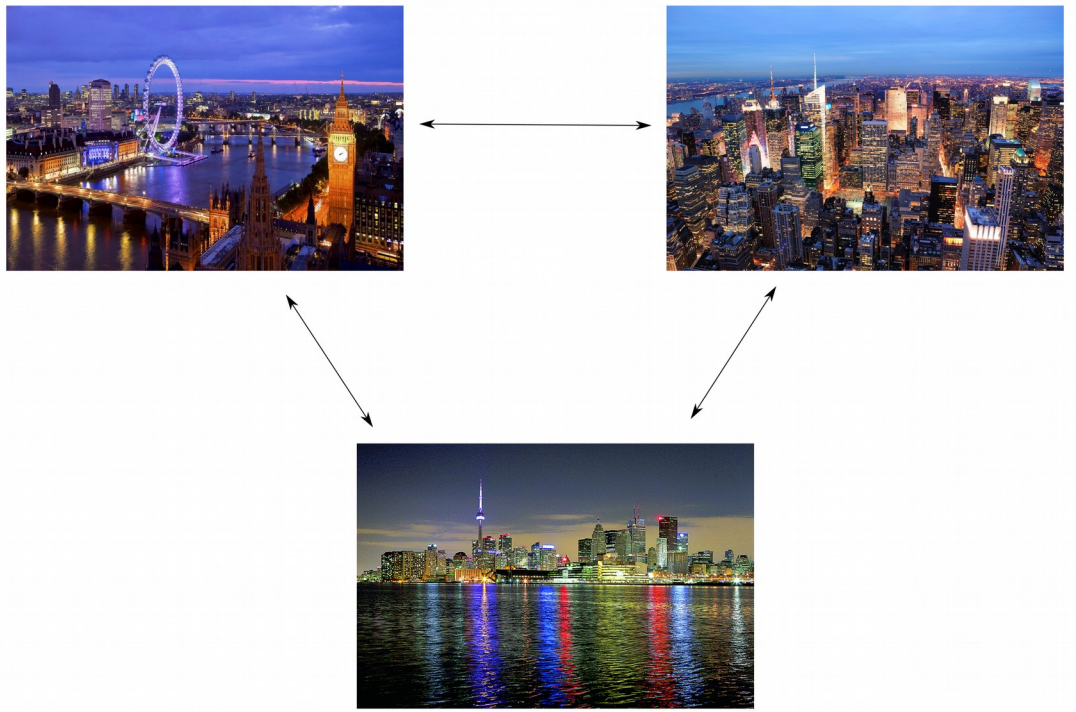
Comparing cities: Toronto, New York and London

Lucas Barbosa Gama

Comparing cities

- Toronto
- New York
- London



Comparing cities
Why is this useful?

- **Recommendation Systems: a way to quantify the similarity of items/users**
- **Travel agency**
- **Real estate agency**
- **Expand a business**

Comparing cities
What data?

Local markets or venues tell a lot about places, they are related to people’s preferences.

Comparing cities
Data

- **Venues from Foursquare**

```
toronto.head()
```

	Venue Category	Counts
70	Coffee Shop	257
56	Café	161
218	Park	134
226	Pizza Place	115
242	Restaurant	88

```
new_york.head()
```

	Venue Category	Counts
276	Pizza Place	431
194	Italian Restaurant	253
104	Deli / Bodega	238
79	Coffee Shop	230
73	Chinese Restaurant	213

```
london.head()
```

	Venue Category	Counts
290	Pub	884
84	Coffee Shop	779
64	Café	685
165	Grocery Store	456
182	Hotel	384

Comparing cities
How to compare dataframes?

- **Jaccard Index for venue categories wouldn’t take into account quantities, which is important to describe people’s preferences...**

  ...but it adjustable.

Comparing cities
How to compare dataframes?

```
def jaccard_weighted(city1, city2):
    set1 = set()
    set2 = set()
    sum_in = 0
    total1 = city1['Counts'].sum(axis=0)
    total2 = city2['Counts'].sum(axis=0)

    for i in range(0,city1.shape[0]):
        set1.add(city1['Venue Category'][i])
    for i in range(0,city2.shape[0]):
        set2.add(city2['Venue Category'][i])
        if city2['Venue Category'][i] in set1.intersection(set2):
            sum_in = sum_in + min(int(city1.loc[city1['Venue Category'] == \
                                                city2['Venue Category'][i]]['Counts'])*100/total1,\
                                  city2['Counts'][i]*100/total2)

    return(sum_in)
```

Comparing cities Results

- jw for NY and Toronto = 66.59696845303091
- jw for NY and London = 57.397012450538696
- jw for London and Toronto = 67.07940505333352

- This indicates that people's preferences in New York are more similar to people's preferences in Toronto than in London.

Comparing cities
Discussion

- It uses only venues, depending on the situation, other data would be crucial.
- Improve model with weighted averages
- Some categories could be put together, depending on the goals.
- Split categories in subcategories like: “bread”, “cakes” in category “bakery”.

Comparing cities
Conclusion

- **It was possible to see that that New York is more similar to Toronto than London, for example.**
- **It describes a similar taste of people, which is a good indicator to someone who wants to expand a business.**
- **The model is useful for recommendation systems for a travel agency without enough user data.**

Thank you!