CLUSTERING POLICE STATIONS OF LIMA, PERÚ

1. INTRODUCTION

This work is about Clustering applied to the police stations of my country, Peru, according to their characteristics (infrastructure of the police station, police officers involved, accidents registered, people intervened, etc).

Every day, the streets and avenues of Peru are the scene of a bloody and silent war, caused, incredibly, by road insecurity, which leaves more than 50,0000 affected per year, among dead, injured and injured, higher than the amount generated by the citizen insecurity. Last year alone 2,700 people died, approximately 8 per day. (http://www.elperuano.pe/noticia-los-accidentes-viales-dejan-mas-50000-afectados-anualmente-59707.aspx, 2017)

It highlights the fact of the large number of traffic accidents that occur and that according to the "Global Status Report on Road Safety 2015" Peru ranks 105 out of 181 in traffic accident victims worldwide. In South America it is in seventh place in nine countries.

2. DATA

The data used are microdata from INEI (National Institute of Statistics and Information Technology), specifically from the 2017 national census of police stations.

The data is national, and we have 1495 police stations, but in the analysis we only took into account the police stations that belong to the region of Lima, which is the capital of the country.

Taking into account the aforementioned and cleaning the data, we have 183 observations and 63 variables, which will be described below and the complete data and official sources will be mentioned in the annexes.

A) VARIABLES

Infrastructure of the police station:

NAME	*DESCRIPTION*
ID	NUMBER OF IDENTIFICATION OF THE POLICE STATION / SPECIALIZED UNIT
F100	HOW MUCH POPULATION DOES THE POLICE STATION / UNIT SPECIALIZED WITHIN ITS JURISDICTION?
F101	WHAT IS THE JURISDICTIONAL SCOPE OF THE POLICE STATION / SPECIALIZED UNIT?
F102	HOW MANY REGIONS DO YOU COVER? - TOTAL
F103	HOW MANY PROVINCES DO YOU COVER? - TOTAL
F104	HOW MANY DISTRICTS DOES IT HAVE? - TOTAL
F105	HOW MANY POLICE STAFF ARE ASSIGNED AT THE POLICE STATION? - TOTAL1
F106	HOW MANY POLICE STAFF PERFORM ADMINISTRATIVE WORK? - TOTAL
F107	HOW MANY POLICE CARRIERS PERFORM MOTORIZED PATROLING (CARS, TRUCKS, LINEAR OR OTHER MOTORCYCLES)? - TOTAL
F108	HOW MANY POLICE STAFF PERFORM PATROL? – TOTAL
F109	HOW MANY POLICE CARRIERS INVESTIGATE CRIMES AND FAULTS? - TOTAL
F110	HOW MANY POLICIES CARRY OUT INVESTIGATION OF TRAFFIC ACCIDENTS? - TOTAL
F111	HOW MANY POLICE CARRIERS DO INVESTIGATION OF FAMILY VIOLENCE? – TOTAL
F112	HOW MANY POLICIES PERFORM WORK IN THE OFFICE OF CITIZEN PARTICIPATION (OPC)? - TOTAL
F113	HOW MANY POLICE CARRIERS DO GUARD SERVICE WORK? - TOTAL
F114	HOW MANY POLICE STAFF PERFORM SUPERVISORY AND CONTROL WORK? - TOTAL
F115	HOW MANY POLICE CARRIERS CARRY OUT ANOTHER TYPE OF ACTIVITY? – TOTAL
F116	HOW MUCH CIVIL STAFF WORK AT THE POLICE STATION? - TOTAL
F117	HAS THE COMMISSIONER / SPECIALIZED UNIT CONFORMED EMERGENCY BRIGADES AND CONTINGENCY RESCUE TO NATURAL DISASTERS?
F118	HOW MANY EMERGENCY BRIGADES AND CONTINGENCY RESCUE ARE EQUIPPED?

Infrastructure of the police station:

NAME	*DESCRIPTION*
F200	THE LOCAL OCCUPIED BY THE POLICE STATION / SPECIALIZED UNIT IS:
F201	WHAT IS THE TOTAL AREA OF THE LAND WHERE IS THE POLICE STATION / SPECIALIZED UNIT LOCATED? - AREA OF THE LAND M2
F202	WHAT IS THE TOTAL AREA THAT OCCASES THE POLICE STATION / SPECIALIZED UNIT? - OCCUPIED AREA M2
F203	WHAT IS THE AREA AVAILABLE FOR EXTENSION? - AREA OF APPLICATION M2
F204	WHAT IS THE ROOFED AREA? - ROOF AREA M2
F205	HOW MANY YEARS OF ANTIQUITY DOES THE PREMISES OF THE POLICE STATION / UNIT SPECIALIZED? - Years since its construction
F206	TOTAL ROOFED AREA
F207	¿THE POPULATED CENTER OR DISTRICT WHERE IS THE POLICE STATION / UNIT SPECIALIZED HAS ELECTRICAL ENERGY SERVICE THROUGH PUBLIC NETWORK?
F208	DOES THE POLICE STATION / SPECIALIZED UNIT HAVE ELECTRICAL ENERGY THROUGH PUBLIC NETWORK?
F209	DOES THE COMMUNITY / UNIT SPECIALIZED UNIT HAVE ACCESS TO THE ELECTRICAL SERVICE EVERY DAY OF THE WEEK?
F210	PREDOMINANT CONSTRUCTION MATERIAL ON FRONTIS WALLS:
F211	¿THE POPULAR CENTER OR DISTRICT WHERE IS THE POLICE STATION / UNIT SPECIALIZED HAS THE INTERNET SERVICE?
F300	DOES THE POLICE STATION COUNT ON DUNGEON (S) / ROOM (S) OF MEDITATION: HOW MUCH?
F400	TOTAL FIXED TELEPHONE LINES
F401	TOTAL EQUIPMENT COMMUNICATION: (A + B + C + D) – Telefax
F402	DOES THE POLICE STATION / SPECIALIZED UNIT HAVE ACCESS TO THE DATA INFORMATION OF THE (LOS) / LA (S): National Registry of Identification and Civil Status (RENIEC)
F403	FOR THE RECORD OF COMPLAINTS, DOES THE POLICE OFFICE HAVE THE COMPUTER SYSTEM OF POLICE COMPLAINTS (SIDPOL)?

Activity of the police station:

NAME	*DESCRIPTION*
F500	IN THE CURRENT YEAR, HOW MANY MEETINGS DID THE DISTRICT CITIZEN SECURITY COMMITTEE HAVE AND HOW MANY OF THESE MEETINGS DID THE COMMISSIONER ATTEND ?: TOTAL NUMBER OF MEETINGS - YEAR 2017
F501	HOW MANY QUESTIONS HAVE YOU MADE?
F502	ARE YOU CURRENTLY CONDUCTING POLICE PATROL?
F503	NUMBER OF PATROL ACTIONS PER MONTH (AVERAGE) - MODALITY: ON FOOT - MORNING SHIFT
F504	AMOUNT OF PATROL ACTIONS PER MONTH (AVERAGE) - MODALITY: MOTORIZED - CAR - MORNING SHIFT
F505	AMOUNT OF PATROL ACTIONS PER MONTH (AVERAGE) - MODALITY: MOTORIZED - LINEAR MOTORCYCLE - MORNING SHIFT
F506	NUMBER OF PATROL ACTIONS PER MONTH (AVERAGE) - MODALITY: ON FOOT - AFTERNOON SHIFT
F507	AMOUNT OF PATROL ACTIONS PER MONTH (AVERAGE) - MODALITY: MOTORIZED - CAR - AFTERNOON SHIFT
F508	AMOUNT OF PATROL ACTIONS PER MONTH (AVERAGE) - MODALITY: MOTORIZED - LINEAR MOTORCYCLE - LATE SHIFT
F509	NUMBER OF PATROL ACTIONS PER MONTH (AVERAGE) - MODALITY: ON FOOT - NIGHT SHIFT
F510	AMOUNT OF PATROL ACTIONS PER MONTH (AVERAGE) - MODALITY: MOTORIZED - CAR - NIGHT SHIFT
F511	AMOUNT OF PATROL ACTIONS PER MONTH (AVERAGE) - MODALITY: MOTORIZED - LINEAR MOTORCYCLE - NIGHT SHIFT
F512	HOW MANY NEIGHBORHOOD BOARDS ARE - TOTAL
F513	WHAT IS THE TOTAL NUMBER OF MEMBERS OF ACTIVE NEIGHBORHOOD COMMITTEES? - TOTAL
F514	WHAT IS THE TOTAL OF CRIMINAL BANDS CAPTURED IN YOUR POLICE JURISDICTION DURING THE MONTHS OF: TOTAL - YEAR 2016?
F515	WHAT IS THE TOTAL OF CRIMINAL BANDS CAPTURED IN YOUR POLICE JURISDICTION DURING THE MONTHS OF: TOTAL - YEAR 2017?
F516	HOW MANY PEOPLE WERE INTERVENED BY THE STAFF OF THE POLICE STAFF DURING THE MONTHS OF: TOTAL - YEAR 2016?
F517	HOW MANY PEOPLE WERE INTERVENED BY THE STAFF OF THE POLICE STAFF DURING THE MONTHS OF: TOTAL - YEAR 2017?

Activity of the police station:

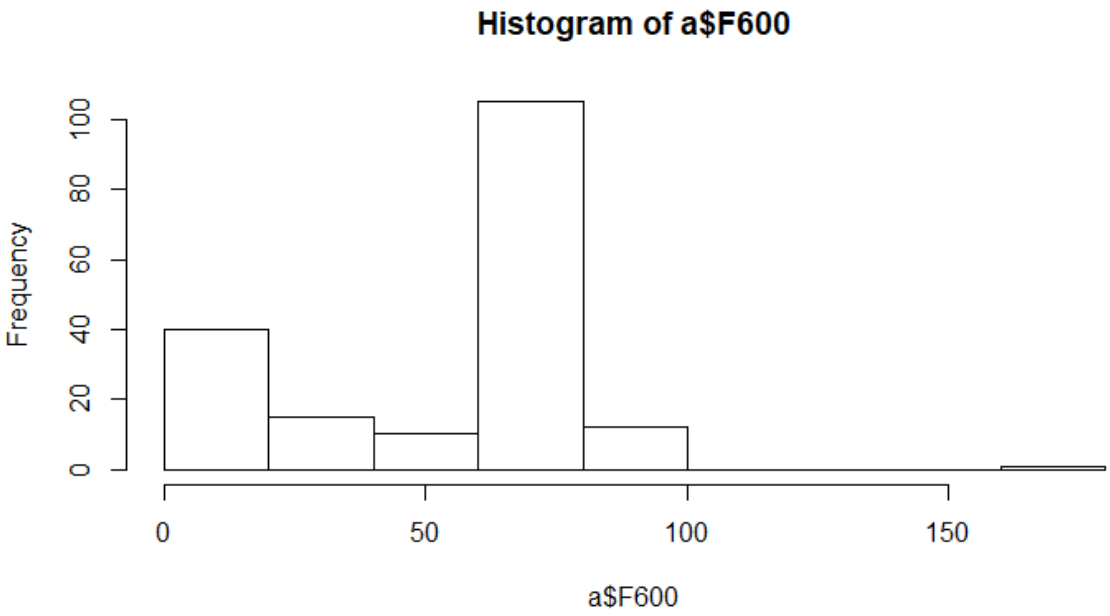
NAME	*DESCRIPTION*
F518	HOW MANY INTERVENED PEOPLE WERE DETAINED BY THE STAFF OF THE POLICE STAFF DURING THE MONTHS OF: TOTAL - 2016?
F519	HOW MANY INTERVENED PEOPLE WERE DETAINED BY THE STAFF OF THE POLICE STAFF DURING THE MONTHS OF: TOTAL - Year 2017?
F520	HOW MANY PERSONS DETAINED WERE POSTED BY THE PUBLIC MINISTRY DURING THE MONTHS OF: TOTAL - Year 2016?
F521	HOW MANY PEOPLE DETAINED WERE MADE AVAILABLE TO THE PUBLIC MINISTRY DURING THE MONTHS OF: TOTAL - Year 2017?

Accidents:

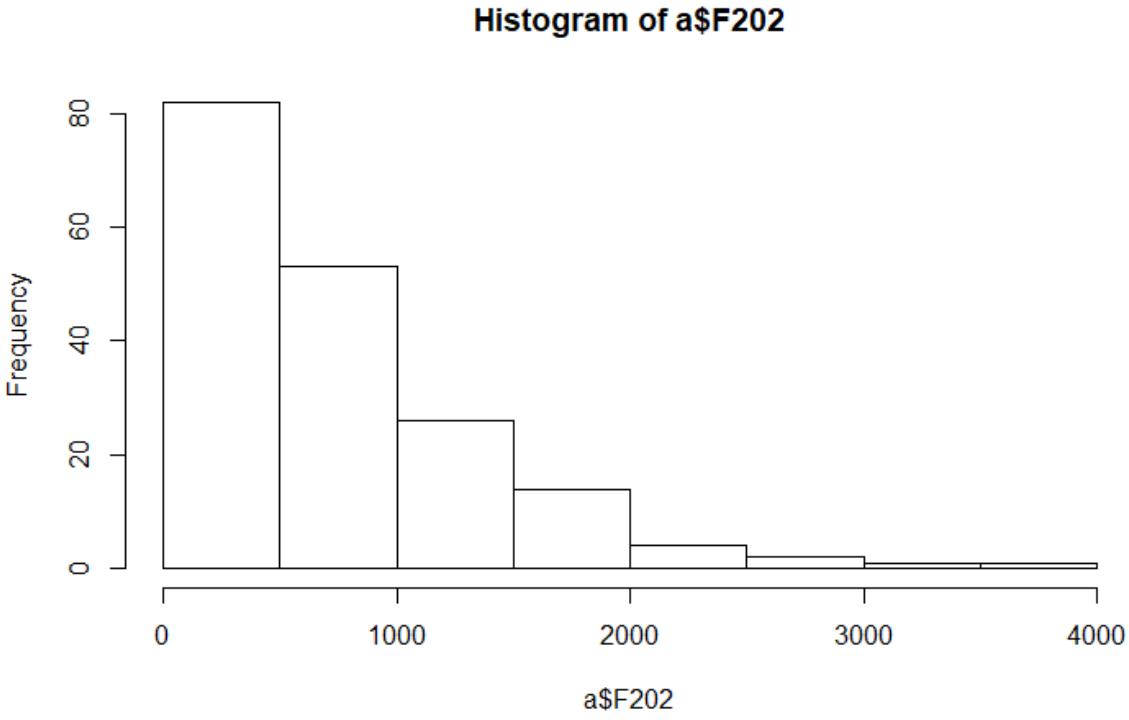
NAME	*DESCRIPTION*
F600	TOTAL NUMBER OF TRANSIT ACCIDENTS
F601	NUMBER OF DECEASED
F602	NUMBER OF INJURED
F603	NUMBER OF ILLUSIONS

3. DESCRIPTIVE ANALYSIS

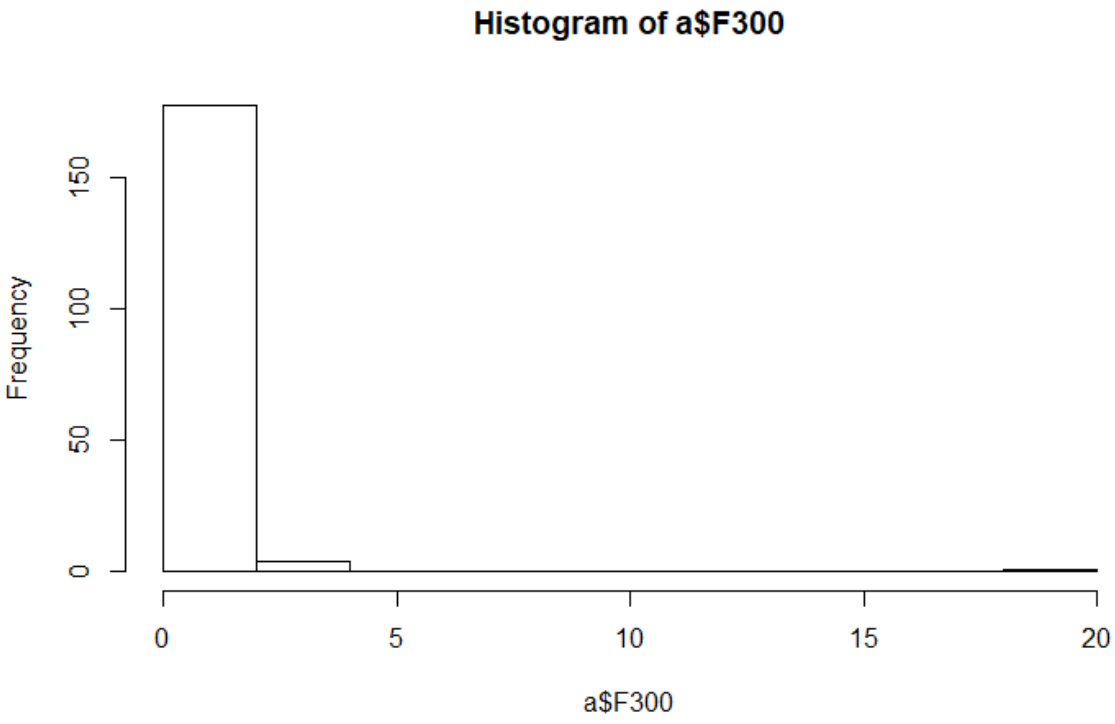
According to the year 2017, we have the variable at $ F600 that measures the total number of traffic accidents that occurred during the year and registered by the police station. If we graph it in a histogram we can see that 70-80 accidents per year and per police station, is the most frequent amount.



We also see a histogram of the variable a $ F202 (area of the commissary) and we can see that most have an area less than 1000 m ^ 2.



According to the histogram of the variable at $ F300 (number of dungeons per commissary), we can see that most have less than 3.



We can use the data, to analyze proportions:

- 8.24% of the police work in administrative activities.
- 85% of the police stations have access to the RENIEC database (National Registry of Identification and Civil Status).
- 18% of the people intervened were arrested by the police station staff.
- 68% of the people arrested were put at the disposal of the Public Ministry.

4. CLUSTERING

Using k-means clustering

A) OPTIMAL k

Creating the function that runs the k-mean algorithm and store the total within clusters sum of squares.

```
kmean_withinss <- function(k) {
   cluster <- kmeans(a, k)
   return (cluster$tot.withinss)
}
```

We will use the sapply() function to run the algorithm over a range of k.

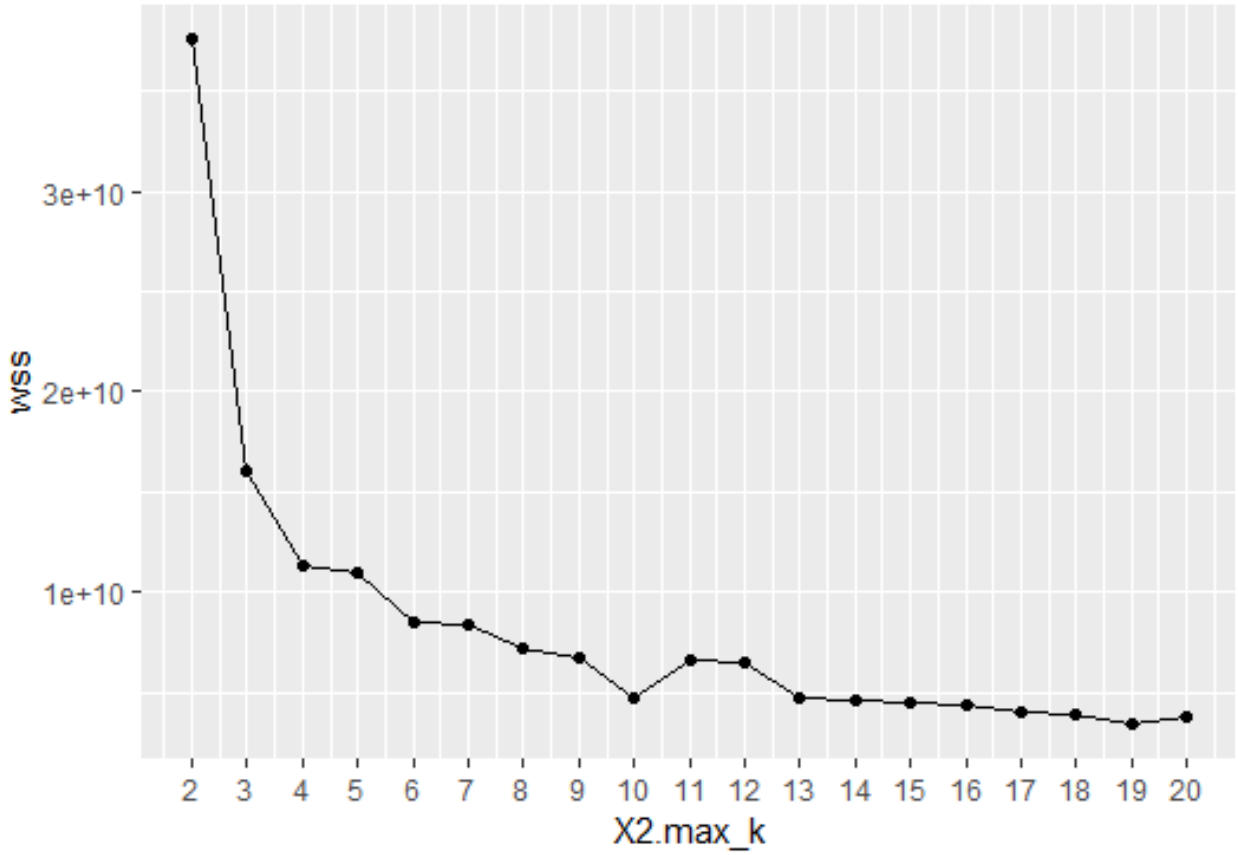
```
# Set maximum cluster
max_k <-20
# Run algorithm over a range of k
wss <- sapply(2:max_k, kmean_withinss)
```

We area ready to run the k-mean algorithm over a range from 2 to 20, and store the tot.withinss values.

```
# Create a data frame to plot the graph
elbow <-data.frame(2:max_k, wss)
```

ploting the graph to visualize where is the elbow point

```
# Plot the graph with gglop
ggplot(elbow, aes(x = X2.max_k, y = wss)) +
   geom_point() +
   geom_line() +
   scale_x_continuous(breaks = seq(1, 20, by = 1))
```



Por lo tanto, la cantidad óptima de K seria 5.

A. Examining the clusters

cluster1 <- kmeans(a, 5)

cluster1$size

[1] 100 34 8 23 18

To table, for better understand:

NUMBER OF CLUSTER	CLUSTER SIZE
1	100
2	34
3	8
4	23
5	18

cluster1$centers

Only taking in consideration important variables:

NAME	DESCRIPTION
F105	HOW MANY POLICE STAFF ARE ASSIGNED AT THE POLICE STATION? - TOTAL1
F116	HOW MUCH CIVIL STAFF WORK AT THE POLICE STATION? - TOTAL
F202	WHAT IS THE TOTAL AREA THAT OCCASES THE POLICE STATION / SPECIALIZED UNIT? - OCCUPIED AREA M2
F300	DOES THE POLICE STATION COUNT ON DUNGEON (S) / ROOM (S) OF MEDITATION: HOW MUCH?
F402	DOES THE POLICE STATION / SPECIALIZED UNIT HAVE ACCESS TO THE DATA INFORMATION OF THE (LOS) / LA (S): National Registry of Identification and Civil Status (RENIEC)
F517	HOW MANY PEOPLE WERE INTERVENED BY THE STAFF OF THE POLICE STAFF DURING THE MONTHS OF: TOTAL - YEAR 2017?
F600	TOTAL NUMBER OF TRANSIT ACCIDENTS

Centers

Cluster	F105	F116	F202	F300	F402	F517	F600
1	50.84000	3.390000	692.3200	1.230000	0.8000000	969.370	39.51318
2	80.52941	5.117647	736.3824	1.823529	0.9411765	8408.882	69.24195
3	94.62500	5.875000	1208.6250	1.875000	0.8750000	51525.375	85.07936
4	78.91304	5.000000	746.4783	2.521739	0.8260870	23804.087	72.14144
5	102.55556	6.055556	1011.7778	1.777778	0.9444444	26586.056	74.94706

5. CONCLUSIONS

There are many police stations (100) which have an average of 35-36 registered traffic accidents, and the others on average have more than 60. Therefore, there should be more specialized police stations to deal with the problems related to traffic accidents.

There are an average of 94-95 policemen (cluster 3) that intervene to more than 50,000 people per year, and have on average 1.87 jails while 78-79 policemen (cluster 4) intervene to 23804 people per year and have 2.52 jails, for therefore there should be more jails for cluster 3 of police stations.

6. SOURCES

INEI microdata (National census of police stations, 2017): http://iinei.inei.gob.pe/microdatos/

R Studio (for cleaning and clustering): https://www.rstudio.com/

Data cleaned: https://drive.google.com/open?id=1BGNnvwNpmyopxLCZ94uucEHtId92DMUR

Watson Studio (for folium maps): https://www.ibm.com/pe-es/marketplace/watson-studio