The Results and Discussion of the BLE Device Sighting Data Analysis In South Africa

Main Objection

Using Statistics methods there is a number of interesting relationships that can be determined in the presented sighting Data. However, due to the presented time constraints, the main objective of this paper is to present the use of statistical methods to determine the most visited area/coordinates (latitude, longitude) and propose a hotspot area/ coordinates for the BLE Device advertising purposes in South Africa.

Approach

To fully answer the forementioned question above, the challenge was divided into 3 stages presented below:

1. **Data Capturing and Processing**
   The first step in solving this question was processing the raw sighting data into a more usable data. This was achieved by:
   - From the HeatMap and the Scatter plot presented in Figure 2, it can be seen that the GPS coordinates are distributed to form 3 data clusters. With two located in South Africa and One in the tropical waters south of Ghana.
   - Using the statistics methods on a widely distributed data like this might affect the reliability of the conclusions. As a result, the sighting data was then divided into three data clusters.
   - For each cluster, the number of visits per coordinate were then calculated by checking the number of times a coordinate (latitude, longitude) appeared on the data. See Figure (3 and 5).
   - From the above step, it was established that most coordinates were visited more than 2 times. To get the most visited coordinates, a filter was applied to get the coordinates that were visited at least 3 times. See Figure (4 and 6).

   This data was then taken to the next stage for analysis and presentation.

2. **Data Analysis and Presentation**
   The data of the number of times each coordinate was visited was then used to determine mean, mode and standard deviation of each cluster.
   - The Mean presents the average number of times a coordinate was visited within the one-month period.
   - The Mode presents the number of visits that appeared most in the overall cluster dataset.
   - The Standard Deviation presents the statistical analysis that measures the spread of data around the mean.

3. Results Discussion

The clusters were divided into two clusters about the line of latitude (-26.1164). This was because looking into the scatter plot in Figure 2, it can be seen that the two clusters seem to be separating along that line.

From the data analysis presented in Data Analysis and Presentation.xlsx, the following results were obtained. Please see the Figure presented in Figure 1 for the cluster reference of the two clusters in South Africa.

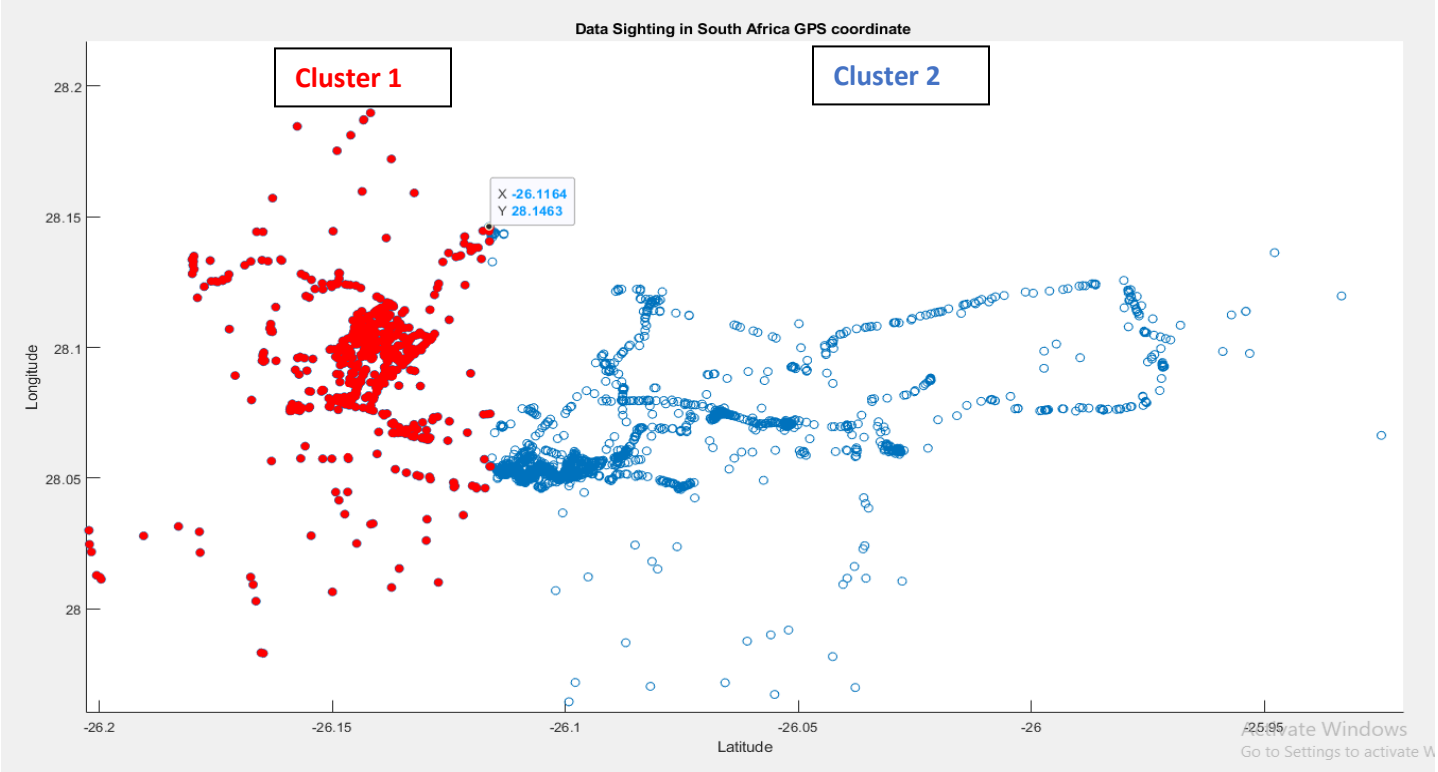

Figure 1: The Two Data Distribution Clusters in South Africa

Comparing the mean of the number of visits for each cluster, Cluster 1 has a mean of 1.276220977 and Cluster 2 has 1.214573938. Therefore, it can be concluded that Cluster 1 has the greatest number of visits per coordinate within the one-month period compared to Cluster 2. However, comparing the mean of the number of the most visited coordinates, Cluster 1 has 3.25 and Cluster 2 has 7.318181818. Therefore, it can be concluded that Cluster 2 has the greatest number of visits per coordinate within the one-month period compared to Cluster 1.

Comparing the Mode of the number of visits for each cluster, both clusters are associated with a mode of 1. As a result, it can be concluded that the number of visits that appeared the most in both clusters is only once. And, Comparing the mode of the most visited coordinates, both clusters also have an equal mode of 3.

Comparing the Standard Deviation of the number of visits per cluster, Cluster 1 is associated with a standard deviation of 0.468215849 and a standard error of 0.94% and Cluster 2 has a standard deviation of 0.917802348 and standard error of 1.41%. Therefore, it can be concluded that most of the number of visits per coordinate spread in Cluster 1 is in line with the average number of visits per coordinate of the cluster.

Conclusion

Using the data analysis and discussion above, it can be concluded that Cluster 1 is associated with the most number of visits per coordinate with high reliability. However, even though it is associated with the most number of visits, Cluster 2 seem to be getting the most highest number of visits for a few coordinates with a minimum reliability. Therefore, it can be concluded that Cluster 1 is the best cluster/location for BLE Device advertising purposes.

Appendix

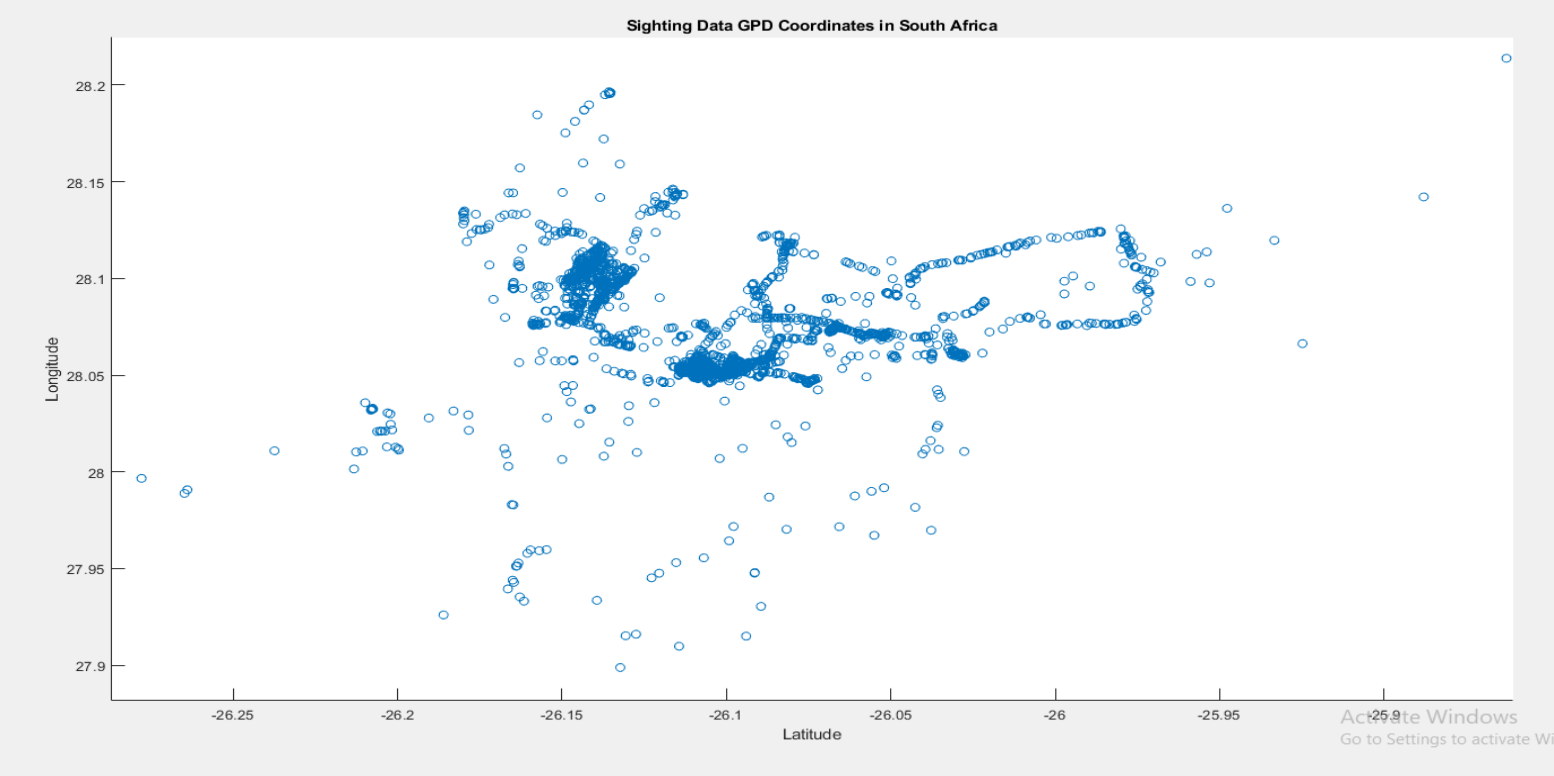

Figure 2: Data set in South Africa.

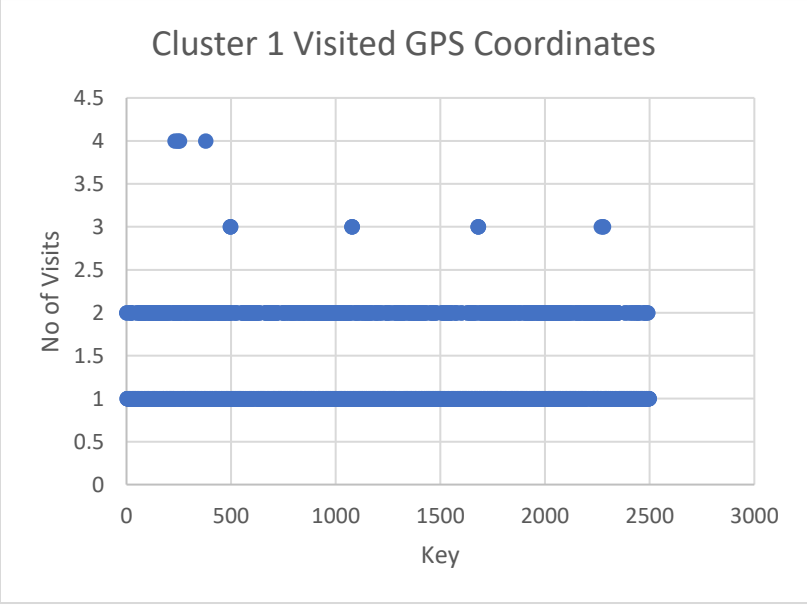

Figure 3: Cluster 1 Visited GPS Coordinates

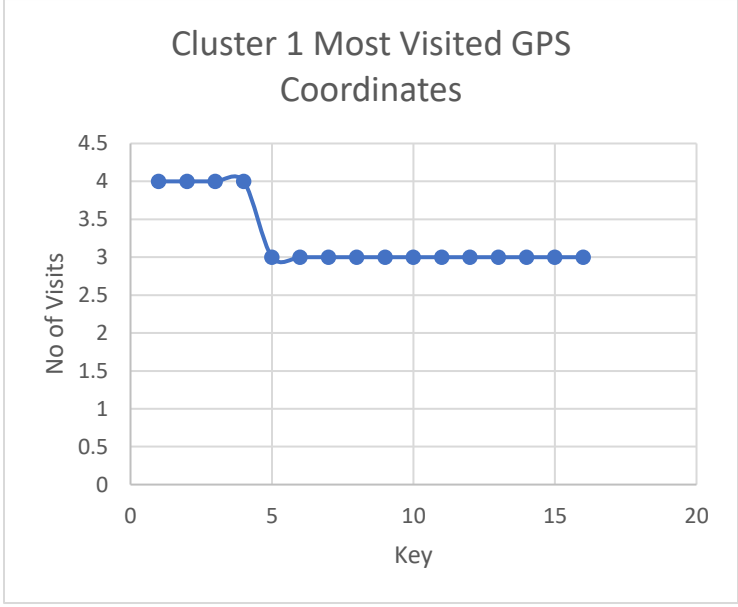

Figure 4: Cluster 1 Most Visited GPS Coordinates

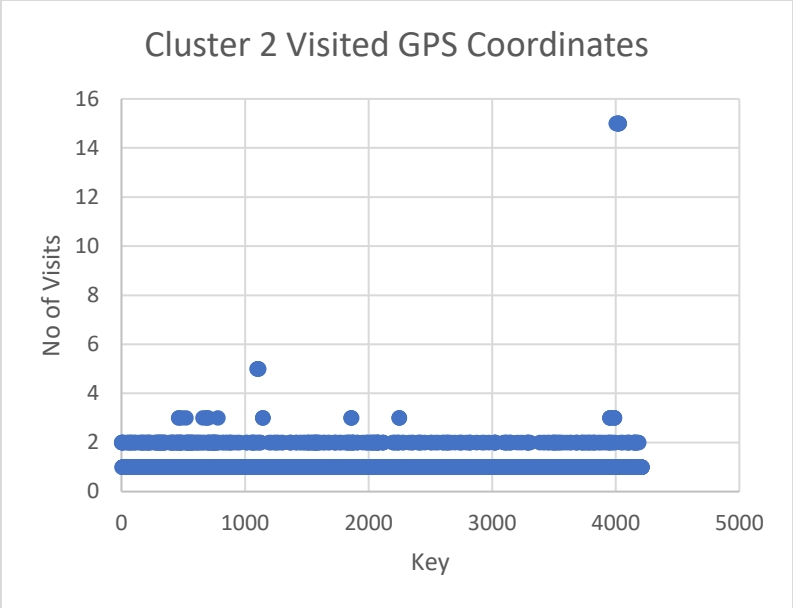

Figure 5: Cluster 2 Visited GPS Coordinates

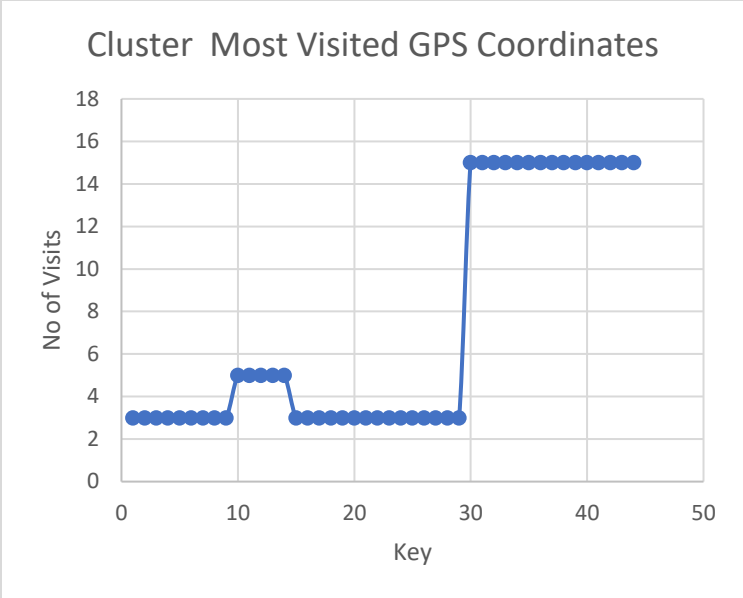

Figure 6: Cluster 2 Most Visited GPS Coordinates