**Introduction/Business Problem**

On Monday morning, the guy at the travel agency told me “Computer says, uh!” A fancy statement actually meaning “I don’t know what you are talking about, but I can offer you a 3 weeks tour as a good start. Good damn it! well without knowing it, that’s where my AI journey begins… Sometimes in life, opportunities pop-up in front of you and you face you yourself with a “should I go ahead and take it” moment. Indeed, starting a new life in Manhattan is by far quite a new story for a “didn’t know there were other cities than Paris on earth” type of guy. Manhattan is one of the 5 boroughs of New York City with over 1,6 M of inhabitants, a bit smaller than Paris “Big Picture” which stands at more than 2,1M then – I knew I would have my revenge, but represents only 1/5 of NYC- OK I give up…

Paris on the other hand is split into 20 boroughs and I do love the charming environment of the one I live in (12th but don’t tell anyone) and as my new life in NYC was about to start I was looking for someone who can help me finding a little Paris in Manhattan as I am about to stay for a while… well you know how it ended, and for sure, I haven’t met the future Dr Spence Reid (Criminal Minds). Therefore, I quickly realize that it was a more DIY job posing the question of finding a place likewise mine in that other environment or in other words how I can find a neighborhood in Manhattan that is similar to mine. Furthermore this could be also expanded to neighborhood mapping helping other Paris folks finding their dream land in Manhattan - So welcome to manathan.predict(myParis) journey.

**Data Description**

Because the goal is to find similarities, we should define what similarity is and it should be understood as areas sharing characteristics considered as being equivalent. For that purpose, data will be used as measurement units. In this context data required can be divided into 2 categories: Geo localization data and characteristics data of the concerned areas.

Geo localization data: (neighborhoods, latitude, longitude)

* Regarding NYC and particularly Manhattan, data file from course 3 is readily available via <https://cocl.us/new_york_dataset> .
* Concerning Paris, geo localization data is available via an xls files from <https://opendata.paris.fr/explore/embed/dataset/arrondissements/table/> .

Data regarding characteristics is available via Foursquare. In each area characteristics of top spots can be retrieved.

| **New York City's five boroughs** | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Jurisdiction** | | **Population** | **Gross Domestic Product** | | **Land area** | | **Density** | |
| *Borough* | *County* | *Estimate (2018)*[[3]](https://en.wikipedia.org/wiki/Boroughs_of_New_York_City#cite_note-3) | *billions (US$)*[[4]](https://en.wikipedia.org/wiki/Boroughs_of_New_York_City#cite_note-4) | *per capita (US$)* | *square miles* | *square km* | *persons / sq. mi* | *persons / km2* |
| [**The Bronx**](https://en.wikipedia.org/wiki/The_Bronx) | Bronx | 1,432,132 | 42.695 | 29,200 | 42.10 | 109.04 | 34,653 | 13,231 |
| [**Brooklyn**](https://en.wikipedia.org/wiki/Brooklyn) | Kings | 2,582,830 | 91.559 | 34,600 | 70.82 | 183.42 | 37,137 | 14,649 |
| [**Manhattan**](https://en.wikipedia.org/wiki/Manhattan) | New York | 1,628,701 | 600.244 | 360,900 | 22.83 | 59.13 | 72,033 | 27,826 |
| [**Queens**](https://en.wikipedia.org/wiki/Queens) | Queens | 2,278,906 | 93.310 | 39,600 | 108.53 | 281.09 | 21,460 | 8,354 |
| [**Staten Island**](https://en.wikipedia.org/wiki/Staten_Island) | Richmond | 476,179 | 14.514 | 30,300 | 58.37 | 151.18 | 8,112 | 3,132 |
| [**City of New York**](https://en.wikipedia.org/wiki/New_York_City) | | **8,398,748** | **842.343** | **97,700** | **302.64** | **783.83** | **28,188** | **10,947** |
| [State of New York](https://en.wikipedia.org/wiki/New_York_(state)) | | 19,745,289 | 1,701.399 | 85,700 | 47,214 | 122,284 | 416.4 | 159 |
| *Sources:*[[5]](https://en.wikipedia.org/wiki/Boroughs_of_New_York_City#cite_note-5)*and see individual borough articles* | | | | | | | | |