**COVID19 VACCINATION CENTERS IN LIMA, PERU**

Angel Maldonado

September, 2021

# **1. Introduction**

## **1.1 Background**

The covid-19 pandemic has negatively impacted the health, economy, education, and other aspects of our society since the beginning of 2020. Peru has been one of the countries with the most infections and deaths per million inhabitants in the world, reaching almost 200k deaths since the start of the pandemic. Below are the accumulated indicators related to the impact of the disease throughout the country according to the Ministry of Health as of August 29, 2021.

|  |  |
| --- | --- |
| Measure | Result |
| Tests performed | 16 733 426 |
| Positive cases | 2 149 591 |
| Deaths | 198 263 |

**Source:** <https://www.gob.pe/institucion/minsa/noticias/514015-minsa-casos-confirmados-por-coronavirus-covid-19-ascienden-a-2-149-591-en-el-peru-comunicado-n-662>

## **1.2 The problem**

The Ministry of Health has installed many vaccination centers in most districts of the country, and since February 9th, 2021, the national vaccination strategy has been carried out, scheduling appointments for each inhabitant of the country by age groups, starting with the elderly and decreasing the age range over the weeks.

Recently, other risk groups are being included in the vaccination schedule (rare and orphan diseases and neurodevelopmental disorders) without age restriction (they must only be 12 years or older) in addition to the possibility of being vaccinated in a district other than their residence. This is causing alterations in the flow of people traveling to vaccination centers, resulting in two possible scenarios:

* Long queues outside the vaccination centers
* Empty vaccination centers

This project's objective is to suggest the number and location of vaccination centers based on the density of population in Lima. This can be used to redirect staff from vaccination centers to appropriate locations to avoid wasting health professional resources.

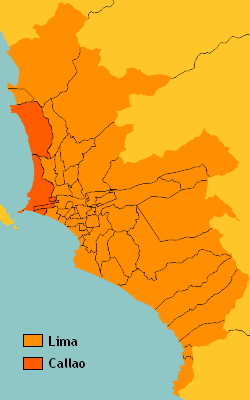
## **1.3 Intended audience**

* Directors and management staff of:
* Ministry of Health
* EsSalud (social insurance)
* District municipalities
* Private health centers

## **1.4 Geographic scope**

This project will cover the districts from Lima and Callao provinces.

* Image on the left: Lima location within the country
* Image on the right: Districts of Lima and Callao

# **2. Data Requirements**

The data to be used in this project comes from two principal sources:

* Open Data National Platform <https://www.datosabiertos.gob.pe/>
* Foursquare

## **2.1 Description of the data**

The following data will be required for this project:

* List of ubigeo equivalences (codes assigned to each district in Peru) available in the Open Data National Platform
* List of venues of Lima (extracted with Foursquare, from all districts in Lima)
* List of vaccination centers with location data (longitude/latitude) available in the Open Data National Platform

## **2.2 How the data will be used**

**List of ubigeo equivalences:** First it will be filtered with the location data from districts of Lima, these coordinates will be used to search venues with Foursquare. This dataset will also be merged with the list of vaccination centers. It has the codes given to each district, which are the data to filter the vaccination centers in this project's scope. This dataset has 17 columns from which the following are of interest for this project:

|  |  |  |
| --- | --- | --- |
| Column | Type | Description |
| ID Ubigeo | Number | Code assigned to districts |
| Department | Text | 1st level of local government, most of times called regional |
| Province | Text | 2nd level of local government |
| District | Text | 3rd level of local government |
| Region | Text | Administrative division, in most cases the same that Department |
| Surface | Number | District area in square kilometers |
| Latitude | Number | Location data |
| Longitude | Number | Location data |

Link: <https://www.datosabiertos.gob.pe/dataset/codigos-equivalentes-de-ubigeo-del-peru>

**List of vaccination centers:** This is a list of places where the vaccination process is being taken. The purpose to use this data is to plot the locations to analyze if the vaccination centers are near to a point of high population density. Some of the columns that will be used from this dataset are:

|  |  |  |
| --- | --- | --- |
| Column | Type | Description |
| ID Ubigeo | Number | District code where the vaccination center has been installed |
| Vaccination Center | Text | Name of the place that hosts the vaccination center |
| Latitude | Number | Location data |
| Longitude | Number | Location data |

Link: <https://www.datosabiertos.gob.pe/dataset/centros-de-vacunacion>

**List of venues of Lima:** It will be extracted with Foursquare to determine the concentration of dwellings by clustering the venues. For this project, due to the nearness between Lima and Callao, the districts in these two provinces will be included. This dataset will have the following columns:

|  |  |
| --- | --- |
| Column | Type |
| Venue Name | Text |
| Category | Text |
| Longitude | Number |
| Latitude | Number |

It will be deduced that each cluster of venues or dwellings is a point of high population density.

# **3. Data Acquisition and Preparation**

This section is divided in the following steps:

* Creating a dataframe with district data
* Creating a dataframe with vaccination centers data
* Creating a dataframe with venues data using Foursquare

First, we will download the list of *ubigeos* and vaccination centers from <https://www.datosabiertos.gob.pe/>. Then these files will be processed to get the dataframes to work.

## **3.1 Districts dataset**

Considerations:

* Lima and Callao provinces are LIMA PROVINCIA and CALLAO values in **region** column.
* A new column **num\_requests** will be created comparing the district surface to an area in a 3Km radius circle (~ 28.27 Km2).
* Column num\_requests will tell the number of times we will query to Foursquare for venues near to a 3Km radius area. If the division between district surface and 28.27 Km2 is more than 1, then num\_requests will be equal to 2, that means all the venues in the result of the first query will be used as central coordinates for new requests.

| id\_ubigeo | department | province | district | region | surface | latitude | longitude | num requests |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 690 | CALLAO | CALLAO | CALLAO | CALLAO | 46.0 | -12.0631 | -77.1469 | 2 |
| 691 | CALLAO | CALLAO | BELLAVISTA | CALLAO | 5.0 | -12.0625 | -77.1292 | 1 |
| 692 | CALLAO | CALLAO | CARMEN DE LA LEGUA REYNOSO | CALLAO | 2.0 | -12.0394 | -77.0903 | 1 |
| 693 | CALLAO | CALLAO | LA PERLA | CALLAO | 3.0 | -12.0658 | -77.1081 | 1 |
| 694 | CALLAO | CALLAO | LA PUNTA | CALLAO | 18.0 | -12.0728 | -77.1633 | 1 |
| 695 | CALLAO | CALLAO | VENTANILLA | CALLAO | 70.0 | -11.8772 | -77.1278 | 2 |
| 696 | CALLAO | CALLAO | MI PERU | CALLAO | 3.0 | -11.8550 | -77.1250 | 1 |
| 1281 | LIMA | LIMA | LIMA | LIMA PROVINCIA | 22.0 | -12.0453 | -77.0308 | 1 |
| 1282 | LIMA | LIMA | ANCON | LIMA PROVINCIA | 285.0 | -11.7739 | -77.1764 | 2 |
| 1283 | LIMA | LIMA | ATE | LIMA PROVINCIA | 78.0 | -12.0264 | -76.9214 | 2 |
| 1284 | LIMA | LIMA | BARRANCO | LIMA PROVINCIA | 3.0 | -12.1492 | -77.0217 | 1 |
| 1285 | LIMA | LIMA | BREÑA | LIMA PROVINCIA | 3.0 | -12.0589 | -77.0461 | 1 |
| 1286 | LIMA | LIMA | CARABAYLLO | LIMA PROVINCIA | 303.0 | -11.8903 | -77.0269 | 2 |
| 1287 | LIMA | LIMA | CHACLACAYO | LIMA PROVINCIA | 40.0 | -11.9753 | -76.7689 | 2 |
| 1288 | LIMA | LIMA | CHORRILLOS | LIMA PROVINCIA | 39.0 | -12.1769 | -77.0164 | 2 |
| 1289 | LIMA | LIMA | CIENEGUILLA | LIMA PROVINCIA | 240.0 | -12.1203 | -76.8142 | 2 |
| 1290 | LIMA | LIMA | COMAS | LIMA PROVINCIA | 49.0 | -11.9572 | -77.0494 | 2 |
| 1291 | LIMA | LIMA | EL AGUSTINO | LIMA PROVINCIA | 13.0 | -12.0483 | -77.0006 | 1 |
| 1292 | LIMA | LIMA | INDEPENDENCIA | LIMA PROVINCIA | 15.0 | -11.9972 | -77.0547 | 1 |
| 1293 | LIMA | LIMA | JESUS MARIA | LIMA PROVINCIA | 5.0 | -12.0756 | -77.0433 | 1 |
| 1294 | LIMA | LIMA | LA MOLINA | LIMA PROVINCIA | 66.0 | -12.0781 | -76.9167 | 2 |
| 1295 | LIMA | LIMA | LA VICTORIA | LIMA PROVINCIA | 9.0 | -12.0650 | -77.0308 | 1 |
| 1296 | LIMA | LIMA | LINCE | LIMA PROVINCIA | 3.0 | -12.0844 | -77.0303 | 1 |
| 1297 | LIMA | LIMA | LOS OLIVOS | LIMA PROVINCIA | 18.0 | -11.9914 | -77.0708 | 1 |
| 1298 | LIMA | LIMA | LURIGANCHO | LIMA PROVINCIA | 236.0 | -11.9358 | -76.6972 | 2 |
| 1299 | LIMA | LIMA | LURIN | LIMA PROVINCIA | 180.0 | -12.2747 | -76.8703 | 2 |
| 1300 | LIMA | LIMA | MAGDALENA DEL MAR | LIMA PROVINCIA | 4.0 | -12.0917 | -77.0672 | 1 |
| 1301 | LIMA | LIMA | PUEBLO LIBRE | LIMA PROVINCIA | 4.0 | -12.0781 | -77.0625 | 1 |
| 1302 | LIMA | LIMA | MIRAFLORES | LIMA PROVINCIA | 10.0 | -12.1217 | -77.0292 | 1 |
| 1303 | LIMA | LIMA | PACHACAMAC | LIMA PROVINCIA | 160.0 | -12.1872 | -76.8667 | 2 |
| 1304 | LIMA | LIMA | PUCUSANA | LIMA PROVINCIA | 37.0 | -12.4817 | -76.7975 | 2 |
| 1305 | LIMA | LIMA | PUENTE PIEDRA | LIMA PROVINCIA | 73.0 | -11.8667 | -77.0769 | 2 |
| 1306 | LIMA | LIMA | PUNTA HERMOSA | LIMA PROVINCIA | 120.0 | -12.3336 | -76.8242 | 2 |
| 1307 | LIMA | LIMA | PUNTA NEGRA | LIMA PROVINCIA | 131.0 | -12.3653 | -76.7956 | 2 |
| 1308 | LIMA | LIMA | RIMAC | LIMA PROVINCIA | 12.0 | -12.0422 | -77.0269 | 1 |
| 1309 | LIMA | LIMA | SAN BARTOLO | LIMA PROVINCIA | 45.0 | -12.3892 | -76.7808 | 2 |
| 1310 | LIMA | LIMA | SAN BORJA | LIMA PROVINCIA | 10.0 | -12.1072 | -76.9989 | 1 |
| 1311 | LIMA | LIMA | SAN ISIDRO | LIMA PROVINCIA | 11.0 | -12.0978 | -77.0272 | 1 |
| 1312 | LIMA | LIMA | SAN JUAN DE LURIGANCHO | LIMA PROVINCIA | 131.0 | -12.0297 | -77.0100 | 2 |
| 1313 | LIMA | LIMA | SAN JUAN DE MIRAFLORES | LIMA PROVINCIA | 23.0 | -12.1636 | -76.9636 | 1 |
| 1314 | LIMA | LIMA | SAN LUIS | LIMA PROVINCIA | 3.0 | -12.0756 | -76.9936 | 1 |
| 1315 | LIMA | LIMA | SAN MARTIN DE PORRES | LIMA PROVINCIA | 37.0 | -12.0300 | -77.0575 | 2 |
| 1316 | LIMA | LIMA | SAN MIGUEL | LIMA PROVINCIA | 11.0 | -12.0922 | -77.0794 | 1 |
| 1317 | LIMA | LIMA | SANTA ANITA | LIMA PROVINCIA | 11.0 | -12.0439 | -76.9714 | 1 |
| 1318 | LIMA | LIMA | SANTA MARIA DEL MAR | LIMA PROVINCIA | 10.0 | -12.4019 | -76.7733 | 1 |
| 1319 | LIMA | LIMA | SANTA ROSA | LIMA PROVINCIA | 21.0 | -11.7872 | -77.1569 | 1 |
| 1320 | LIMA | LIMA | SANTIAGO DE SURCO | LIMA PROVINCIA | 36.0 | -12.1450 | -77.0050 | 2 |
| 1321 | LIMA | LIMA | SURQUILLO | LIMA PROVINCIA | 3.0 | -12.1186 | -77.0217 | 1 |
| 1322 | LIMA | LIMA | VILLA EL SALVADOR | LIMA PROVINCIA | 35.0 | -12.2133 | -76.9372 | 2 |
| 1323 | LIMA | LIMA | VILLA MARIA DEL TRIUNFO | LIMA PROVINCIA | 71.0 | -12.1625 | -76.9436 | 2 |

## **3.2 Vaccination centers dataset**

Considerations:

* The dataset with vaccination centers will be merged with the district dataframe to get only the vaccination centers in Lima and Callao provinces.

This is the resulting dataset after process.

| id\_ubigeo | vaccination\_center | latitude | longitude | district |
| --- | --- | --- | --- | --- |
| 1317 | Vacunatorio San Isidro labrador | -12.043744 | -76.946264 | SANTA ANITA |
| 1317 | Estadio Municipal de Santa Anita | -12.034220 | -76.965259 | SANTA ANITA |
| 1292 | Clínica Jesus Del Norte | -11.989580 | -77.058795 | INDEPENDENCIA |
| 1292 | Plaza Norte | -12.007614 | -77.058953 | INDEPENDENCIA |
| 1292 | Coliseo de la Amistad Perú - Japón | -11.978520 | -77.050480 | INDEPENDENCIA |
| 1281 | Centro De Vacunación Aljovin | -12.057287 | -77.032330 | LIMA |
| 1281 | Universidad Nacional Mayor de San Marcos - UNMSM | -12.056170 | -77.084240 | LIMA |
| 1281 | Clínica Internacional | -12.058217 | -77.038457 | LIMA |
| 1281 | Parque La Exposición | -12.062962 | -77.035168 | LIMA |
| 1281 | Visita Domiciliaria | -12.049450 | -77.037760 | LIMA |
| 1281 | Punto de Vacunación 1 | -12.059328 | -77.041171 | LIMA |
| 1311 | Complejo Deportivo - Municipalidad De San Isidro | -12.108021 | -77.054500 | SAN ISIDRO |
| 1311 | Clínica Ricardo Palma | -12.090467 | -77.018228 | SAN ISIDRO |
| 1311 | Clínica Angloamericana | -12.109345 | -77.039282 | SAN ISIDRO |
| 1311 | Clinica Aliada | -12.092212 | -77.013812 | SAN ISIDRO |
| 1312 | Clínica San Miguel Arcangel (Sjl) | -11.970352 | -77.001719 | SAN JUAN DE LURIGANCHO |
| 1312 | Estadio Monumental Anexo 22 San Antonio | -11.921603 | -76.955551 | SAN JUAN DE LURIGANCHO |
| 1312 | I.E Niños Del Paraiso | -12.002772 | -76.987548 | SAN JUAN DE LURIGANCHO |
| 1312 | I.E. Antenor Orrego | -11.878979 | -76.952165 | SAN JUAN DE LURIGANCHO |
| 1312 | Clínica San Juan Bautista - Grupo San Pablo | -12.008310 | -77.003556 | SAN JUAN DE LURIGANCHO |
| 1312 | Parque Zonal Huiracocha | -12.008310 | -77.003556 | SAN JUAN DE LURIGANCHO |
| 1312 | Universidad César Vallejo - Sede Sjl | -11.981377 | -77.000231 | SAN JUAN DE LURIGANCHO |
| 1312 | IPD COMPLEJO DEPORTIVO | -11.956181 | -76.987990 | SAN JUAN DE LURIGANCHO |
| 1312 | ESTADIO MUNICIPAL ANEXO 22 SAN ANTONIO | -11.919156 | -76.949800 | SAN JUAN DE LURIGANCHO |
| 1312 | I.E. ANTENOR ORREGO | -11.878979 | -76.952165 | SAN JUAN DE LURIGANCHO |
| 1312 | I.E NIÑOS DEL PARAISO | -12.002772 | -76.987548 | SAN JUAN DE LURIGANCHO |
| 1312 | INPE SJL | -11.992415 | -76.996926 | SAN JUAN DE LURIGANCHO |
| 1312 | IE. SANTO DOMINGO | -11.900260 | -76.960730 | SAN JUAN DE LURIGANCHO |
| 1313 | Complejo Del Ipd | -12.162643 | -76.971942 | SAN JUAN DE MIRAFLORES |
| 1313 | Módulo Deportivo Siglo Xxi | -12.163024 | -76.962612 | SAN JUAN DE MIRAFLORES |
| 1315 | Parque Zonal Mayta Capac | -12.009103 | -77.086599 | SAN MARTIN DE PORRES |
| 1315 | Gran Unidad Escolar Jose Granda | -12.026996 | -77.074761 | SAN MARTIN DE PORRES |
| 1315 | Instituto Nacional de Salud Mental "Honorio De... | -12.020600 | -77.056240 | SAN MARTIN DE PORRES |
| 1320 | I.E.F.A.P Manuel Polo Jiménez | -12.148859 | -76.992935 | SANTIAGO DE SURCO |
| 1320 | Base Aérea Las Palmas | -12.157931 | -77.000259 | SANTIAGO DE SURCO |
| 1320 | Clínica San Pablo | -12.100092 | -76.971475 | SANTIAGO DE SURCO |
| 1320 | Campo Deportivo Las Malvinas | -12.142089 | -77.002749 | SANTIAGO DE SURCO |
| 1320 | Jockey Club del Perú | -12.097407 | -76.978125 | SANTIAGO DE SURCO |
| 1320 | Monte de Los Olivos | -12.138890 | -76.987171 | SANTIAGO DE SURCO |
| 1298 | Explanada De Vacunacion Del Establecimiento D... | -12.007212 | -76.955770 | LURIGANCHO |
| 1298 | Loza Deportiva Valle Triunfo | -11.985359 | -76.949727 | LURIGANCHO |
| 1298 | Estadio Municipal de Chosica | -11.941486 | -76.695885 | LURIGANCHO |
| 1298 | Ie Manuel Fernando Cabrel Nicho - Cajamarquilla | -11.986643 | -76.899017 | LURIGANCHO |
| 1298 | EXPLANADA DE VACUNACIÓN DEL ESTABLECIMIENTO ... | -12.007212 | -76.955770 | LURIGANCHO |
| 1298 | LOZA DEPORTIVA VALLE TRIUNFO | -11.985359 | -76.949727 | LURIGANCHO |
| 1290 | Parque Zonal Sinchi Roca | -17.651580 | -71.326050 | COMAS |
| 1290 | Explanada de la Municipalidad de Comas | -11.948411 | -77.059325 | COMAS |
| 1290 | Plaza Cívica Año Nuevo | -11.923830 | -77.038489 | COMAS |
| 1290 | Colegio La Libertad | -11.920800 | -77.039466 | COMAS |
| 1290 | Colegio Estados Unidos, Comas | -11.963481 | -77.056450 | COMAS |
| 1293 | Campo de Marte | -12.068334 | -77.041380 | JESUS MARIA |
| 1293 | Padomi | -12.078530 | -77.037276 | JESUS MARIA |
| 1304 | Centro de Salud Pucusana | -12.482862 | -76.796187 | PUCUSANA |
| 1294 | Centro Integral del Adulto Mayor de la Molina | -12.084382 | -76.957371 | LA MOLINA |
| 1294 | Universidad Agraria La Molina | -12.087757 | -76.944166 | LA MOLINA |
| 1302 | Clínica Delgado | -12.113587 | -77.033227 | MIRAFLORES |
| 1302 | Estadio Manuel Bonilla | -12.109900 | -77.052697 | MIRAFLORES |
| 1288 | Colegio Pedro Ruiz Gallo | -12.159471 | -77.022435 | CHORRILLOS |
| 1288 | Complejo Deportivo Nro 1 Chorrillos | -12.168426 | -77.019007 | CHORRILLOS |
| 1288 | Vacunacar 2 Costa Verde (Acudir Solo En Carro) | -12.160598 | -77.025793 | CHORRILLOS |
| 1318 | ESTACIONAMIENTO MUNICIPAL - DISTRITO SANTA MAR... | -12.405311 | -76.773775 | SANTA MARIA DEL MAR |
| 1318 | PS VILLA MERCEDES | -12.393512 | -76.774408 | SANTA MARIA DEL MAR |
| 1299 | Complejo Deportivo Jose Fuentes Chapeda - ZONA A | -12.300145 | -76.846963 | LURIN |
| 1299 | P.S VILLA JARDIN CUCUYA | -12.221510 | -76.842840 | LURIN |
| 1323 | Complejo Deportivo VMT | -12.148745 | -76.952639 | VILLA MARIA DEL TRIUNFO |
| 1323 | Polideportivo César Vallejo - Nueva Esperanza | -12.178129 | -76.941618 | VILLA MARIA DEL TRIUNFO |
| 1305 | Complejo Deportivo y Estadio de Puente Piedra | -11.863082 | -77.073995 | PUENTE PIEDRA |
| 1310 | ESSALUD San Borja | -12.109031 | -77.006364 | SAN BORJA |
| 1284 | Estadio Galvez Chipoco | -12.138062 | -77.023497 | BARRANCO |
| 1282 | Estadio José Balta | -11.774470 | -77.167521 | ANCON |
| 1286 | Estadio Lolo Fernández de Carabayllo | -11.871133 | -77.017740 | CARABAYLLO |
| 1283 | Estadio Monumental | -12.055785 | -76.936086 | ATE |
| 1283 | Estadio Municipal Ollantaytambo | -12.028547 | -76.908508 | ATE |
| 1321 | Estadio Municipal de Surquillo | -12.114060 | -77.009993 | SURQUILLO |
| 1321 | INEN | -12.112441 | -76.998277 | SURQUILLO |
| 1303 | Estadio Municipal del Cercado de Pachacamac | -12.229641 | -76.856222 | PACHACAMAC |
| 1303 | Complejo Deportivo Del Secto B De Manchay | -12.116720 | -76.874025 | PACHACAMAC |
| 696 | Estadio Municipal Mi Perú | -11.857395 | -77.119570 | MI PERU |
| 1287 | IE Felipe Santiago Estenos | -11.978240 | -76.778586 | CHACLACAYO |
| 1287 | Estadio Pachacutec | -11.984779 | -76.809623 | CHACLACAYO |
| 1308 | IE Tte Crnl Alfredo Bonifaz | -12.025861 | -77.040826 | RIMAC |
| 1308 | Centro de Salud Militar Rímac | -12.023179 | -77.037300 | RIMAC |
| 1308 | Solidaridad sede Flor de Amancaes (SISOL) | -12.017228 | -77.029577 | RIMAC |
| 695 | Liceo Naval Capitán de Corbeta Manuel Clavero ... | -11.887991 | -77.128324 | VENTANILLA |
| 695 | Plaza Civica de Ventanilla | -11.876043 | -77.126999 | VENTANILLA |
| 695 | Plaza Cívica De Pachacutec | -11.831571 | -77.154359 | VENTANILLA |
| 1289 | Loza Deportiva Municipal de Cieneguilla | -12.111816 | -76.816246 | CIENEGUILLA |
| 1316 | Parque de las Leyendas | -12.072001 | -77.086890 | SAN MIGUEL |
| 1296 | Playa Miller | -12.080349 | -77.036501 | LINCE |
| 690 | Plaza Casanave | -12.060463 | -77.140697 | CALLAO |
| 690 | Real Felipe | -12.062775 | -77.149134 | CALLAO |
| 1301 | Polideportivo Torre Tagle | -12.076601 | -77.057726 | PUEBLO LIBRE |
| 1322 | Polideportivo Villa El Salvador | -12.199949 | -76.957750 | VILLA EL SALVADOR |
| 1322 | Parque Zonal Huáscar | -12.231678 | -76.928208 | VILLA EL SALVADOR |
| 1306 | Puesto de Salud Punta Hermosa | -12.337502 | -76.825984 | PUNTA HERMOSA |
| 1307 | Puesto de Salud Punta Negra | -12.363736 | -76.795856 | PUNTA NEGRA |
| 1307 | LOCAL COMUNAL PAMPAPACTA | -12.279900 | -76.648790 | PUNTA NEGRA |
| 1309 | Puesto de Salud San Bartolo | -12.387249 | -76.777162 | SAN BARTOLO |
| 1291 | Regiminento de Caballería Mariscal Domingo Nie... | -12.056708 | -77.008142 | EL AGUSTINO |
| 1314 | Videna | -12.079690 | -77.003504 | SAN LUIS |
| 1285 | Colegio Salesiano San Francisco De Sales | -12.062014 | -77.043618 | BREÑA |
| 692 | Plaza Miguel Grau | -12.039257 | -77.090733 | CARMEN DE LA LEGUA REYNOSO |
| 1300 | Vacunacar 1 Costa Verde (Acudir Solo En Carro) | -12.096428 | -77.077998 | MAGDALENA DEL MAR |
| 1300 | HOSPITAL MILITAR | -12.087420 | -77.062723 | MAGDALENA DEL MAR |
| 1300 | Hospital Larco Herrera | -12.098196 | -77.065646 | MAGDALENA DEL MAR |
| 693 | Estadio Campolo Alcalde | -12.065040 | -77.107876 | LA PERLA |
| 1297 | Universidad Cesar Vallejo - Lima Norte | -11.955241 | -77.068470 | LOS OLIVOS |
| 1297 | Explanada de la Municipalidad de Los Olivos | -11.991584 | -77.071019 | LOS OLIVOS |
| 1297 | Parque Nº 1 Las Palmeras | -11.988582 | -77.072232 | LOS OLIVOS |
| 1319 | Club Zonal Santa Rosa | -11.788432 | -77.159211 | SANTA ROSA |

## **3.3 Venues dataset**

Considerations:

* To get this dataset we will use Foursquare
* Each request will have 100 venues as limit and 3 Km as radius
* The district coordinates (lat/long) will be used as central points for each request
* If the district has num\_requests = 2 then all the venues in the result of the query to Foursquare will be used as central points for a second query. This is to cover a greater surface that posibly has more population density

This is the resulting dataset.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| venue\_name | category | lat | lng | origin\_district |
| CasaCor Callao | Public Art | -12.060004 | -77.147282 | CALLAO |
| Casa Fugaz | Art Gallery | -12.060145 | -77.147335 | CALLAO |
| Fortaleza del Real Felipe | Monument / Landmark | -12.062062 | -77.148059 | CALLAO |
| Cabos Restaurante de Puerto | Seafood Restaurant | -12.060282 | -77.150105 | CALLAO |
| Panadería Olcese | Bakery | -12.061041 | -77.143544 | CALLAO |
| ... | ... | ... | ... | ... |
| Mercado "Señor de la Divina Misericordia" | Shopping Mall | -12.152731 | -76.991412 | VILLA MARIA DEL TRIUNFO |
| Urb. Los Parrales | Park | -12.167568 | -76.992087 | VILLA MARIA DEL TRIUNFO |
| Parque Julio C. Tello | Park | -12.148285 | -76.991862 | VILLA MARIA DEL TRIUNFO |
| El Hueco, barra cebichera | Seafood Restaurant | -12.165424 | -76.992030 | VILLA MARIA DEL TRIUNFO |
| Ciclo via | Trail | -12.195898 | -76.957242 | VILLA MARIA DEL TRIUNFO |

2723 rows × 5 columns

# **4. Methodology**

## **4.1 Algorithm and hyperparameters**

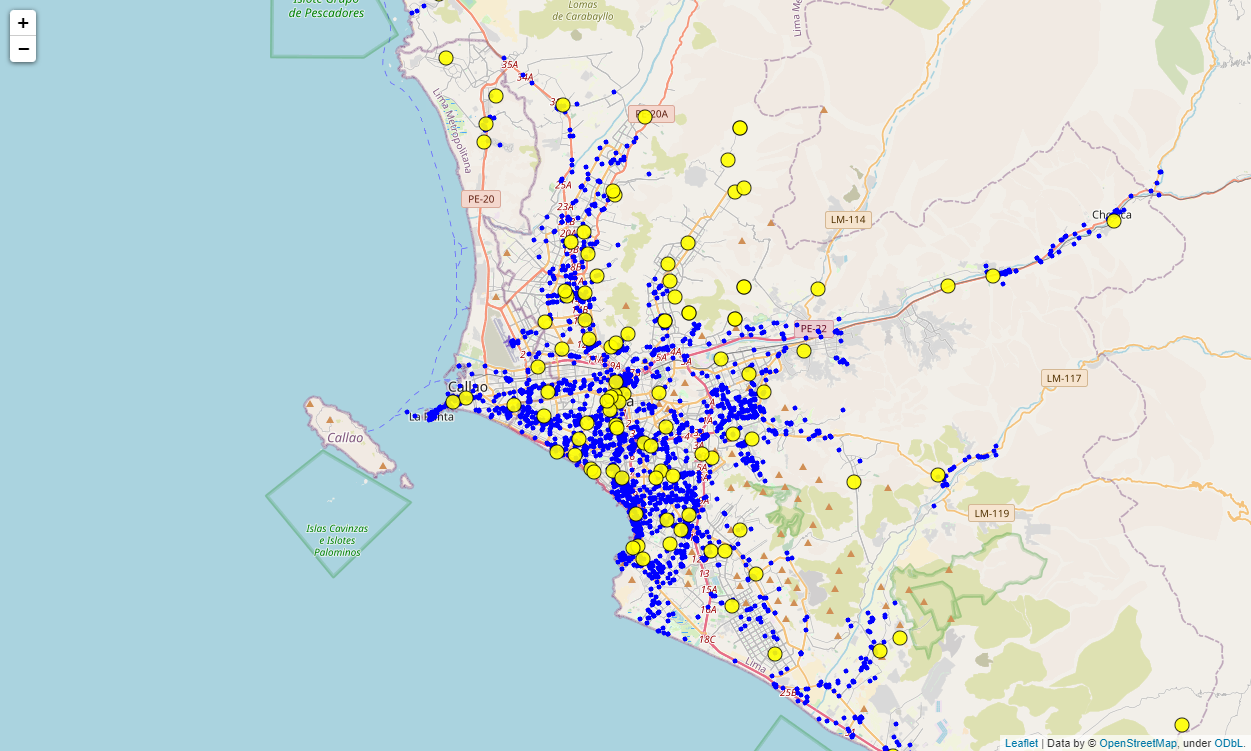
Since we want to plot high density population clusters, the algorithm to use is DBSCAN. To estimate the hyperparameters we will use the recommendations by Tara Mullin in <https://medium.com/@tarammullin/dbscan-parameter-estimation-ff8330e3a3bd>.

* **Minimum samples:** Dimensions multiplied by 2 (2 x 2 = 4). For the venues dataset, as this has almost 2000 rows, we're incrementing the minimum samples to 5.
* **Epsilon:** Average distance between each point and its k nearest neighbors, where k is equal to the minimum samples.

## **4.2 Taking a look into the data**

The following code is to take a look to the data about venues and vaccination centers. To do this, we will plot both datasets centers and venues.

* Venues: Small blue circles
* Vaccination centers: Large yellow circles



## **4.3 The definitive dataframe**

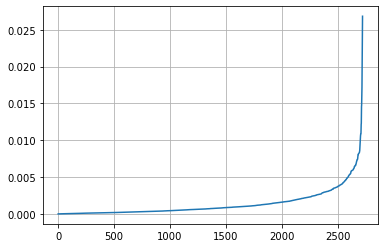
In this step the venues data will join the vaccination centers data. The resulting dataframe will be used to compute the clusters with DBSCAN and to know which clusters have or not vaccination centers within them.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| venue\_name | category | lat | lng | origin\_district |
| CasaCor Callao | Public Art | -12.060004 | -77.147282 | CALLAO |
| Casa Fugaz | Art Gallery | -12.060145 | -77.147335 | CALLAO |
| Fortaleza del Real Felipe | Monument / Landmark | -12.062062 | -77.148059 | CALLAO |
| Cabos Restaurante de Puerto | Seafood Restaurant | -12.060282 | -77.150105 | CALLAO |
| Panadería Olcese | Bakery | -12.061041 | -77.143544 | CALLAO |
| ... | ... | ... | ... | ... |
| Estadio Campolo Alcalde | Vaccination Center | -12.065040 | -77.107876 | LA PERLA |
| Universidad Cesar Vallejo - Lima Norte | Vaccination Center | -11.955241 | -77.068470 | LOS OLIVOS |
| Explanada de la Municipalidad de Los Olivos | Vaccination Center | -11.991584 | -77.071019 | LOS OLIVOS |
| Parque Nº 1 Las Palmeras | Vaccination Center | -11.988582 | -77.072232 | LOS OLIVOS |
| Club Zonal Santa Rosa | Vaccination Center | -11.788432 | -77.159211 | SANTA ROSA |

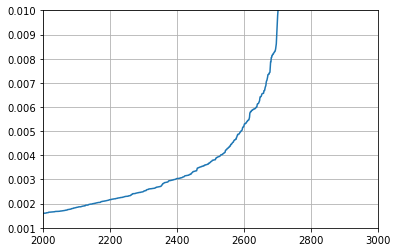
2833 rows × 5 columns

## **4.4 Epsilon hyperparameter estimation with KNN**

Here we estimate the epsilon hyperparameter using k nearest neighbors. First, we calculate the average distance between each point in the data set and its 5 nearest neighbors.



Now, we zoom the elbow area. The ideal value for epsilon will be equal to the point of maximum curvature, for this **epsilon = 0.005**.



## **4.5 First run of DBSCAN algorithm**

Here we estimate the epsilon hyperparameter using k nearest neighbors. First, we calculate the average distance between each point in the data set and its 5 nearest neighbors.

| cluster | count of venue\_name |
| --- | --- |
| -1 | 2359 |
| 0 | 5 |
| 1 | 15 |
| 2 | 10 |
| 3 | 21 |
| 4 | 14 |
| 5 | 5 |
| 6 | 11 |
| 7 | 6 |
| 8 | 7 |
| 9 | 4 |
| 10 | 7 |
| 11 | 5 |
| 12 | 30 |
| 13 | 12 |
| 14 | 8 |
| 15 | 5 |
| 16 | 10 |
| 17 | 5 |
| 18 | 6 |
| 19 | 6 |
| 20 | 24 |
| 21 | 8 |
| 22 | 5 |
| 23 | 9 |
| 24 | 12 |
| 25 | 34 |
| 26 | 6 |
| 27 | 11 |
| 28 | 25 |
| 29 | 5 |
| 30 | 9 |
| 31 | 5 |
| 32 | 9 |
| 33 | 5 |
| 34 | 24 |
| 35 | 22 |
| 36 | 6 |
| 37 | 8 |
| 38 | 5 |
| 39 | 21 |
| 40 | 7 |
| 41 | 5 |
| 42 | 6 |
| 43 | 5 |
| 44 | 6 |

**Number of clusters with vaccination centers:** We get only two clusters with vaccination centers within it (clusters 25 and 26).

**Number of clusters without vaccination centers:** Here we get that 43 clusters don't have any vaccination center.

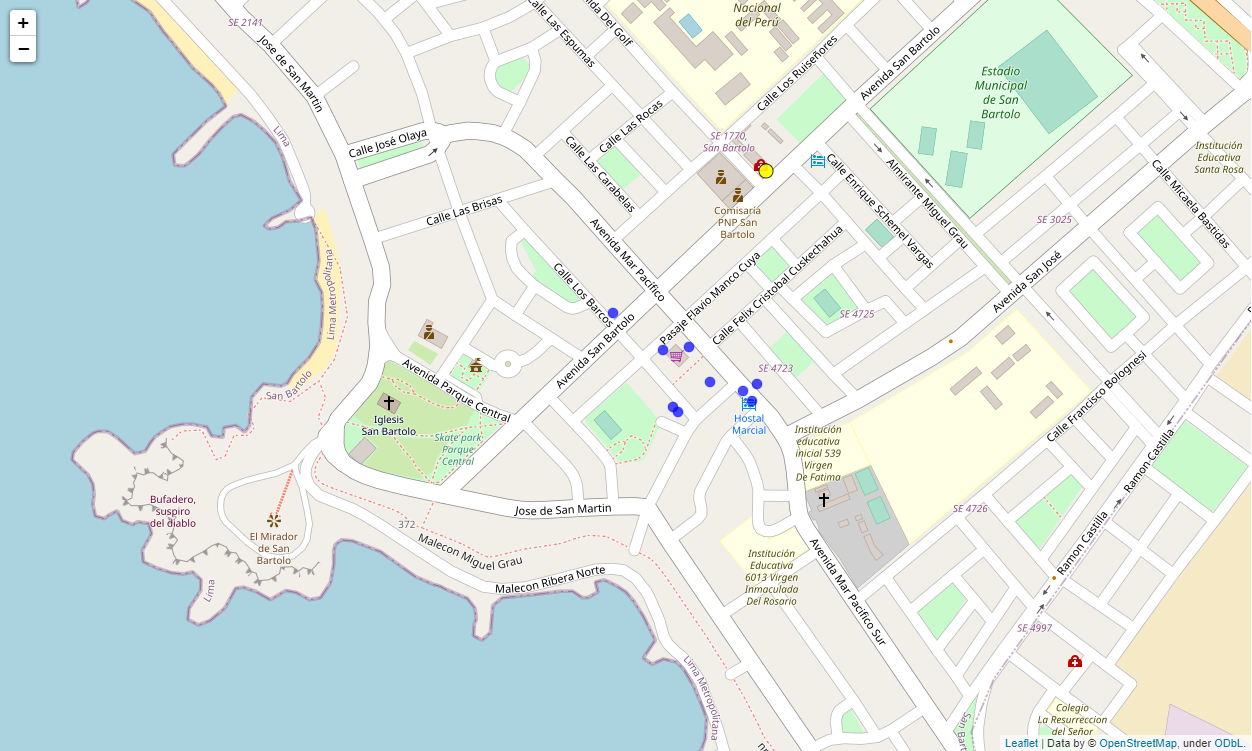
**Vaccination centers that belong to clusters:** As we saw before, only one vaccination center.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| venue\_name | category | lat | lng | origin\_district | cluster |
| Plaza Norte | Vaccination Center | -12.007614 | -77.058953 | INDEPENDENCIA | 25 |
| Parque Nº 1 Las Palmeras | Vaccination Center | -11.988582 | -77.072232 | LOS OLIVOS | 26 |

## **4.6 A problem with first DBSCAN run**

Some vaccination centers are close to clusters but not part of them. In the following example, there is a cluster near a vaccination center, but this vaccination center is not part of the cluster. To fix this, we will increase the epsilon value from 0.005 to 0.05.

* Venues: Small blue circles
* Vaccination centers: Large yellow circles



## **4.7 Second run of DBSCAN algorithm**

Here we change the epsilon hyperparameter, now **epsilon = 0.05**.

| cluster | count of venue\_name |
| --- | --- |
| -1 | 230 |
| 0 | 87 |
| 1 | 1436 |
| 2 | 17 |
| 3 | 26 |
| 4 | 192 |
| 5 | 5 |
| 6 | 10 |
| 7 | 18 |
| 8 | 15 |
| 9 | 8 |
| 10 | 10 |
| 11 | 7 |
| 12 | 12 |
| 13 | 23 |
| 14 | 10 |
| 15 | 17 |
| 16 | 16 |
| 17 | 32 |
| 18 | 8 |
| 19 | 11 |
| 20 | 10 |
| 21 | 204 |
| 22 | 11 |
| 23 | 10 |
| 24 | 6 |
| 25 | 6 |
| 26 | 14 |
| 27 | 7 |
| 28 | 10 |
| 29 | 8 |
| 30 | 56 |
| 31 | 5 |
| 32 | 12 |
| 33 | 44 |
| 34 | 18 |
| 35 | 11 |
| 36 | 9 |
| 37 | 19 |
| 38 | 5 |
| 39 | 10 |
| 40 | 10 |
| 41 | 7 |
| 42 | 12 |
| 43 | 7 |
| 44 | 8 |
| 45 | 58 |
| 46 | 11 |
| 47 | 10 |
| 48 | 5 |
| 49 | 10 |
| 50 | 12 |
| 51 | 7 |
| 52 | 6 |
| 53 | 5 |

**Number of clusters with vaccination centers:** We now get 22 clusters with vaccination centers within it.

**Number of clusters without vaccination centers:** Here we get that 32 clusters don't have any vaccination center.

**Vaccination centers that belong to clusters:** 74 vaccination centers are installed within 22 clusters.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| venue\_name | category | lat | lng | origin\_district | cluster |
| Vacunatorio San Isidro labrador | Vaccination Center | -12.043744 | -76.946264 | SANTA ANITA | 21 |
| Clínica Jesus Del Norte | Vaccination Center | -11.989580 | -77.058795 | INDEPENDENCIA | 4 |
| Plaza Norte | Vaccination Center | -12.007614 | -77.058953 | INDEPENDENCIA | 4 |
| Coliseo de la Amistad Perú - Japón | Vaccination Center | -11.978520 | -77.050480 | INDEPENDENCIA | 4 |
| Centro De Vacunación Aljovin | Vaccination Center | -12.057287 | -77.032330 | LIMA | 1 |
| ... | ... | ... | ... | ... | ... |
| HOSPITAL MILITAR | Vaccination Center | -12.087420 | -77.062723 | MAGDALENA DEL MAR | 1 |
| Hospital Larco Herrera | Vaccination Center | -12.098196 | -77.065646 | MAGDALENA DEL MAR | 1 |
| Estadio Campolo Alcalde | Vaccination Center | -12.065040 | -77.107876 | LA PERLA | 1 |
| Explanada de la Municipalidad de Los Olivos | Vaccination Center | -11.991584 | -77.071019 | LOS OLIVOS | 4 |
| Parque Nº 1 Las Palmeras | Vaccination Center | -11.988582 | -77.072232 | LOS OLIVOS | 4 |

# **5. Results**

We have two results to analyze:

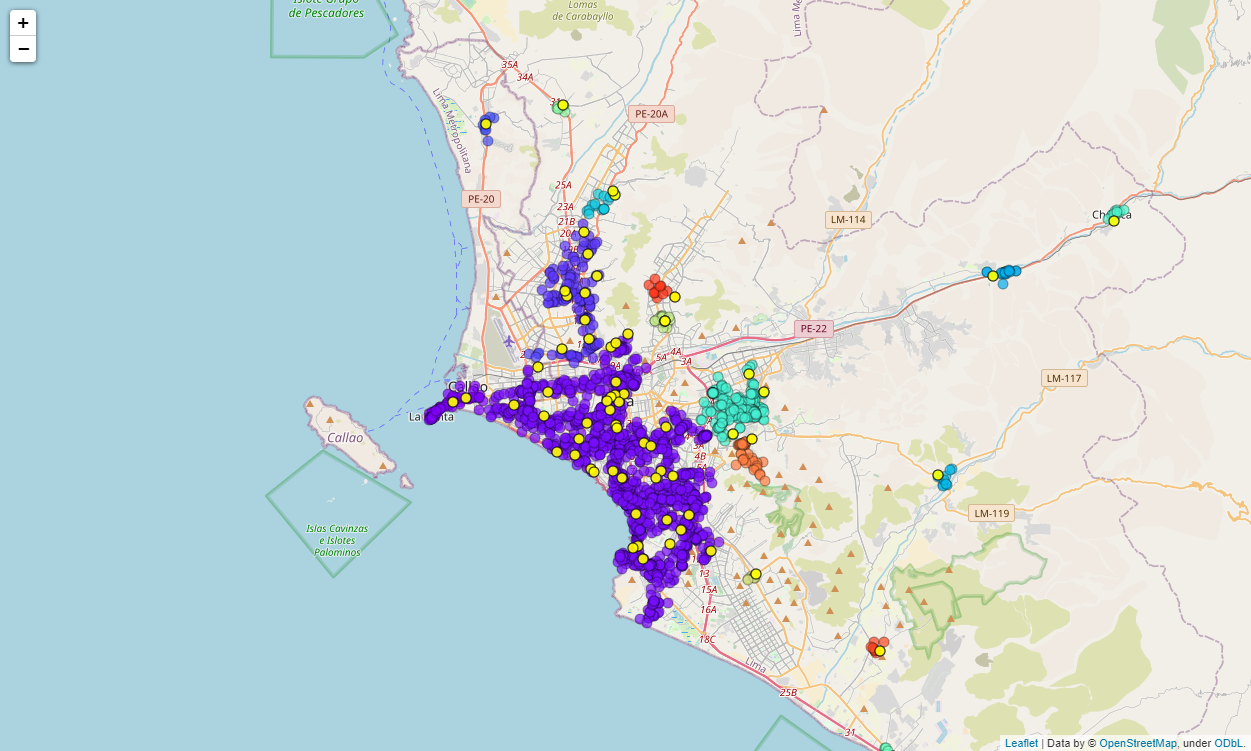
* 5.1 Clusters with vaccination centers
* 5.2 Clusters without vaccination centers

## **5.1 Clusters with vaccination centers**

Here we plot:

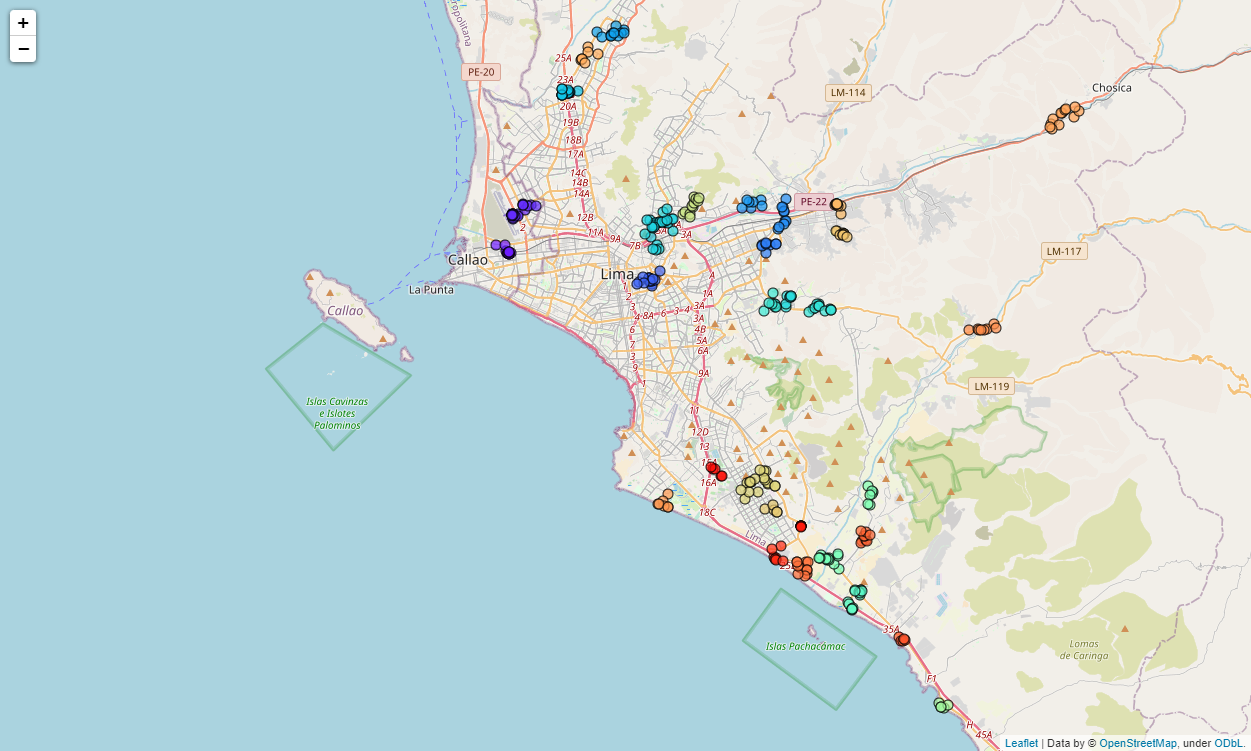
* Clusters (22) with vaccination centers within (colorful circles)
* Vaccination centers (74) belonging to those clusters (yellow circles with black border)

As we can see, **the problem we had in section 5.6 is not present now**.



## **5.2 Clusters without vaccination centers**

Here we plot the clusters without any vaccination center. **These clusters need vaccination centers to be installed**.



# **6. Discussion**

As a result of the DBSCAN run, we have the following data:

* There are 54 clusters of high population density in Lima and Callao.
* 74 vaccination centers out of 110 belong to 19 clusters of high population density.
* 36 vaccination centers are spread in the noise zone.
* 32 clusters don't have or are far from any vaccination center.

To reach the conclusions, here are some information that may help:

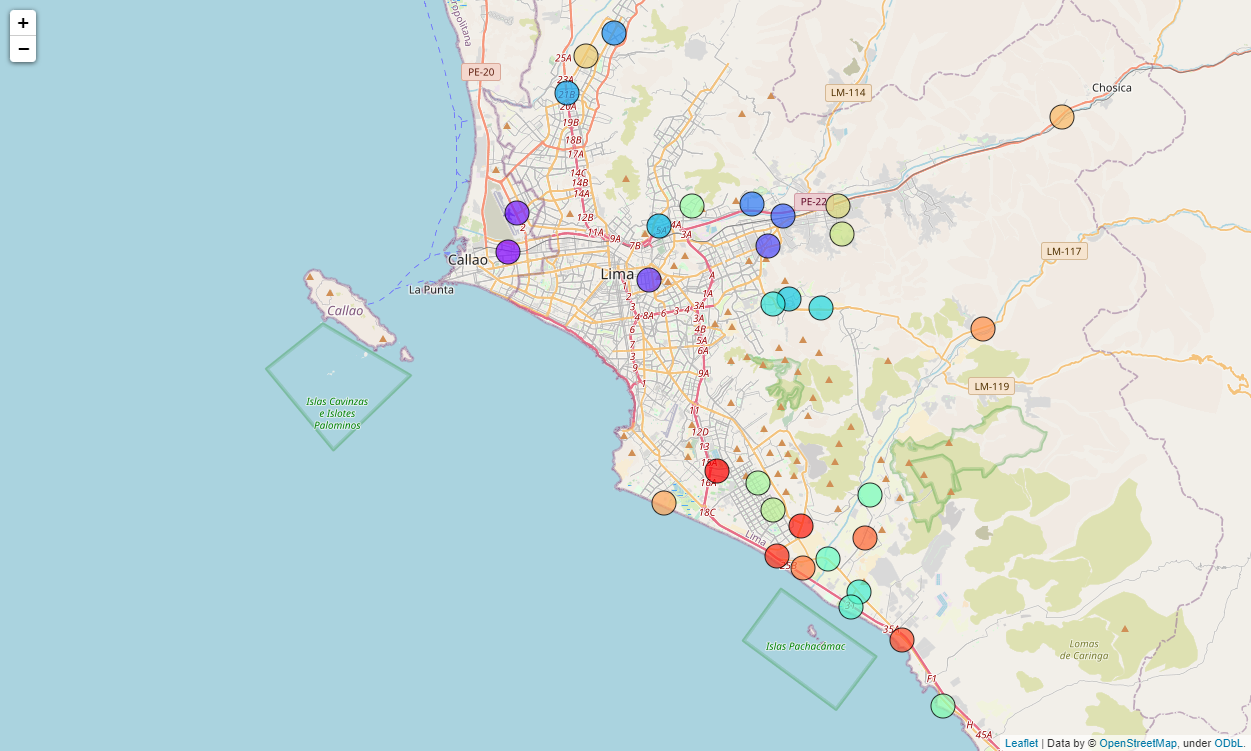
* Table of clusters and suggested coordinates to install vaccination centers: It has the clusters and coordinates calculated with average latitude/longitude from points in the clusters. This could be useful to query Foursquare or another source to find near venues where to install the new vaccination centers.
* Vaccination centers needed: Map with referential coordinates for the new vaccination centers. This could help to find a middle point to install shared vaccination centers (inter clusters) in case the logistic cost exceeds budget.
* Table of vaccination centers by district: Could help to estimate costs based in the number of vaccination centers to install by district.

**Table of clusters and suggested coordinates to install vaccination centers:** The following table has:

* Clusters
* District to which the cluster belongs
* Referential latitude: Average latitude from points in the clusters
* Referential longitude: Average longitude from points in the clusters
* Number of venues within the cluster

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| cluster | district | ref\_lat | ref\_lng | venues |
| 1 | BELLAVISTA | -12.047422 | -77.111977 | 17 |
| 2 | CARMEN DE LA LEGUA REYNOSO | -12.020992 | -77.105494 | 26 |
| 3 | EL AGUSTINO | -12.066292 | -77.014773 | 18 |
| 4 | ATE | -12.043131 | -76.933174 | 8 |
| 5 | ATE | -12.023028 | -76.922556 | 10 |
| 6 | ATE | -12.014940 | -76.944074 | 7 |
| 7 | CARABAYLLO | -11.900432 | -77.039199 | 12 |
| 8 | COMAS | -11.940270 | -77.071323 | 17 |
| 9 | EL AGUSTINO | -12.030044 | -77.007799 | 32 |
| 10 | LA MOLINA | -12.078778 | -76.918640 | 8 |
| 11 | LA MOLINA | -12.085069 | -76.896899 | 11 |
| 12 | LA MOLINA | -12.082512 | -76.929903 | 10 |
| 13 | LURIN | -12.275285 | -76.870384 | 10 |
| 14 | LURIN | -12.285656 | -76.876173 | 6 |
| 15 | LURIN | -12.253413 | -76.891913 | 14 |
| 16 | PACHACAMAC | -12.210810 | -76.862853 | 7 |
| 17 | PUNTA HERMOSA | -12.352125 | -76.813290 | 5 |
| 18 | SAN JUAN DE LURIGANCHO | -12.016550 | -76.985180 | 11 |
| 19 | VILLA EL SALVADOR | -12.202330 | -76.939918 | 19 |
| 20 | VILLA EL SALVADOR | -12.220434 | -76.929596 | 5 |
| 21 | ATE | -12.035214 | -76.882367 | 10 |
| 22 | ATE | -12.016283 | -76.885371 | 10 |
| 23 | CARABAYLLO | -11.916020 | -77.057913 | 7 |
| 24 | CHACLACAYO | -11.956716 | -76.731164 | 12 |
| 25 | CHORRILLOS | -12.216027 | -77.004545 | 7 |
| 26 | CIENEGUILLA | -12.098917 | -76.785255 | 8 |
| 27 | LURIN | -12.259440 | -76.909243 | 11 |
| 28 | LURIN | -12.239058 | -76.866735 | 10 |
| 29 | LURIN | -12.307674 | -76.840946 | 5 |
| 30 | VILLA EL SALVADOR | -12.251240 | -76.927203 | 7 |
| 31 | VILLA EL SALVADOR | -12.231239 | -76.910651 | 6 |
| 32 | VILLA EL SALVADOR | -12.194641 | -76.968378 | 5 |

**Vaccination centers needed:** This map presents the vaccination centers needed to install. The clusters around these points in the map are far from another vaccination centers.



**Table of vaccination centers by district:** This table shows the number of vaccination centers each district should install to cover all the clusters of high population density.

|  |  |
| --- | --- |
| district | vaccination centers needed |
| LURIN | 6 |
| VILLA EL SALVADOR | 5 |
| ATE | 5 |
| LA MOLINA | 3 |
| EL AGUSTINO | 2 |
| CARABAYLLO | 2 |
| CHACLACAYO | 1 |
| CIENEGUILLA | 1 |
| CARMEN DE LA LEGUA REYNOSO | 1 |
| SAN JUAN DE LURIGANCHO | 1 |
| BELLAVISTA | 1 |
| CHORRILLOS | 1 |
| PUNTA HERMOSA | 1 |
| PACHACAMAC | 1 |
| COMAS | 1 |

# **7. Conclusion**

Thirty-two vaccination centers should be installed in Lima and Callao to reach to all high population density clusters and, in this way, improve the national vaccination strategy. This could be beneficial if there's a need to address vaccination on demand, without appointments. The fifteen districts in focus to implement the new vaccination centers need to plan the installations and get the appropriate budget to do so, mainly Lurín, Villa El Salvador and Ate, these districts need more than five new vaccination centers.

In near future, this approach to get the number and location of the vaccination centers should have a change based on the number of vaccinated people. A weighing process to apply to the venues before compute with DBSCAN may be considered.