SDG: WASTE MANAGEMENT

GAURAV PANDEY
Department of Computer Science
Ramakrishna Mission Vivekananda Educational and Research Institute
Belur Math, Howrah
Pin – 711202, West Bengal

A report submitted to
Ramakrishna Mission Vivekananda Educational and Research Institute in partial fulfilment of the requirements for the degree of
MSc in Big Data Analytics
2021

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my faculties for providing their invaluable guidance, comments and suggestions throughout the course of the project. I would specially thank Dr. Sudipta Das (Assistant Professor) and Riya Arora (Mentor) for helping me and constantly motivating me to work harder and because of that I am able to finish my term project at iTech Mission Pvt. Ltd.

Also, I would like to thank my Head of Department Swamy Dhyanagamyananda for his assistance and support.

Ramakrishna Mission Vivekananda Educational and Research Institute, Belur Math, Wet Bengal

Date: Jun 17, 2021

Gaurav Pandey

CERTIFICATE

This is to certify that the report entitled 'SDG: Waste Management' submitted by Mr Gaurav Pandey, who has been registered for the award of MSc in Big Data Analytics degree of Ramakrishna Mission Vivekananda Educational and Research Institute, Belur Math, Howrah, West Bengal is absolutely based upon his/her own work under the supervision of Mrs. Riya Arora of Lead of Data Team, iTM,Delhi and that neither his/her report nor any part of the report has been submitted for any degree/diploma or any other academic award anywhere before.

Date: Jun 17, 2021

Mrs. Riya Arora
Sr. Project Manager
iTech Mission Pvt. Ltd.

TABLE OF CONTENTS

ABSTRACT

This report is about using the data fetched from the world bank, iTM data server and Foursquare API. From this data we will try to find out the cleanest cities around the world because in the world bank SDG (Sustainable development Goals) Climate change and Environment protection is one of the major indicators and issues currently facing many countries. We will discuss further about the climate change, waste management at various geo-locations and the logic that I used was — if there exists a clean city then it is has a highly efficient waste management facilities and there will be strong and successful management practices behind it and those management practices will be governed by strong pro-environmental and socially-committed rules and policies.

Calgary is one of example of such cities, which is considered as world's cleanest city by the United Nations and various international reports.
They are getting the number one position in the lieu of this category because of their waste management practices performed by the authorities.

We will adopt such a model to conduct the similar search and find the neighbourhood location which can give us the scientific locations to set up the waste management units and process the cities waste without/minimal causing any harm to the city.

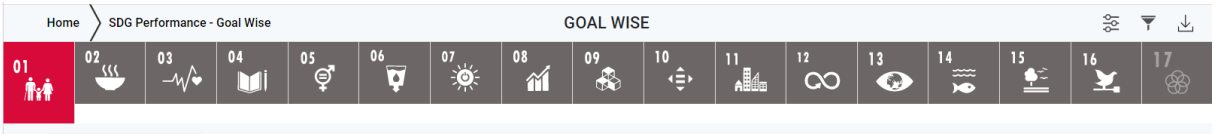
LIST OF FIGURES

INTRODUCTION

In this project, the data is originally recorded by the team of researchers from the world bank. The dataset has so many missing values and the reason provided by the researchers for missing values is the no proper facilities to collect data, data loss and bad situation (war like situation) in the countries.
The data covers every aspect of sustainable development goal with the target of achieving these goals by 2030.

There are 17 SDG Goals target provided by the world bank, UNDP and other international organisation on which mostly countries follow and prepare the reports accordingly.

In India, NITI Aayog (a subsidiary of GoI) prepares this report. We have taken substantial portion of these goals to work on this project.



When we are dealing with the global index, SDGs Goal are the important features to consider for development perspective. This year on National Statistics Day government of India realises the importance of these goals and made a commitment to work towards these goals and even they themed SDG-2 (Eradicate Hunger, Achieve Food Security and Improved nutrition and Promote Sustainable Agriculture) of the NSD.

Amongst one of the many goals Environment Protection and Improvisation in Standard of Living (which is SDG:11) is also a major concern. We are going to focus on this topic.

To focus on this topic, we will use the Geo-spatial Data Analysis technique because we do not have much data to work on and whatever data we have can not be used due to technical reason. So, to perform our task we will use a different technique to create dataset (using the Geospatial maps).

Our major area of focus on two cities i.e., East Calgary (Canada) and Kolkata (India).

East Calgary- This city considered as the cleanest city in the world and its waste management facilities are so well regulated which improves the standard of living of the people living in the neighbourhood of East Calgary compare to the rest of the cities in the world.

Kolkata- This city has the one of the highest population densities in all over Indian cities. The city ranked 3rd dirtiest city in India amongst the major metro cities of India. There are many complaints has been seen in the news regarding the waste management systems of this city which degrades the quality of life of the people living in this city.

PROBLEM DEFINITION

We aim to find an ideal spot for the waste management and disposal for Kolkata city, one of highest population dense city in India (growing city means growing waste). An ideal plant should be able to accommodate the input of waste (Only commercial) and then find a optimal way to treatment. The city's social and economic conditions should not be interrupted. The impact of this project would increase the living standards of the residents and it will save a lot of monetary spending by WB of govt. in long term. Meanwhile the environmental benefits would encourage the other cities to adopt the similar model fot the waste management and disposal system.

DESCRIPTION OF METHOD

METHODOLOGY:

The first step in this project is to find the best neighbourhood that Kolkata city would like to simulate. For that purpose, a simple Google search was done and it was found that the city of Calgary, Canada has consistently been ranked as one of the world's cleanest cities and it continues to be till 2025 [3]. It was found that Calgary city has two kinds of waste management facilities - one that deals with commercial waste and the other with industrial one. This study mainly focuses on commercial waste associated with Kolkata city. The City of Calgary has two facilities that manage commercial waste - East Calgary and Spyhill. This study takes into consideration the East Calgary facility's neighbourhood based on simple convenience.

FOURSQUARE API

Created a developers account on the Foursquare's website, follow the link below to create the similar developers' account. https://developer.foursquare.com.

My credential for the account would be as followed which I attached to the codes to fetch the latitude and longitude and create a dataset.

Credentials of Foursquare:
{"access_token":"ZL2M3AYM5YNSOD0GNHFFP1RSTZZXE1YNBGEEFOAOVJXEOKBL"}
Client ID=
ZKOAEG30WMUY1OKP0P5PLWXYFEBI3NY5IWXPYGIKIYI00NZS
Client Secret=
VLIYA3WIUACVHTCY5ZKOSDUNF1LGGUNGMX514ZFML3ZJ3N3F

K-MEANS

***k*-means clustering** is a method of vector quantization, originally from signal processing, that aims to partition *n* observations into *k* clusters in which each observation belongs to the cluster with the nearest mean (cluster centers or cluster centroid), serving as a prototype of the cluster. This results in a

partitioning of the data space into Voronoi cells. *k*-means clustering minimizes within-cluster variances (squared Euclidean distances), but not regular Euclidean distances, which would be the more difficult Weber problem: the mean optimizes squared errors, whereas only the geometric median minimizes Euclidean distances. For instance, better Euclidean solutions can be found using k-medians and k-medoids.

CLUSTERING OF NEIGHBOURHOOD

Geospatial clustering is the method of grouping a set of spatial objects into groups called "clusters". Objects within a cluster show a high degree of similarity, whereas the clusters are as much dissimilar as possible. The goal of clustering is to do a generalization and to reveal a relation between spatial and non-spatial attributes. K-means clustering is one of the simplest and popular unsupervised machine learning algorithms.
In other words, the K-means algorithm identifies *k* number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible.
The *'means'* in the K-means refers to averaging of the data; that is, finding the centroid.

WEBSCRAPPING

Web scraping is an automated method used to extract large amounts of data from websites. The data on the websites are unstructured. Web scraping helps collect these unstructured data and store it in a structured form. There are different ways to scrape websites such as online Services, APIs or writing your own code.
When we run the code for web scraping, a request is sent to the URL that you have mentioned. As a response to the request, the server sends the data and allows you to read the HTML or XML page. The code then, parses the HTML or XML page, finds the data and extracts it.
Libraries used for scraping: Pandas, Selenium, BeautifulSoup

Tools used:

A. Pandas - data frame creation
B. NumPy - vectorized computation
C. Matplotlib - simple visualization, colour palette selection
D. Requests - obtain webpages
E. Folium - Geospatial visualization
F. Sci-kit Learn - Machine learning
G. BeautifulSoup - Web-scraping
H. Geopy - mining geolocation data

ASSUMPTIONS

- East Calgary is an ideal location to choose for waste management.

- East Calgary facility does not disrupt the surrounding in any aspect (Social or Economic).

- This study assumes geographical, political, climatic, and demographically difference between the both cities (i.e., east Calgary and Kolkata) are negligible.

- Population density and environment hotspots are not considered in dataset.

These assumptions are very important to follow and considering these assumptions correct we are proceeding further in the project work.

EXPERIMENTAL EVALUATION

DATA:

This project will make use of Foursquare API for two data collection purposes. One, to find the venues surrounding the waste management plant of one of the world's cleanest cities and two, to generate a dataset that comprises the neighbourhoods and venues in the city of Kolkata, West Bengal, India.

East Calgary Facility location will be obtained from the *geopy* package and Foursquare API will be then used to mine the nearby venues within 5000 metres radius. Then the East Calgary Facility datapoint will be created by representing the facility using the venue categories of each venue mined in the form of average value of one-hot encoded score. The East Calgary Facility datapoint is shown in Figure.1 (due to space constraint only few columns are depicted):

	Neighborhood	Fast Food Restaurant	Accessories Store	Asian Restaurant	Pizza Place	Italian Restaurant	Vietnamese Restaurant	Korean Restaurant	Gym / Fitness Center	Falafel Restaurant	Movie Theater	Brewery	Warehouse Store
0	East Calgary Facility	0.071429	0.071429	0.071429	0.071429	0.071429	0.214286	0.071429	0.071429	0.071429	0.071429	0.071429	0.071429

Figure 1.1: East Calgary Neighbourhood

Kolkata city will be defined using the neighbourhoods of district and the neighbourhood's will be web scraped using BeautifulSoup from the following website: 'https://worldpostalcode.com/india/west-bengal/kolkata/'. The neighbourhood names scraped will be then used to mine the geolocation data of each of these using the *geopy* package. A small excerpt from the scraped data is shown in Figure.1.2. The web scraped dataset will be then processed to eliminate any missing values and then fed to Foursquare API to obtain nearby venues for each neighbourhood. Similar to the East Calgary Facility datapoint, neighbourhoods of Kolkata will also be processed and few parts of it is represented in Figure.1.3.

	Neighborhood	lat	lng
0	A.j.c.bose Road	23.5735	87.3276
7	Archana	25.3461	55.3848
9	Asylum Lane	39.2641	-76.7297
11	Baghajatin	22.484	88.3756
12	Baghbazar	27.7058	85.3193
13	Bakery Road	40.6669	-76.191
14	Ballygunge	22.5205	88.3722
18	Barabazar H.o	22.5814	88.3528
19	Baranagar	22.6531	88.3806
20	Barisha	36.1656	36.636

Figure 1.2: Web scrapped data of Kolkata's Neighbourhood

	Neighborhood	Accessories Store	Asian Restaurant	Brewery	Falafel Restaurant	Fast Food Restaurant	Gym / Fitness Center	Italian Restaurant	Korean Restaurant	Movie Theater	Pizza Place	Vietnamese Restaurant	Warehouse Store
0	A.j.c.bose Road	0.5	0.000000	0.000000	0.000000	0.500000	0.000000	0.000000	0.000000	0.0	0.000000	0.00	0.0
1	Archana	0.0	0.285714	0.000000	0.000000	0.571429	0.000000	0.000000	0.000000	0.0	0.142857	0.00	0.0
2	Asylum Lane	0.0	0.083333	0.000000	0.000000	0.083333	0.166667	0.166667	0.083333	0.0	0.166667	0.25	0.0
3	Baghajatin	0.0	0.200000	0.000000	0.100000	0.100000	0.000000	0.000000	0.000000	0.2	0.400000	0.00	0.0
4	Baghbazar	0.0	0.500000	0.000000	0.000000	0.142857	0.000000	0.285714	0.000000	0.0	0.071429	0.00	0.0
5	Bakery Road	0.0	0.000000	0.058824	0.000000	0.411765	0.117647	0.058824	0.000000	0.0	0.352941	0.00	0.0
6	Ballygunge	0.0	0.111111	0.111111	0.111111	0.111111	0.000000	0.222222	0.000000	0.0	0.333333	0.00	0.0
7	Barabazar H.o	0.0	0.100000	0.000000	0.000000	0.400000	0.000000	0.100000	0.000000	0.0	0.400000	0.00	0.0
8	Baranagar	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	1.000000	0.00	0.0
9	Bartala	0.0	0.111111	0.000000	0.000000	0.444444	0.000000	0.000000	0.000000	0.0	0.444444	0.00	0.0

Figure 1.3: Kolkata's District Neighbourhood data

The East Calgary Facility datapoint will be concatenated with the Kolkata's district neighbourhood data frame and K-means clustering with 4 clusters will be performed to find the suitable location choices for Kolkata city waste management.

DATA- PREPROCESSING:

- Fetching the nearby venues of Kolkata similar to the East Calgary. We got 855 similar venues as we have in the East Calgary data.
- Visualizing some of these venues by plotting on the Kolkata map.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	A.j.c.bose Road	23.573490	87.327609	KFC	23.533700	87.322900	Fast Food Restaurant
1	A.j.c.bose Road	23.573490	87.327609	World of Titan	23.574926	87.304960	Accessories Store
2	Archana	25.346113	55.384793	Eat Sol	25.334909	55.375804	Fast Food Restaurant
3	Archana	25.346113	55.384793	Novo Cinemas	25.330886	55.375792	Movie Theater
4	Archana	25.346113	55.384793	The Pizza Company	25.366012	55.402368	Pizza Place
...	...	...	...	...	...	...	...
850	Writer's Building	34.090189	-118.346595	Angelini Osteria	34.076286	-118.349140	Italian Restaurant
851	Writer's Building	34.090189	-118.346595	Jon & Vinny's	34.078633	-118.361351	Italian Restaurant
852	Writer's Building	34.090189	-118.346595	Terroni	34.076204	-118.355488	Italian Restaurant
853	Writer's Building	34.090189	-118.346595	Cinerama Dome at Arclight Hollywood Cinema	34.097512	-118.328131	Movie Theater
854	Writer's Building	34.090189	-118.346595	Joe's Pizza - Hollywood Blvd	34.101493	-118.331331	Pizza Place

855 rows × 7 columns

Figure 1.4: Pre-Processed Data of Kolkata

DATA DESCRIPTION:

Table 1

VARIABLES	DESCRIPTION
Neighbourhood	Districts of Kolkata (PIN code-wise)
Neighbourhood Latitude	Latitude of the Area (District)
Neighbourhood Longitude	Longitude of the Area (District)
Venue	Location/ shops or famous points in the district of the city where waste can be generated.
Venue Category	Type of the location from waste is generated like restaurant, salon, etc.
Venue Latitude	Latitude of the particular venue/ location
Venue Longitude	Longitude of the particular venue/ location

DATA ANALYSIS:

The entire project was then split into 5 steps:

1. **Generate East Calgary Facility datapoint** - Get East Calgary Facility location data and represent East Calgary landfill with respect to the surrounding venues using Foursquare explore option and generate a pandas data frame of single row representing the East Calgary Facility. This is depicted in Figure.1

2. **Generate Kolkata neighbourhood dataset** - Firstly, the list of neighbourhoods of Kolkata is web scraped using *BeautifulSoup library* from the *bs4* package (Figure.1.2). Then using geopy, the location data for these neighbourhoods are extracted. The neighbourhood data is then fed to Foursquare API to mine 100 venues from each of the neighbourhoods. This step will generate a pandas data frame that represents Kolkata district in terms of neighbourhoods. This dataset is represented in Figure.1.3

3. **Prepare the clustering dataset** - Calgary and neighbourhoods of Kolkata will have differences in venue categories and hence the datasets must be pre-processed for clustering application. This pre-processing involves elimination of those neighbourhoods with missing geolocation data and removal of non-overlapping venue categories to ensure the neighbourhood of East Calgary and that of Kolkata are a done on a common ground. After pre-processing, East Calgary Datapoint and Kolkata neighbourhood dataset are concatenated to develop the clustering dataset. Then clustering will be performed using Sci-kit learns implementation of K-Means API.

4. **Prepare the clustered dataset of Kolkata** - The cluster labels generated in the previous step will be inserted into the clustering data frame.

5. **Filter out the neighbourhoods similar to East Calgary Facility** - the cluster label for East Calgary Facility will be determined first and using that label, neighbourhoods of Kolkata having the same label will be sliced from the data frame.

We will discuss about these steps in further details. These steps are just provided to give the overview of the project.

Now, let's discuss about the results we get after performing the above data analysis.

RESULTS:

Initially, Foursquare API explore endpoint option revealed 14 venues within 5000 metres of East Calgary Facility. The visualization of those 14 venues done using the *Folium* package is given in Figure 2.

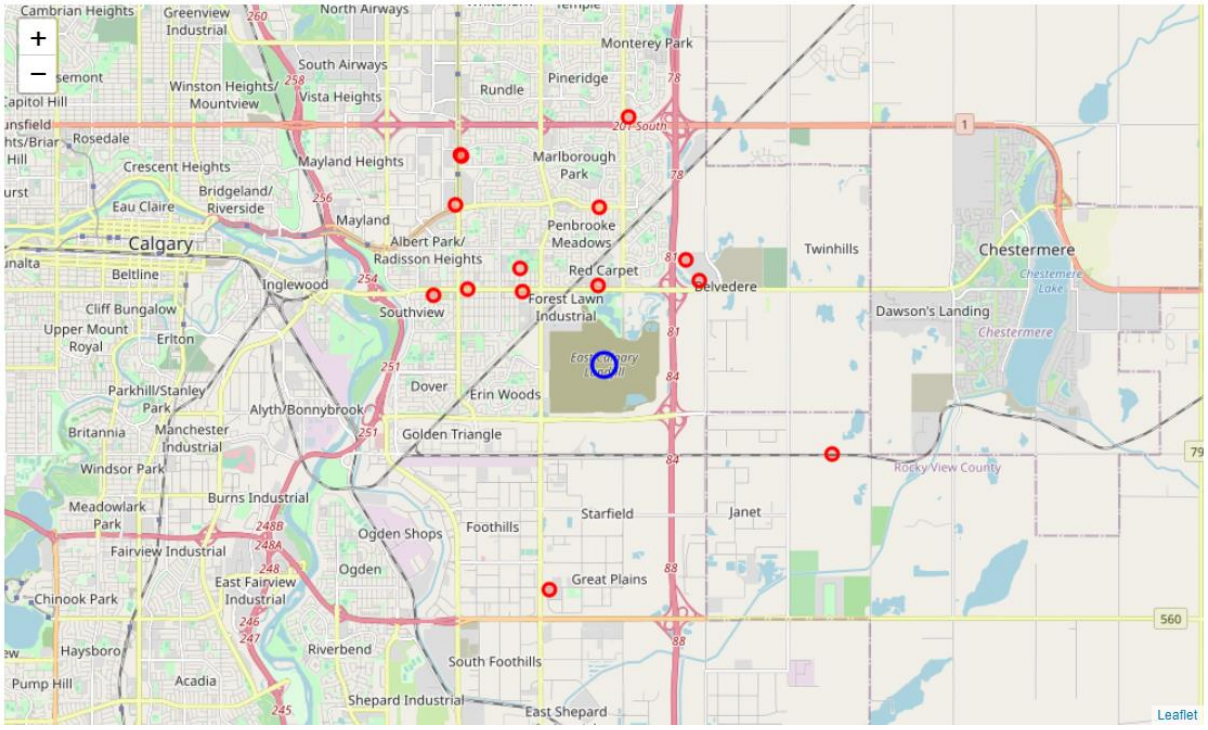


Figure 2: Nearby Venues of East Calgary

Kolkata neighbourhood geolocation data extraction using *geopy* initially resulted in 218 venues but was reduced to 155 after elimination of some neighbourhoods without geolocation data was ignored. Using these 155 neighbourhoods of Kolkata, Foursquare API was used to extract venues from each of them and yielded a pandas data frame with 2730 venues in total. To generate the clustering dataset, both East Calgary datapoint as well Kolkata neighbourhood dataset had to be refined into having identical venue categories. This operation resulted in 855 venues in Kolkata neighbourhood dataset as there were only 122 common venues between the two places. The visualization of the refined Kolkata neighbourhood dataset is given in Figure.2.1.

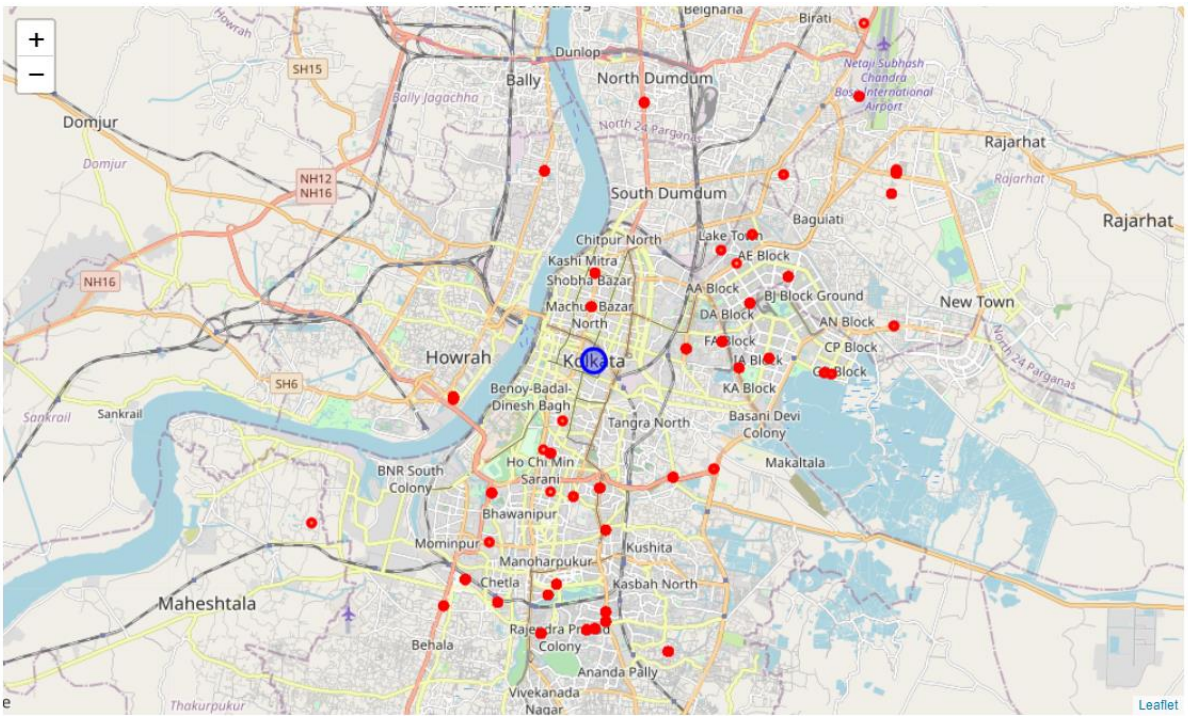


Figure2.1: Some Neighbourhood Venues of Kolkata

The two pre-processed datasets (i.e., East Calgary and Kolkata) were combined to generate a clustering dataset. This then immediately followed an Elbow Curve analysis to obtain the optimal number of clusters to be considered. The plot for the analysis is given in Figure.2.2. As it became evident from the plot that cluster number beyond 4 does not cause significant reduction in segmentation loss, cluster count was fixed as 4.

Clustering was then performed and its outcome is depicted in Figure.2.4. A standard Folium marker (with green colour) is used to pinpoint the current waste management facility at Dhapa. Finding out to which cluster did East Calgary Facility fall was the next step and it was found to be the second cluster (labelled as 0). There were 36 neighbourhoods of Kolkata that showed similarity to that of East Calgary Facility.

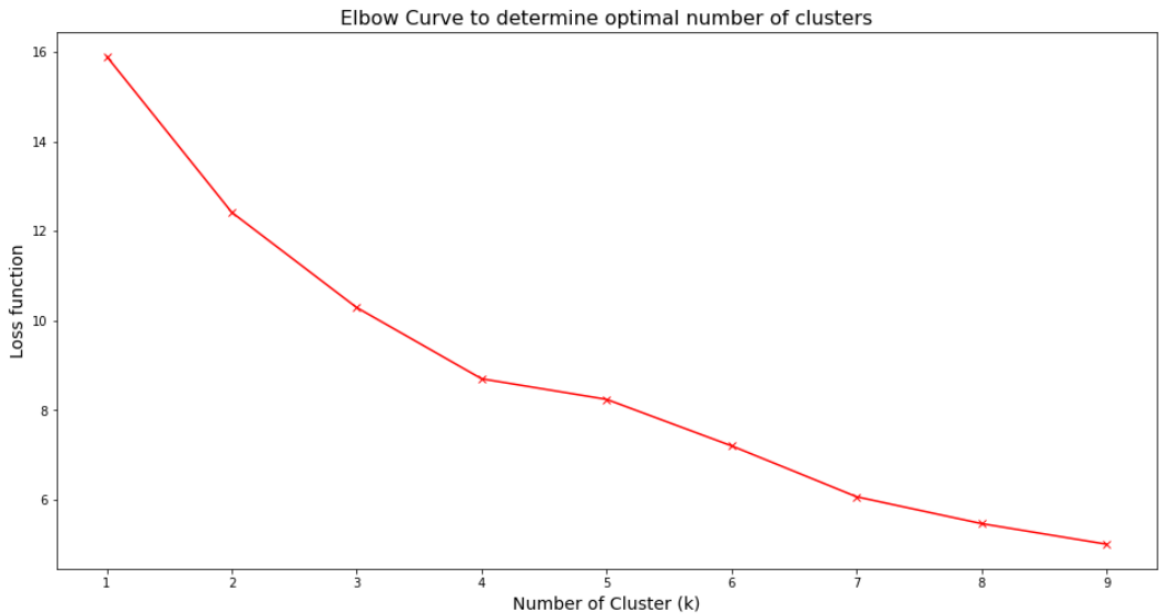


Figure 2.2: Elbow Curve Plot

Clusters we have created by choosing k=4. We have marked each cluster on the Kolkata City map with the different colours so that it is easy to identify the cluster similar to the East Calgary city.

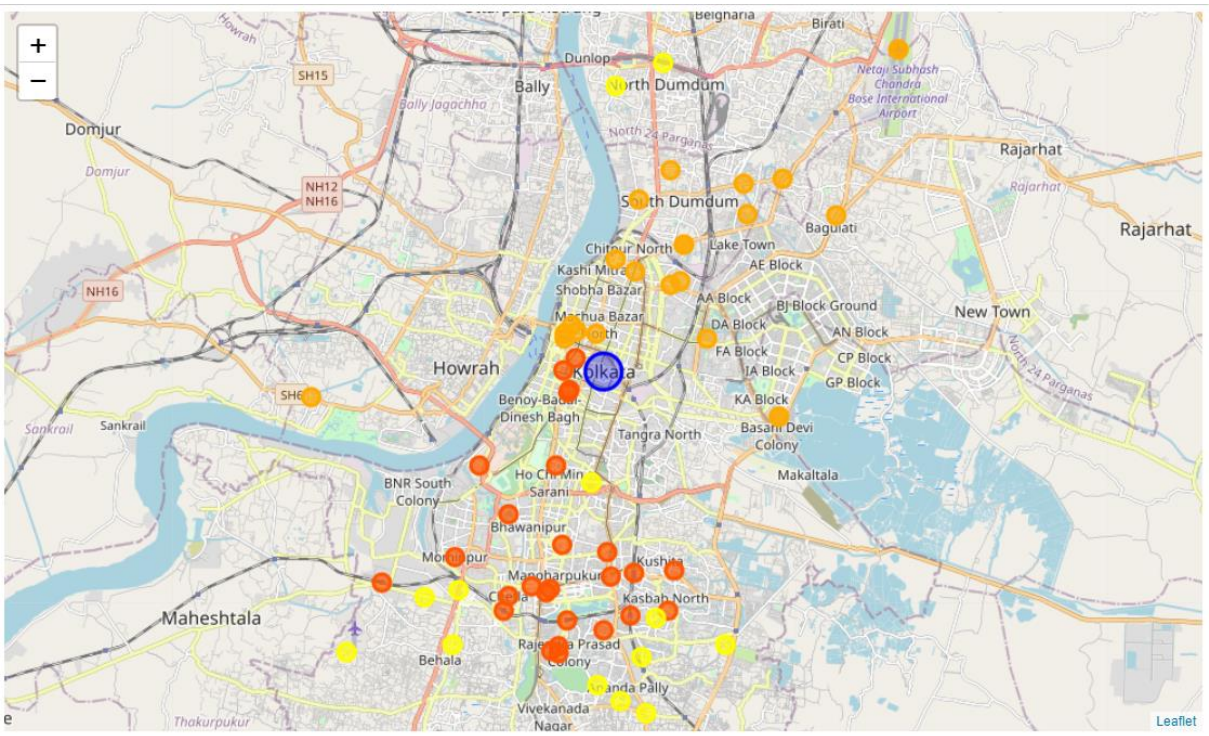


Figure 2.3: Cluster of Kolkata Neighbourhood

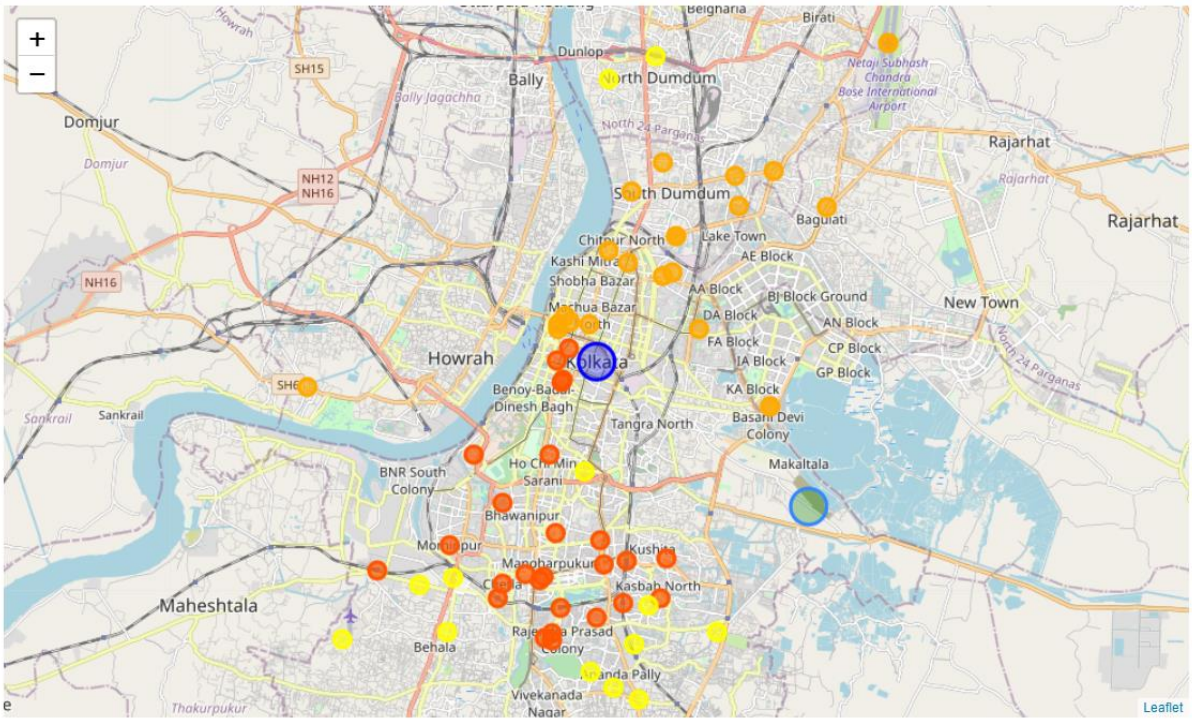


Figure 2.4: Cluster with Current largest Waste Management Facility of Kolkata

In the Figure 2.4, The green marker on the Kolkata city's map shows the current largest facility of the Kolkata city which is facing the issues at various levels and many reporting's of landfills are being registered by the Kolkata Municipal Corporation.

```
Number of suitable waste management facility locations for Kolkata city is :  36
```

	Cluster_labels	Neighborhood	lat	lng
0	0	Asylum Lane	39.2641	-76.7297
1	0	Baghajatin	22.484	88.3756
2	0	Baranagar	22.6531	88.3806
3	0	Beadon Street	51.6816	-3.38153
4	0	Bijoygarh	22.4873	88.3686
5	0	Bosepukur Road	22.4346	88.4152
6	0	Brahma Samaj Road	52.6469	-1.12441
7	0	Council House Street	52.4802	-1.90281
8	0	Customs House	-33.8621	151.211
9	0	Deshbandhu Road	22.6469	88.3673

Figure 2.5: Most Suitable locations to build waste Management Facility in Kolkata

In Figure 2.5, we have shown the suitable locations where the waste management facility set-up would be fruitful and environment friendly as well as cost effective for the government of West Bengal.

Narrowing down these locations to few of the best locations and markings on the Kolkata map. If the distance seems to be the very close but the actual distance from visualization varies a lot. Don't be confused with the distance visualized in map and the actual distance between the locations.

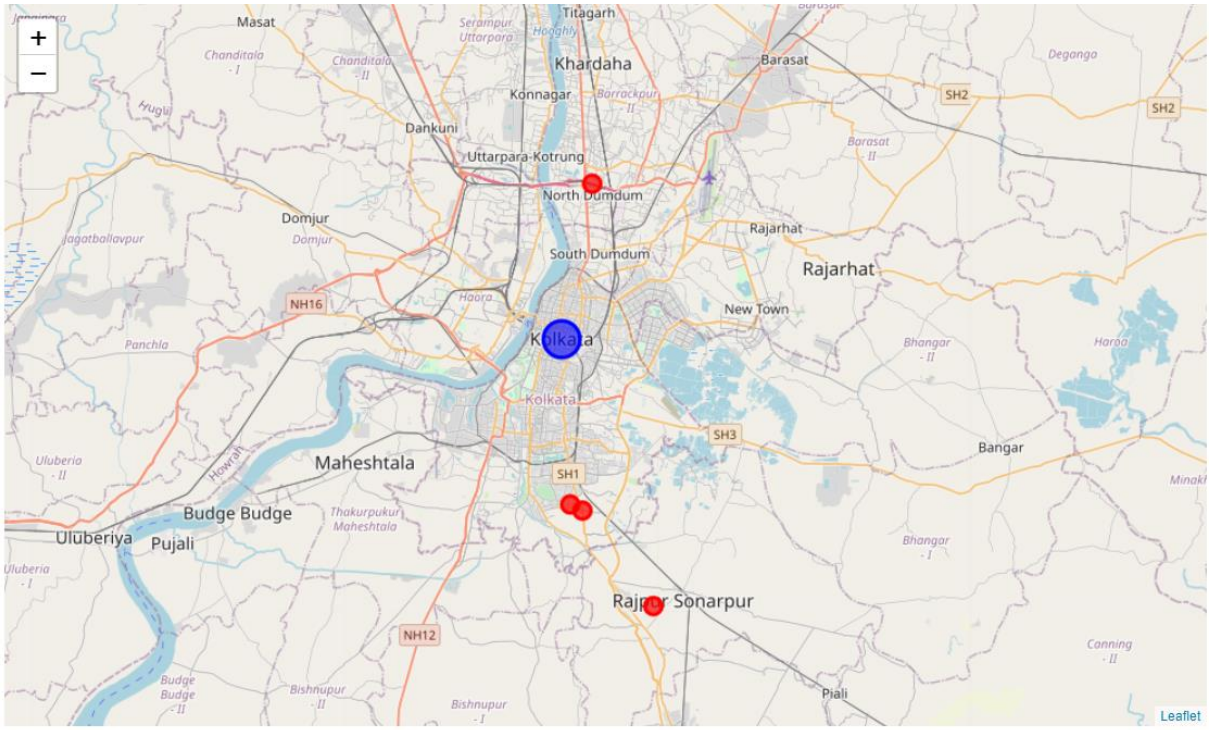


Figure 2.6: Some of the locations plot on the map

These red dot shows the some the best suitable locations from the locations we have found above in Figure. 2.5.
Now, we haven't taken all the 36 locations on the map and there is no such logical reasoning behind that it is just show the particular locations openly on the map.

Now, the main question arises that can we narrow down the the suitable locations we have found after clustering the dataset using the K-means algorithm?
So, the answer is Yes, we can narrow down locations. But in order to do it we need to consider all the factors such as economic conditions of the surroundings as well as the social conditions of the surroundings of each location only after

that it is possible to narrow down the locations for the suitable waste management facility. We can not do this here in this project because for that we require the more information or data regarding the factors we need to consider.

CRISIS REPORTED AT THE CURRENT WASTE MANAGEMENT FACILITY (DHAPA):

As we have shown above in the Figure 2.4, with the green marked circle it is showing the current waste management facility at Dhapa in Kolkata and many problems reported at this facility. Some of the newspaper cuttings of ToI attached in the image below which shows the disastrous conditions of the current waste management facility at Dhapa.

Kolkata's Landfill Crisis: The City's Dependency On The Sole Landfill Of Dhapa May Soon Result In An Unresolvable Crisis

At Kolkata's sole landfill, more than 4,000 metric tonnes of waste go in every day, despite the landfill's lifespan being over decades ago

Landfill Crisis | Written By: **Saptarshi Dutta** | Edited By: **Sonia Bhaskar** | September 08, 2017 11:23 AM | 0

As the KMC is determined to reduce health hazards from the Dhapa dumping ground, a search for an alternative land is also on. "Going by the present rate of reclamation, it will take much long to reclaim the entire 23 hectares of Dhapa. We need an alternative land that could be set up with a sanitary landfill facility to take the load off Dhapa," a KMC official said. The civic body has stopped dumping waste on 12 hectares of land in the crore Dhapa area. The portion has been beautified and a green zone created there.

A civic official pointed out that the capacity of the compost plant would be enhanced soon. "In the first phase, we are recycling 300 tonnes of garbage a day. The capacity of the compost plant will be increased to 900 tonnes a day in three months," he said.

The official said 100 tonnes of manure were produced from 300 tonnes of wastes every day. This apart, the compost plant, set up on a PPP model, receives 500 tonnes of biodegradable waste collected from Kolkata households every day.

This newspaper cutting shows the current condition of the waste management facility at Dhapa which affects the daily livelihood of the people of Kolkata.

DISCUSSIONS:

Kolkata city is the capital of West Bengal, India. Kolkata city is facing a serious crisis in the area of city waste management and the current facility is being criticized by domain experts, environment as well as the residents of Dhapa for the unscientific location of the plant, the waste management practices and the catastrophes awaiting the state like ground-water contamination, climate change and damage to ecosystem. This study recognized the requirement of the city and implemented machine learning to extract the best possible locations to set up a new waste management plant- one that mitigates all the aforementioned problems. In short, this project demonstrates a scientific way to easily identify suitable locations for meeting city waste management needs using Foursquare API and simulating the neighbourhoods of one of the world's cleanest city's waste management plants, East Calgary Facility of the city of Calgary, Canada. This method thus minimizes the time required for city managers to narrow down possible locations by doing individual site visits which in turn, reduces initial plant set up costs and saves considerable time spent on making policies that govern the location and vicinity of waste management plants - the city managers can easily borrow the policies from one of the cleanest cities and tune it somewhat based on the requirements. And this solution can be replicated easily and is the greatest strength of this method- this means that this method can be implemented anywhere regardless of districts, states or countries. Hence, this solution will do justice to society, government and to nature. Now with this scientific method of locating the waste management plant, the authorities can focus all the efforts and resources onto identifying the best possible way of managing city waste - one step less towards making the world a better place.

The ML approach used in this project relies on the assumption that the geographical, political and social climate of both Calgary city and Kolkata city are same and thus the regulations which operate the plant at Calgary city is expected to run the Dhapa facility of Kolkata city with the same efficiency. This may not be that realistic. In addition to this, the population density and vegetation patterns of two cities vary and the final recommendation can change based on those considerations and the pro-environment regulations enforced at Kolkata city. Another caveat is the use of Foursquare API. If the rules dictating erection of commercial or residential ventures nearby the waste management facility are not strict, Foursquare venues will be a poor indicator of any neighbourhoods. In short, the results of this study must be subjected to further filtering to arrive at an optimal solution.

CONCLUSION

This study will help city managers and city planners to make the best decision regarding the first step towards city waste management - optimal location of the plant that does not interfere with social or economic aspects of the city. This endeavour will allow them to focus onto the more pressing issue - identify or research the best possible way to treat the city waste and hence mitigate the collateral damage done to nature and social life. The present study demonstrates the possibilities of Machine Learning based problem solving for better city administration.

We can further continue this project to do the similar analysis on the other cities of India. Kolkata is not the only city which is considered the dirtiest one even though at world's Index out top 50 dirtiest cities in world most of them are from India, China, Pakistan and Bangladesh.

POSSIBLE FUTURE EXTENSIONS TO THE PROJECT

This study implemented K-Means clustering approach while Density Based Spatial Clustering for Applications with Noise (DBSCAN) methodology can be utilized which might yield different and improved results from this study. Future extensions can include a study conducted with different location data vendors or inclusion of population density or local government regulations or local geography. Another interesting aspect of this project is that this solution can be transformed into a solution for other domains, where users can decide their use-case and the application can suggest suitable localities which are served to users via a web-based or smartphone app based big data platform that can deliver even better results.

REFERENCES

Top 10 cleanest cities in the world 2020:

- https://kenyaprime.com/top-10-cleanest-cities-in-the-world-2020-latest-ranking-things-you-need-to-know-about-the-cities-2/

Doing Business Group Report

- https://www.doingbusiness.org/en/reports/subnational-reports/india

United Nations Sustainable Development Programme

- https://unstats.un.org/sdgs/

Foursquare API- For developers

- https://developer.foursquare.com