

The background is a dark, abstract composition. It features a blurred financial candlestick chart with red and green bars. Overlaid on this are glowing blue and green binary digits (0s and 1s) and faint, dashed white lines that suggest a grid or data paths. The overall aesthetic is high-tech and digital.

**CAPSTONE
PROJECT**

**BATTLE OF THE
NEIGHBORHOODS
FINAL PROJECT**

DANIEL E. *RIVERO MENDOZA*

JUNE 2020

BACKGROUND

This Capstone Project - *Battle of the neighborhoods* aims to leverage the tools and techniques learned over the past few months in the IBM Data Science Professional Certificate to answer if Data Science tools can be used to help in some of the aspects of having to move. The aim is to suggest moving to the “*right place*” by providing an approach and methodology to find neighborhoods with desired characteristics and attributes, assuming the more familiar and comfortable it feels @ destiny the better-off anybody will be. The case of study is a hypothetical move from Boulder (CO, United States) to Sydney (NSW, Australia), two cities halfway across the world, with very different landscapes, but perhaps more similar than anybody would think at first glance.



PROBLEM TO BE SOLVED

The problem to solve is finding a neighborhood in Sydney (NSW, Australia) with similar characteristics to one in downtown Boulder (CO, United States), plus some extra attributes. Thus, to set the basis for comparison, the following are applicable to the neighborhood at destination:

Must haves:

- ✓ Surroundings with amenities and venues like the ones found in Boulder (CO), USA
- ✓ Located within 2 km from a train or light rail station in greater Sydney (NSW), Australia

Desirable, pending availability of time and reliable data:

- ✓ Rent price in the 500 AUD/week range for a unit with at least 2 bedrooms, 1 bathroom, 1 parking spot, and 75 sqm.



DATA REQUIRED

To establish compliance of neighborhood(s) @ destination with the list of attributes in Section 1.2., the following data is needed.

- Coordinates for downtown Boulder, CO - Source: Nominatim tool from geopy.geocoders
- List of top venues in downtown Boulder, CO. - Source: API call to access Foursquare data (name, coordinates, type, etc.) on venues @ given location
- List of suburbs in Sydney, NSW. - Source: Scraped from the website <http://www.walksydneystreets.net/suburbssydneyall.htm>
- Coordinates for the suburbs in Sydney, NSW. - Source: Nominatim tool from geopy.geocoders., based mostly on the list of neighborhoods outlined above.
- GEOJSON data for suburbs in Sydney, NSW. - Source: File with all the 4000+ neighborhoods in New South Wales, Australia, downloaded from federal government site <https://data.gov.au/data/dataset/91e70237-d9d1-4719-a82f-e71b811154c6>
- List of Train Stations in Sydney, NSW. - Source: File with all the train and light rail stations in New South Wales, Australia, downloaded from the OpenData portal from the local government site <https://opendata.transport.nsw.gov.au/dataset/train-station-entries-and-exits-data>
- Coordinates of Train Stations in Sydney, NSW. - Source: See above.

Time and reliability of data permitting, the following dataset is needed to establish compliance of neighborhoods @ destination with desired attributes for rental units.

- Average rental price and characteristics of units per neighborhood in Sydney, NSW. - **NOTE** No reliable data was found in time for this.



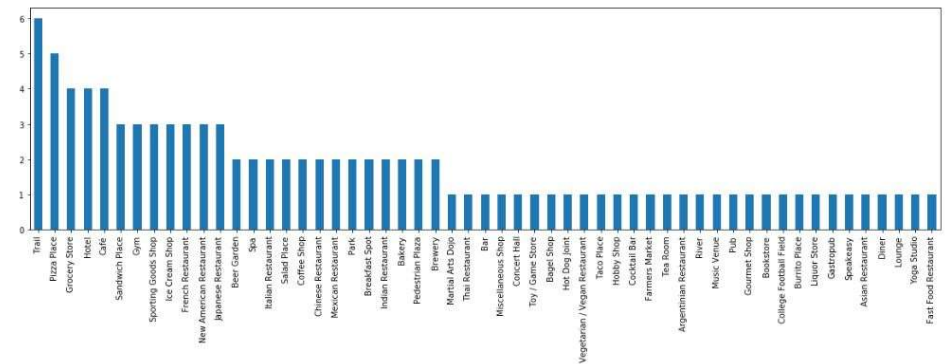
METHODOLOGY

The data outlined above will be used as follows:

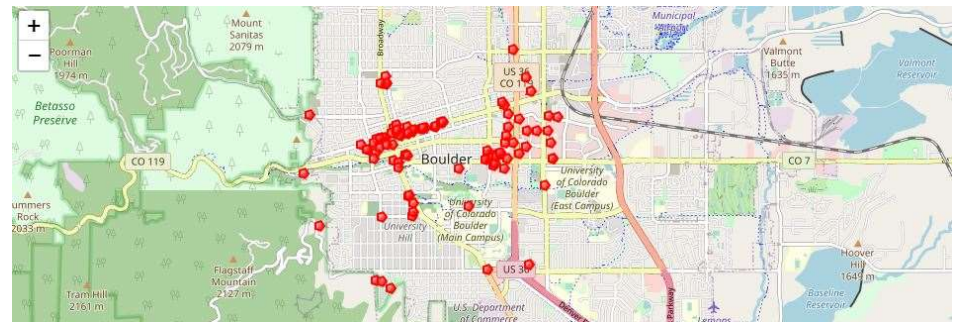
- Coordinates for downtown Boulder will be used to retrieve top venues in 5 km radius _via_ FourSquare.
 - Charts and maps to be used to visualize the data.
- List of suburbs in Sydney will be used to retrieve the geographic coordinates _via_ geopy, which will be complemented with data from the GEOJSON.
- List of Train Stations and their coordinates in Sydney, NSW, will be used to filter out suburbs(s) with centres further than 2 km.
 - Maps to be used to visualize the data.
- List of venues in Boulder and Sydney will be used to establish similarities between their corresponding suburbs.
 - Charts to be used to explore the data and look for correlations.
 - Maps to be used to depict the data.
 - kMeans algorithm to be used in order to categorize the suburbs in Sydney and Boulder, which will pair the latter with one or several suburbs in the former based on the venues available close by.
- Additional metrics to be used as fit to figure out the list of suburbs that are suitable for moving in Sydney.

RESULTS AND ANALYSIS

- ✓ List and category breakdown of venues in Boulder, CO.

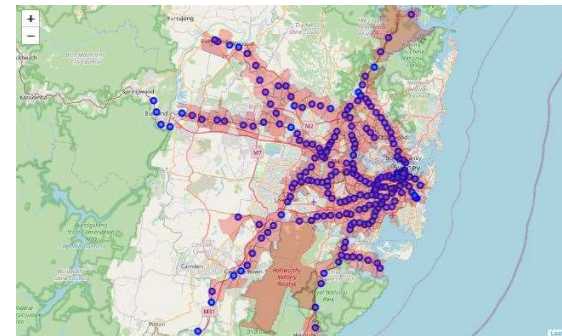
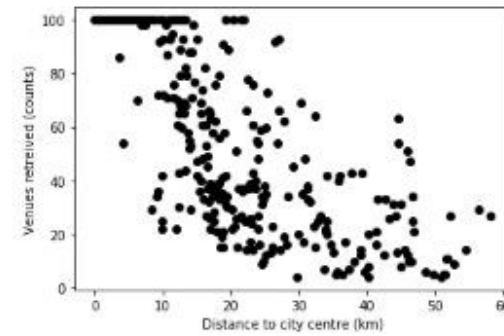


- ✓ Geographical distribution of venues retrieved in Boulder, CO.



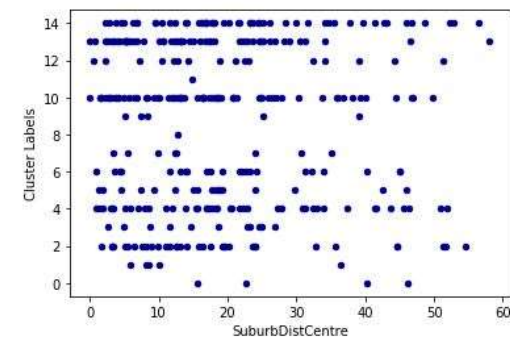
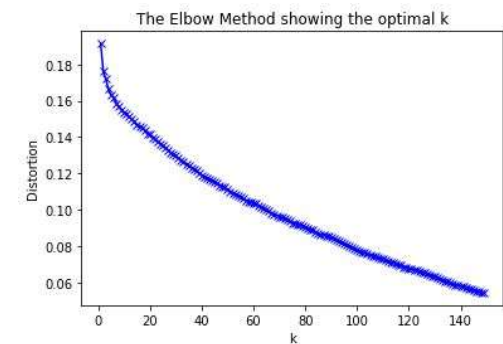
RESULTS AND ANALYSIS

- ✓ Number of places retrieved via FourSquare for Sydney, NSW. vs. distance to Sydney centre.
 - ✓ No correlation observed.
- ✓ Map of Sydney with train stations (blue) and suburbs < 2km from a train station.



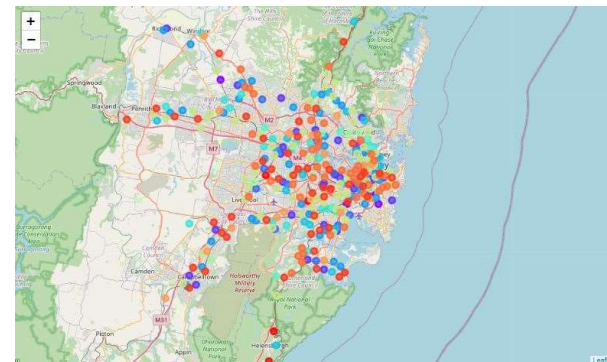
RESULTS AND ANALYSIS

- ✓ kMEANS distortion plot as a function of cluster number
 - ✓ Optimum cluster number set at 15
- ✓ Cluster number vs. distance to city centre
 - ✓ Clustering not dependent on distance



RESULTS AND ANALYSIS

- ✓ Map of Sydney with suburbs color coded per cluster
 - ✓ Cluster 13 with similar attributes as Boulder, CO – parks and gyms featuring in most common venues, along with cafes, bars and restaurants



- ✓ Table with Sydney suburbs like Boulder, CO, and within 2km of a train station
 - ✓ Darlington, Redfern, Rushcutters Bay, Camperdown and Rozelle are suggested as places to move.

	Suburb	SuburbDistCentre	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Train Stations < 2km	Train Stations
2	DARLINGTON	2.830000	Coffee Shop	Café	Thai Restaurant	Bookstore	Bar	26.0	Bridge Street LR, Capitol Square LR, Central, ...
47	REDFERN	4.400000	Café	Pub	Japanese Restaurant	Bakery	Bar	13.0	Capitol Square LR, Central, Central Chalmers S...
50	RUSHCUTTERS BAY	2.470000	Café	Park	Australian Restaurant	Italian Restaurant	Pub	8.0	Edgecliff, Kings Cross, Martin Place, Museum, ...
13	CAMPERDOWN	5.120000	Café	Bar	Ice Cream Shop	Pub	Coffee Shop	8.0	Ersineville, Glebe LR, Jubilee Park LR, Macdon...
3	ROZELLE	4.040000	Café	Park	Italian Restaurant	Bar	Seafood Restaurant	7.0	Fish Market LR, Glebe LR, John Street Square L...
16	ANNANDALE	5.130000	Café	Italian Restaurant	Park	Ice Cream Shop	Bakery	7.0	Glebe LR, Jubilee Park LR, Leichhardt North LR...

CONCLUSION

- ✓ To conclude, the data analysis was performed to identify the suburbs in Sydney with train or light rail stations within 2km and venues/amenities similar to downtown Boulder, CO. During the analysis, several suburb and venues/amenities features were explored and visualized. Furthermore, clustering helped to find a list of suitable moving suburbs

