IBM Data Science Professional Certificate
Capstone Project

Improving quality of living in cities

A comparative study of Singapore,
New York, Manila and Mexico City
focusing on transportation and recreation

Jhoanna Chan-Punzalan
14 August 2020

Introduction / Business Problem

Due to the rise of globalization, most of the world's population – as many as two out of three people by 2050[1] - are going to reside in cities. In order to sustain a healthy population and economic growth, these urban areas must adapt to maintain their attractiveness as a place to live, work and visit.

According to the United Nations, many governments will "face challenges in meeting the needs of their growing urban populations, including for housing, transportation, energy systems and other infrastructure, as well as for employment and basic services such as education and health care." Providing a good quality of life for residents is therefore an urgent need for the world's cities.

According to Mercer, a consultancy firm which publishes the annual Quality of Living Ranking and works with city governments and planners on assessing factors to improve their rankings, it is "more necessary than ever for city leaders to have a clear understanding of how well they are doing to realize the goal of being competitive economically, socially, culturally and environmentally, for now and for the future."[2]

Ranking 231 cities on 39 factors, the Mercer Quality of Living Ranking provides recommendations to companies on where to deploy mobile workers and set up new offices. The index shows that the best cities understand that the quality of living is an essential component of a city's attractiveness for businesses and mobile talent. In the 2019 survey, all US cities covered in the analysis fell in the rankings with the exception was New York (ranked 44), rising one place. In Asia, Singapore (ranked 25) has the highest quality of living followed by Tokyo. [3]

Two of the key performance metrics used by Mercer to analyze living conditions in a city are:

- Public services and transportation (electricity, water, public transportation, traffic congestion)
- Recreation (restaurants, theatres, cinemas, sports and leisure)

Targeting decisionmakers and stakeholders such as urban planners and city governments, I propose to use Location Based Social Networks (LBSN) such as Foursquare to determine the developments needed for a top-tier, future-oriented city with a fast-growing population. Due to the wide scope of factors which determine the attractiveness of a city, I shall limit my review to public transportation and recreation. The objective is to provide city planners and governments evidence-based guidelines so they can make informed decisions on how they should invest in and develop their cities based on forecasted population sizes and density. This study may also support them in locating sites – such as a new hospital or park - for future development.

I will use the industrialized cities of Singapore (ranked 25) and New York City (ranked 44) as benchmarks due to their high ranking. The developing cities of Manila (ranked 137) and Mexico City (ranked 129) shall be used to compare and to find out how similar and dissimilar the higher-ranked cities and lower-ranked cities.

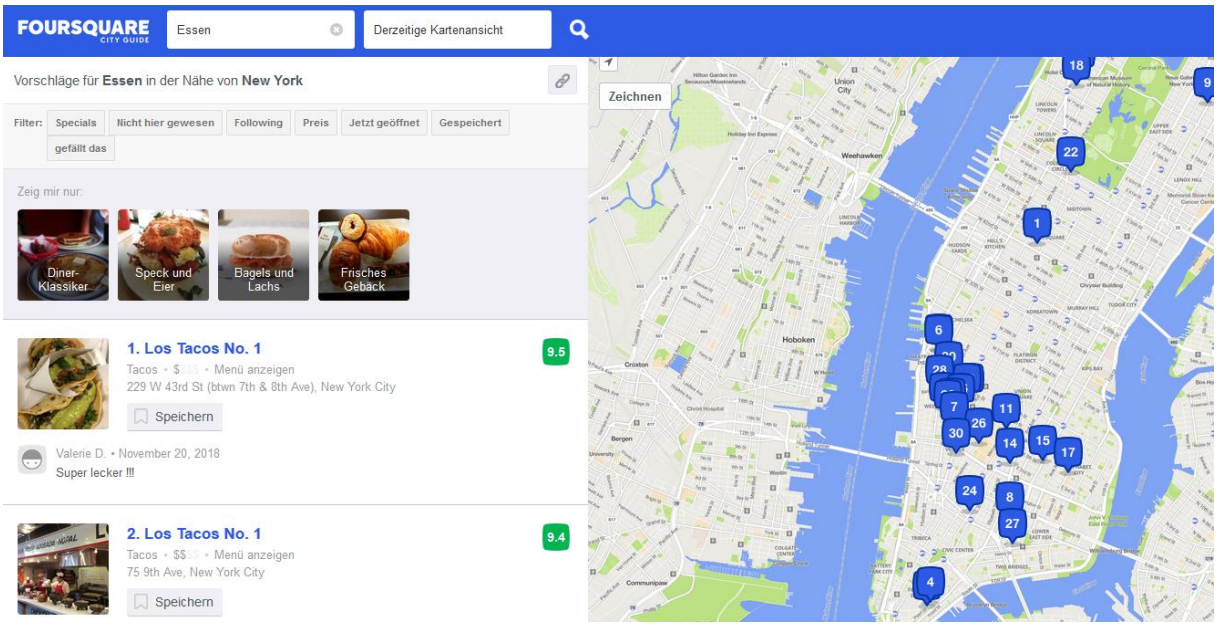
[1] "68% of the world population projected to live in urban areas by 2050, says UN"
16 May 2018, New York https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html
[2] "City Attractiveness Index", Mercer https://www.mercer.com/what-we-do/workforce-and-careers/city-attractiveness-index.html
[3] "Vienna Tops Mercer's 21st Quality of Living Ranking", 13 March 2019
https://www.mercer.com/newsroom/2019-quality-of-living-survey.html

Data

To work on this comparative study, I shall utilize data from Foursquare to get information about venues in Singapore, New York, Manila and Mexico City.

I shall leverage on Foursquare's location data[4] on the two key indicators, public transportation and recreation, to determine the presence and availability level of venue categories such as:

- Train stations
- Bus stations
- Restaurants and cafes
- Theaters
- Concert halls
- Cinemas
- Shopping malls
- Supermarkets
- Sports stadiums
- Parks
- Gyms
- Bike trails



Extract from a Foursquare search for restaurants in New York City. Foursquare.com.

[4] "Venue Categories", Foursquare https://developer.foursquare.com/docs/build-with-foursquare/categories/

These will be supplemented by data taken from Wikipedia. For example, the information on population sizes and area of each city will support in normalizing the data during analysis.

Largest cities

There are 81 cities with a population over 5,000,000 people according to the United Nations 2018 estimates.[13]

City[a]	Country	Skyline	UN 2018 population estimates (metro area) [13]	City proper[b]			Metropolitan area[c]		Urban area (Demographia)[12]	
				Definition	Population	Area (km²)	Population	Area (km²)	Population	Area (km²)
Tokyo	Japan		37,400,068	Metropolis prefecture	13,515,271[14]	2,191[14]	37,274,000[15]	13,452[15]	38,505,000	8,223[d]
Delhi	India		28,514,000	National capital territory	16,753,235[16]	1,484	29,000,000[17]	3,483[17]	28,125,000	2,240[e]
Shanghai	China		25,582,000	Municipality	24,183,000[18]	6,341	N/A	N/A	22,125,000	4,015[f]
São Paulo	Brazil		21,650,000	Municipality	12,252,023[19]	1,521	21,734,682[20]	7,947	20,935,000	3,043[g]
Mexico City	Mexico		21,581,000	City-state	8,918,653[21]	1,485	20,892,724[22]	7,854	20,395,000	2,370

"List of largest cities", Wikipedia https://en.wikipedia.org/wiki/List_of_largest_cities

This is a project that will make use of data science skills: from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to statistical models (Multiple Linear Regression Model (MLRM)) and map visualization (Folium).