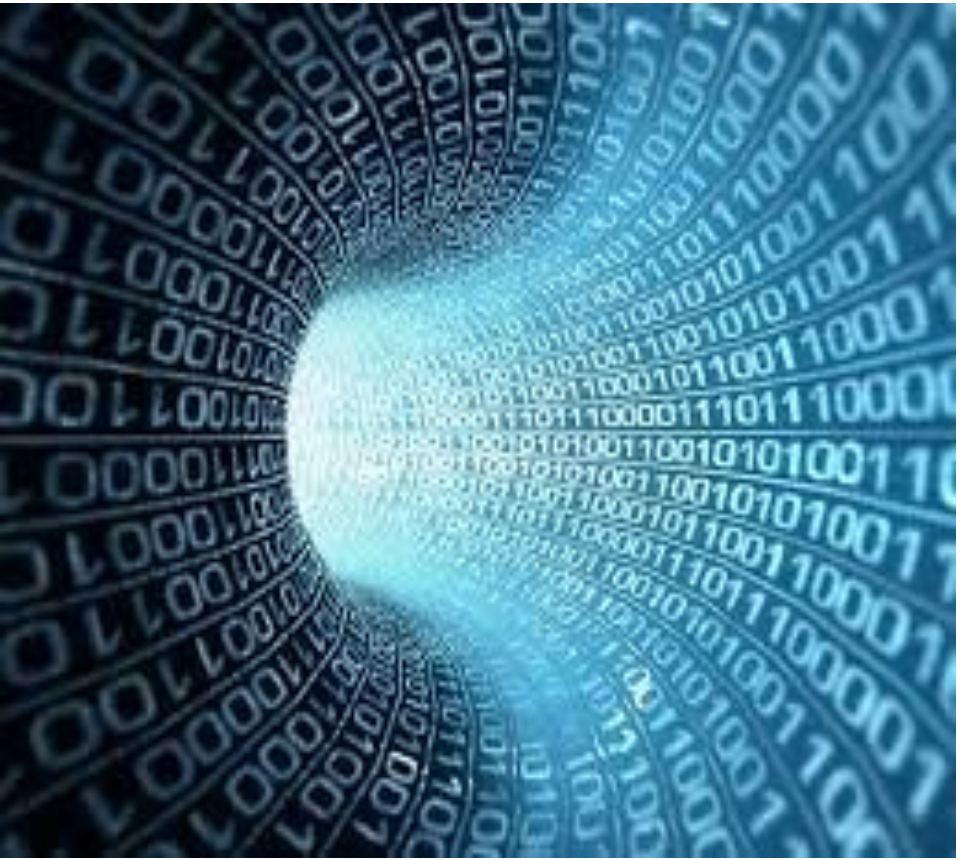
Recommendation System for travellers

The Problem: where to stay???

- Normally, when it comes where to stay we take in consideration which venues there are nearby
- We will use Foursquare API to not only get the venues nearby, but also which are the most common ones per neighborhood
- We will limit the problem to Toronto neighborhoods

The Data



- Imagine that neighborhood "A" have one coffee shop, and "B" 30 coffee shops. If we use purely one hot encoding to indicate whether or not a neighborhood has coffee shops, both will have the same weight. To solve this, we will put some weights according to the amount of venues of each type. For example, in the coffee shop column, the neighborhood "A" will have a less weight than "B".

Methodology

I will use the Foursquare API to get the data about the nearby venues for each neighborhood, and then sort them according to the 10 most popular ones in each neighborhood. Then, with this subset of most common venues, I will create another data set and put some weights according to the position of the venue in the respective neighborhood.

First line of the data set

Neighborhood	Berczy Park
1st Most Common Venue	Coffee Shop
2nd Most Common Venue	Cocktail Bar
3rd Most Common Venue	Bakery
4th Most Common Venue	Beer Bar
5th Most Common Venue	Seafood Restaurant
6th Most Common Venue	Farmers Market
7th Most Common Venue	Cheese Shop
8th Most Common Venue	Steakhouse
9th Most Common Venue	Café
10th Most Common Venue	Greek Restaurant

Steakhouse	4
Café	2
Cocktail Bar	18
Bakery	15
Greek Restaurant	1
Seafood Restaurant	10
Coffee Shop	25
Farmers Market	8
Cheese Shop	6
Beer Bar	12
Neighborhood	Berczy Park

Results

Lets get user preferences and then normalize it

```
user_preferences = dict()
```

```
user_preferences['Rental Car Location'] = 7
user_preferences['Italian Restaurant'] = 10
user_preferences['Sports Bar'] = 3
user_preferences['Indian Restaurant'] = 8
user_preferences['Recording Studio'] = 10
user_preferences['Japanese Restaurant'] = 10
user_preferences['Dance Studio'] = 10
user_preferences['Harbor / Marina'] = 10
total =  np.sum(list(user_preferences.values() ) )
for key in user_preferences.keys():
    user_preferences[key] = user_preferences[key]/float(total)
```

```
user_preferences
```

```
[7]: {'Rental Car Location': 0.10294117647058823,
 'Italian Restaurant': 0.14705882352941177,
 'Sports Bar': 0.04411764705882353,
 'Indian Restaurant': 0.11764705882352941,
 'Recording Studio': 0.14705882352941177,
 'Japanese Restaurant': 0.14705882352941177,
 'Dance Studio': 0.14705882352941177,
 'Harbor / Marina': 0.14705882352941177}
```

system recommendations:

Neighborhood
Berczy Park
Grange Park, Chinatown, Kensington Market
The Junction South, High Park
The Beaches West, India Bazaar
The Beaches

Further work

- Pickup venues ratings rather than it’s percentage
- Instead of getting user preferences directly, we can ask the user to rate each neighborhood he already visited around the world and then let the system guess which venue types the user likes most.
- To recommend hotels, not only neighborhoods to stay
- Consider tourist hotspots

Thank you for your time :)