



# THE BATTLE OF NEIGHBORHOODS

COURSERA CAPSTONE PROJECT



# Introduction



- Well in each city in the world has a lot of Neighborhood, and each one has its structure based on their avenues. Having a neighborhood doesn't mean that you can know it's city, and this is the major question we're going to ask. Can we predict the city of a neighborhood based on its avenues? Our study will be based on 3 cities: New York, Toronto and Paris.
- To understand globalization and the impact of the cultural exchange in this world we need to understand the structure of a city and its neighborhoods and how avenues sometimes new ones may impact the way of living.
- For people who are studying cities and their avenues and want to understand the relationship between them, for people who are interested to know to which city their neighborhood would fit. and how can we make an avenue looks like another to one based on their avenues?



# Data

- Data used in this project is exported from various sources:
- New York : [https://geo.nyu.edu/catalog/nyu\\_2451\\_34572](https://geo.nyu.edu/catalog/nyu_2451_34572)
- Toronto : [https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)
- Paris  
: [https://opendata.paris.fr/explore/dataset/quartier\\_paris/download/?format=json&timezone=Africa/Lagos&lang=fr](https://opendata.paris.fr/explore/dataset/quartier_paris/download/?format=json&timezone=Africa/Lagos&lang=fr)
- We will use the Foursquare API to explore neighborhoods in New York, Toronto and Paris. We will use the explore function to get the most common venue categories in each neighborhood



# Extracting essential columns

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

# Introducing Venues on data

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Regent Park , Harbourfront	43.65426	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
1	Regent Park , Harbourfront	43.65426	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
2	Regent Park , Harbourfront	43.65426	-79.360636	Morning Glory Cafe	43.653947	-79.361149	Breakfast Spot
3	Regent Park , Harbourfront	43.65426	-79.360636	Cooper Koo Family YMCA	43.653249	-79.358008	Distribution Center
4	Regent Park , Harbourfront	43.65426	-79.360636	Body Blitz Spa East	43.654735	-79.359874	Spa

# Transforming Venue Category to dummy variables

	Neighborhood	Accessories Store	Afghan Restaurant	African Restaurant	Alsatian Restaurant	American Restaurant	Antique Shop	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Arts & Entertainment	R
0	Amérique	0.0	0.0	0.00000	0.0	0.00	0.0	0.0	0.00	0.00	0.00	0.000000	0.0	
1	Archives	0.0	0.0	0.00000	0.0	0.00	0.0	0.0	0.00	0.04	0.01	0.000000	0.0	
2	Arsenal	0.0	0.0	0.00000	0.0	0.00	0.0	0.0	0.00	0.00	0.00	0.000000	0.0	
3	Arts-et-Métiers	0.0	0.0	0.00000	0.0	0.00	0.0	0.0	0.01	0.02	0.00	0.000000	0.0	
4	Auteuil	0.0	0.0	0.00000	0.0	0.00	0.0	0.0	0.00	0.00	0.00	0.000000	0.0	
5	Batignolles	0.0	0.0	0.00000	0.0	0.01	0.0	0.0	0.00	0.00	0.00	0.000000	0.0	
6	Bel-Air	0.0	0.0	0.00000	0.0	0.00	0.0	0.0	0.00	0.00	0.00	0.000000	0.0	
7	Belleville	0.0	0.0	0.02439	0.0	0.00	0.0	0.0	0.00	0.00	0.00	0.000000	0.0	
8	Bercy	0.0	0.0	0.00000	0.0	0.00	0.0	0.0	0.00	0.00	0.00	0.013333	0.0	
9	Bonne-Nouvelle	0.0	0.0	0.00000	0.0	0.00	0.0	0.0	0.00	0.00	0.00	0.000000	0.0	



# Getting the most common Venue on each neighborhood

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Amérique	French Restaurant	Supermarket	Café	Bakery	Plaza	Park	Bed & Breakfast	Smoke Shop	Bistro	Theater
1	Archives	French Restaurant	Hotel	Coffee Shop	Clothing Store	Italian Restaurant	Bar	Art Gallery	Japanese Restaurant	Cocktail Bar	Tapas Restaurant
2	Arsenal	French Restaurant	Hotel	Plaza	Park	Tapas Restaurant	Boat or Ferry	Italian Restaurant	Cocktail Bar	Gastropub	Pedestrian Plaza
3	Arts-et-Métiers	French Restaurant	Hotel	Cocktail Bar	Wine Bar	Italian Restaurant	Chinese Restaurant	Restaurant	Moroccan Restaurant	Bar	Japanese Restaurant
4	Auteuil	Tennis Court	Museum	Garden	Botanical Garden	Bike Rental / Bike Share	Racecourse	French Restaurant	Stadium	Sporting Goods Shop	Outdoors & Recreation

That will allow us to understand more our data and what's common venue in different neighborhood

# Creating Final Dataset

After understanding the data in the 3 datasets we will create our final data set which contains data with dummy variables with the label 'City'

	Neighborhood	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	Airport	Airport Food Court	Airport Lounge	Airport Service	Airport Terminal	Alsatian Restaurant	American Restaurant	Animal Shelter	Antique Shop	Aquari
0	Amérique	0.0	0	0.0	0.0	0	0	0	0	0	0.0	0.0	0	0.0	
1	Archives	0.0	0	0.0	0.0	0	0	0	0	0	0.0	0.0	0	0.0	
2	Arsenal	0.0	0	0.0	0.0	0	0	0	0	0	0.0	0.0	0	0.0	
3	Arts-et-Métiers	0.0	0	0.0	0.0	0	0	0	0	0	0.0	0.0	0	0.0	
4	Auteuil	0.0	0	0.0	0.0	0	0	0	0	0	0.0	0.0	0	0.0	





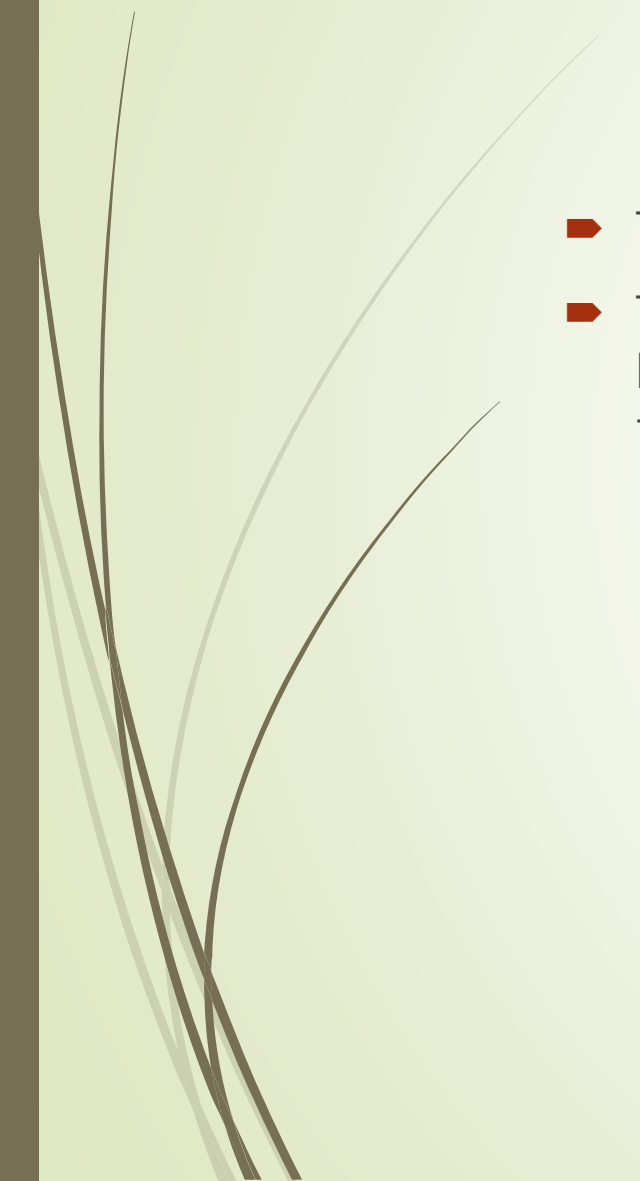
# Methodology



- To get for each neighborhood we will use 2 algorithms:
- Logistic regression.
- Decision Tree
- We will use them both to compare our prediction and also to understand with the decision tree why a neighborhood is most likely to belong to a city.

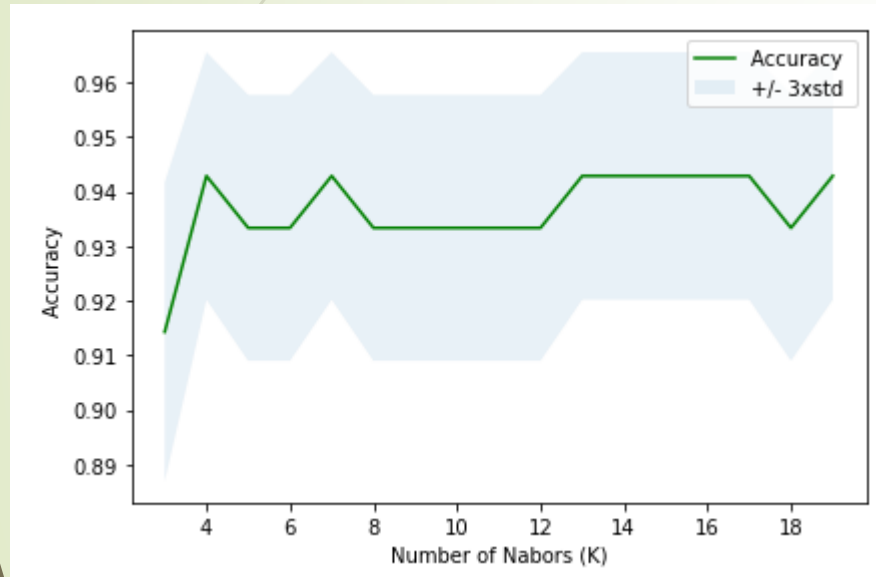


# Train Test Split

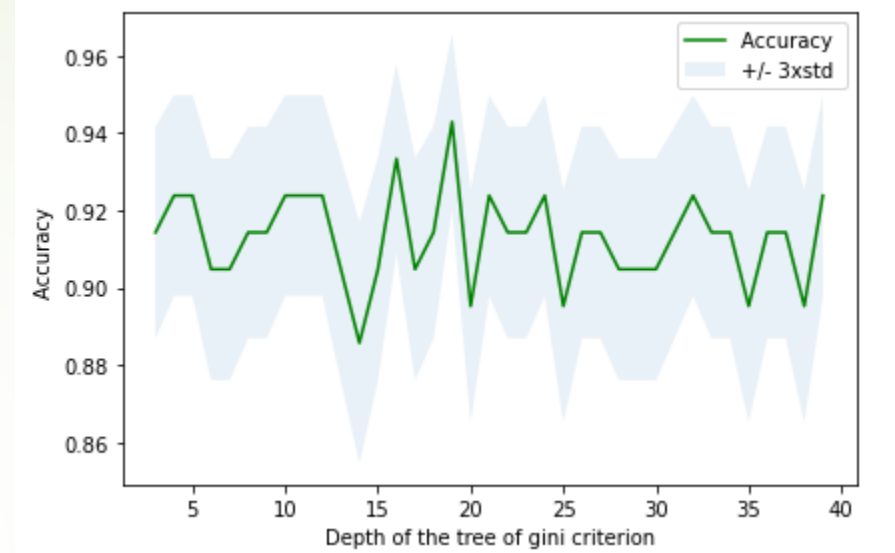
- To do so first we will do the Train Test Split.
  - This will provide a more accurate evaluation on out-of-sample accuracy because the testing dataset is not part of the dataset that have been used to train the data. It is more realistic for real world problems.
- 

# Tuning our prediction

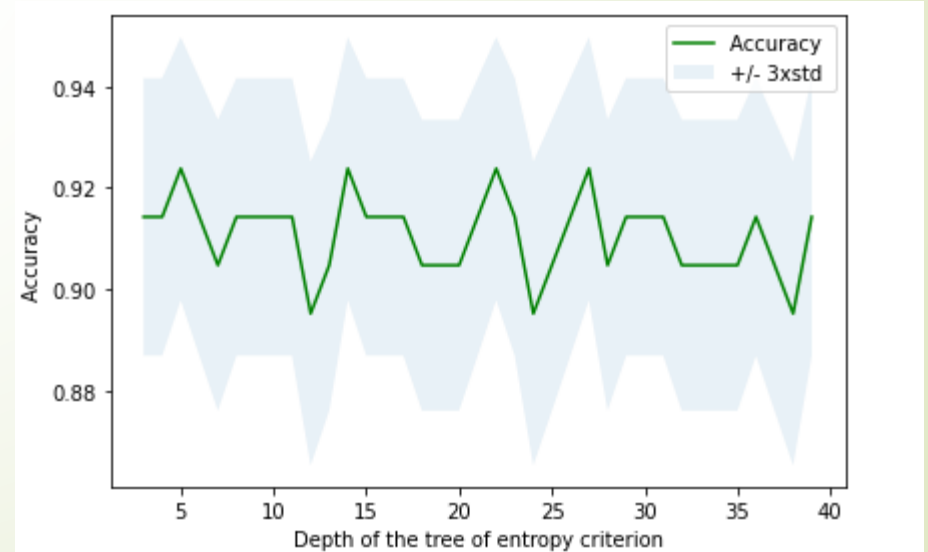
## ➤ KNN



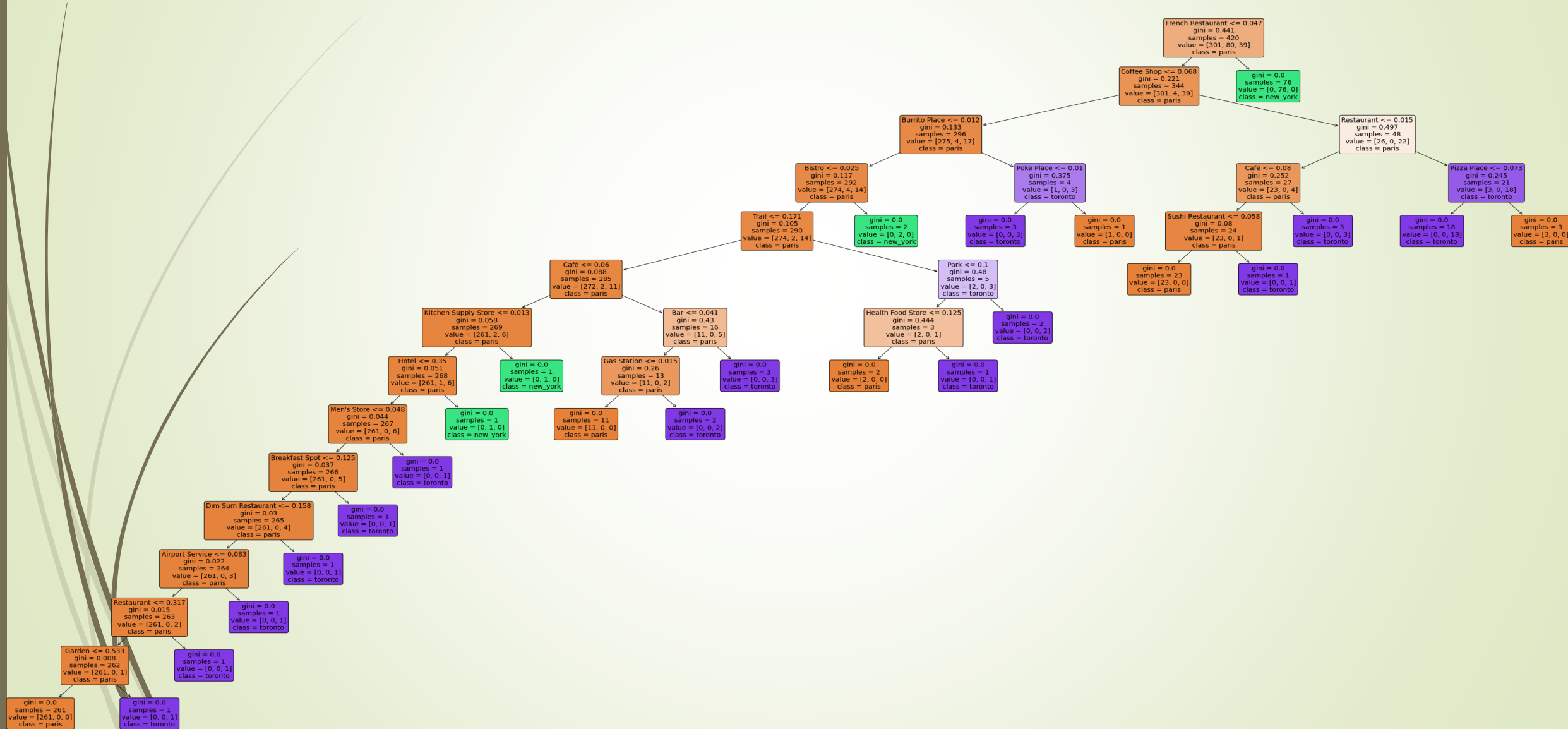
## ➤ Decision Tree gini Criterion



## ➤ Decision Tree gini Criterion



# Decision Tree Result





# Conclusion and Discussion



- This method can be used for different goals:
  - we can predict the city based on the avenues
  - the error of prediction can give us idea about how similar an avenue to others from in another city
  - From this Tree we can understand more how we can make a neighborhood similar to one in a country. If this method used with many cities we can go further and why not make a city that have neighborhood that would make you feel like if you're living in another one.
- If we add more cities we will have more understanding of how cities are constructed.