Report:
Recommendation engine for location of ice cream trucks during summer in Toronto

1. **Introduction:**
   An ice cream company having outlets all over North America is planning to enter Toronto. But due to high rental costs in the city and a perceived seasonal demand for ice cream, the company has decided to try out a new model of mobile ice cream trucks only during summer. The company has invested in 3 trucks.

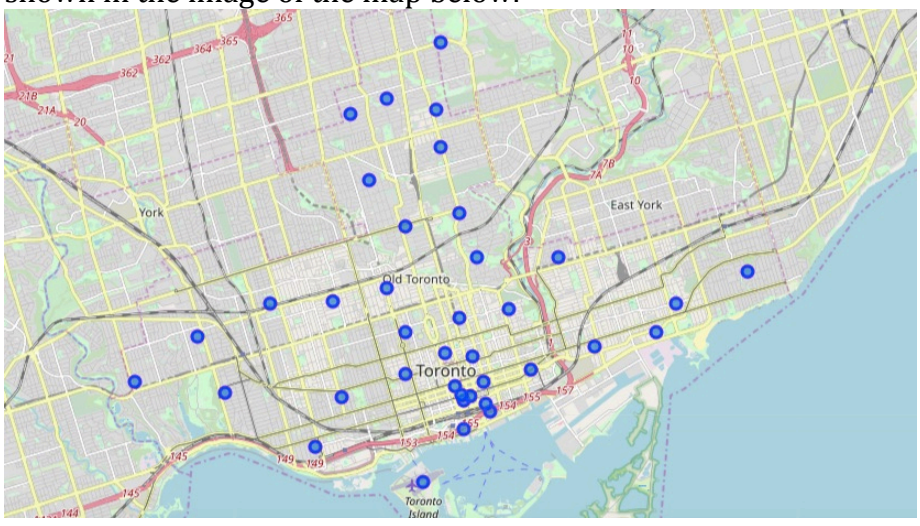
   The business question which needs to be answered is where to optimally station these trucks to maximize footfalls and sales.

2. **Data to solve the problem:**
   There are two data sources which I will be using to answer this question.

   (i) The FourSquare Venues API to understand the spread of existing ice cream shops and how the distribution is spread out by each neighborhood

   (ii) The city of Toronto's open data portal (https://open.toronto.ca/) has lots of datasets about the city. I will be using a dataset of list of public airconditioned spaces which is put out in public interest during the summer months. This dataset can be found at https://open.toronto.ca/dataset/air-conditioned-and-cool-spaces-heat-relief-network/. I will be using this dataset to determine how each neighborhood ranks with respect to these spaces and understand how many people are expected to use these spaces by neighborhood in the summer months.

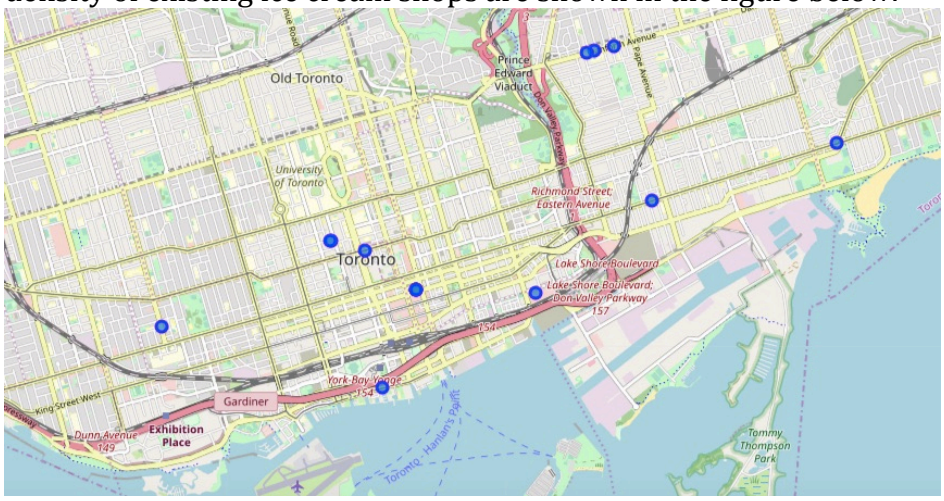
3. **Methodology:**
   The first step of the process is to extract the list of neighborhoods in Toronto along with the corresponding latitude and longitudes. I get this for all pincodes in Canada from Wikipedia and further filter out only pincodes of neighborhoods in Toronto. I get a dataframe with 40 rows as shown in the image of the map below.



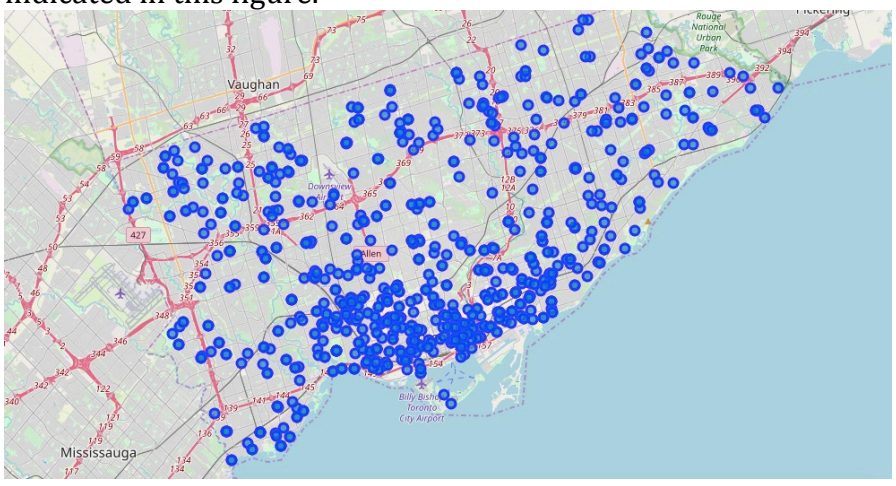
Foursquare API: Now I fetch the venues nearby each venue within a radius of 500m and up to 100 venues per neighborhood. While we loop through the 40 rows in the neighborhood dataframe and create a new row for each venue returned for every neighbourhood.

This new dataframe can be grouped by 'Venue Category' which shows us that there are 15 venues in the dataframe with category 'Ice cream Shop'. There might be duplicate entries but since we are only concerned about the density of ice cream shops in any given neighbourhood, we need not filter out the duplicates.

Existing Ice cream shop density: Now from the dataframe of all venues in all neighborhoods, I just extract the rows which are ice cream shops. Once extracted into a new data frame, I group this data frame by neighbourhood and join this into our neighbourhood data frame. The density of existing ice cream shops are shown in the figure below.



Airconditioned summer spaces:
Now to move to the other half of the data needed to estimate public day time gathering, I use the public airconditioned spaces data as mentioned in the data section. This data set has 800+ spaces along with longitude and latitude. We load this data into a new data frame. Once loaded, I loop through this data frame and find the closest neighborhood latitude and longitude to the air conditioned space latitude and longitude using the haversine distance formula. I add the closest neighborhood latitude and longitude to each row in this data frame. All airconditioned spaces indicated in this figure.

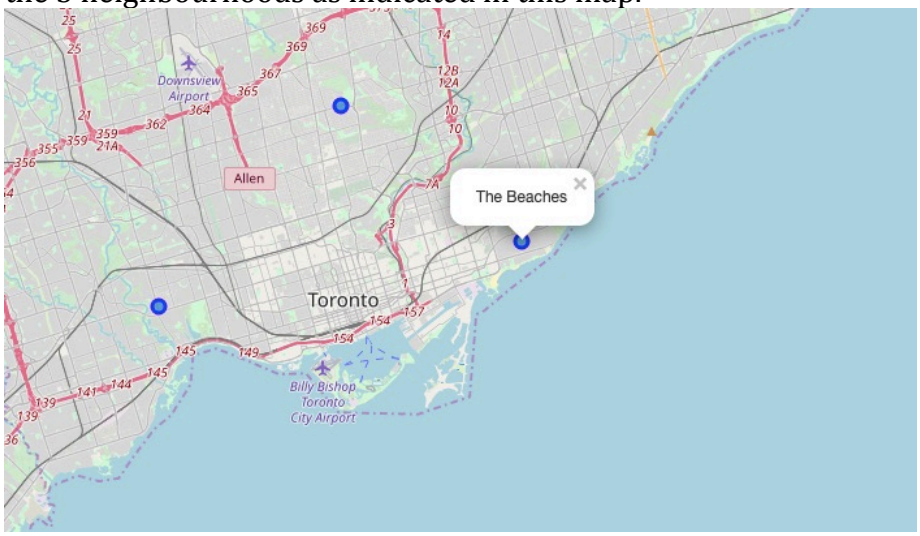


Merging all the data frames:
In the airconditioned spaces dataframe, I group by closes neighbourhood latitude and longitude which effectively gives me the density of airconditioned spaces in each neighborhood. Now I add this density count to the original neighborhood data frame which also has density of existing ice cream shops.

4. **Results:**
The results are very unambiguous and as desired. We have found the 3 most dense neighborhoods by public air conditioned spaces which also do not have any ice cream shops nearby. The locations are as below
    1) The Beaches (139)
    2) Runnymede, Swansea
    3) Lawrence Park

5. **Discussion:**
For the 3 most dense neighborhoods based on air conditioned space density, there are no ice cream shops in any of the neighborhoods. Furthermore, the 4th most dense neighborhood has 40 fewer airconditioned spaces than the 3rd which makes our recommendation more straightforward to make.

For now, we can recommend that the 3 ice cream trucks are stationed in the 3 neighbourhoods as indicated in this map.



6. **Conclusion:**

Though the results and recommendation drawn are clear, they are based on the the assumption that airconditioned space density will lead to footfall and potential customer will not be tempted by non exclusive outlets which sell ice cream.

But none the less, it is a good starting inference which can be further validated after the trucks are commissioned. The model can be improved with newer list of air conditioned spaces and any additional details about them which may be observed and disclosed.