COURSERA CAPSTONE PROJECT
Week 4 - on Homelessness
By Ira de Guzman
San Jose, California, USA

* * *

In a city of your choice, if someone is looking to open a restaurant, where would you recommend that they open it? Similarly, if a contractor is trying to start their own business, where would you recommend that they set-up their office?

These are just a couple of many ideas and problems that can be solved using location data in addition to other datasets. No matter what you decide to do, make sure to provide sufficient justification of why you think what you want to do or solve is important and why would a client or a group of people be interested in your project.

Review criteria

This capstone project will be graded by your peers. This capstone project is worth **70%** of your total grade. The project will be completed over the course of 2 weeks.

Week 1 submissions will be worth 30% whereas
week 2 submissions will be worth 40% of your total grade.

* * *

For **the first week**, you will be required to submit the following:

1) A description of the problem and a discussion of the background. (15 marks)

San Jose has consistently landed as among the most expensive housing markets in the country. One unmissable downside is the ever surging rise of the homeless population. "The reasons for homeless are many and varied." said Joe Simitian, President of the Santa Clara County Board of Supervisors. "The high cost and shortage of housing are making a bad problem worse."

This presentation offers a glimpse of what the City of San Jose is doing to curb the seemingly inexorable increase of homelessness. Sam Liccardo, the Mayor of San Jose, wants to "double down on homeless prevention, and in turn, alleviate the human misery."

"Homelessness" is Councilman Raul Peralez's number one priority. According to him, and others from San Jose Housing and Santa Clara County Supportive Housing, they have "solutions for almost 100% of the homeless on the streets."

"But the problem is far from over." Peralez admits. San Jose voters approved of almost a billion dollars for affordable housing.

2) A description of the data and how it will be used to solve the problem. (15 marks)

Affordable housing is one of many LONG-TERM solutions the City of San Jose is embarking on. According to Peralez, "we need more of these units, and we need them to get in the pipeline." There are already projects in the San Jose downtown, and several more interspersed in the city.

**Information from data extracted from the City of San Jose public website will be discussed using Python and Github. Because the addresses that were extracted from the website were (1) not businesses, but considered "residences," this category does not fall among the usual entities that one would likely find in social-driven map apps such as Foursquare. (2) Furthermore the addresses obtained were incomplete, i.e. there is no city and zip codes provided.*

I have considered using a Visual-Basic macro in Excel (see sample of the coding below), but that too demands COMPLETE addresses (number, street, city, zip code). The other option is to input these incomplete addresses to a location finder, that will generate the corresponding coordinates (latitude, longitude, and even the point of elevation). But that too requires inputting the data manually, and individually. An arduous task. **(Nonetheless, it was completed for this project!)**

For **the second week**, the final deliverables of the project will be:

1) link to your Notebook on your **Github** repository, showing your code. (15 marks)
2) full report consisting of all of the following components (15 marks):

a) Introduction where you discuss the business problem and who would be interested in this project.

The data presented here are public information. The created maps derived from these data could be of interest to contractors and architects who may wish to study the distribution of the current (and future) affordable housing units interspersed more equally, so they could plan accordingly.

But the recipients who will most benefit, of course, are the non-profit organizations whose focus is homelessness.

b) Data where you describe the data that will be used to solve the problem and the source of the data.

The data was obtained from an accessible website (**data.sanjoseca.gov**)*. Of the 42 datasets available to the public, none contains any latitude- and longitude*

coordinates. Foursquare, albeit a functional resource, does not emphasize non-business entities, much less affordable-housing units.

The coordinates (for the addresses of the aforementioned units) were in fact obtained manually, using Google-driven **mapcoordinates.net**. The latitudes and longitudes are then logged into a separate table that was later merged with the public-accessed dataset.

c) Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.

The usual methodology of using Foursquare may not have been the most efficient as Foursquare's cornerstone (pardon the pun) is mainly businesses and other attractions, such as restaurants, bars, shops, parks, and most importantly, their auxiliary comments. Residential areas are not included in the listings.

As mentioned earlier, obtaining the coordinates of affordable housing units from Foursquare is nonviable. The other option, albeit arduous, is to acquire the pertinent information from websites that generates the coordinates of specific, non-business addresses. The site that I chose for this project was **mapcoordines.net***.*

d) Results section where you discuss the results.

Once the coordinates of the addresses were obtained and placed into a proprietary table, that table is then merged into a dataset obtained from **data.sanjoseca.gov** *("Affordable Family Housing Units-2020").*

The resulting merged dataframe is then mapped and clustered using FOLIUM.

e) Discussion section where you discuss any observations you noted and any recommendations you can make based on the results.

The affordable housing units are mostly converted apartments, the majority of which are built on the eastside of San Jose. A few are nestled on the west of San Jose, in the city of Campbell. But none in the more affluent, adjacent cities of Santa Clara, Los Gatos, Saratoga, and Cupertino.

The majority of the homeless make their "homes" in San Jose, where various organizations supporting the homeless -- and "transients" -- are located, so it makes sense to "reside" near these places.

f) Conclusion section where you conclude the report.

There are no silver bullets in solving the challenges that surround homelessness. Long-term solutions will alleviate these challenges in the future, but the more pressing need stresses on the present. Short-term solutions, like building the de-rigeur 'tiny homes' proved to be expeditious; albeit the practice of converting and repurposing old apartments seems to be the sound choice because it bypasses new construction to be undertaken.

3) Your choice of a presentation or blogpost. (10 marks)
**This is the presentation (completed on Word.doc / Google docs).*

/end of report