Capstone Project
“Finding your next vacation spot”

Franziska Metge

Problem Description

We live in a world of travelers but we also live in a world of individuals with divers interests. It would be preposterous to assume that everybody enjoys exactly the same cities. Some people might enjoy cities with a good night life, while others want to travel to cities with a diverse cultural landscape.

Solution

A program which is able to find cities a user might find enjoyable based on an input city and venue landscape.

Results:

- table listing the similar cities,
- heat-map visualizing the similarity of these cities by venue occurrence
- scatter-plot using the cities coordinates

This material will enable the user to confidently select their next travel destination.

Data

- Population data from UN
- Coordinates pulled using geopy
- Venue information pulled using Foursquare API
- One-hot encoding using pandas.get_dummies () and pandas.groupby()

Population data from UN

City	Population Size	Country
Adrar	200834.0	Algeria
Ain Defla	450280.0	Algeria
Ain Temouchent	299341.0	Algeria
ALGIERS	2712944.0	Algeria

Final input table (one-hot encoding summarized by city

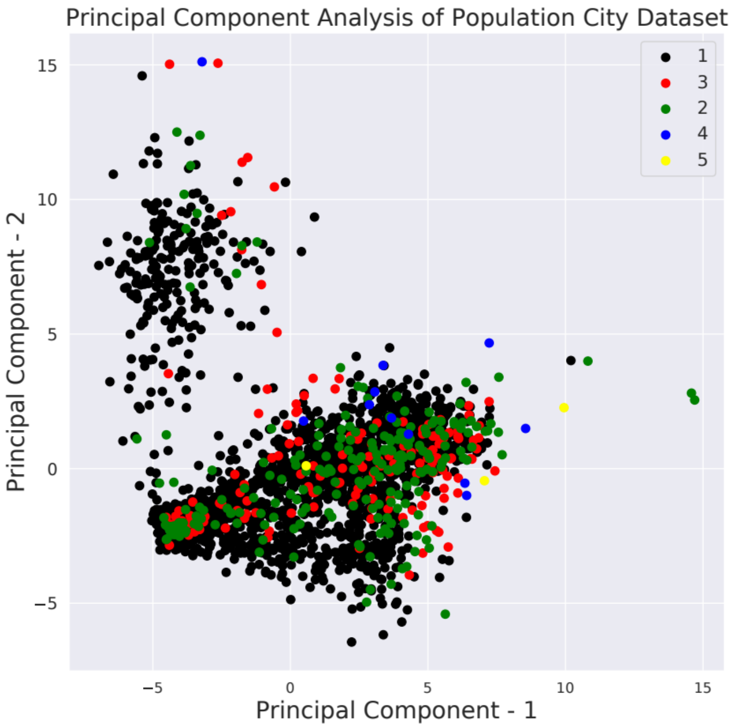
City	Population Size	Country	Latitude	Longitude	ATM	...	Fast Food	...	Zoo Exhibit
Ain Temouchent	299341.0	Algeria	35.26	-1.14	0		1		0
ALGIERS	2712944.0	Algeria	36.77	3.06	0	...	0	...	0
Annaba	442230.0	Algeria	36.89	7.75	0		0		0
Bejaïa	559981.0	Algeria	36.75	5.06	0		0		0

Methods

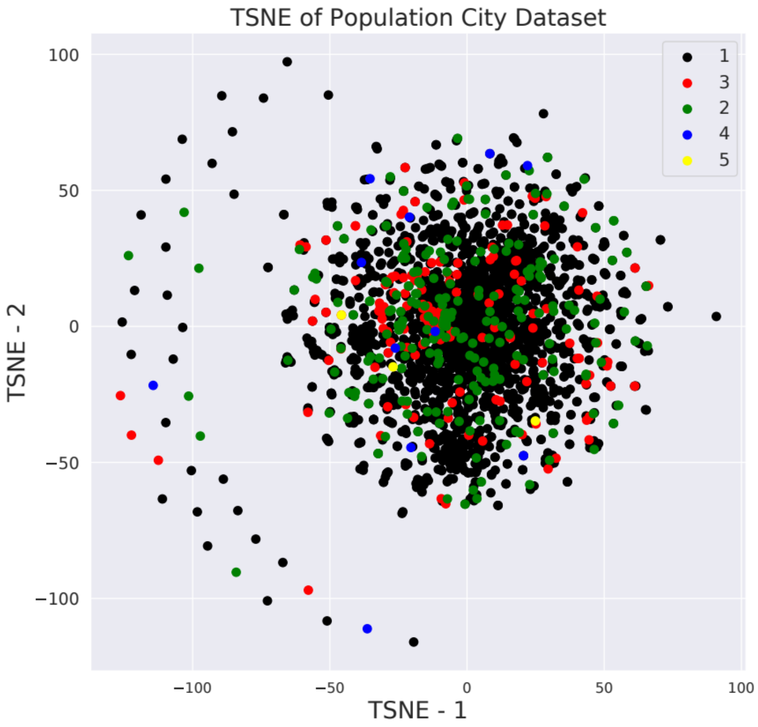
- Exploratory data analysis
  - View city landscape in lower dimensonal space: PCA, t-SNE
  - Look at number of venues per city: Histogram, Scatter-plot
- find_next_vacation()
  - Calculates distance matrix using the pdist from the scipy.spatial.distance package
  - Returns k cities most similar in table format
  - Visualizes results in heat map and scatter plot
- get_information_on_next_vacation()
  - Uses Foursquare API to look up venues of a selected city
  - Returns list of venues in table format

Results (PCA and t-SNE Embedding)

A

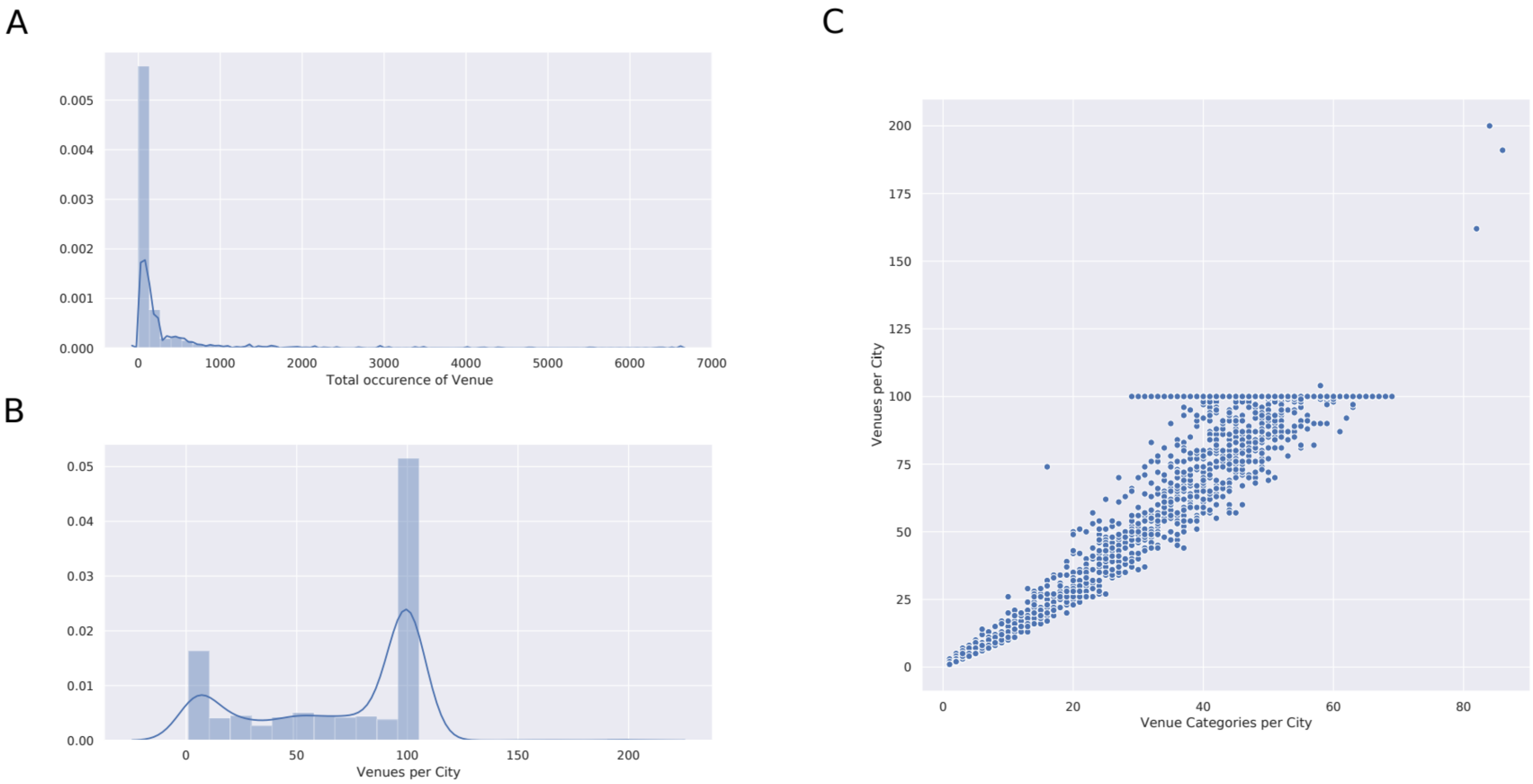


B



- Both embedding show one dense cluster and a sparse cluster of cities
- Clustering is not driven by population size

Results (Histogram and Scatter plot)



A) Majority of venue types occur seldom, while few categories appear more than 6000 times, such as coffee shops and hotels

B) Some cities have no reported venues. Foursquare limit of 100 venues per city becomes evident

C) Larger number of venues correlates with higher diversity of types of venues (R = 0.95)

Results (return table)

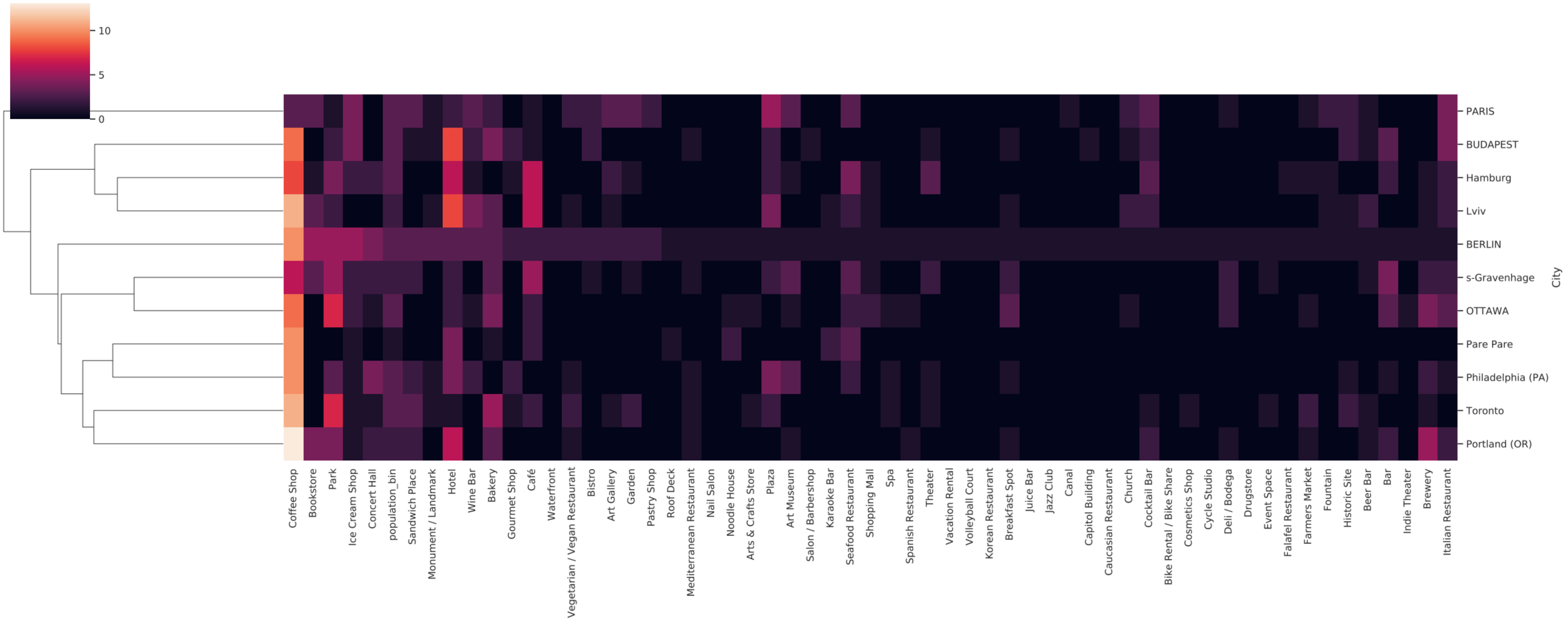
- Function call:

```
future_vacation_destination_Berlin = find_next_vacation(’BERLIN’,
population_venue_data, 10, ’euclidean’)
```

- Return table:

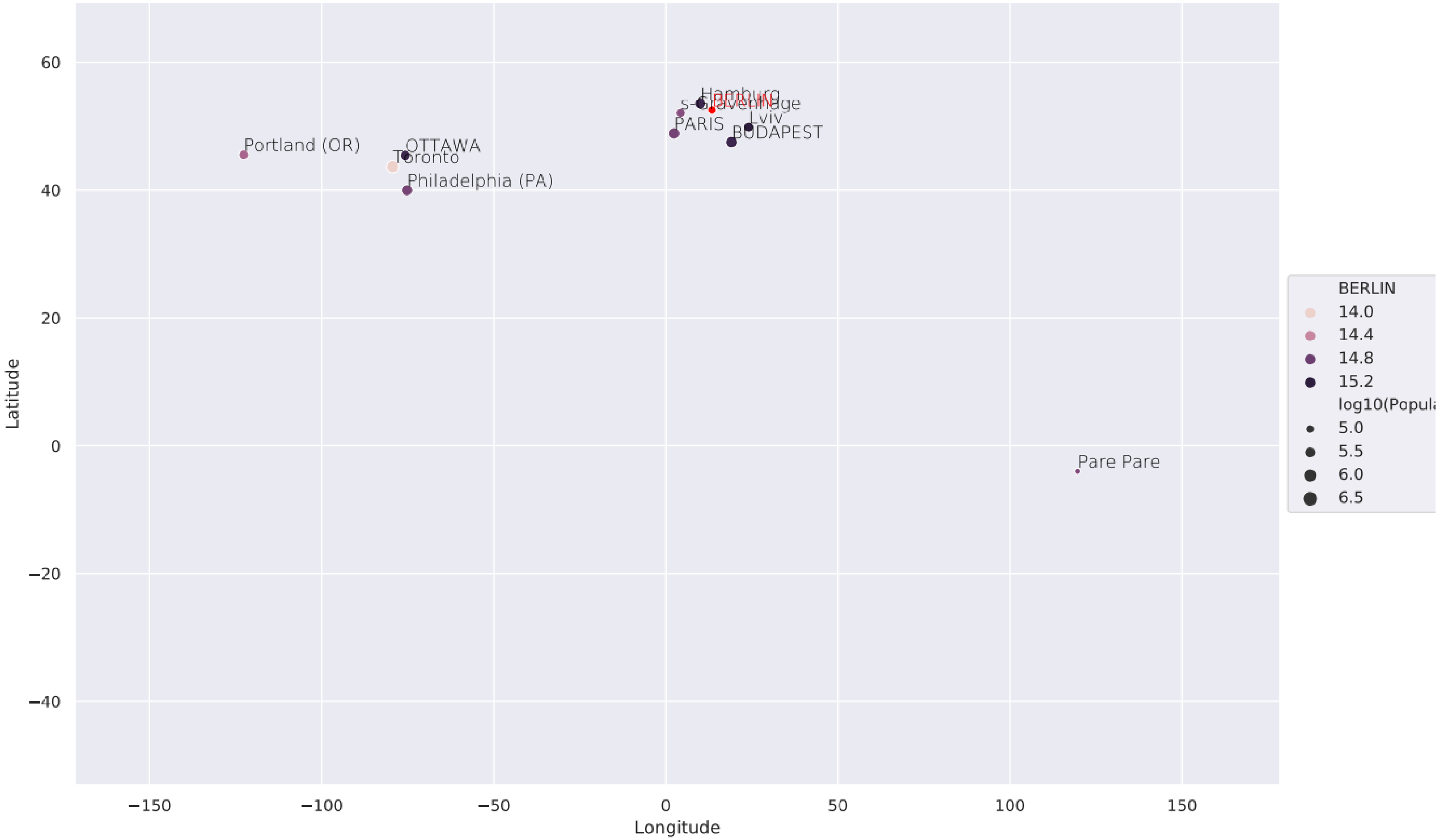
City	Population Size	Country	Latitude	Longitude	Art Gallery	...	Coffee Shop	...	Park	BERLIN
BERLIN	3613495	Germany	52.517	13.3888	2		10		5	0.0
Toronto	2956024	Canada	43.653	-79.3839	1		11		7	14.071
Portland (OR)	639863	United States of America	45.520	-122.674	0	...	13	...	4	14.560
s-Gravenha	514861	The Netherlands	52.074	4.26968	0		6		5	14.696
Pare Pare	142391	Indonesia	-4.005	119.623	0		10		0	14.730

Results (Heat map)



- Heat map of number of venues per category per city
- Rows are ordered based on hierarchical clustering
- Columns are ordered from most to least abundant venue category in Berlin

Results (Scatter plot)



- “BERLIN” result cities form two cluster; one European and one North American

Results (return table 2)

- Function call:

```
future_vacation_destination_info = get_information_on_next_vacation(
    ’s-Gravenhage’, future_vacation_destination_Berlin, 20)
```

- Return table:

City	City Latitude	City Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
s-Gravenhage	52.07494555	4.26968022053	De Kaasspeciaalzaak Ed Boele	52.07856369	4.269423625	Deli / Bodega
s-Gravenhage	52.07494555	4.26968022053	cafe madeleine	52.07875093	4.276190053	Tea Room
s-Gravenhage	52.07494555	4.26968022053	Zondag lunchroom	52.07963034	4.268465460	Café
s-Gravenhage	52.07494555	4.26968022053	Restaurant Rakang Thai	52.07710976	4.278414389	Thai Restaurant
s-Gravenhage	52.07494555	4.26968022053	Emma	52.07672963	4.282775738	Bar

Conclusion

- In this Capstone project I developed a method to recommend future travel destinations to a user based on the similarity of their input city
- In the return tables, the user obtains a comprehensive overview of similar cities as well as the venue landscape in the similar cities
- The heat map representation of that venue landscape is more condensed and thus easier to interpret
- The scatter plots representation of the similar cities helps the user to easily place the cities in a geographical relationship

- I demonstrated that it is possible to find potential vacation destinations based on their venue landscape
- With a little bit more work this could easily become a valuable app for travel agencies or travel websites