COURSERA CAPSTONE

IBM Applied Data Science Capstone

RESTAURANTS NEAR ME, CHENNAI

By Swetha Mohan

INTRODUCTION:

Chennai, also known as Madras, is the capital of the Indian state of Tamil Nadu. It is one of the largest cultural, economic and educational centres of south India.

According to the Indian census, it is the sixth-most populous city and fourth-most populous urban agglomeration in India. The traditional and de facto gateway of South India, Chennai is among the most-visited Indian cities by foreign tourists.

Being one of the most populous and frequently visited cities in India, Chennai has a large number of restaurants to cater to the needs of all of the people who are living in or visiting this place. Since there are large number of restaurants and people usually prefer to go to locations nearby, in this project, solving this problem is the main focus.

TARGET AUDIENCE:

This business idea mainly targets the audience who have just moved to the city and need to get to know places in and around. Since a lot of students from outside to Chennai for joining in their colleges, and a lot of people relocate to Chennai for their jobs and stuff, this is mainly intended for them to have a better experience in finding nearby restaurants.

This can also be used by residents when they move from one location to another, since Chennai is a vast city with multiple restaurants/dining places available.The restaurants may also get benefitted from this project, and the areas which do not have many restaurants nearby, can also be chosen to construct more restaurants by business people.

DATA:

The data which will be needed for this project is the coordinates of the various areas in Chennai. Web Scraping is done from any page having the areas of Chennai, using the Beautiful Soup package available in python. Beautiful Soup is a Python package for parsing HTML and XML documents. It creates a parse tree for parsed pages that can be used to extract data from HTML, which is useful for web scraping.
The coordinates of the areas will then be used to find out the various venues near the areas, using the FourSquare API.

The Foursquare Places API provides location based experiences with diverse information about venues, users, photos, and check-ins. The API supports real time access to places, Snap-to-Place that assigns users to specific locations, and Geo-tag. Additionally, Foursquare allows developers to build audience segments for analysis and measurement. JSON is the preferred response format.

With further processing techniques using Pandas, which is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool, built on top of the Python programming language, we can separate the restaurant venues, using "Venues Category".

Folium is a very powerful python library which lets you create several kinds of Leaflet maps. As Leaflet/folium maps are interactive, they are ideal for making dashboard building.

This library is used to display the 5 nearest venues in any random area selected, and also for visualising all of the areas and all of the venues in Chennai, simultaneously.

METHODOLOGY:

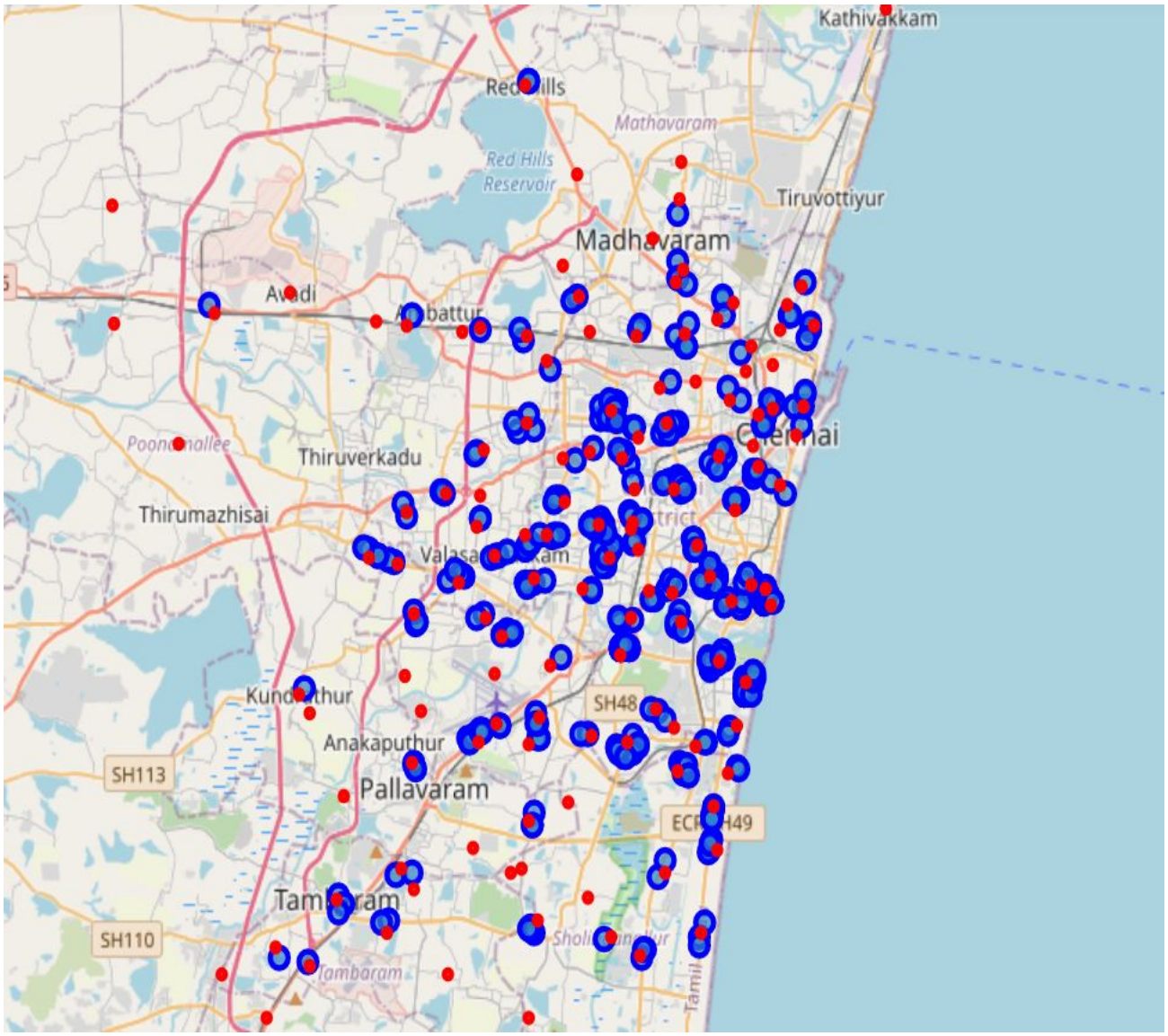
When the name of the areas in Chennai is extracted by utilizing the bs4 package, the latitude and longitude coordinates are obtained by using the Nominatim package.
The latitude and longitude are obtained so that FourSquare API can return venues near to each area in Chennai.
The venues, their latitude and longitude, and the venues' category, were returned. Now after reading the data in a pandas dataframe, the unique values in the venues category were returned in order to extract the venues which are restaurants/dining places.
For that, first, to make the selection of venues category, all of the venues category having the word 'restaurant' were added to another data frame. Then, the venues having the name 'store' were removed, because, as per observation, not a single one of them was a restaurant/dining place. Then the rest of the unique values in the venues category were studied and slowly each venue was removed, until all the venues categories left were only dining places/restaurants.
A map was used to display the areas and all the restaurants/dining places. The areas were marked in red while the restaurant/dining places were marked in blue.

The next step was to select a random area in Chennai, and display the nearest 5 restaurants. The euclidean distance between the latitude,longitude of the given area and the latitude,longitude of all of the venues were calculated. After this, the distances were sorted in index, in order to select the venues which had the smallest euclidean distance from the area.

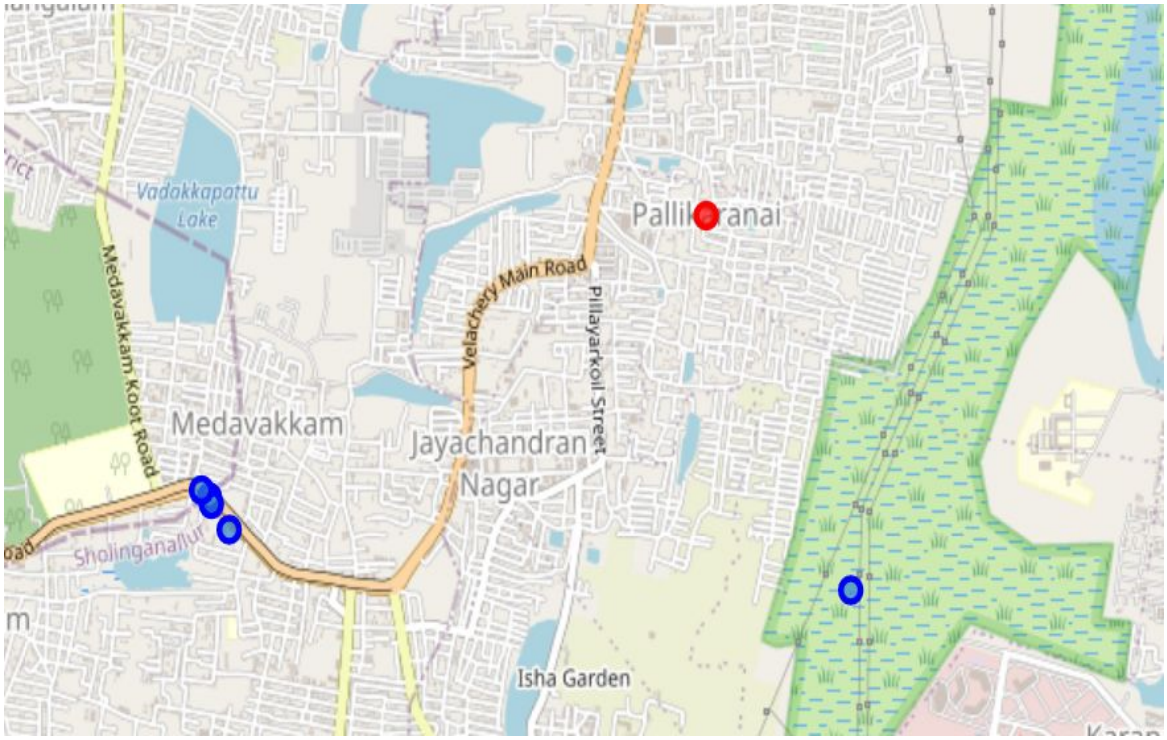
A folium map, centred at Chennai’s coordinates, was used to display the area and then the nearest 5 restaurants/dining places.

RESULT:

The result was 5 nearest restaurant/dining places to the given area. The results were visualized by using a folium map, and the venues were displayed using blue markers. The area was displayed using a red marker, to show the location of the user, so that they could easily visualize the destinations from their own location.



All the areas(red) in Chennai and the restaurant/dining places(blue).



Pallikaranai(area) in Chennai and the nearest restaurant/dining places(blue).

DISCUSSION:

An observation which was seen was, central Chennai tends to have more restaurants than the other parts. In some unpopular areas, the nearby restaurants also were located at a distance, like the map showing the example of places near Pallikaranai. This can be a motivation to business people to start closer restaurants. The best venue out of the nearest 5 restaurants could also be found, according to the user's preferences, like they could decide based on type of cuisine, since this map shows all types of cuisines, and this broadens their options.

CONCLUSION:

An introduction to the problem statement was given, and the reasons behind choosing this problem and the audience it targets was also given. Then the data which was used to solve the problem statement was also introduced, and the various methods of data analysis techniques were walked through. Finally, the result, and the method in which it was obtained, were discussed. This solution can also be developed further by considering the ratings and stuff, which were not taken into account, here. Overall, it is a useful product developed for the target audience discussed above.

REFERENCES:

1. https://en.wikipedia.org/wiki/Areas_of_Chennai
   Data source
2. https://developer.foursquare.com/
   To retrieve venues near the given areas from FourSquare API
3. https://pandas.pydata.org/docs/
   To perform data cleaning and analysis on given data
4. https://python-visualization.github.io/folium/#
   To display the results obtained