**Soft Skill problem set**

- How will you schedule work in GoodAI Consulting to meet the deadlines?

To optimize the solution in proper way with meeting the condition with deadline. I will allow all my employee to read the task carefully and come up all solutions in a day and will write it down together. Every person has different problem-solving skills set and it would be good to listen everyone’s opinion with some reasons. Then there can be voting to choose right way. Best way to approach to any big task is to cut down the task in smaller task with more deadlines.

- Which particular solutions/directions will you suggest to our employees so that they don’t explore too many wrong methods?

First, I will ask database experts to filter down Important data, so I don’t get lost during optimization. Then I will ask my visualization expert to present data in forms of graphs, so it can be more easily understood by my software engineer and ML experts. ML experts will try to figure out the best algorithm suitable for the task and the best solution. Finally, the software developer will develop the solution with proper professional standards along with documentation optimizing the code.

- Write your solution in one page (max). Add only important and relevant information there.

The whole task consists of 3 small tasks which can be solved alone with different approaches

Location of the Rescue station: We have data and we need place the rescue station effectively in map. First, we can filter out the ems cases from data and make our new dataset because ambulance will only serve to emergency cases. Now when we have filtered data we can visualize it graphically to get proper idea how is our data distributed. My solution was to use clustering algorithm for 5 rescue station and the centroid of my cluster will give the location of new rescue station, so they can be served easily.

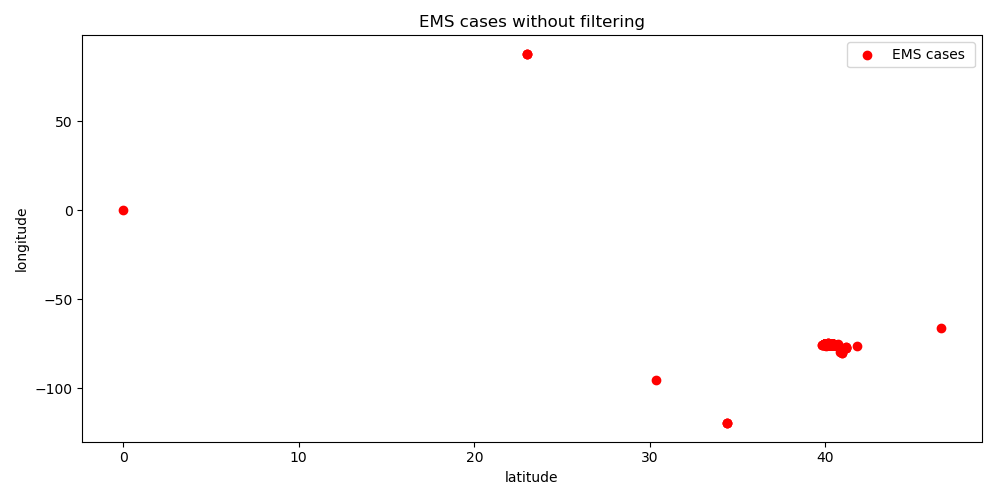
Find the call center working hour: again we can filter out our data first in form of histogram, I am assuming real time situation that cases occur differently through out the year due to weather condition .Then filtering out data again making histogram of data for months then finally each hour .Now we have visual data with high times and low times .We can form general simple equation to find the number of employee required each hour given 2 min to process the problem .Now we can use our ML technique to find no of shifts , per month and during which quarter of year to save money along with fulfilling our demand.

Navigation: I think data is not enough to answer the problem. I can not comment until I know how much rush will be there, how is other way going to save with time. I would ask for road statistics with time needed to arrive from one location to another with different path. It would be again NP-Hard problem, but of course with more statics we can formulate some conclusions and solution.

**Technical problem report**

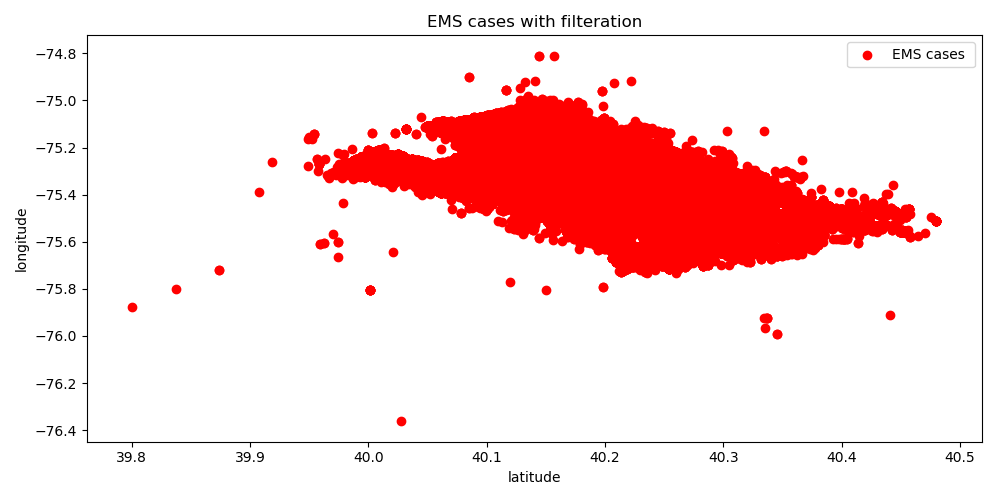
**Rescue station**

First, I filtered out the EMS cases from the data cause the ambulance will be only serve the EMS cases. Put in into figure to see the diversity of data points then. Then I saw that the all EMS cases cannot be served with the 5-rescue station. Even the solution looks virtually possible to serve the all the EMS case with nearby rescue station but it’s not practically possible cause distance are far away from each other.

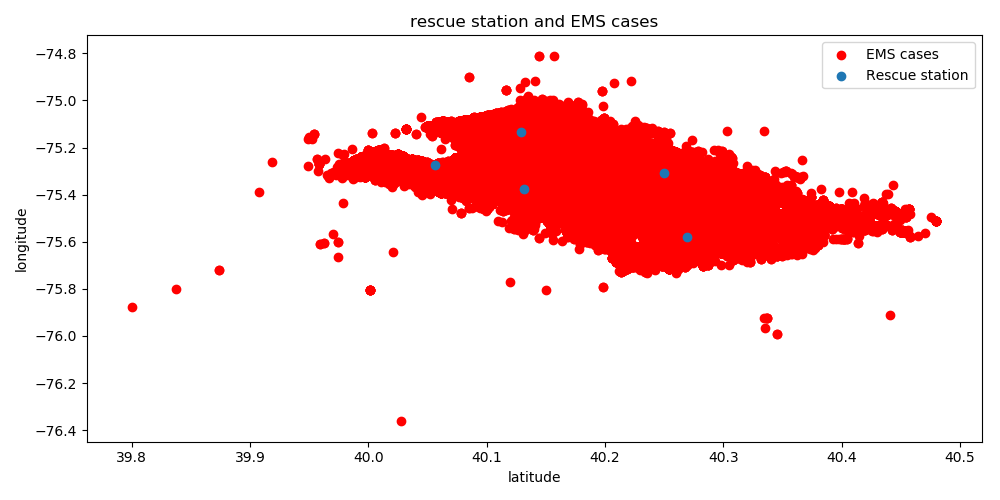


Example in figure above: most of data points are in between ~(40,-75) I conclude to point these data points maybe noise /false reading or human error .Making rescue center with these reading will be the wrong decision.

Assuming the data is from same city these noisy data points are not possible practically. I filtered out the noise just using general static statistics. I used data inside 3 times standard deviation with upper and with lower limit in which I was able to eliminate 22 false reading with 99.7% of data. Result of datapoints after filtration can be visualized from image below



My approach is to make cluster using Kmeans which will converge and gives the centroid with coordinates which can be used to for making new rescue centers. So, considering the data points to serve most cases algorithm converges to 5 points which can be seen in figure below.



With centroids [ 40.13166619 -75.37713108]

[ 40.05610948 -75.27301505]

[ 40.26925621 -75.58154617]

[ 40.2498393 -75.30768708]

[ 40.12892178 -75.13533935]

In figure above, Blue points are represented by rescue stations and red point are the emergency cases.

If I had more time I would try other ML algorithms with different hyperparameters. I consider this task to be best solved using unsupervised learning specially with clustering.

Considering the data points which I consider as noisy or false which can be possibly true readings which can lead to wrong decision. Also clustering algorithm is based no number of cases per took in area so the EMS cases occurred far from the rescue center will need more time to process.