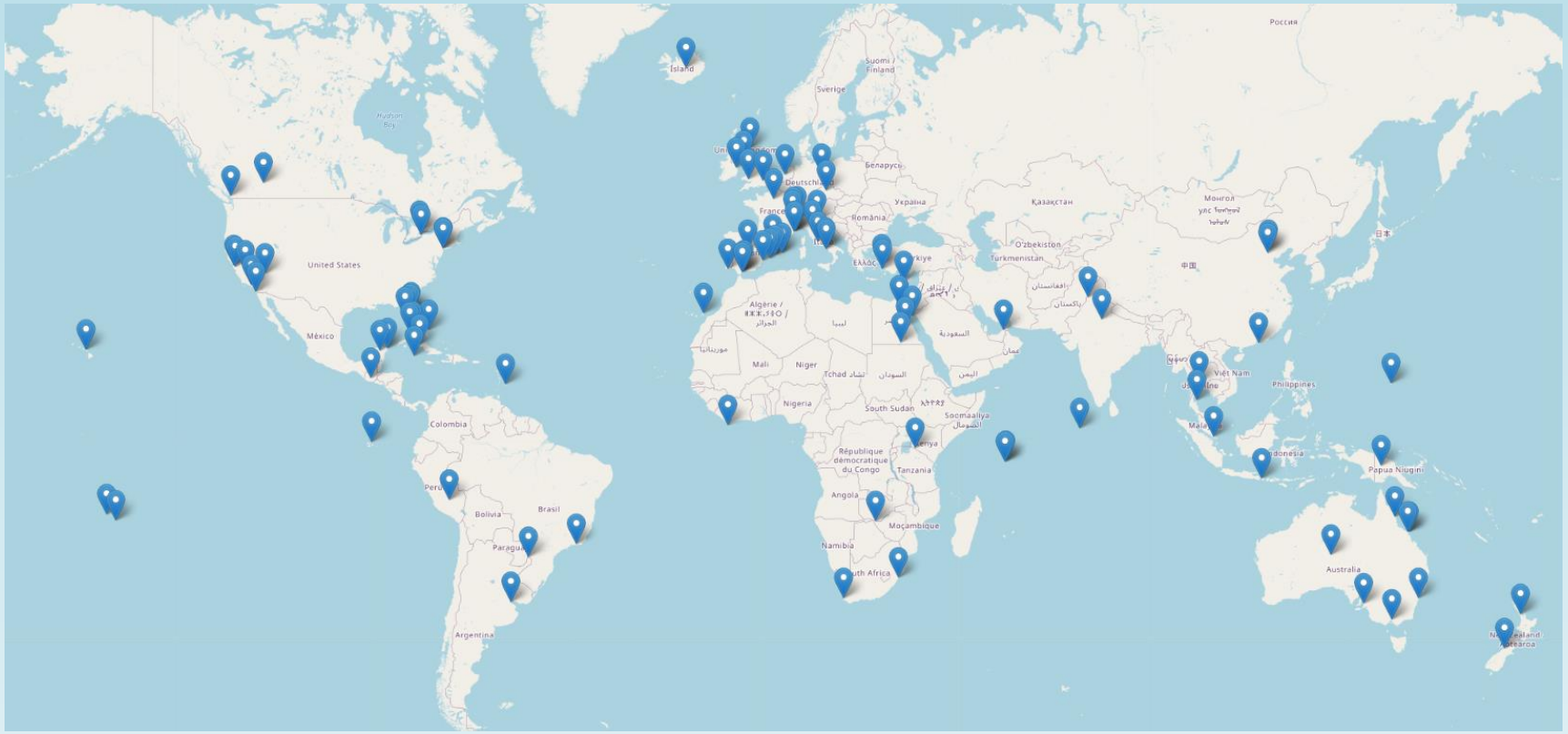
- Most people go on vacation every year, spending an average of $1,536 for household in USA.
- Lots of possible destinations, and impossibility of studying them all without wasting a lot of time.
- Need for a beater solution, which tells the user which are the top destinations for him/her based on him/her past experiences.

Data acquisition:

- Data will be extracted from two main sources:
  - Foursquare
  - Geopy.geocoders

Cities and features selection

- We will use a total of 100 cities as vacation candidates, which are:



- For the features, we will use the total number of venues of each category a city has. We will search for venues from a total of 100 different categories

Methodology

- We will search for every feature of every city, and save the returned venues from the foursquare API. We will have a table like this:

	City	Category	Name	Latitude	Longitude
0	Aberdare	Cinema	Vue	51.738475	-3.377418
1	Aberdare	Night club	Aberdare Constitutional Club	51.713557	-3.447650
2	Aberdare	Night club	Aberdare Rugby Club	51.709252	-3.432933
3	Aberdare	Night club	Aberdare golf club	51.716892	-3.429162
4	Aberdare	Night club	Cwmdare Club	51.715940	-3.469329
...	...	...	...	...	...
150771	Zermatt	Metro station	Taxi Metro	46.068560	7.776482
150772	Zermatt	Metro station	Riffelalp Station	46.004908	7.753993
150773	Zermatt	Metro station	Bahnhof Zermatt	46.023864	7.748048
150774	Zermatt	Metro station	Green Motion Charging Station	46.067318	7.775392
150775	Zermatt	Metro station	Blauherd Station	46.017076	7.785827

Methodology

- The next step is to sort all this data, and create a new table with every city in one row, and 100 columns, each with one feature. Each cell will contain the number of venues of that category that exist in that city:

	Aquarium	Arcade & Bowling	Casino	Cinema	Night club	Disco	Music	Art	Stadium	Theme Park	Water Park	Zoo	American Restaurant	African Restaurant	Italian Restaurant
Aberdare	0	0	0	4	80	0	0	4	0	34	0	0	12	12	12
Abu Simbel	0	0	0	0	0	0	0	0	0	0	0	0	4	4	4
Adelaide	4	140	40	96	200	92	132	200	36	100	0	28	200	200	200
Airlie Beach Queensland	0	0	0	0	44	4	4	0	0	16	0	0	44	44	44
Algarve	0	0	0	4	200	16	4	52	0	28	0	8	200	200	200
Amritsar	0	0	0	20	24	0	0	16	0	20	0	0	72	68	68
Amsterdam	20	28	88	136	200	132	200	200	36	100	0	104	200	200	200
Antigua	0	0	0	0	28	0	0	88	4	4	0	4	200	200	200
Auckland	8	80	20	64	200	64	124	200	24	100	0	44	200	200	200
Ayers Rock	0	0	0	0	0	0	0	8	0	6	0	0	4	4	4
Bahamas	0	0	24	4	180	48	28	36	8	62	0	4	200	200	200
Bali	0	0	0	0	0	0	0	4	4	0	0	0	8	8	8
Banff National Park	0	0	0	4	24	8	8	20	0	42	0	0	92	88	88
Bangkok	56	52	48	200	200	200	200	200	200	100	0	128	200	200	200
Barbados	0	0	0	0	12	4	0	12	0	6	0	0	16	20	16

Methodology

- The next step is to see if there are any cities that should be discarded because it has few sites. This may be because Foursquare has little data about a site, because we put the coordinates wrong (for example, we put the name of a country or region instead of a city) or that the place simply has few sites. An example of a misnomer is for example the place "Hawaii". However, when looking for the coordinates in geocoders, putting the name of Hawaii instead of the name of its capital (Honolulu) returned us a location far from any city.

- After dropping invalid places we have 78 places left.

Methodology

- Now we normalize our data set.

	Aquarium	Arcade & Bowling	Casino	Cinema	Night club	Disco	Music	Art	Stadium	Theme Park	Water Park	Zoo	American Restaurant	African Restaurant	Italian Restaurant
Aberdare	0.000000	0.000000	0.000000	0.002730	0.054608	0.000000	0.000000	0.002730	0.000000	0.023208	0.0	0.000000	0.008191	0.008191	0.008191
Adelaide	0.000390	0.013643	0.003898	0.009355	0.019489	0.008965	0.012863	0.019489	0.003508	0.009745	0.0	0.002729	0.019489	0.019489	0.019489
Airlie Beach Queensland	0.000000	0.000000	0.000000	0.000000	0.023170	0.002106	0.002106	0.000000	0.000000	0.008425	0.0	0.000000	0.023170	0.023170	0.023170
Algarve	0.000000	0.000000	0.000000	0.000816	0.040800	0.003264	0.000816	0.010608	0.000000	0.005712	0.0	0.001632	0.040800	0.040800	0.040800
Amritsar	0.000000	0.000000	0.000000	0.008617	0.010340	0.000000	0.000000	0.006894	0.000000	0.008617	0.0	0.000000	0.031021	0.029298	0.029298
Amsterdam	0.001605	0.002247	0.007061	0.010913	0.016049	0.010592	0.016049	0.016049	0.002889	0.008024	0.0	0.008345	0.016049	0.016049	0.016049
Antigua	0.000000	0.000000	0.000000	0.000000	0.007646	0.000000	0.000000	0.024031	0.001092	0.001092	0.0	0.001092	0.054615	0.054615	0.054615
Auckland	0.000748	0.007478	0.001870	0.005982	0.018695	0.005982	0.011591	0.018695	0.002243	0.009348	0.0	0.004113	0.018695	0.018695	0.018695
Ayers Rock	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.023599	0.000000	0.017699	0.0	0.000000	0.011799	0.011799	0.011799
Bahamas	0.000000	0.000000	0.003306	0.000551	0.024793	0.006612	0.003857	0.004959	0.001102	0.008540	0.0	0.000551	0.027548	0.027548	0.027548
Bali	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.027397	0.027397	0.000000	0.0	0.000000	0.054795	0.054795	0.054795
Banff National Park	0.000000	0.000000	0.000000	0.001401	0.008403	0.002801	0.002801	0.007003	0.000000	0.014706	0.0	0.000000	0.032213	0.030812	0.030812

Custom client part

- The next thing is defining a user's score to previous cities to see where their next vacation should be.
- For this we will invent a user, whose scores will be:

city	rating
New York	2.0
Barcelona	2.5
Bora Bora	5.0
Melbourne	5.0
Bangkok	3.0
Barbados	4.0
Airlie Beach Queensland	5.0
Cancun	5.0
Berlin	2.5
Vancouver	3.0
San Francisco	4.0
Las Vegas	3.5
Cairo	4.0

Custom client part

- From the user ratings and the Cities features, we can compute how much the user likes each feature. For this user we get:

Beach Bar	1.804576
Sport bar	1.531414
Bar	1.516815
Cocktail Bar	1.516815
Seafood Restaurant	1.006117
Asian Restaurant	0.980840
African Restaurant	0.980840
French Restaurant	0.977399
Mediterranean Restaurant	0.970454
American Restaurant	0.966242
Italian Restaurant	0.966242
Turkish Restaurant	0.966242
Mexican Restaurant	0.966242
Gift shop	0.850035
Antique shop	0.845822

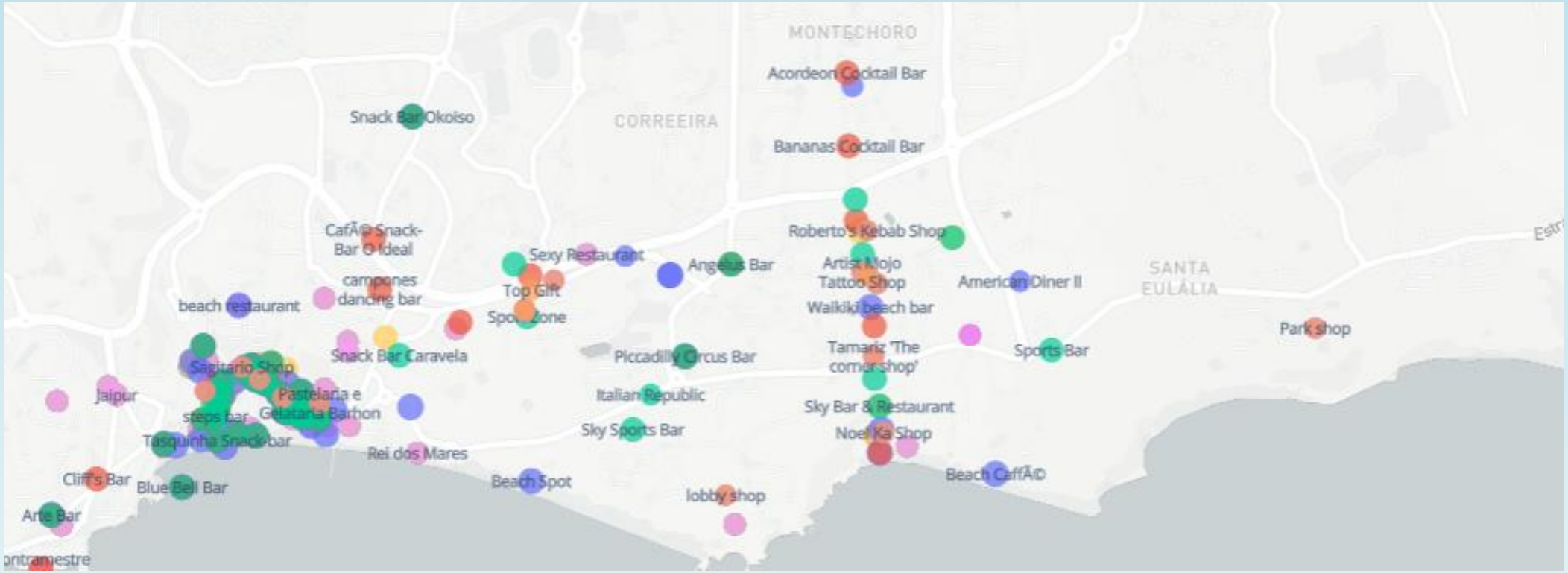
User recomendations

- Once we have the user predilection for each feauter, we can compute the user favourites cities, which are:

	Match	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Maldives	0.024405	Sport bar	Beach Bar	Bar	Antique shop	Cocktail Bar
La Digue Island	0.023480	Beach Bar	Cocktail Bar	Bar	Sport bar	Island
Isle of Man	0.022069	Bar	Beach Bar	Cocktail Bar	Sport bar	Mountain
Puerto del Rosario, Canary Islands	0.021413	Sport bar	Beach Bar	Cocktail Bar	Bar	Cafeteria
Zermatt	0.021052	Asian Restaurant	Sport bar	American Restaurant	Mediterranean Restaurant	African Restaurant
Antigua	0.020885	Turkish Restaurant	Hotel	American Restaurant	African Restaurant	Italian Restaurant
Courchevel	0.020550	Sport bar	Ski Area	Hotel	Cocktail Bar	Beach Bar
Chamonix	0.020474	Hotel	Sport bar	Cocktail Bar	Beach Bar	Bar
Sorrento	0.020099	Cocktail Bar	Sport bar	Bar	Beach Bar	Hotel
Luxor	0.019727	African Restaurant	Hotel	Italian Restaurant	Turkish Restaurant	American Restaurant
Machu Picchu	0.019577	Seafood Restaurant	Hotel	Italian Restaurant	Asian Restaurant	American Restaurant
Victoria Falls	0.019413	Hotel	Beach Bar	African Restaurant	Bar	Cocktail Bar
Algarve	0.019077	French Restaurant	Turkish Restaurant	American Restaurant	African Restaurant	Italian Restaurant
Dubai	0.019050	Sport bar	Turkish Restaurant	Beach Bar	Bar	Cocktail Bar
Chichen Itza	0.018828	Mexican Restaurant	American Restaurant	French Restaurant	Turkish Restaurant	Italian Restaurant

User recomendations

- Toghether with the table showing the best matches and the top venues of each city, we provide a dashboard where the user can select a city from the recommended list, and the city and it’s venues will show in a map:



Results, discussion and conclusion

- The result of this project is the recommended sites for a particular user.
- As has been seen and explained, the scores entered clearly correspond to a person with little interest in big cities, someone who enjoys relaxing vacations much more, in quiet places and especially with the beach. We can see, as of the 15 recommended cities, except Dubai, the other cities are relatively quiet cities, and most are beach, so it seems that the algorithm works quite well, and the recommendations are good.
- In conclusion, we can say that this program is a good tool when planning a vacation, since due to the wide range of places to go, and how quickly they all change, it is difficult to know where to go, and where you will find what you are looking for
- The final decision of where to go will be up to the client, but good recommendations are made.