

A PROJECT REPORT ON

**CLUSTERING APPROACH TO DATA MINING FOR
IDENTIFYING LOCALITIES IN PUNE,INDIA WHERE
THE HEALTH CARE INFRASTRUCTURE NEEDS TO
BE IMPROVED**

SUBMITTED BY

Name: VEDANT R. ASAWALE
(Data Science Enthusiast)

Under the guidance of

Alex Aklson



**IBM DATA SCIENCE PROFESSIONAL
CERTIFICATION**

Coursera Learning Platform

ACKNOWLEDGEMENT

I sincerely thank Coursera and IBM for providing students a platform to learn new skills and enhance their knowledge.

I also sincerely convey my gratitude to the course teacher Alex Aklson and all other teachers for their support, providing all the help, motivation and knowledge from beginning till end to make this learning path a grand success.

Above all I would like to thank my parents for their wonderful support and blessings, without which I would not have been able to accomplish any of my goals.

Contents

1	INTRODUCTION	1
1.1	Lack of Sufficient Health Care Infrastructure	1
1.2	Improving Health Care Infrastructure	1
2	PROBLEM STATEMENT	2
2.1	Statement	2
2.2	Target Audience	2
3	MOTIVATION	3
4	DATA	4
4.1	Importance of Data	4
4.2	Data used in the Project	4
4.2.1	Finding Localities in Pune	4
4.2.2	Getting the Geographical Coordinates	5
4.2.3	Fetching Hospital data	5

1 INTRODUCTION

1.1 Lack of Sufficient Health Care Infrastructure

The Corona virus outbreak in 2020 has exposed the lack of sufficient health care infrastructure throughout the World. Even the most developed economies such as the United States and Italy could not handle the unprecedented crisis. The virus outbreak not only lead to numerous deaths but also resulted in huge financial losses. Moreover it has given us a wake up call regarding how under prepared the humanity is for such kind of a crisis. There were situations wherein the Hospitals were completely filled and unfortunately no more people could be admitted for treatment eventually resulting in loss of human lives.

1.2 Improving Health Care Infrastructure

The society should take lessons from this crisis and prepare ourselves for any such disaster beforehand. This project will helps us in improving the Health care facilities by identifying localities where the number of hospitals are less so that a number of new hospitals can be constructed in these regions. In this project, analysis is done for regions from one city (Pune,India) only. However, the same analysis can be carried out for different regions as long as we can gather the right data.

2 PROBLEM STATEMENT

2.1 Statement

Identify the localities or regions in Pune city,India where the number of hospitals are less and hence there is urgent need to improve the Health care infrastructure to prepare for any epidemic or disaster crisis which might arise in future.

2.2 Target Audience

This project can be used by the Government or social workers to identify regions for creation of new Health Care facilities.



3 MOTIVATION

The motivation for creating such a project is driven by two main factors:

- 1.The damage caused by Covid-19.
- 2.As a Data science enthusiast its my responsibility to solve problems and make human life better using data.

The data being generated by various sources has increased multi fold over the past few years. Thus it has given rise to a completely new domain called Data Science which involves understanding and analyzing data to solve everyday as well as major problems. This technique can be used in various domains such as Business,Online Marketing,Spam Detection and many other.

The problems caused by Covid-19 were disastrous and we need to find ways to make sure that we are prepared for any such crisis.We can use data science in many different ways to do so and this project is just one such attempt.

4 DATA

4.1 Importance of Data

A model or any analysis is as good as the data which means that data is the prime and most important component of any analysis. The result produced wouldn't be on expected lines even by using the best of techniques and algorithms if the data provided is not accurate. Identifying the best sources of data, collecting it, extracting it and transforming it as per our needs are the major and highly important steps associated with Data Science. The data as well as the sources of data have increased exponentially over the years. Thus, for leveraging the value of this data, it's necessary to select the right sources and thereafter select the right features for better analysis.

4.2 Data used in the Project

The current project identifies localities in the Pune city, India where the Health Care facilities are not adequate.

Thus the following data is needed for our analysis:-

1. List of all the localities in Pune city.
2. The geographical coordinates for all these localities.
3. The number of hospitals present in each locality.

4.2.1 Finding Localities in Pune

List of all localities present in Pune is readily available on the Internet.

The URL "<https://www.mapsofindia.com/pune/localities/>" is used to get the list of all localities in Pune using the technique of web scraping via the BeautifulSoup API.

Pune Locality Maps		
▶ Alandi Road	▶ Hadapsar	▶ Navi Peth
▶ Ambegaon Budruk	▶ Hadapsar Industrial Estate	▶ Padmavati
▶ Anandnagar	▶ Hingne Khurd	▶ Parvati Darshan
▶ Aundh	▶ Jangali Maharaj Road	▶ Pashan
▶ Aundh Road	▶ Kalyani Nagar	▶ Paud Road
▶ Balaji Nagar	▶ Karve Nagar	▶ Pirangut
▶ Baner	▶ Karve Road	▶ Prabhat Road
▶ Baner road	▶ Kasba Peth	▶ Pune Railway Station
▶ Bhandarkar Road	▶ Katraj	▶ Rasta Peth
▶ Bhavani Peth	▶ Khadaki	▶ Raviwar Peth
▶ Bibvewadi	▶ Khadki	▶ Sadashiv Peth
▶ Bopodi	▶ Kharadi	▶ Sahakar Nagar
▶ Budhwar Peth	▶ Kondhwa	▶ Salunke Vihar
▶ Bund Garden Road	▶ Kondhwa Budruk	▶ Sasson Road
▶ Camp	▶ Kondhwa Khurd	▶ Satara Road
▶ Chandan Nagar	▶ Koregaon Park	▶ Senapati Bapat Road
▶ Dapodi	▶ Kothrud	▶ Shaniwar Peth
▶ Deccan Gymkhana	▶ Law College Road	▶ Shivaji Nagar
▶ Dehu Road	▶ Laxmi Road	▶ Shukrawar Peth
▶ Dhankawadi	▶ Lulla Nagar	▶ Sinhadgad Road
▶ Dhayari Phata	▶ Mahatma Gandhi Road	▶ Somwar Peth
▶ Dhole Patil Road	▶ Mangalwar peth	▶ Swargate
▶ Erandwane	▶ Manik Bagh	▶ Tilak Road
▶ Fatima Nagar	▶ Market yard	▶ Uruli Devachi
▶ Fergusson College Road	▶ Model colony	▶ Vadgaon Budruk
▶ Ganesh Peth	▶ Mukund Nagar	▶ Wadgaon Sheri

4.2.2 Getting the Geographical Coordinates

Geographical coordinates in ready form is difficult to find.

However,many APIs are available which can fetch the latitude and longitude of a region.

The Nominatim module from Geopy.Geocoders library was used to fetch coordinates for the fetched localities.

	Locality	Latitude	Longitude
0	Alandi Road	18.5523	73.8733
1	Ambegaon Budruk	18.4511	73.8376
2	Anandnagar	18.5083	73.8152
3	Aundh	18.5619	73.8102
4	Aundh Road	18.5619	73.8102
...
76	Viman Nagar	18.5214	73.8545
77	Wagholi	18.5806	73.9833
78	Wanowrie	18.4884	73.8987
79	Warje	18.482	73.8002
80	Yerawada	18.5656	73.8866

81 rows × 3 columns

4.2.3 Fetching Hospital data

There are many location data providers available.

We will use the FourSquare API to fetch the venue data for our localities.