EDG (group 5) A4 Write-up

1. How will setting `MIN_DISTANCE = 0` (in `LocationService`) affect the clusters? Try it and see! What is a reason to choose 0?

   Setting MIN_DISTANCE to 0 makes the program track data when a person is sitting still and the phone doesn't change location. The DBScan algorithm could create clusters using all the data at one place because it needs multiple data points within a range.

   This also could be used to determine how long a person is staying in a certain location. For example, it could track how much time a student spends at the library, in class, or in their room, since they won't be moving too much in those locations.

2. The DBScan code infers generic type `T` so that the clustering algorithm can work on other data points, not only GPS coordinates. But under what condition will it work? In other words, what must `T` define in order for the list to be a valid parameter for `DBScan`?

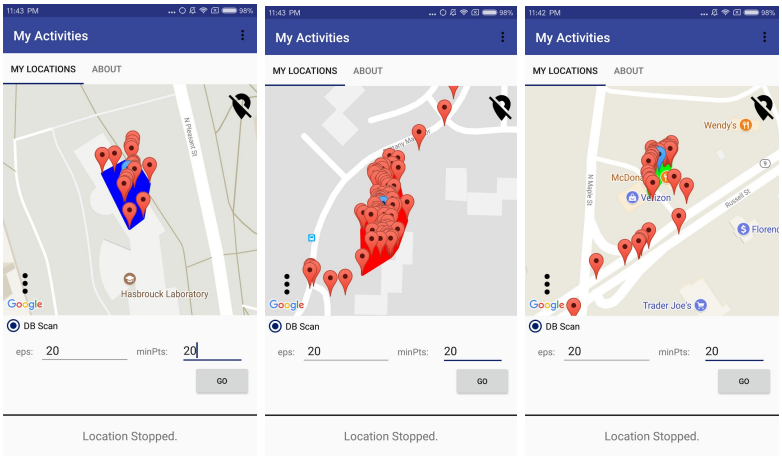
   T extends the interface Clusterable<T> and DBScan uses Clusterable' s distance() method. For our assignment, this method is implemented in GPSlocation.java to calculate how far two points are away from each other. If the distance between points can't be calculated, then finding what points have a distance less than the epsilon is impossible. Clusters cannot be formed if there is no distance to be found. Therefore, T must define a distance method to be a valid parameter for DBScan.

3. Why do we use a `HashMap` to store the state of each point in the DBScan algorithm?

   A HashMap is a data structure for storing exact key value pairs that look up values using a given key. The HashMap is used to determine whether a point has been visited or not. An array or linked list would not be as effective since it is unlikely the points will be in such a pattern that we can iterate across them. As a cluster expands its search from a centroid, it doesn't need to iterate if it can immediately look up where the point is in the HashMap.

4. Collect data as described for the equivalent of at least one day. (It can be spread out over several data-collection sessions on different days). Make sure that you cover a decent-sized area -- around campus, or ideally a combination of campus and other areas.
   a. Visualize the map and submit a screenshot of the DBScan clusters for at least two choices of `epsilon` and `minPts` parameters. For most Android devices, holding the power button and home button simultaneously will take a screenshot.

   Unfortunately, when we rebuilt the app using a different computer, we accidentally deleted the data already on the phone so we don't have a screen shot of cases 2 and 3 to show and we don't remember the exact numbers used for case 3.(we posted this on piazza at Wed, 9 : 46).



Case 1:

Eps = 20, minPts = 20. In this setting, our DBScan worked well on this set of data. It can clearly determine whether it's a meaningful place we stayed at (with a cluster) or just a road we took(single noise point)

Case 2:

eps = 40 minpts = 40, this setting will make some clusters disappear. For example, the short stay in McDonald's doesn't have a cluster with these settings since it doesn't have enough data points.

Case 3:

Very small eps and minpts: This will result some unmeaningful stay with a cluster on map for example a very short conversation on the way to class.

b. Briefly describe how each parameter changes the resulting clusters.

As Eps increases, the size of each cluster grows around the given centroid. Say a cluster with a given eps encompasses the Du Bois library. Another cluster given an eps with ten times the size of the previous one could encompass the student union as well.

As minPts increases, the number of clusters decreases, vice versa.McDonald example.

c. Describe the data-collection process. Where did you go? How well did your final clusters describe your day? Were there any surprises when you saw where you spent your time? Was it a typical day, or did you spend time in any locations that are out of the ordinary?

1.Where: Home, Hasbrouck, McDonalds, W.E.B Du Bois Library. LGRT

2. How well: pretty good

3.Surprise: I spent a lot of time at home.

4.Typical?: It is not a typical day, I drove to McDonalds since it was too late to eat in dining common.

5. What is one other possible application of clustering you could imagine (not location data)?

Facial recognition: This could cluster pigments of color together to determine distinct features of individuals and the differences between different individual faces to try to match a name to a face.

Spread of cancer in a person's body: Once data differentiating between cancerous cells and normal cells is implemented, it is possible to cluster and map where cancer cells are and

how they are spreading so that chemotherapy or any other treatments can be more effective in their targeting.

6. Any difficulties? Comments? Ideas for improvement?

Difficulties 1: When the phone's screen goes black, the application stops collecting data after a few minutes. However, by changing the settings such that the screen was on at all times, there was no interference in data collection.

Difficulties 2: In buildings, the gps data was not precise and sometimes never registered at all. It was not precise when collecting data in a building like the LGRT and Bolder apartments. It never recorded data in the library, bluewall and the cafe in the student union basement.

Difficulties 3: Data lost. MAKE SURE U DON'T UNINSTALL THE PROGRAM. We lost two days worth of data for that.