The Battle of Neighborhood

FY 2020

“Smart Cities used to be about technology and governance, but future cities pays greater attention on citizen. Thus Smart city needs a new definition and metric to measure Citizen Centric Elements”

Azman Ali

Tel [Telephone]
Fax [Fax]

Kuala Lumpur
Malaysia

[Website]
[Email]

Contents

“ Location and Spatial Analytic are keys in addressing many challenges in managing urban infrastructure and community”

Executive Summary

The notion of Smart City has evolved from being technology centric to more citizen centric. That means, our perception on the places has changed and evolved from the traditional notion of smart infrastructure to the like of ‘Livable City” ,”Happy City” or “Green City” instead. While City infrastructure will always be one of the indicator for city posture, in the end, Citizen has to be significantly measured. This reports aims to provide a framework for measuring and comparing scores between neighborhoods for to assist stakeholders in making strategic decision or action.

The Importance of Measure-ability

One could not improve on what one could not measure. To date, city governors are adopting different metric for measuring city performance.

Feature Highlights

This report will be based on our own assumption on features that in our opinion representing some basic parameters to measure score of city performance. These features are discussed in greater detail in the Data section.

Who will benefit

City managers, governors or stakeholder may adopt the approach build a more sustainable city dashboard for their own users or publish public.

At citizen level, the public or tax payers are expecting more transparent government and information that involve public interest

Business Users can also use the framework to make assisted decision in either Business Planning or operation. Businesses can understand more about the neighborhood as the target market and optimize their resources and business offering.

Azman Ali

Senior CONSULTANT

The Report Structure

- **Introduction** where we discuss the problem statements, hypothetical solutions, objectives and who would be interested in this project.
- **Data** where we describe the data that will be used to solve the problem, their sources and particular parameters selected..
- **Methodology** section which represents the main component of the report where we discuss and describe any algorithm, workflow, exploratory data analysis that we did, statistical testing performed, if any, and what machine learnings that were used and why.
- **Results** section where we discuss the results and draw any business values
- **Discussion** section where we discuss any observations made and any recommendations based on the results.
- **Conclusion** section where we conclude the report

Introduction

Background

Future governance of city infrastructure and it's community/neighborhood demands for structured approach to assist in decision making. As such, the ability to measure how a city or community is progressing as compared to others will be handful. Managing cities and their community are complex job and thus requires integrated approach that measures various aspects. On the other hand, with the advent of digitalization, infrastructure operation and citizen activities can now generate more data in real time that can be useful, which can contribute towards new insights and knowledge.

Today, Citizen's awareness are increasing more than ever and city governance are expected to be more transparent especially in the matter involving public interest. As such any analysis and measurement of city health posture will be useful. In the lights of outbreak such as COVID-19 for example, information are sought (by the public) at a rate higher than ever as people need to know how their city/neighborhood is coping with the situation. This project aims to conduct a comparative study on neighborhoods within a selected city (New York) based on integrated location data and demography. While doing so, we believed that analytical framework developed for this study would be applicable to other cities or further developed into a proper data product.

Statement of Problems

- Understanding analyzing location and or spatial data are key to urban planning and to address many different problem within community and neighborhood. The perspective of any issues on the hand are based on unique stakeholders' interest. Here the stakeholder can either be the institution (the government), the citizen (public), or businesses.
- A city governors might want an overall view of all neighborhood to strategically plan any future development or keeping safe and vibrant environment. As a citizen, the concern is over information that is of public interest such as education, economic activity, property investment etc. Businesses are more concerned over making investment, increase productivity, generating revenue and keeping tab on the market segments and competition.
- The overall posture or profiling metric for city and its neighborhood is lacking and often measured in silo, prompting for the needs of structured analysis, tools or framework.

Introduction

Research Questions

- Can we profile a city or neighborhood based on set of preferred or personalized criteria? Ie based on different stakeholders interest?
- Can we establish a metric or structured analytic framework that will allow us to measure current state vs future state of a city (or against other neighborhood)?
- Can we analyze (using the framework), city infrastructure readiness taking into account various aspect needs (ie health, education) and demographic study?

The Solution Hypothesis

- A metric or scheme to measure and profiling a neighborhood would be useful for stakeholders. The measurement will help establishing current state, identify gap and compared with other neighborhood.
- The metric is aimed to be comprehensive by looking into multiple integrated criteria. As a very basic requirement, it should measure readiness of relevant infrastructure that reflects the needs of citizen including economy, health, education, social/recreation as well as other demographic profile.
- Other than report, the outcome can be in the form of live dashboard or tools specializing in performing exploratory or comparative analysis between different city based on integrated and customized metric..

Who Will Benefit

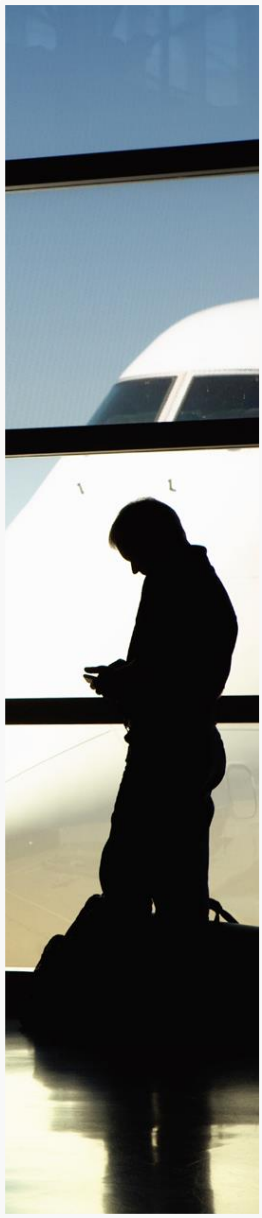
- City managers, governors or stakeholder may use the framework to start profiling their cities and neighborhood as part their planning exercise. They can explore to build a more sustainable city dashboard for their own use or published for public engagement (open data platform)
- At citizen level, the public or tax payers are expecting more transparent government and information that involve public interest. If used for public consumption, it can help city to increase citizen engagement and participation in city management which promote transparency in governance.
- Business Users can also use the framework to make assisted decision in either Business Planning or operation. Businesses can understand more about the neighborhood such as the target market and optimize their resources and business offering accordingly.

Objective

- Comparative analysis between neighborhood within a City (ie New York , Toronto) on integrated aspect of criteria including healthcare, education, economic, and social.
- Metric or measurement framework that match user input profile with the city based on personalized preference (ie based on recommender system approach)

Introduction

Note: The analysis and metric proposed are based on infrastructure readiness and demographic information which might not be sufficient to address complexity of the community being measured. The main ideas of this project is to establish a basic metric and framework that measure integrated aspect of city/neighborhood.



Location Data and Demographic Data will be used in this project

Data

Data is the most important aspect of this project. However, before any implementation or commencement of study, a good understanding of business case is crucial to ensure we will engage and acquire the right set of data and information. This is to eliminate 'garbage in garbage out' issue at the later stage. While there are many methods of measuring city performance and benchmarking existed, our approach aims to strike balance in the way we evaluate our city/neighborhood. As such we believed that the life elements such as health, social, economy and some demographic aspect would be worth experimenting or studied, especially when comparing places for living or work. Driven by these different needs we can then decide on the data that is needed, their sources as well as analytical and presentation approaches.

Data Requirement & Hypothesis

Data Selection is made based on our Hypothesis of the following critical parameters to be measured grouped by 5 key pillars (we can call this with acronym HEEDS as follows :

c	Indicator	Data Required
[H]ealth	Health Infrastructure	POI {Hospital/Clinic) Count/population
[E]conomy	Business Establishment	POI Business Count (Categories), POI/CENCUS Property Price/Type
[E]ducation	Education Infrastructure	POI {School/University) per population CENCUS Education level
[D]emography	Population, Age, Income Distribution	CENCUS Age/Gender (mean), Income Distribution, Population Size,
[S]ocial/ Recreation	Green infrastructure, Recreation POI	POI Count Green Infrastructure/Coverage, USERS Check in, Tips,

Data Sources and Description

Based on the table above we have then identify the Data Sources as follows

FOURSQUARE is a location data providers that have served business and developers that require location information for their application**.** *Location Data that we needs and stated in the table above such as POI, Categories, Users (check in, tips) are data endpoints that are accessible as free and limited API calls. (Referrence : www.foursquare.com)*

CENSUS DATA: The other set of data such as Demographic Data (Age, Gender, Population, Income, Education Level can be extracted from US Census data that can be obtained openly from cencus.gov website. Due to the static nature of the information we can use the recent report or cencus data downloadable from the official website (www.census.gov)

Scope

For the data analysis, the scope of the study will cover Neighborhood within New York City.

Recommender Systems and Scoring

Recommender Systems will take user input as initial user profile (PROFILE). Neighborhoods understudied will be compared based on HEEDS indicators above and assigned score based on their relative ranking compared to the other neighborhood (Normalized) (we can call it HEEDS score). PROFILE and HEEDS score will be multiplied together to get the FINAL SCORE for the ‘Battle of Neighborhood study”

Methodology

As stated in our objectives, our deliverables will include comparative analysis between neighborhood, employing various statistical and machine learning approaches which will be reported using relevant visualization approaches. Basic data preprocessing and analysis (such as neighborhood ranking) will be the dominant part of this report with some application of basic regression, prediction and recommender systems included.

In general, our approach is aspired by several other established methodology. One of the example is for predicting neighborhood (Target) based on independent variables such as income range, age, gender or education level

Another example of deliverable is an adoption of Recommender Systems (in particular the content based approach) which has been widely used in e-commerce and entertainment services such as Youtube and Netfix. Similar scheme of recommender system will be applied (adjusted) to measure neighborhood scoring based on key parameters have proposed (similar to genre)

While the scope of the predictor may not be as comprehensive (due to scope and timeline), our novelty relies on the selection of parameters used for the scoring and visualization These parameters are unique aspects of our decision making systems.

In general the framework will be implemented based on the following flow:

1. Identify key features representing 5 parameters {Health, Social/Recreational, , Economy, Education and Population Demographic}. For the recommender implementation, these feature are comparable to the genre in Netflix recommender systems.
2. Identify relevant data and sources. The data will be Extracted, Transferred and preprocessed accordingly.
3. Basic analysis will be performed on Data including ranking table and chart based on pillar (comparative study on multiple neighborhood), development of neighborhood profile, the weighted table and scoring for recommender system.
4. The result will be presented using appropriate techniques including map, radar, barchart and word clouds
5. We will discuss our finding based on the recommender scoring and other comparative analysis

Data Preprocessing

The following table outlines the indicator for respective pillar and respective data endpoint needed (for example particular point of interest data to be retrieved)

Pillar	Indicator	Data Required
[H]ealth	Health Infrastructure	POI {Hospital/Clinic) Count/population
[E]conomy	Business Establishment	POI Business Count (Categories), POI/CENCUS Property Price/Type
[E]ducation	Education Infrastructure	POI {School/University) per population CENCUS Education level
[D]emography	Population, Age, Income Distribution	CENCUS Age/Gender (mean), CENCUS Income Distribution, CENSUS Population Size,
[S]ocial/ Recreation	Green infrastructure, Recreation POI	POI Count Green Infrastructure/Coverage, USERS Check in, Tips,

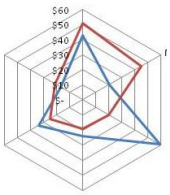
Proposed Method For Comparative Analysis

The following table maps the indicator and data to be analyzed, It specify the proposed data presentation or output for each pillar as well as for integrated metric (overall result)

Pillar	Indicator	Data Analysis	Result Presentation	Overall Result
[H]ealth	Health Infrastructure	POI {Hospital/Clinic) Count/population	Ranking Table & Bar Chart	Radar /Spider HEEDS Score based on Neighborhood Regression/Predictor (Age/Gender/ Income) vs Neighborhood
[E]conomy	Business Establishment	POI Business Count (Categories), POI/CENCUS Property Price/Type	Folium Map (POI Distribution) -Categories -Property Price	
[E]ducation	Education Infrastructure	POI {School/University) per population CENCUS Education level	-Bar chart Education level -Ranking/count -Map with labels	
[D]emography	Population, Age, Income Distribution	CENCUS Age/Gender (mean), CENCUS Income Distribution, CENSUS Population Size,	-Folium Map with label, Demographic -Choropleth Demographic Distribution-Bar Chart	
[S]ocial/ Recreation	Green infrastructure, Recreation POI	POI Count Green Infrastructure/Coverage, USERS Check in, Tips,	Folium Map (Green infra Distribution)	

Proposed Methods for calculating Score

The following table outlines the indicator for respective pillar and HOW recommender systems generally work. By. Combining User input (user specify preference based on pillar) with indicator weightage , the final overall score (HEEDS score) and radar chart can be produced

Pillar	Indicator	User Input Profile (Rank by Users)	Indicator Weightage Score (Recommended Features)	HEEDS SCORE & RADAR
[H]ealth	Health Infrastructure	Rank {int 1-5 } , Normalized	⊗ POI_Count per population size (Normalized Max-Min)	Σ 
[E]conomy	Business Establishment	Rank {int 1-5 } Normalized	⊗ POI_Count (Normalized Max-Min)	
[E]ducation	Education Infrastructure	Rank {int 1-5 } Normalized	⊗ POI_Count per population size (Normalized Max-Min)	
[D]emography	Population, Age, Income Distribution	Rank {int 1-5 } Normalized	⊗ Population_Size per squarefeet (Normalized Max-Min)	
[S]ocial/ Recreation	Green infrastructure, Recreation POI	Rank {int 1-5 } Normalized	⊗ POI_Count per population size (Normalized Max-Min)	

Results

Discussion

Conclusion