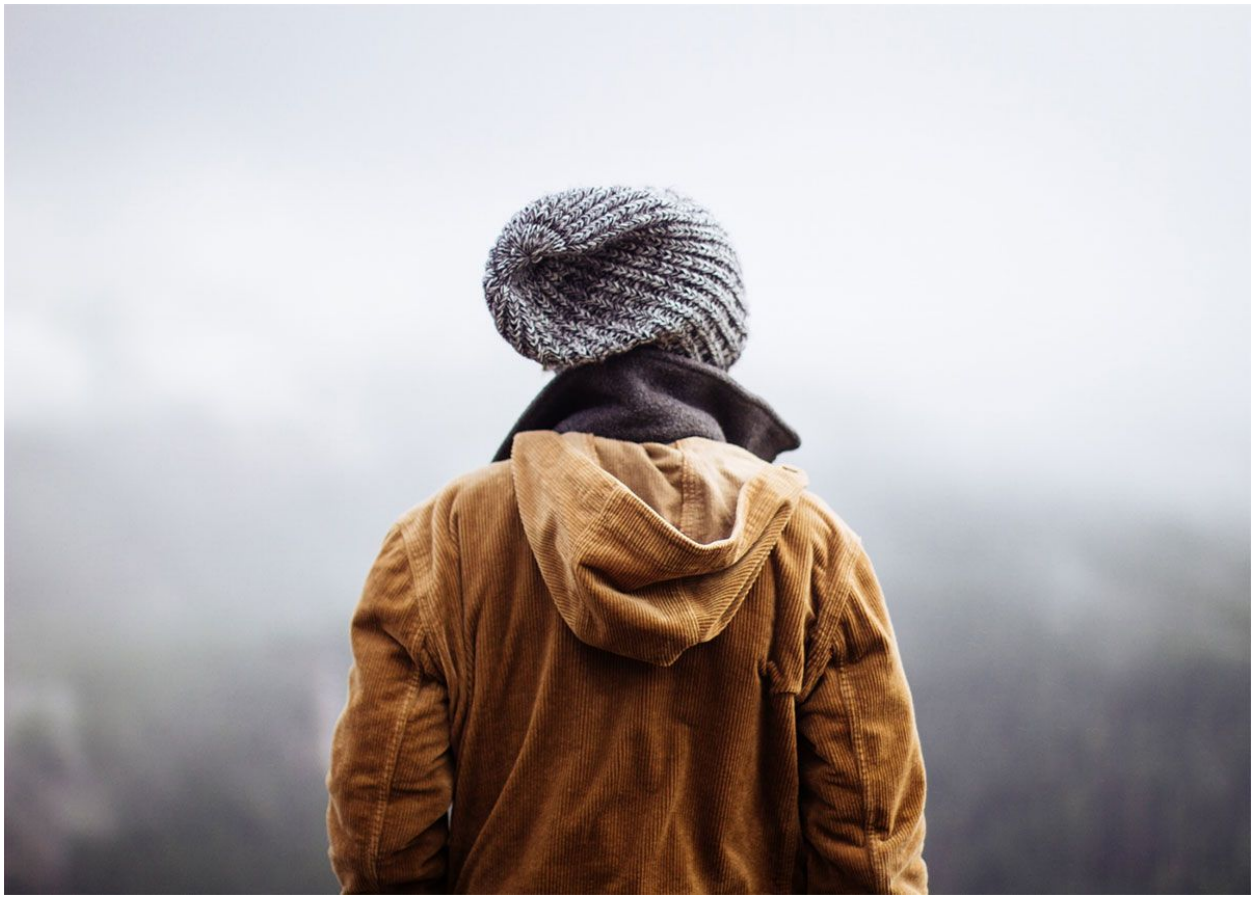
IBM DATA SCIENCE CAPSTONE

Supercities: London and New York
Melting Pot or Unique Diversity?

Juan Trujillo



Introduction

London and New York tend to be described as two of the most cosmopolitan cities in the world.[1] They also compete for the title of world's financial hub,[2] and are often high on the polls around the best place to live.[3]

Both have a very similar population (ca. 8.5 million inhabitants) and similar surfaces (London has 1572 square kilometers vs New York 1214 square kilometers).

We can certainly assert that they are two of the most vibrant and thriving urbs. They share a common official language, but how similar are they if we look at the granular distribution of amenities, shops, bars and other elements of their urban orography? How similar are their neighbourhoods if we would combine them into a single super-city? Are their city areas and their distribution of services and venues homogeneous to a point where they can be traceable in their counterpart?

This study should also propose an empirical approach to a theory of gentrification that tends to see a continuous trend towards the amalgamation and blending of highly developed cities into 'melting pots' that compose diversity into homogeneity.[4]

[1] https://www.worldatlas.com/articles/the-most-cosmopolitan-cities-in-the-world.html
[2] https://uk.reuters.com/article/uk-britain-banks/brexit-helps-new-york-take-top-finance-spot-from-london-survey-idUKKCN1SY08L
[3] https://www.businessinsider.com/new-york-city-vs-london-comparison-2015-5?r=US&IR=T
https://www.bloomberg.com/opinion/articles/2019-06-04/why-london-is-better-than-new-york
[4] https://en.wikipedia.org/wiki/Melting_pot

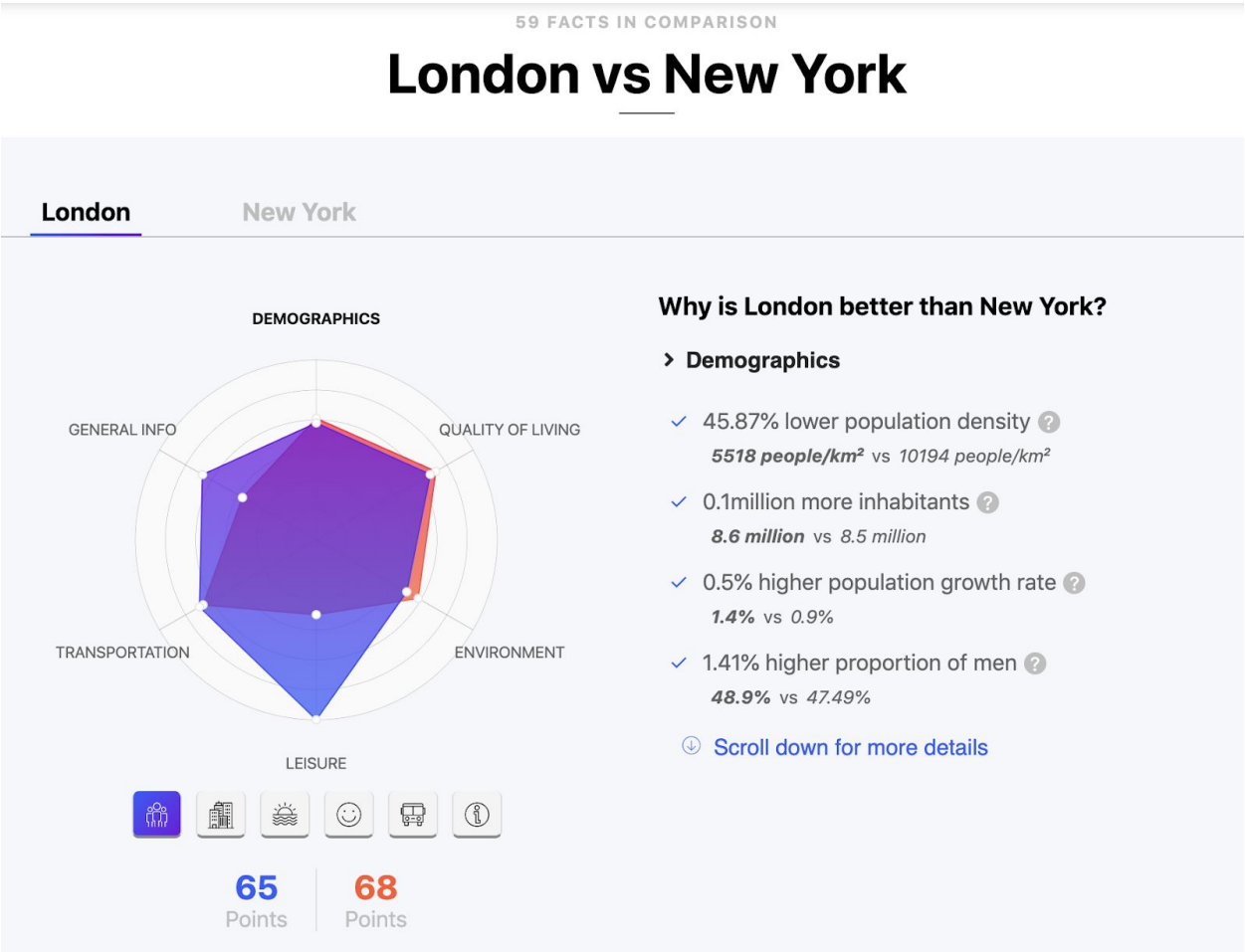


Fig.1 Comparison of both cities across 6 dimensions (Demographics, Quality of Living, Environment, Leisure, Transportation, General Info) https://versus.com/en/london-vs-new-york

Problem Description

As part of this project, we plan to combine data obtained from the Foursquare Places API surveying the places and offerings in the different neighbourhoods in New York and London as if they were a single super-city.

We will obtain information about all the neighbourhoods in New York and all tube stations in London in order to group the similarities across the Atlantic between both cities.

We want to identify the commonalities and singularities in terms of activities that these cities offer in each one of their neighbo(u)rhoods. We want to understand if there are neighbourhood traits that are unique to one of them or there are commonalities in certain areas. Is the Soho in London similar in its places to its eponymous in New York?

Can you obtain the same services in Chelsea and the Upper East Side? Are hip cafés more pervasive in Shoreditch or in the Meatpacking District?

A friend of mine is planning to move to New York and has her eyes in moving to Battery Park. She is currently living in Lewisham in London. Is there much resemblance in the physiognomy of these two areas? What are the types of places she will be finding more often in this new neighbourhood? Are there similarities in the predominant cuisines in both areas?

Data Description

In the case of New York, we have a well-documented list of neighbourhoods grouped in community boards.[5] We will utilise the data provided by Cognitive Class[6] that already includes geo-location information to determine the data points that will be used to cluster the information.

London is composed of 32 wards, however, these are large geographical areas that can span over a hundred square kilometres (e.g. Bromley or Hillingdon). To be more granular we decided to use the pervasive and entrenched nominal usage of tube stations in London. With over 300 stations, Londoners tend to refer to their closest station when describing their neighbourhood. Transport For London (TFL) provides a data set containing information about all their stations including their geolocation. The information is provided in the form of kml and xds files.[7] It is true that this choice brings some bias to the study since there are parts of London that are not well connected to the tube, particularly the south. On the other hand, this approach will offer additional granularity and will cover the

[5] https://en.wikipedia.org/wiki/Neighborhoods_in_New_York_City
[6] https://cocl.us/new_york_dataset
[7] https://api-portal.tfl.gov.uk/docs

most densely populated and representative areas in the nowadays old Roman settlement of Londinium.[8]

We will use the Foursquare Places API[9] to describe the sites and venues that surround these neighbourhoods and we will use them to cluster the neighbourhoods across both cities as if we were handling a contiguous territory. We will limit the number of venues due to limitations in the API to 200 within 1,000 meters of each neighbourhood based on its geolocation.

With these three sets of data (COCL New York neighbourhoods, TFL London tube stations, and Foursquare Places API) we will review the most common venues in a cluster of this super-city to review how much similarity there is across the neighbourhoods of both cities.

[8] https://en.wikipedia.org/wiki/Londinium
[9] https://developer.foursquare.com/places-api