IBM Data Science: Capstone Project

SYDNEY WATER BUBBLER POSITIONING

Madhur Prashant Bhandari

Table of Contents

Introduction *2*

Business Problem *2*

Methodology *3*

Data *3*

Neighbourhoods Sydney Water Bubbler Data 3

Geocoding 3

Venue Data 4

Hot Spot Encoding *4*

Clustered Regions *5*

Top 10 most common places *5*

Results and Discussions *5*

Conclusion *6*

Introduction

It is only when something is taken away that we realise just how importance it was to our lives, and it is safe to say that the loss of outdoor drinking fountains and indoor water dispensers would have a significant impact on our lives. We expect to see them in schools, offices and even hotel lobbies and luckily, nowadays we are rarely disappointed.

Different locations will prefer different taps because of the type of customer they have. School drinking fountains and gym fountains will have a constant need for small quick bursts of water, and this is exactly what the bubbler tap is designed for. While offices and professional areas are unlikely to want their staff to be dipping down to catch a quick mouthful of water. It hardly gives off the best professional image. The bubbler tap allows the user to quickly dip down, filling their mouth with refreshment and then get on with their day. Even those who are not being active might be in need of a quick drink, such as in between classes or after lunch.

There is a large network of drinking fountains (water bubblers) dotted around the city of Sydney where one can get fresh water while they are walking or cycling. Most are in parks and playgrounds while some can also be found along main roads or near tourist attractions. Sydney water fountains are not only useful, they are something of a tourist attraction in themselves - with some more than 100 years old.

Business Problem

A research to evaluate the access to and supply of water in a variety of settings, such as open spaces and sports and recreation centres could be carried out so that places with higher demand have adequate and continuous supply of water. This guide is based on the research findings and a review of drinking water fountains.

Consider location:

- Map out locations of existing water fountains and identify opportunities for installation of new water fountains/refill stations
- Research indicates that if drinking water sources are not in prominent areas and blend in with surroundings, they are less likely to be used
- Units in poor locations or not installed on appropriate (flat) surfaces can make it difficult to access the water source

For new installation sites:

- identify high pedestrian traffic areas, such as next to a playground
- open spaces where there are opportunities to do physical activity

- open spaces where there are planned picnic tables/BBQ facilities.
- Presence of a water bottle refill station

Methodology

This report helped my detect areas of Sydney that have low water bubbler density, particularly those around crowded CBD areas.

In first step I have collected the required **data: location and type (category) of every water bubbler in Sydney CBD area**. I have also **identified venues near those bubblers** (according to Foursquare categorization).

In final step we will focus on most promising areas and within those create **clusters of bubbler locations that meet some basic requirements** established in discussion with stakeholders: we will take into consideration locations with **venues around these bubblers**, and we want locations **so that these venues could be addressed with appropriate number of bubblers**. We will present map of all such locations but also create clusters (using **k-means clustering**) of those locations to identify general zones / neighbourhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

Data

The data for this project has been retrieved and processed through multiple sources, giving careful considerations to the accuracy of the methods used.

Neighbourhoods Sydney Water Bubbler Data

The data of the neighbourhoods in Kolkata can be extracted out by downloading from City of Sydney site. This downloaded data is then read through panda data frame.

Geocoding

The latitude and longitude of Sydney are retrieved using Google Maps Geocoding API. The geometric location values are then stored into the initial data frame.

5.13 Finding Georagraphical coordinates of Sydney

```
from geopy.geocoders import Nominatim
address = 'Sydney, NSW'

geolocator = Nominatim(user_agent="Sydney_explorer")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of Sydney are {}, {}.'.format(latitude, longitude))
```

Venue Data

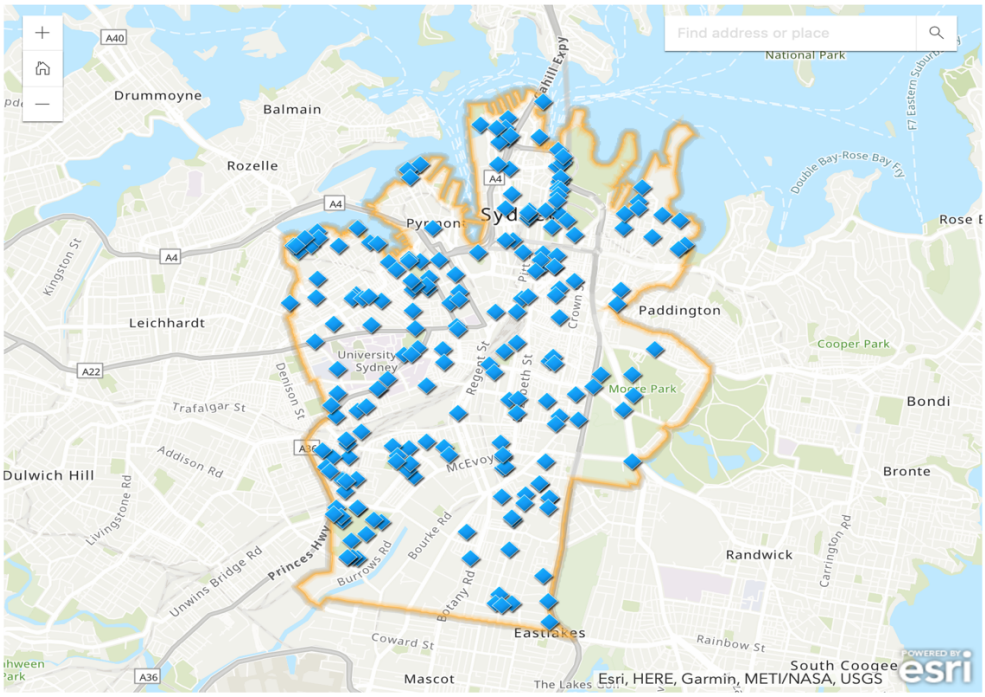
From the location data obtained after Web Scraping and Geocoding, the venue data is found out by passing in the required parameters to the FourSquare API and creating another Data Frame to contain all the venue details along with the respective neighbourhoods.

5.4 Creating a function to Four Square API Call

```
def getNearbyVenues(names, latitudes, longitudes, radius=500):

    venues_list=[]
    for name, lat, lng in zip(names, latitudes, longitudes):
        print(name)

        # create the API request URL
        url = [redacted].format(
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION,
            lat,
            lng,
            radius,
            LIMIT)

        # make the GET request
        results = requests.get(url).json()["response"]['groups'][0]['items']

        # return only relevant information for each nearby venue
        venues_list.append([(
            name,
            lat,
            lng,
            v['venue']['name'],
            v['venue']['location']['lat'],
            v['venue']['location']['lng'],
            v['venue']['categories'][0]['name']) for v in results])

    nearby_venues = pd.DataFrame([item for venue_list in venues_list for item in venue_list])
    nearby_venues.columns = ['Neighborhood',
                  'Neighborhood Latitude',
                  'Neighborhood Longitude',
                  'Venue',
                  'Venue Latitude',
                  'Venue Longitude',
                  'Venue Category']

    return(nearby_venues)
```

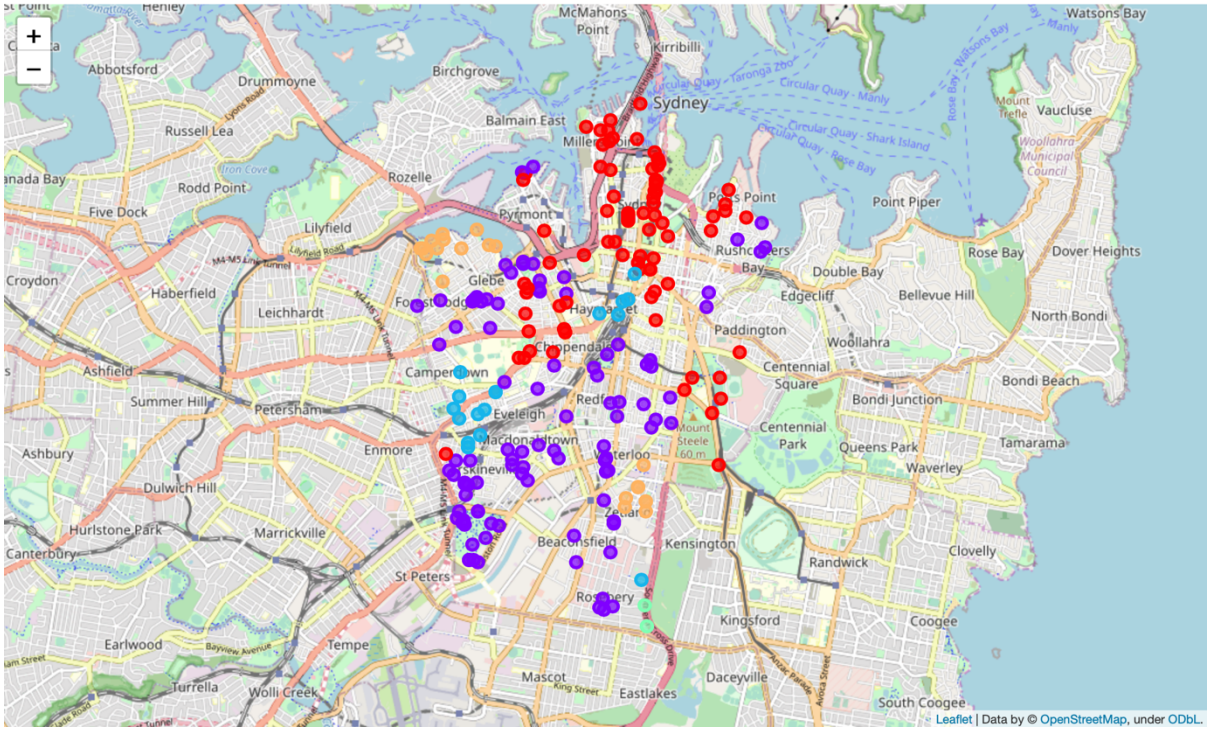
Hot Spot Encoding

Below is the original hot spot encoding that has been sourced from Sydney Water Department.



Clustered Regions

We used k-means clustering and grouped them to identify the zones that could be used to better position and the need of water bubbler.



Top 10 most common places

FourSquare API returns lot of venues closed to a place so the results were restricted to only 10 results using below code:

5.11 Function to return most common venues

```
def return_most_common_venues(row, num_top_venues):
    row_categories = row.iloc[1:]
    row_categories_sorted = row_categories.sort_values(ascending=False)

    return row_categories_sorted.index.values[0:num_top_venues]

num_top_venues = 10

indicators = ['st', 'nd', 'rd']

# create columns according to number of top venues
columns = ['Neighborhood']
for ind in np.arange(num_top_venues):
    try:
        columns.append('{}{} Most Common Venue'.format(ind+1, indicators[ind]))
    except:
        columns.append('{}th Most Common Venue'.format(ind+1))

# create a new dataframe
neighborhoods_venues_sorted = pd.DataFrame(columns=columns)
neighborhoods_venues_sorted['Neighborhood'] = bubbler_grouped['Neighborhood']

for ind in np.arange(bubbler_grouped.shape[0]):
    neighborhoods_venues_sorted.iloc[ind, 1:] = return_most_common_venues(bubbler_grouped.iloc[ind, :], num_top_venues)

neighborhoods_venues_sorted.head()
```

Results and Discussions

Our analysis shows that although there is a great number of bubblers in Sydney, there are pockets of water bubblers density fairly close to famous venues.

Result of all this is 5 zones containing largest number of potential new water bubblers locations based on number of and distance to existing venues - both restaurants and cafes in general. This, of course, does not imply that those zones are actually optimal locations for a new water bubbler! Purpose of this analysis was to only provide info on areas close to Sydney CBD but not crowded with existing water bubblers. Recommended zones should therefore be considered only as a starting point for more detailed analysis.

Conclusion

Purpose of this project was to identify Sydney CBS area with low number of water bubblers in order to aid Sydney water department in narrowing down the search for optimal location for a new water bubbler. By calculating venues density distribution from Foursquare data, we have first identified general neighbourhood and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby bubblers. Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations) and addresses of those zone centres were created to be used as starting points for final exploration by Sydney Water Department.

Final decision on optimal restaurant location will be made by Sydney Water Department based on specific characteristics of neighbourhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or restaurant).