Travel Suggestion Scorecard

Leveraging on Foursquare Data and Individual Preferences to Create Travel Recommendations

Background

- Travel agencies among the **hardest hit** in this pandemic.
- We can leverage on Foursquare data and user preferences to create a scorecard that can make recommendations to future vacationers.
- This will hopefully help bolster the economy of the target countries.

Data

List of Countries, Capital Cities and their coordinates

- This will be used to query from Foursquare API

Top Venues per Location

- This will be used to cluster countries

Happiness Index of each country

- To establish that environment impacts mood

Preferred types of venues and relative importance for each venue type

- To help us get scores for each country and make a recommendation to the client

Methodology and Results

Univariate Analysis on Top Venues Distribution vs. Happiness Index

K-Means Clustering

Scorecard

Methodology-*Univariate Analysis*

- Correlations were inconclusive
  - None were particularly high or low

Venue Category	Correlation with HI
restaurant	-0.406501
cafe/desserts	0.145805
amusement_nature	0.349484
alc_bev	0.028303
hotel	-0.490189
amusement_cultural	0.226268
entertainment	-0.146013

Venue Category	Correlation with HI
fitness	0.011613
grocery	0.112568
bookstore	0.177417
spa	0.030326
airport	-0.200702
multiplex	-0.110114
castle	0.241804
hostel	-0.097842

Methodology-*Univariate Analysis*

- So were plots
  - No relationship can be visually observed
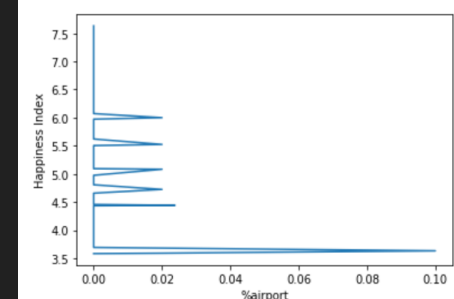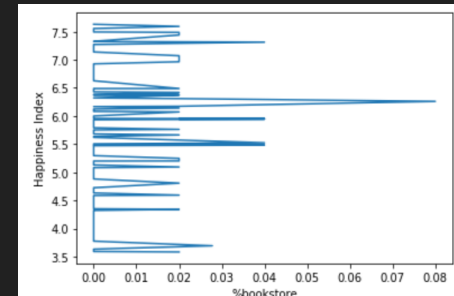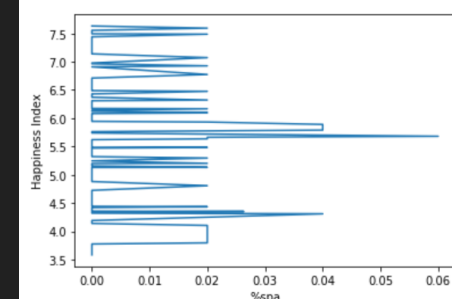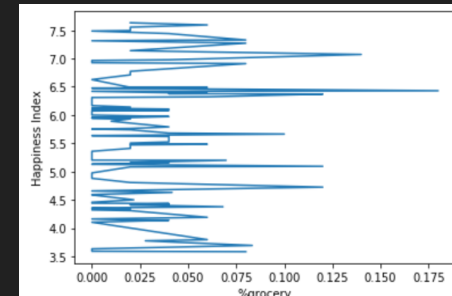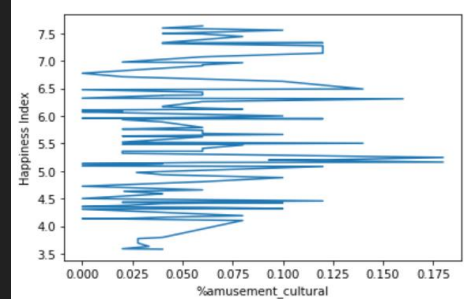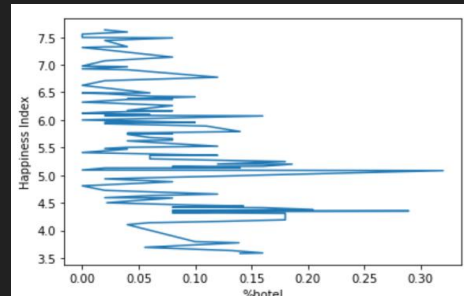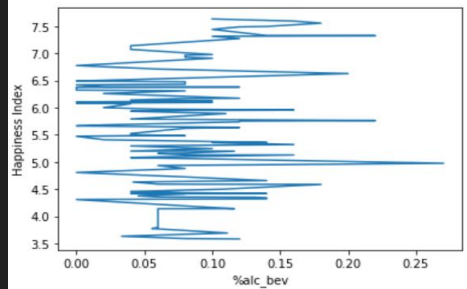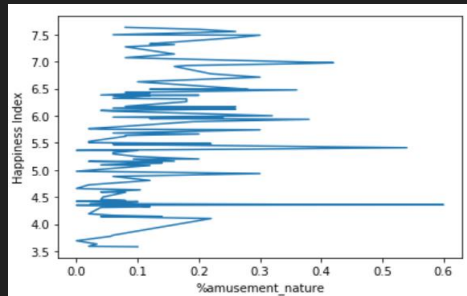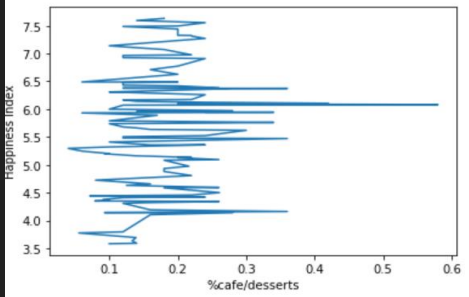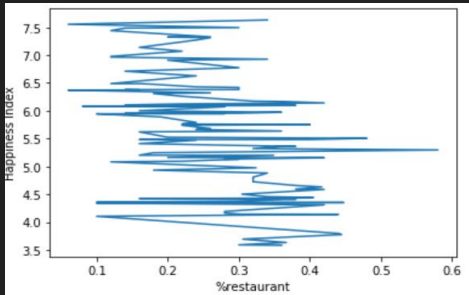
Methodology-*K-means Clustering*

- However, employing K-means clustering, an unsupervised ML method, clusters formed using data on top venue categories produced statistically significant different average happiness index levels as shown by the p-values below.

	1	2	3	4
1	1.000000	0.941399	0.004050	0.016212
2	0.941399	1.000000	0.003849	0.016724
3	0.004050	0.003849	1.000000	0.354098
4	0.016212	0.016724	0.354098	1.000000

- Cluster 1 is significantly different from clusters 3 and 4. Cluster 2 is significantly different from clusters 3 and 4. Clusters 3 is significantly different from clusters 1 and 2. Cluster 4 is significantly different from clusters 1 and 2.

Methodology-*K-means Clustering*

Cluster 1:

- lowest overall average happiness at 5.29

- majority of the venues are composed of restaurants, cafes and dessert places and places that serve alcoholic beverages.

Cluster 1

```
cluster1=group[group['Cluster Labels']==0]
print('# of members:', cluster1.shape[0])
print('average happiness index:', cluster1['HI'].mean())
cluster1
# average happiness, Lots of restaurants, cafes/dessert places, and places that serve alcohol
```

```
# of members: 33
average happiness index: 5.2920606060606055
```

	Country	HI	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Finland	7.632	0	restaurant	cafe/desserts	alc_bev	amusement_nature	amusement_cultural
3	Iceland	7.495	0	restaurant	cafe/desserts	alc_bev	fitness	amusement_nature
6	Canada	7.328	0	restaurant	cafe/desserts	amusement_nature	alc_bev	grocery
14	Belgium	6.927	0	restaurant	amusement_nature	cafe/desserts	fitness	alc_bev
31	Slovakia	6.173	0	restaurant	cafe/desserts	alc_bev	amusement_nature	hotel
32	El Salvador	6.167	0	restaurant	amusement_nature	cafe/desserts	hotel	alc_bev
33	Nicaragua	6.141	0	restaurant	cafe/desserts	amusement_cultural	amusement_nature	clothing store
36	Uzbekistan	6.096	0	restaurant	cafe/desserts	alc_bev	hotel	entertainment
40	Ecuador	5.973	0	restaurant	cafe/desserts	hotel	alc_bev	amusement_nature
43	Slovenia	5.948	0	restaurant	cafe/desserts	amusement_nature	entertainment	alc_bev
49	Bolivia	5.752	0	restaurant	cafe/desserts	alc_bev	hotel	amusement_cultural

Methodology-*K-means Clustering*

Cluster 2:
- overall slightly better happiness level than cluster 1 at 5.31 average

- majority of the venues are composed of restaurants, amusement_nature and hotels

Cluster 2

```
cluster2=group[group['Cluster Labels']==1]
print('# of members:', cluster2.shape[0])
print('average happiness index:', cluster2['HI'].mean())
cluster2

# average happiness index, lots of restaurants and hotels
```

```
# of members: 29
average happiness index: 5.310999999999999
```

	Country	HI	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
10	Austria	7.139	1	amusement_nature	restaurant	entertainment	amusement_cultural	hotel
16	United Arab Emirates	6.774	1	restaurant	amusement_nature	cafe/desserts	hotel	fitness
22	Panama	6.430	1	restaurant	grocery	cafe/desserts	amusement_nature	alc_bev
23	Brazil	6.419	1	restaurant	cafe/desserts	amusement_nature	hotel	fitness
25	Uruguay	6.379	1	amusement_nature	restaurant	cafe/desserts	fitness	alc_bev
28	Malaysia	6.322	1	restaurant	cafe/desserts	entertainment	grocery	amusement_nature
29	Spain	6.310	1	restaurant	amusement_nature	amusement_cultural	entertainment	cafe/desserts
38	Thailand	6.072	1	restaurant	hotel	amusement_nature	cafe/desserts	buddhist temple

Methodology-*K-means Clustering*

Cluster 3:
- highest average happiness level at 6.22

- majority of the venues are of the amusement_nature category. These include beaches, trail parks, gardens, mountains, surfing sports, waterfalls, and other scenic viewpoints.

Cluster 3

```
cluster3=group[group['Cluster Labels']==2]
print('# of members:', cluster3.shape[0])
print('average happiness index:', cluster3['HI'].mean())
cluster3

#highest happiness, Lots of amusement_nature venues
```

```
# of members: 18
average happiness index: 6.223055555555556
```

	Country	HI	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1	Norway	7.594	2	amusement_nature	alc_bev	cafe/desserts	restaurant	grocery
2	Denmark	7.555	2	amusement_nature	cafe/desserts	alc_bev	amusement_cultural	restaurant
4	Switzerland	7.487	2	amusement_nature	restaurant	cafe/desserts	hotel	alc_bev
5	Netherlands	7.441	2	amusement_nature	cafe/desserts	restaurant	alc_bev	amusement_cultural
12	Ireland	6.977	2	amusement_nature	cafe/desserts	restaurant	alc_bev	grocery
13	Germany	6.965	2	amusement_nature	restaurant	cafe/desserts	amusement_cultural	palace
17	Czech Republic	6.711	2	amusement_nature	cafe/desserts	restaurant	fitness	alc_bev
19	France	6.489	2	amusement_nature	castle	amusement_cultural	restaurant	hotel
21	Chile	6.476	2	amusement_nature	restaurant	cafe/desserts	fitness	alc_bev
39	Italy	6.000	2	amusement_nature	restaurant	amusement_cultural	cafe/desserts	entertainment

Methodology-*K-means Clustering*

Cluster 4:
- higher happiness level than clusters 1 and 2

- primary difference appears to be that the top category is cafes and dessert places as opposed to restaurants.

Cluster 4

```
cluster4=group[group['Cluster Labels']==3]
print('# of members:', cluster4.shape[0])
print('average happiness index:', cluster4['HI'].mean())
cluster4

# high happiness
# primarily cafes/dessert places as opposed to restaurants from Cluster1
```

# of members: 26
average happiness index: 5.940615384615385

|	Country	HI	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
7	New Zealand	7.324	3	cafe/desserts	restaurant	alc_bev	amusement_nature	amusement_cultural
8	Sweden	7.314	3	restaurant	cafe/desserts	amusement_nature	alc_bev	bookstore
9	Australia	7.272	3	restaurant	cafe/desserts	amusement_cultural	alc_bev	amusement_nature
11	Costa Rica	7.072	3	cafe/desserts	restaurant	grocery	fitness	amusement_nature
15	Luxembourg	6.910	3	cafe/desserts	restaurant	amusement_nature	alc_bev	grocery
18	Malta	6.627	3	restaurant	alc_bev	cafe/desserts	amusement_cultural	amusement_nature
20	Mexico	6.488	3	cafe/desserts	restaurant	amusement_cultural	amusement_nature	grocery
24	Guatemala	6.382	3	restaurant	cafe/desserts	entertainment	grocery	amusement_cultural
26	Qatar	6.374	3	cafe/desserts	restaurant	grocery	entertainment	hotel
27	Saudi Arabia	6.371	3	cafe/desserts	grocery	amusement_nature	farm	amusement_cultural

Methodology-*K-means Clustering*

○ The results indicate that the environment does play a factor into happiness. In particular, the existence of **nature** as a top venue in a country appears to have the strongest positive impact. These include beaches, mountains, surfing spots, skiing spots, forests, among others.

We can view the clusters using folium.
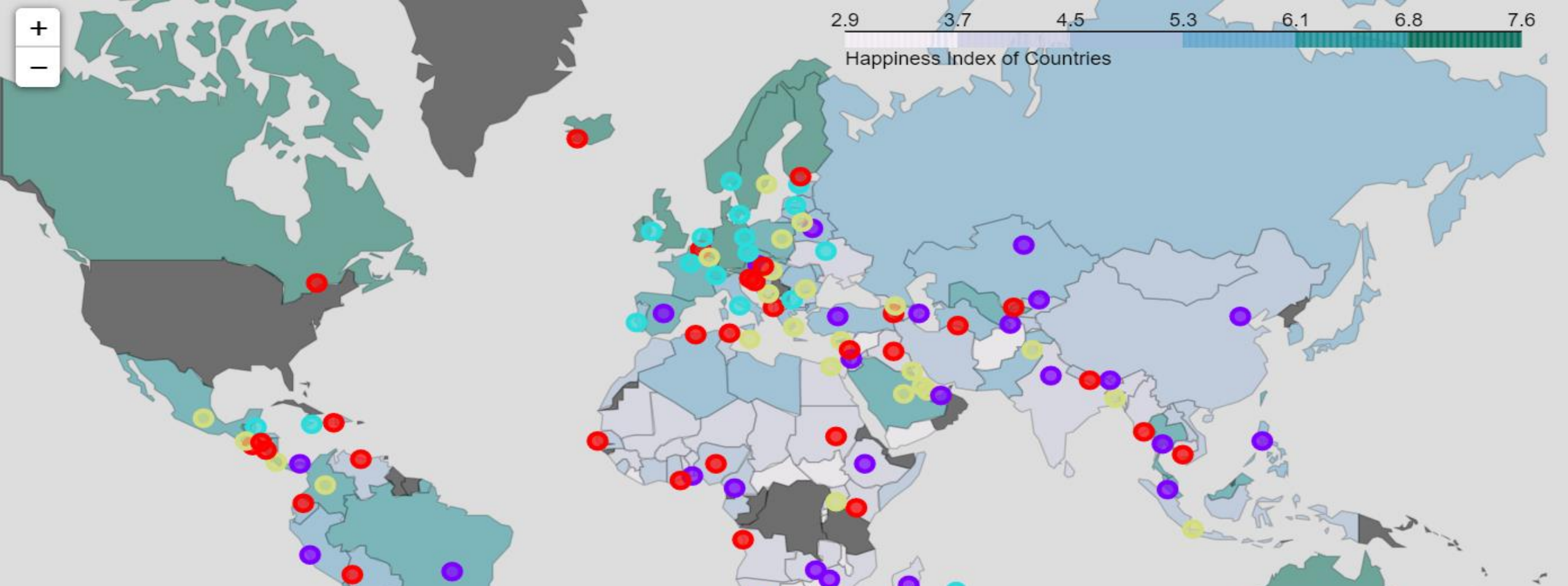
Or view them against a choropleth map.

Discussion

Results can initially be used as marketing material.

Discussion

And when clients are ready to travel again, we can request for user inputs and give them customized recommendations:

STEP 1:
Choose the most relevant categories for you

```
Choose 5 categories from the dropdown list below.
Press Ctrl key to select multiple.

Venue   restaurant
        cafe/desserts
        amusement_nature
        alc_bev
        hotel
        amusement_cultural
        entertainment
        fitness
        grocery
        spa
```

STEP 2a:
Enter the wights/relative importance of each category

```
Enter the weights for each category chosen above. The weights must sum to 100.
20 30 10 20 20
```

STEP 2b:
If correct, see a confirmation of the inputted weights:

```
Enter the weights for each category chosen above. The weights must sum to 100. 20 30 10 20 20

weight for restaurant is 20 %
weight for amusement_nature is 30 %
weight for alc_bev is 10 %
weight for amusement_cultural is 20 %
weight for entertainment is 20 %
```

Discussion

We can then compute a score and show the ranked countries. These will be our recommendations.

$$Score = \sum_{i=1}^{n\ categories\ listed} weight_i \times number\ of\ venues_i$$

Country	restaurant	amusement_nature	alc_bev	amusement_cultural	entertainment	score
Jamaica	19	20	11	4	2	12.1
Norfolk Island	19	20	11	4	2	12.1
Saint Lucia	10	28	3	0	2	11.1
Portugal	9	26	1	3	2	10.7
Aruba	17	22	4	1	0	10.6
Mauritania	5	30	3	0	1	10.5
Andorra	18	19	4	1	0	9.9
United States of Virgin Islands	12	20	12	0	0	9.6
Martinique	10	21	5	1	2	9.4
British Virgin Islands	11	20	12	0	0	9.4

Recommendations

Conclusion

- Despite the simplistic scorecard approach, one can create something of use because we were able to leverage on free Foursquare data.

- While the initial univariate analysis did not reap anything useful and one can initially prematurely conclude that the environment does not factor into happiness (assuming we did not read the methodology for this index), an unsupervised machine learning method (k-means clustering) was able to somehow group the countries based on their top venues and lead to select clusters with significantly different happiness index levels.