Using Foursquare data to explore green infrastructure potential in urban planning – case study in Dubai

1. Introduction

There is an emerging branch of academic literature interested in using Foursquare as a complementary data source for urban studies and urban development related public policy. Here I will principally contentrate on a paper by Martí et al. (2020, [Green infrastructure planning: Unveiling meaningful spaces through Foursquare users' preferences](), Land Use Policy, Vol 97, where the Foursquare is applied on researching green infrastructure.

Martí and others argue in their paper that Foursquare is an interesting tool for urban studies, as it allows for greater understanding of existing urban tissue and citizen preferences, allowing for deeper understangin of social value of different *venues*. In their paper, Martí et al. research urban infrastructure in Valencia, Spain, to identify relevant units of green infrastructure and guide policy decisions.

As such, Foursquare data can leverage crucial insights for officials in national and sub-national governments in charge of urban development. However, use of Foursquare is still in its infancy in urban studies, and authors enphasize that further study is still necessary. Rather, their purpose is to explore different uses of Foursquare and assess its potential as a complementary tool. Identifying and understanding Foursquare venues could help identify priority areas both thematically and regionally, to better direct investment and interventions and prioritizing decisions.

In this context I intend to replice parts of Martí and others' study on a different case, that of Dubai, United Arab Emirates. Dubai is an uncontested regional hub and a significant laboratory for urban development initiatives in the Middle East – and a significant component of UAE's national branding and soft power. UAE's current national policies on environmental, social, and financial sustainable development are proggressive both on regional and international level, as described in this National Geographic exclusive from 2017.

This policy context makes Dubai an interesting case study for exploiting Foursquare data for urban planning.

2. Data

Dubai is not divided into neighbourhoods and does not use zip codes. Rather, the most basic administrative sub-division is a community, identified by a three-digit number. Wikipedia has a list of Dubai communities with some demographic data, available here.

Based on this data, it is possible to extract venues data per community. The *venues* search will result in a data set of relevant venues in Dubai. The Foursquare divides its venues into a large quantity of categories and sub-categories based on their purpose. The API allows for automatically filtering only the relevant main category for this study, "Outdoors and Recreation".

Foursquare also records user-generated data on user activity of the venues, through *check-ins*, where users register their presence at a venue. This data is often used as a proxy for popularity and relevance of venues in longtitudinal studies – but is not freely accesible.

3. Methodology

3.1. Mining and wrangling the data

First step was to scrape data on communities from a list on Wikipedia, which allows retrieving and classifying venues possible in later steps. The data was built into a dataframe, which contains community code, name in English and Arabic, and some demographic data for most communities. The data was complemented with location data for each community through ArcGis Geocoder.

Graph 1: Dubai communities

Second step was retrieving the relevant venues for each community through the Foursquare developer API. The API allows to recover name, coordinates, and category for a limited number of venues for each community, filtered by relevant categories. The paper by Maní et al. restricts its analysis to the “Outdoors and Recreation” main category; I restricted my results to this category as well.

Accessing Foursquare user data and venue popularity is not easy for the public. Cumulative check-in data is no longer readily available. The API allows for extracting data on trending venues, but this data is very limited in terms of time, and likely not representative of long term usage patterns due to the COVID-19 travel and movement restrictions.

3.2. Grouping venues

I divided venues into three principal groups based on the framework suggested by Maní et al. They based thei division into already accepted green infrastructure notions “green spaces”, “activity hotspots” and “links”, which stood as a model for classifying venue subcategories into functional groups that share some similar traits. I used broadly the same division, and grouped my venues into groups A, B, and C.

Here A represents venues comprising green spaces and natural areas, B represents action oriented venues like sports spaces, and C represents venues that link other spaces together. I generated a list of unique values in the "Venue Categories" column the data and distributed these categories manually into the thematic groups.

Graph 2: Distribution of venues into groups

3.3. Descriptive analysis of Dubai venues

Prior to any predictive analysis or designin any complex algorithms, I will explore the data for some key indicators, such as total number of venues and number of venues in the "Outdoors and Recreation" main category; count of venues in each sub-category that the API provides; count of venues per community; and count of venues in each main category. I will also visualize some of these insights on a map.

This exploratory analysis will help to form a better understanding of the city, how it is built, and how its different venues are organized. This will provide a general parting point for further research and any policy planning.

3.4. Clustering Dubai communities

The main substance of the analysis will comprise a k-means clustering analysis of the Dubai communities based on the venues found in each one. For data access reasons and for simplicity, I will not replicate the segments on user preferences of Martí and others' research, as check-in data is not readily available for public use and I did not consider trending venues data as an equally relevant source of data.

I will first establish the most common venues in each community, and use them as a basis for designing a k-means model. I will use the model to assign the communities into qualitatively different groups for further analysis. This segment was not present in the paper by Martí et al., and I hope it would yield some additional insight as to the structure of different communities.

3.5. Visualizing venues and communities

I will end my analysis by visualizing my central results in maps. I will use the Folium library for creating the maps and presenting communities and venues in an easier to digest format. I will present the distribution of venues, venue groups, and community clusters.

4. Results

Based on the Foursquare API query, I recovered a total of 6 152 venues for the city. Of these, 1 332 or approximately 21,7 % fell within the relevant "Outdoors and Recreation" main category. This percentage seems significant in comparison to the study by Martí et al., where only 7 % of all venues in Valencia, Spain, fell in this category.

In terms of the three main groups of venues – green *spaces*, activity *hotspots* and *links*, the activity related venues were clearly over-represented. 939 venues or some 70 % of all venues fell in this category. Only 229 venues were considered as green or natural spaces, and 128 were considered links or connecting zones.

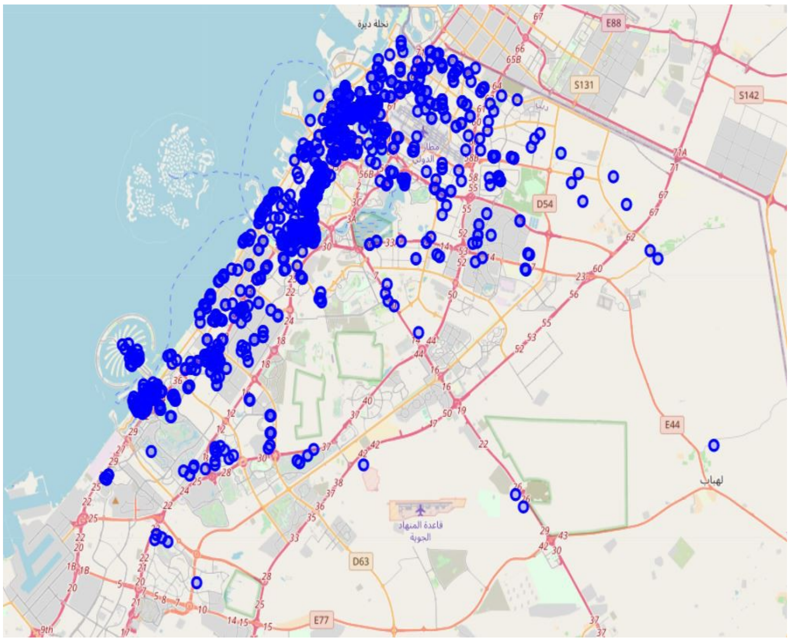
Of these venues, the most common ones were gyms and gyms/fitness centres, followed by an assortment of venues described in the graph below.

Gym	345
Fitness centre	217
Park	120
Pool	87
Beach	84
Plaza	53
Harbor	47
Athletics and sports	38
Soccer field	23

The communities with most venues are described in the graph below. These were located, rather unsurprisingly, close to the coast and in the densely built areas in the city centre.

Marsa Dubai	75
Downtown	56
Trade Centre 2	48
Za'abeel Second	48
Trade Center 1	39
Al Satwa	38
Al Safouh Second	37
Rigga Al Buteen	34
Business Bay	32
Al Rigga	28

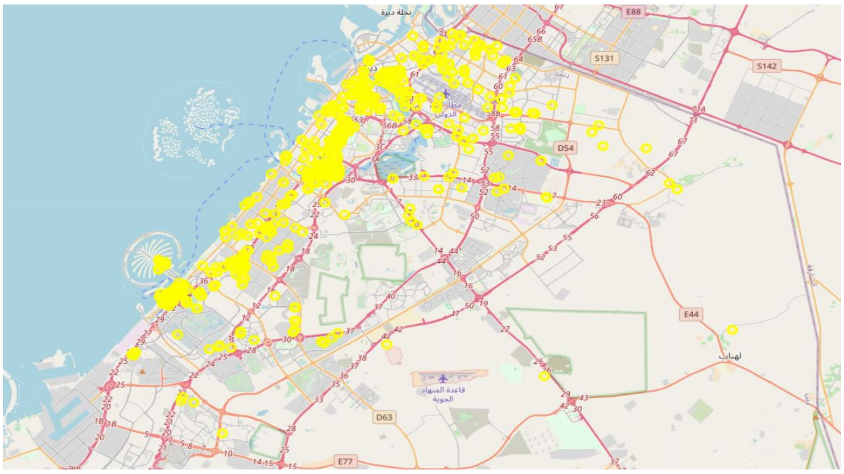
The below map demonstrates the geographic distribution of the venues. Again unsurprisingly, most venues are concentrated in the densely built coastal zones. However, there are also significant clusters of venues inland, often correlating well with green and rural zones with little buildings and low population, but with green spaces and sports venues that benefit from calm, such as yoga or martial arts studios.



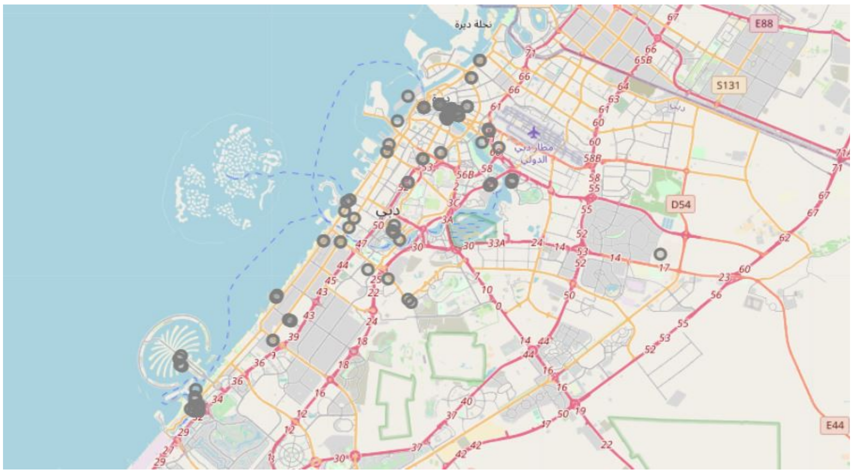
Graph 3: Distribution of "Outdoors and Recreation" venues

The last three maps demonstrate the distribution of venues by functional groups A, B, and C. The green icons represents distribution of green or natural spaces of the groups A, the yellow icons represent the activity hotspots of the group B, and the grey icons represent the connecting zones of the group C.

Graph 4: Group A: Green spaces



Graph 5: Group B: Activity hotspots



Graph 6: Group C: Connecting zones

K-means clustering analysis resulted in a division in five groups, demonstrated in the below map.

The green cluster seems like an outlier. It only consists of two venues, and they seem to be mostly defined by their proximity to the Mushrif park than anything else.

The yellow cluster is somewhat more interesting. Only present in the outskirts of the densely developed coastal region, it can be described as isolated and rural. The most common venue in this zone is a farm, but there are also several natural spaces and a high number of activity based venues oriented towards well-being and sports, such as yoga and pilates. This cluster could be described as the retreats of Dubai.

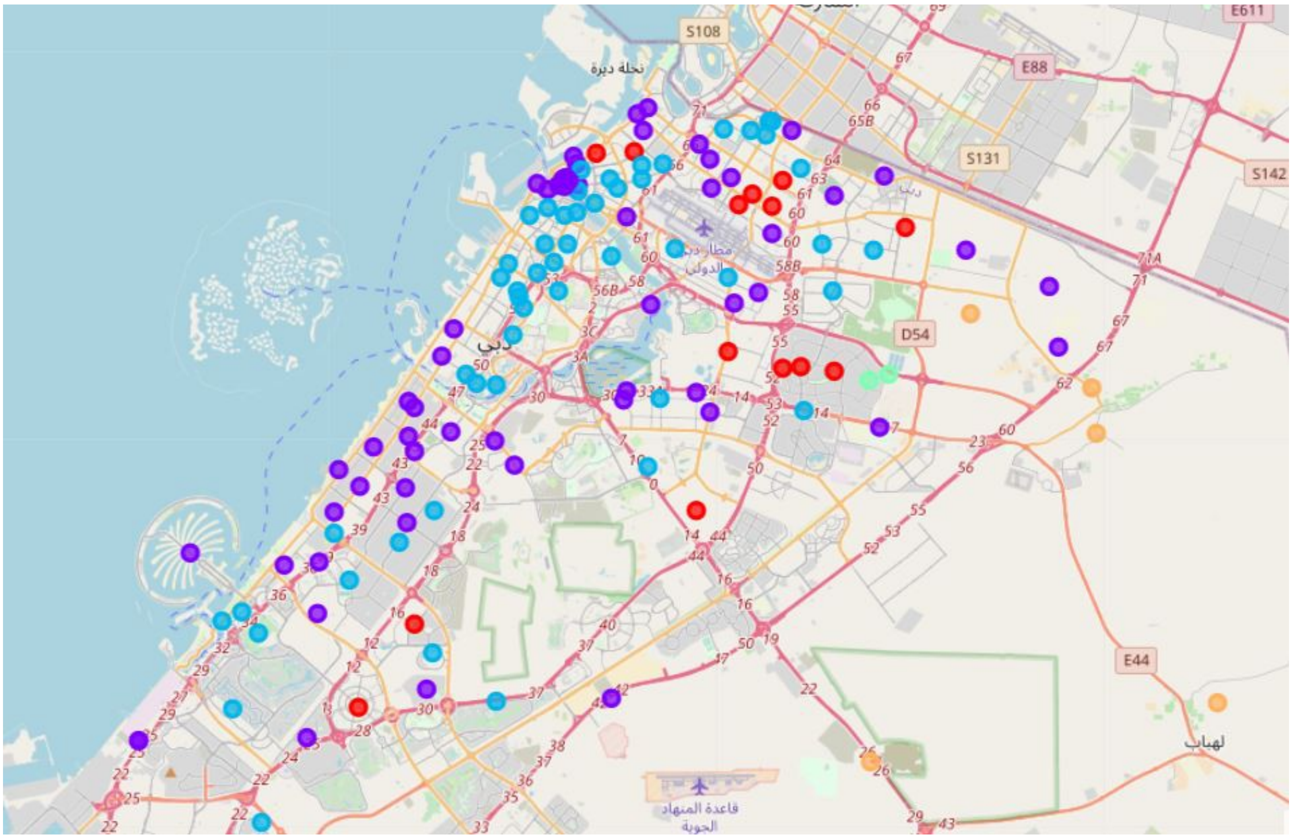
Slightly larger red cluster represents some of the less populous zones, some of which are close to the city centre and some are further away. This cluster is dominated by parks, lakes, springs and pools. The locations correlate well with Dubai's green zones, and the cluster could be described as the oases of calm in a bustling metropolis.

The last two clusters, purple and blue, share many factors. They are by far the largest clusters, containing 59 (purple) and 52 (blue) venues. They contain most of the large and/or populous communities of Dubai, which makes them the centres of the action. They are both located in the coastal and densely developed downtown area.

But at the same time, each cluster has its own distinct profile, which can be perceived by a glance at the map.

The purple cluster is concentrated at the coastline. It is more oriented towards beaches and outdoor venues. This cluster could be described as the Dubai beach zone.

The blue cluster is only slightly further inland. It lacks beaches and is more densely built, making it a hub for indoor activities. Most gyms are within this cluster, that could be described as Dubai's heart, beating with activity.



Graph 7: Clusters of communities

5. Discussion

This excercises purpose has been more exploratory than practical, in that I aimed to assess the potential of Foursquare as a research tool rather than answer a concrete business or policy problem.

This excercise has attempted to explore the potential of Foursquare in urban studies and policy planning. The results have contributed towards a clearer understanding of the composition of the city of Dubai, and demonstrated in practice how geodata from an LBSN platform like Foursquare can be used as a complementary data source.

This excercise has been significantly more restricted than the paper by Martí et al. it follows for practical and data access reasons. However, even a restricted excercise like this yilded easy insights over the urban patchwork of Dubai; with more time and resources this framework could be used for significant impact in academic research and policy planning.

Even the restricted resources and narrow methodology of this excercise allowed for a functional segmentation of Dubai and analysis of socially sustainable infrastructure in the city. When combined with sufficient volumes of user-generated data, it would be possible to identify priority areas for research and investment, and execute policy analyses for urban design.

6. Conclusion

This project has attempted to replicate parts of a study by Martí et al. using Foursquare API as a complementary data source in urban research and urban policy planning. I have applied a framework proposed by the authors on a different case, that of Dubai.

Obviously the case researched by Maní and others, Valencia, is quite qualitatively different from the case study of Dubai.

Valencia has greater focus on preservation and balancing historical constructions and existing green spaces as a part of a modern urban pattern due to specific historical trajectory and climate conditions.

Dubai, on the other hand, emerged quickly in a desert zone, making preservation of little concern. Rather, focus is on creating human focused and sustainable urban pattern and thus increase the city's profile.

Deeper understanding of existing infrastructure is no less important, however. Martí et al. concluded their reseach expressing optimism towards use of Foursquare data on urban research.

Due to methodological concerns and access to data issues, only parts of the study were replicated, and for simplicity's sake a framework already established by the authors was used without significant modifications.

However, I was able to contribute a different angle by the cluster analysis of Dubai communities based on venues in each community.

References

Kunzig, 2017, The World's Most Improbable Green City, National Geographic
https://www.nationalgeographic.com/environment/urban-expeditions/green-buildings/dubai-ecological-footprint-sustainable-urban-city/

Marti et al. 2020, Green infrastructure planning: Unveiling meaningful spaces through Foursquare users' preferences, Land Use Policy, Vol 97

https://www.sciencedirect.com/science/article/pii/S0264837719320435

List of communities Wikipedia

https://en.wikipedia.org/wiki/List_of_communities_in_Dubai