Predict where to open a venues in New York

Predict where to open a venues in New York

- New york is one of the world's biggest city with millisons of people live in it. People need different business of venues in the city. For a venue, you have to pick the right place to open the business in order to be successful. It is valuable to predict where is appropiate to open a specific business.
- Data that might contribute to determining where to open a business might include the location, existing other venues, the number of other venues, and the rating of them. This project aims to predict where a neighorhood is good for a certain venue like restaurant.
- It is clear that those who want to open any kind of venue would be interested in where is good to open their business.

Data acquisition and cleaning

- Data downloaded or scraped from both the webpage.
- We don't have the location information from the data, so we need to look up the latitude and longitude from google programatically to fill in the table.
- We use the foursquare to look up top 100 popular venues in one neighborhood and then one-hot-encode them.
- There is still some categories that is so small in all neighborhood. Which I believe they are outliers, I'll just remove them.

Popularity vs Location

Popularity vs Similarity

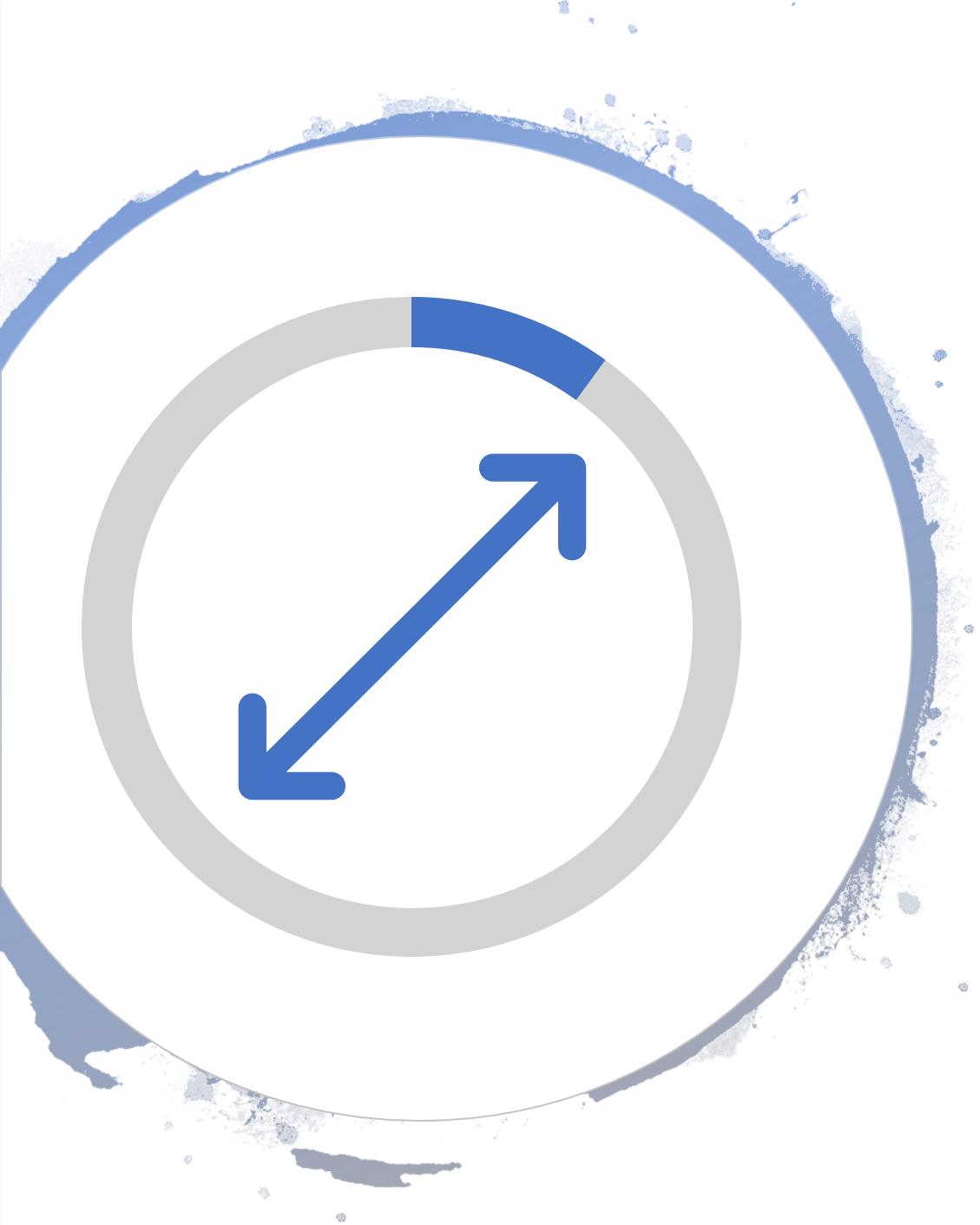
	closeness
Neighborhood	
Bayswater	0.412369
Brownsville	1.000000
Castle Hill	0.350742
Claremont Village	0.343169
Clason Point	0.402737
Concourse	0.353356
East Flatbush	0.382339
Flatlands	0.333383
Fulton Ferry	0.391748
High Bridge	0.418598
Kingsbridge Heights	0.357159
Mount Eden	0.334083
New Lots	0.445068
Norwood	0.462878
Randall Manor	0.396686
Riverdale	0.345237
Rossville	0.353267
Somerville	0.484640
Todt Hill	0.484640
Woodlawn	0.337997

Popularity vs Unsupervised Cluster Label

	Cluster Label
Neighborhood	
Brownsville	1
Midland Beach	1
Annadale	13
Cambria Heights	8
New Dorp Beach	7
Heartland Village	1
Hunts Point	13
Madison	7
Brighton Beach	1
Port Morris	1
Stapleton	13
Claremont Village	7
Old Town	13
Corona	13
Ocean Parkway	1
Co-op City	0
Blissville	1
Eastchester	8
Borough Park	13
Rego Park	13

Predictive Model using Pearson Similarity

- predict rate of 'Office': 8 out of 40 which is 20%
- predict rate of 'Pharmacy': 4 out of 40 which is 10%
- Predict rate of 'Restaurant': 6 out of 40 which is 15%



Conclusion and future directions

- In this study, I analyzed the relationship between popularity of certain venue in a neighborhood and their location, unsupervised cluster label, similarity other than the venue. I noticed that none of them can clearly predict if a venue has already been popular in a neighborhood. But similarity model in recommendation system has the best potential. So I build the predictive model based on similarity. The model's predictive ability is aound 10% - 20%. However, it doesn't mean the model is not useful. Maybe it has uncovered potential good places for the future.
- To further understand if the model is useful or not, I need to further collect more data over years so as to understand if a place would become popular so as to either prove the existing model still has potential to uncover potentials people cannot easily understand. Or create regression models using machine learning so as to predict if a venue would become popular in the future based on its current status.
- More features might be needed like if certain venues' rating is high or low might also need consideration.