

# AI6128 Urban Computing Project 1

Luai Wei Jie Jonathan  
Nanyang Technological University  
G2102392J  
Singapore  
luai0001@e.ntu.edu.sg

Li Xuemeng  
Nanyang Technological University  
G2202226E  
Singapore  
lixu0025@e.ntu.edu.sg

Ma Lingjie  
Nanyang Technological University  
G2202658A  
Singapore  
lma012@e.ntu.edu.sg

Lu Pai  
Nanyang Technological University  
G2102572F  
Singapore  
lupa0002@e.ntu.edu.sg

**Abstract**— Despite the rapid advances in the field of sensorization and locating technologies today, indoor localization remains as an open area of research due to the multitude of complexities and challenges faced during data collection and processing. In this report, the team is presented with a set of spatio-temporal data collected by surveyors for different floors across two buildings. Based on the data, an exploratory data analysis is conducted by the team and the waypoints, Geomagnetic, Wi-fi and iBeacon heat maps are then plotted for each of the respective sites and floors.

**Keywords**— Indoor localization, sensorization

## I. INTRODUCTION

The emergence of the Internet of Things (IoT) has led to the widespread development of IoT devices which are interconnected to networks for communication. These IoT devices are increasingly ubiquitous and various applications of these devices can be seen in our everyday lives. With the increasing interest in this area, these devices are also being applied into indoor localization. Over the recent years, a plethora of technologies have been proposed for indoor localization. While the list of technologies are non-exhaustive, more prevalent examples of such include the use of linear and rotational motion sensors (Inertial Navigation System), the use of radiating magnetic fields for location detection (Magnetic Based Navigation), fingerprinting via Received Signal Strength (Wi-Fi) and also received signal strengths via Bluetooth Low Energy (BLE). For this group project, our team was tasked to pre-process a set of sample data which was provided by Microsoft for their Indoor Localization Competition 2.0 and make use of that data to visualize the waypoints, geomagnetic, wi-fi and iBeacon heat maps for each of the floors across the two sample sites.

## II. DATASET

### A. Data Information

The sample data provided information on the indoor traces of the surveyor on both sites. The traces were provided in the form of text (.txt) files by Microsoft and each file corresponded to a path taken by a site surveyor from point  $p_1$  to point  $p_2$  on both site 1 and site 2. During the survey, the surveyor held on to an Android smartphone which was running a sensor data recording application in the background. Various data readings were then being gathered by the Inertial Measurement Unit (IMU) and Magnetometer which were embedded in the Android smart phone. Accelerometer and

gyroscope readings were extracted from the IMU, while geomagnetic field readings were extracted from the magnetometer into the text file. Data which were provided from Wi-Fi and Bluetooth (iBeacon) signals were also extracted into the same text file. The format of the text files as well as their corresponding values which were collected chronologically via Unix Timestamps are summarized below:

Data Type	Data Values
TYPE WAYPOINT	Coordinate X, Coordinate Y
TYPE_ACCELEROMETER	X axis, Y axis, Z axis, accuracy
TYPE_GYROSCOPE	X axis, Y axis, Z axis, accuracy
TYPE_MAGNETIC_FIELD	X axis, Y axis, Z axis, accuracy
TYPE_ROTATION_VECTOR	X axis, Y axis, Z axis, accuracy
TYPE_ACCELEROMETER_UNCALIBRATED	X axis, Y axis, Z axis, X_b axis, Y_b axis, Z_b axis, accuracy
TYPE_GYROSCOPE_UNCALIBRATED	X axis, Y axis, Z axis, X_b axis, Y_b axis, Z_b axis, accuracy
TYPE_MAGNETIC_FIELD_UNCALIBRATED	X axis, Y axis, Z axis, X_b axis, Y_b axis, Z_b axis, accuracy
TYPE_WIFI	ssid, bssid, RSSI, frequency, timestamp
TYPE_BEACON	UUID, MajorID, MinorID, Tx Power, RSSI, Distance, MAC Address, timestamp

Table 1: Data values for each corresponding data type

Based on the summary in table 1, there were a total of ten different data types which were collected during the survey. Of these ten data types, the ‘TYPE WAYPOINT’, ‘TYPE\_MAGNETIC\_FIELD’, ‘TYPE\_WIFI’ and ‘TYPE\_BEACON’ data were used for the visualization of the waypoints, geomagnetic, Wi-Fi and iBeacon heatmaps respectively. More details of these will be individually elaborated on in the later sections of this report.

### B. Exploratory Data Analysis

To better understand the provided dataset, an exploratory data analysis was performed before the essential tasks. Total counts of unique data records of 7 types of data are presented in the Table 2 for all sites and floors. Uncalibrated data which was not used in this assignment was not included in the EDA.

Site	Floor	Waypoint	ACCEL	GYRO	MAG	ROTATE	WIFI	BEACON
1	B1	1034	265122	265122	265122	265122	333724	31814
	F1	975	290966	290966	290966	290966	868647	31477
	F2	1049	382553	382553	382553	382553	778674	52499
	F3	1012	475461	475461	475461	475461	702449	47882
	F4	1042	356808	356808	356808	356808	703488	20619
2	B1	534	132112	132112	132112	132112	201667	2049
	F1	1006	272351	272351	272351	272351	795937	3900
	F2	362	85189	85189	85189	85189	228229	311
	F3	278	69756	69756	69756	69756	177006	254
	F4	215	54967	54967	54967	54967	122219	231
	F5	298	67935	67935	67935	67935	163887	868
	F6	565	130772	130772	130772	130772	272565	2164
	F7	273	67491	67491	67491	67491	178423	1878
	F8	265	66601	66601	66601	66601	147718	697

Table 2: Unique data record counts for all sites and floors

The first thing needs to be called out is that the number of WAYPOINT records are dramatically less than other data types due to lower sampling rate. Thus, most of records are not paired with recorded ground truth location. In the later session, an estimated location is used for visualization, which was calculated using motion sensors' readings.

ACCEL/GYRO/MAG/ROTATE have equal numbers of records and share the same timestamp, from which we knew the sampling rate was  $\sim 50\text{Hz}$ , providing stable and adequate motion traces for position estimation. WIFI data were collected at a lower rate  $\sim 0.5\text{Hz}$ , but number of records are larger as multiple WIFI Ap's were recorded together at one time stamp. The timestamp recorded in the last entry of WIFI data (known as last seen timestamp) is different from the record timestamp, but it is not used for visualization. Number of BEACON data is lower due to fewer available devices.

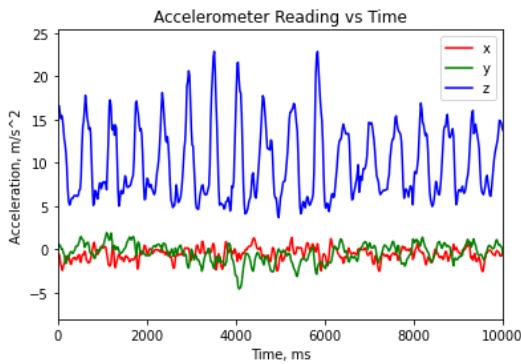


Figure 1: Screenshot of accelerometer reading from dataset

### C. Data Preprocessing

To better visualize the heatmaps of given signals, data pre-processing is required to create more Waypoints for plotting. In this assignment, we use the step position to create waypoint.

When the surveyor walked, the velocity was not constant, but periodically increase and decrease because of the stepping motion, which can be detected by the accelerometer shown in Fig.1. Thus, the peaks and valleys in ACCELEROMETER readings can be used to determine the movement status. Here, we consider the surveyor starts a step when acceleration reached maximum, and 1 stride is between two adjacent valleys. The direction of the step can be calculated from the ROTATION\_VECTOR measured by the gyroscope. The movement distance or the stride length is estimated based on the acceleration readings, capped by the normal stride length of an adult. The step position Waypoints created in this way were used in the following tasks.

Site	Floor	Original	Step Position
1	B1	1034	8468
	F1	975	8483
	F2	1049	9569
	F3	1012	9138
	F4	1042	8982
2	B1	534	4363
	F1	1006	8759
	F2	362	2766
	F3	278	2142
	F4	215	1825
	F5	298	2270
	F6	565	4511
	F7	273	2326
	F8	265	2211

Table 3: Original and step position waypoints for all sites and floors

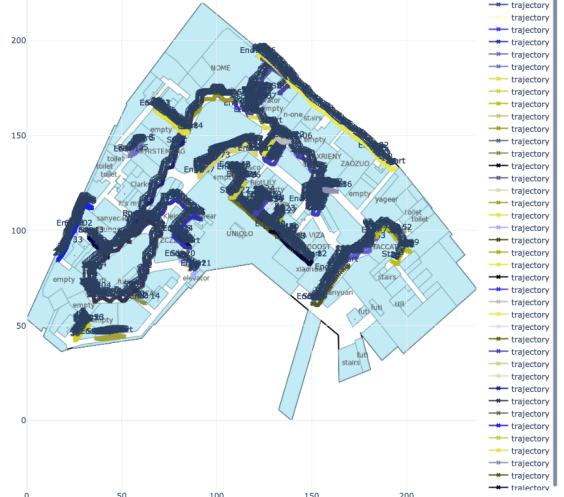


Figure 2: Plot of the combined step position waypoints for site 2, floor 2

### III. ESSENTIAL TASKS

In this part, we show the 4 essential tasks, the visualized way points, geomagnetic heat map, Wi-Fi RSS heat maps of 3 Wi-Fi Ap's and the Ibeacon RSS heat map.

#### a. Visualization of Waypoints

For task 1, TYPE\_WAYPOINT was used. We used the Matplotlib function and an updated `visualize_trajectory()` function to show the whole view of plotting all waypoints for all floors (Fig.1) and the waypoints for each floor individually. Also, from the given function, we got all the sets of Start and End points from the text files for floors and sites.

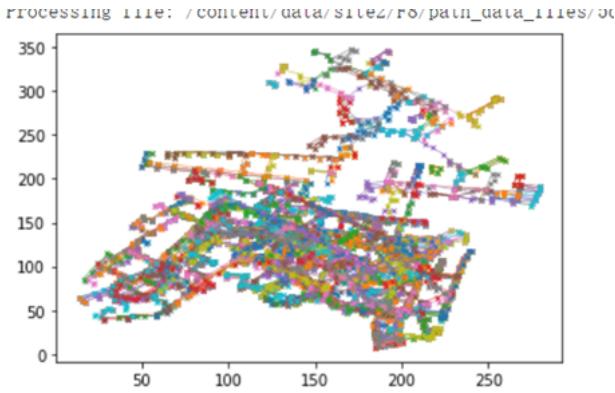


Figure 3: Combined trajectories for all sites and floors

From Figure 3, different trajectories are represented in different colors and connected by ‘-’. It may be difficult to see due to the amount of data. To get them, all paths of text files were loaded, this was reached by changing the loop for the map based on the given ‘Load B1’ function. Meanwhile, in all path files, we gathered and read the waypoint data in advance. Then we extracted the TYPE WAYPOINT from files and split them into x-axis and y-axis directions. All x and y coordinates of the entire dataset are stored in one x array and one y array.

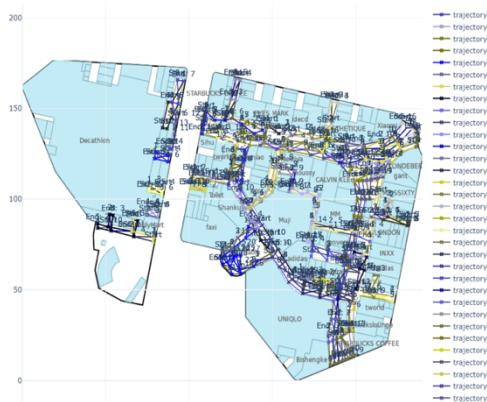


Figure 4: All trajectories for floor F1, Site 1

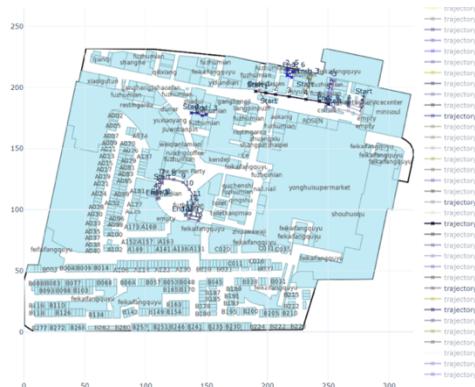


Figure 5: Several chosen trajectories for Floor B1, Site 1

Figure 4 above shows the map of floor F1 from Site 1 with all the trajectories taken from all text files. The other

trajectories over other floors can be checked in Appendix. A. Different colors stand for different trajectories, colors are randomly chosen from (0,255). For each trajectory, it is made up of the waypoints extracted by TYPE WAYPOINT from each file, all of them are represented by nodes(waypoints) and connected by lines. The node is marked as ‘x’, line for ‘-’, And the Start point and End Point are shown with text. Also, the number order of them was also listed. In Figure 5 above, the screenshot of a .html file is taken as an output. In this file, all trajectories in this file are listed on the right, if you double-click on any of them, the chosen waypoints will flash once, and if triple click, all lines will be removed temporary, them just click one or more trajectories you want, they will appear as the shown figure. For recovering the original map, a triple-click again will work. This may help when the user wants to locate any single or triple routes they want, and it is clearer to see. HTML files can be extracted from the Google Drive link in Appendix. E.

During the processing of printing single data from a single file, we found that it took a lot of time, and it is hard to make any conclusion from them, so we output the maps for each floor under site1 and site 2. This step was completed based on the function *visualize\_waypoint()*. For this function, we have updated it from showing each text file’s data to showing all text data on one floor. This was implemented by changing the parameter from the ‘trajectory’ to ‘*trajectory\_sum*’, also a loop was added to make it easier.

From all the results in Appendix. A, it is easy to see that Site 1 has more and unrepeatable data collected than some floors in Site2 like F4, F5, F7 and F8. In our thoughts, maybe there were more collectors walking in floors in Site1, or when they walked in Site 2, maybe they forgot which path they had been on once and did it again, so that causes a repeating trajectory. Meanwhile, as a dataset, more is better (for most situations), so it is more essential to their research so that they could get a more study-worthy conclusion for floors with more waypoints, also may be affect the next tasks.

#### b. Visualization of Geomagnetic Heat Map

The geomagnetic sensor of Android phone measures the geomagnetic field in 3 dimensions and records the directional components (x, y, z). To visualize the geomagnetic heatmap, the geo-magnetic strength was calculated for all timestamps using the following formula.

$$B(P_i) = \sqrt{x_i^2 + y_i^2 + z_i^2}$$

The calculated strength was then assigned to the closest step position  $P_j$  described in previous section, based on its estimated position  $P_i$ .

$$B(P_i) \Rightarrow B(P_j) | j = \underset{j}{\operatorname{argmin}} (|P_j - P_i|)$$

Since we have more geomagnetic records than step position points, there are multiple values assigned to the same step position. Here we simply take the average of all readings as our final geomagnetic strength.

$$B(P_j) = \operatorname{Ave}(\sum B_i)$$

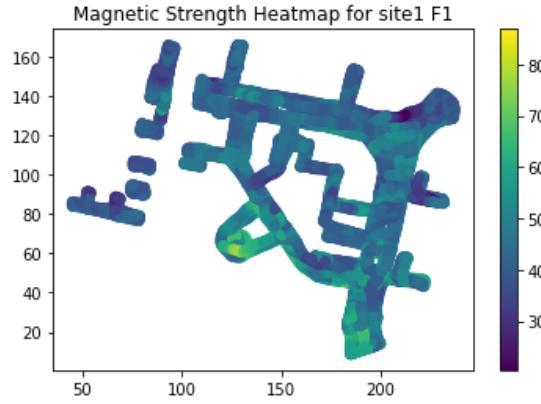


Figure 6: Geomagnetic heatmaps for Site 1 Floor 1

Figure 6 above shows one example of the final strength heatmap of site 1 floor 1. The plots for all sites and floors are attached at the end of the report.

### c. Visualization of Wi-Fi RSSI Heat Map

Wi-Fi is a WLAN technology based on the IEEE 802.11 family of standards. For this essential task, Wi-Fi RSS heat maps of 3 Wi-Fi Ap's will be visualized.

From the dataset, the data with type TYPE\_WIFI are selected. Each row of the data contains the BSSID information (which is the identifier of a particular AP), and RSSI information. These data are then calibrated with the location information. Finally, RSS heat map of 3 Ap's from each floor each site is generated and labelled based on their location and strength.

Below figures are the RSS heat map of the selected Ap's from site1 F1 and site2 F8. All RSS heat maps are also displayed in the Appendix.

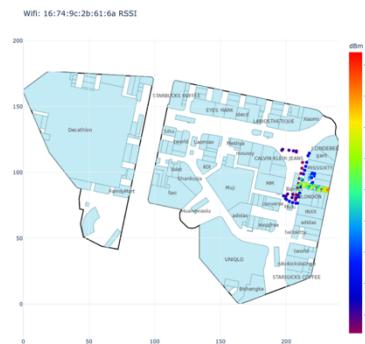


Figure 7: Wi-Fi RSSI Heat Map with BSSID 16-74-9c-2b-61-6a, Site 1, Floor 1

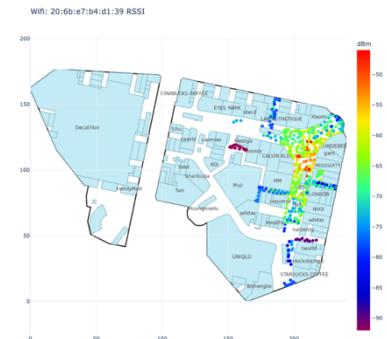


Figure 8: Wi-Fi RSSI Heat Map with BSSID 20-6b-e7-b4-d1-39, Site 1, Floor 1

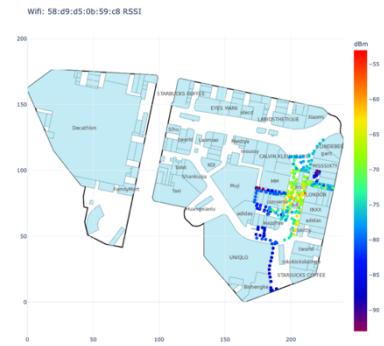


Figure 9: Wi-Fi RSSI Heat Map with BSSID 58-d9-d5-0b-59-c8, Site 1, Floor 1

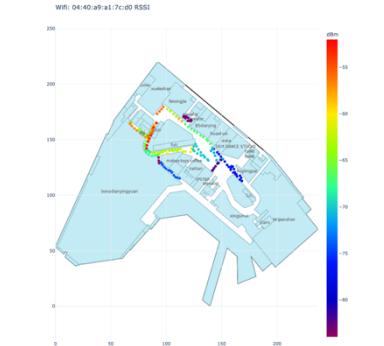


Figure 10: Wi-Fi RSSI Heat Map with BSSID 04-40-a9-a1-7c-d0, Site 2, Floor 8

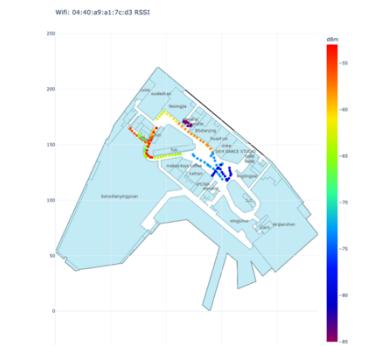


Figure 11: Wi-Fi RSSI Heat Map with BSSID 04-40-a9-a1-7c-d3, Site 2, Floor 8

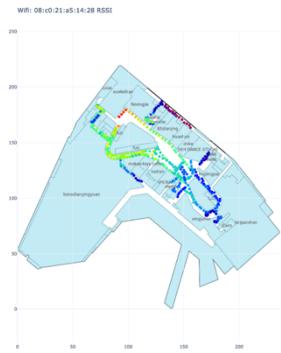


Figure 12: Wi-Fi RSSI Heat Map with BSSID 08-c0-21-a5-14-28, Site 2, Floor 8

#### d. Visualization of iBeacon RSS Heat Map

The iBeacon which was originally developed by Apple Inc back in 2013 is a networking protocol that uses Bluetooth Low Energy (BLE) which is transmitted through a user's mobile device to get information pertaining to their location. In the case of the Indoor Location Competition held by Microsoft Corporation, the iBeacon data was collected by a site-surveyor who was holding an Android smartphone with a sensor data recording application running on the device. Data which was collected under 'TYPE\_BEACON' comprises of mainly the Universally Unique Identifier (UUID), the major ID, the minor ID, the transmitter power (TX), the Received Signal Strength Indication (RSSI), the distance from the surveyor to a deployed iBeacon, the MAC address as well as the padded data to the Unix time. Based on the iBeacon data which was collected, a function to calibrate the iBeacon data to the position of the user was called and this function returns a dictionary called *mwi\_datas*. The dictionary is then input into another function to extract the RSSI readings which are mapped to the closest waypoint data based on the *argmin* function of the closest time stamp. Running the code across both sites and floors, we realized that there were different numbers of iBeacons which were deployed for each of the different floors in both sites 1 and 2. Table 4 below summarizes the total number of iBeacons deployed for each site and floor respectively:

Site	Floor Number	Number of iBeacons
1	B1	55
	F1	20
	F2	16
	F3	14
	F4	12
2	B1	27
	F1	23
	F2	8
	F3	11
	F4	11
	F5	30
	F6	18
	F7	15
	F8	7

Table 4: Summary of iBeacons deployed on each site and floor

From the table above, we can see that there is a significant disparity between the density of the deployed iBeacons for different floors on both sites. The *matplotlib* library from python was then used to plot out the RSSI heat map for the combined iBeacons on each floor and an example of that is shown in Figure 13 below.

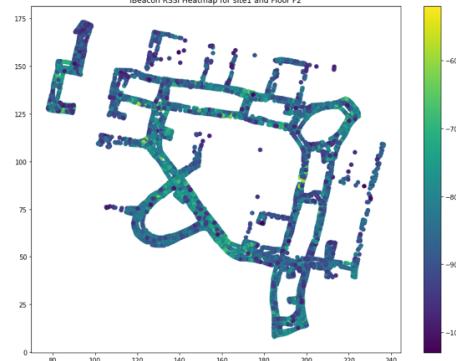


Figure 13: Combined iBeacon RSSI Heat Map, Site 1, Floor 2

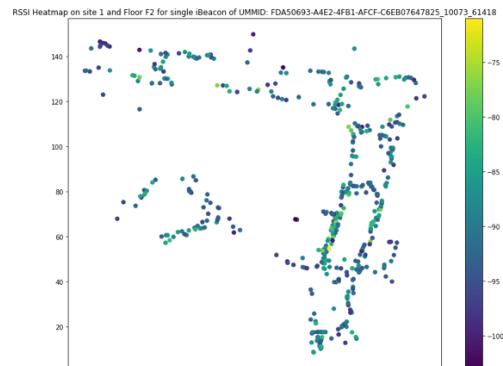
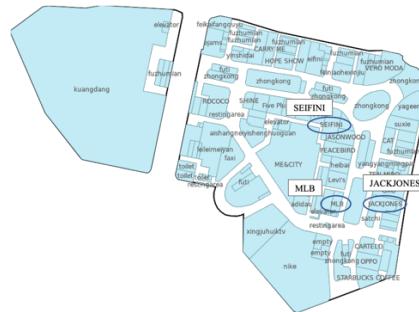


Figure 14: iBeacon RSSI Heat Map for single iBeacon, Site 1, Floor 2



*Figure 15: Annotated map of the layout for Site 1, Floor 2*

In the example of the heat map provided for site 1 floor 2, there were a total of 20 iBeacons being deployed on that level. By comparing the RSSI fingerprints which were collected over the entire floor area, we can see that the range of the recorded RSSI magnitudes across the floor area for site 1, F2 ranged from approximately -100 to -55. The strength of the recorded RSSI gives us an insight on the distribution of the signal strength which was recorded on the device. Based on Figure 13, the recorded iBeacon RSSI signals received from the combined iBeacons seemed to be well-distributed, spanning rather evenly across the floor area, indicating the adequacy of deployments of the iBeacons on floor 2. The plot for the RSSI heat map of a single iBeacon for the same level also showed a difference in RSSI signals received across the floor in Figure 14. Given that the RSSI signal for the single

beacon seemed the strongest when it was near to the *MLB* shop, it is likely that the iBeacon is deployed there. Interestingly, the distribution of the recorded signal strength for this single iBeacon was seemingly inconsistent throughout the varying distances on the floor. One such example can be shown from the presence of several comparatively weak signals clustered near the *JACKJONES* shop which is nearer in distance to *MLB* as compared to *SEIFINI*, a shop which is located further away. One of the reasons for this could be due to the presence of obstructions within the signal path such as walls, or furniture, which prevented the collection of a stronger signal strength despite the nearer distance. This implication also means that in order to optimize the collection of the recorded iBeacon RSSI signals, the placement and density of the iBeacons should take into consideration the presence of different variables such as obstructions, radio noise and even reflections caused by metallic objects.

#### IV. FUTURE WORKS

In this report, we have analyzed and processed the sample data which was given for the waypoints, trajectories, geomagnetic, Wi-fi and iBeacon heatmaps. We have also

identified some of the issues which could be faced when just using a single type of data for indoor localization. In the subsequent work for this data, it is possible to use the different types of data which was processed in the earlier segment of this report to build a deep-learning model and make comparisons of its accuracy against the original waypoints to see if better results can be obtained.

#### V. CONCLUSION

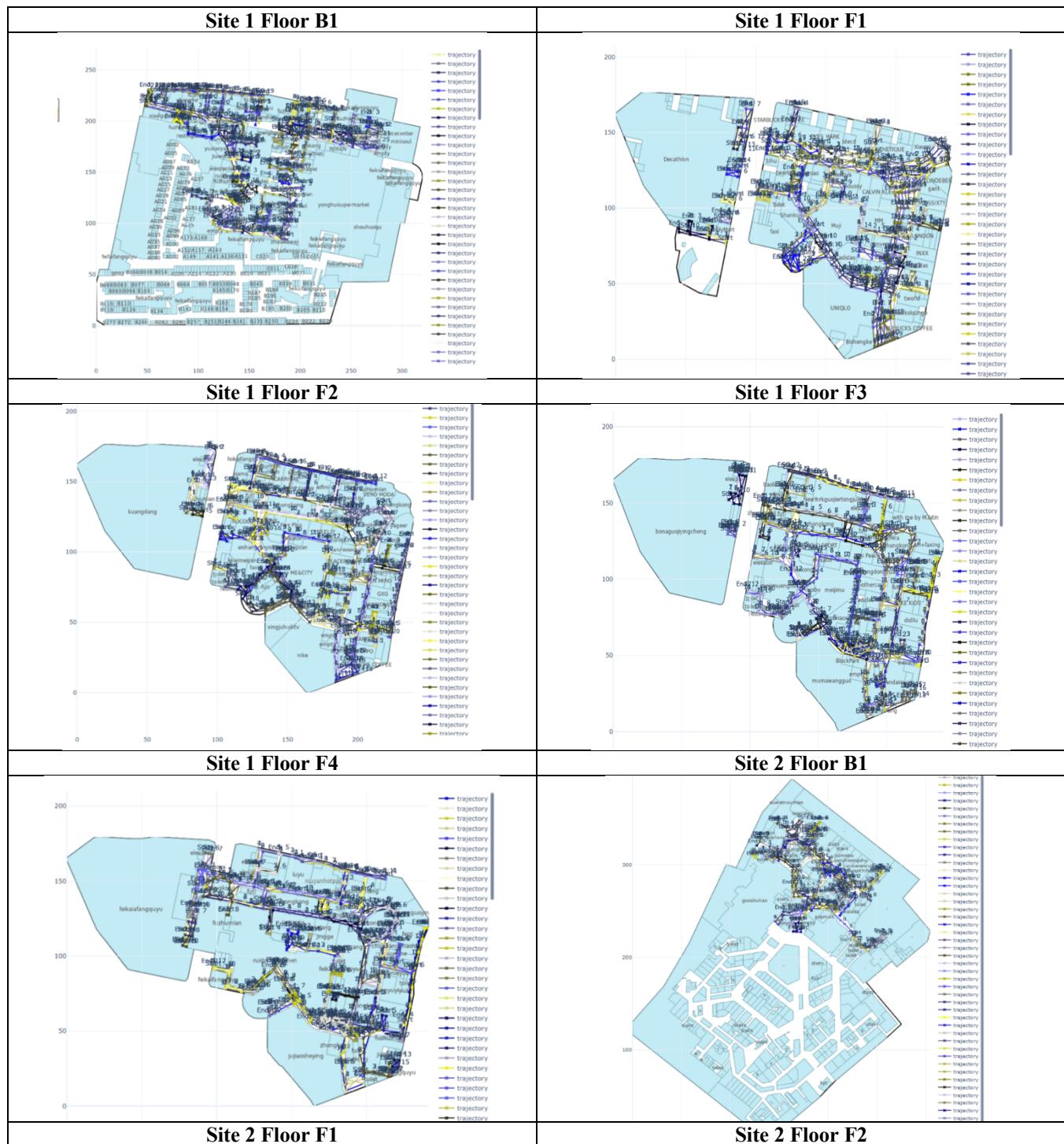
In conclusion, for this Urban Computing assignment, all the group members have done essential research on the dataset and all 4 essential tasks. All source code and visualization output has been included in the Appendix. It is a great learning and exploring experience for us all in regards to the urban computing topics.

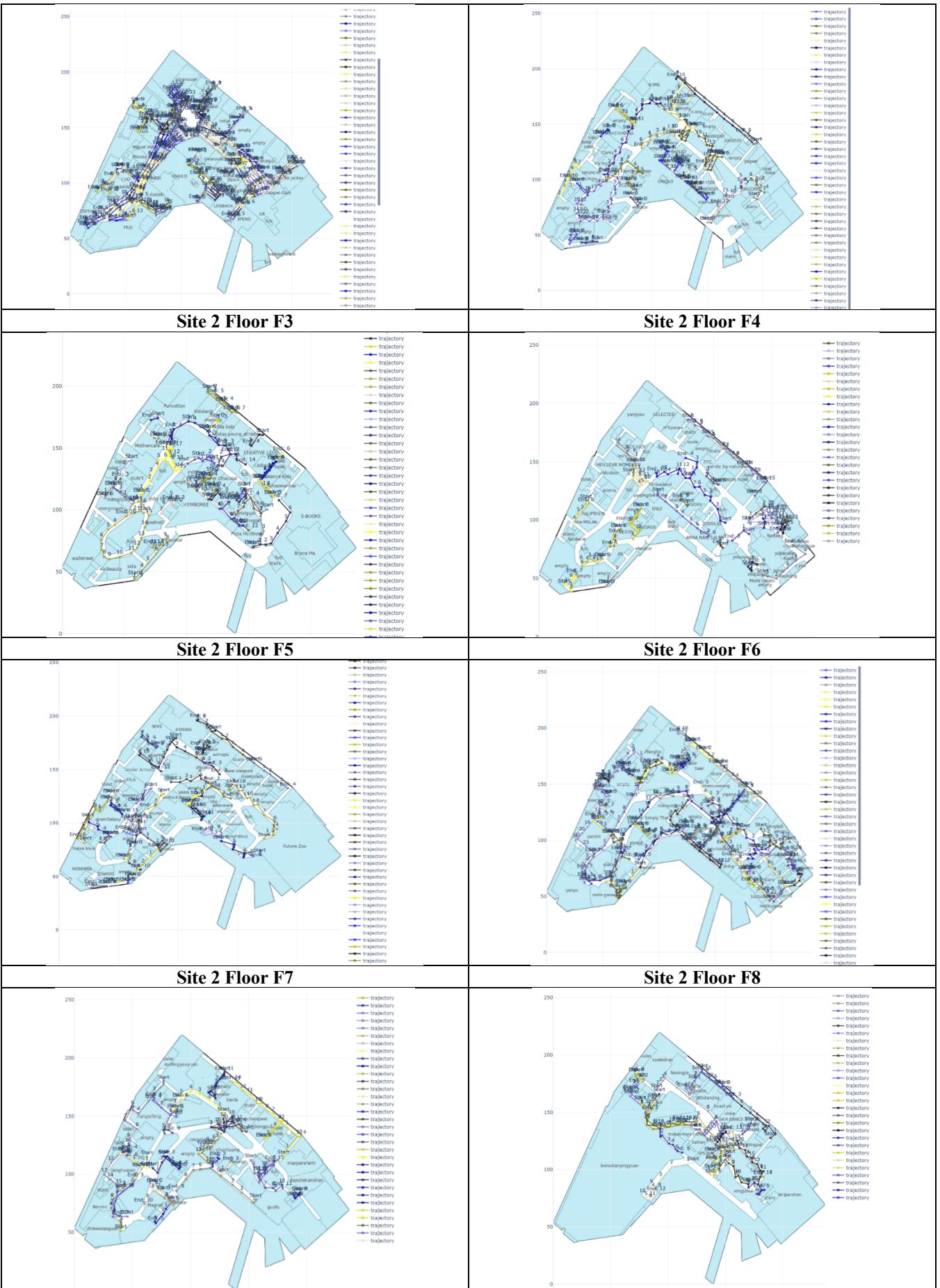
#### GROUP MEMBER CONTRIBUTION

All of us picked one essential task respectively and worked together on this assignment positively. We arranged meetings through Teams and shared ideas through the chatting group channel. Li Xuemeng did the first task, Ma Lingjie worked on the second task, Lu Pai completed the third task and Luai Wei Jie Jonathan focused on the fourth task.

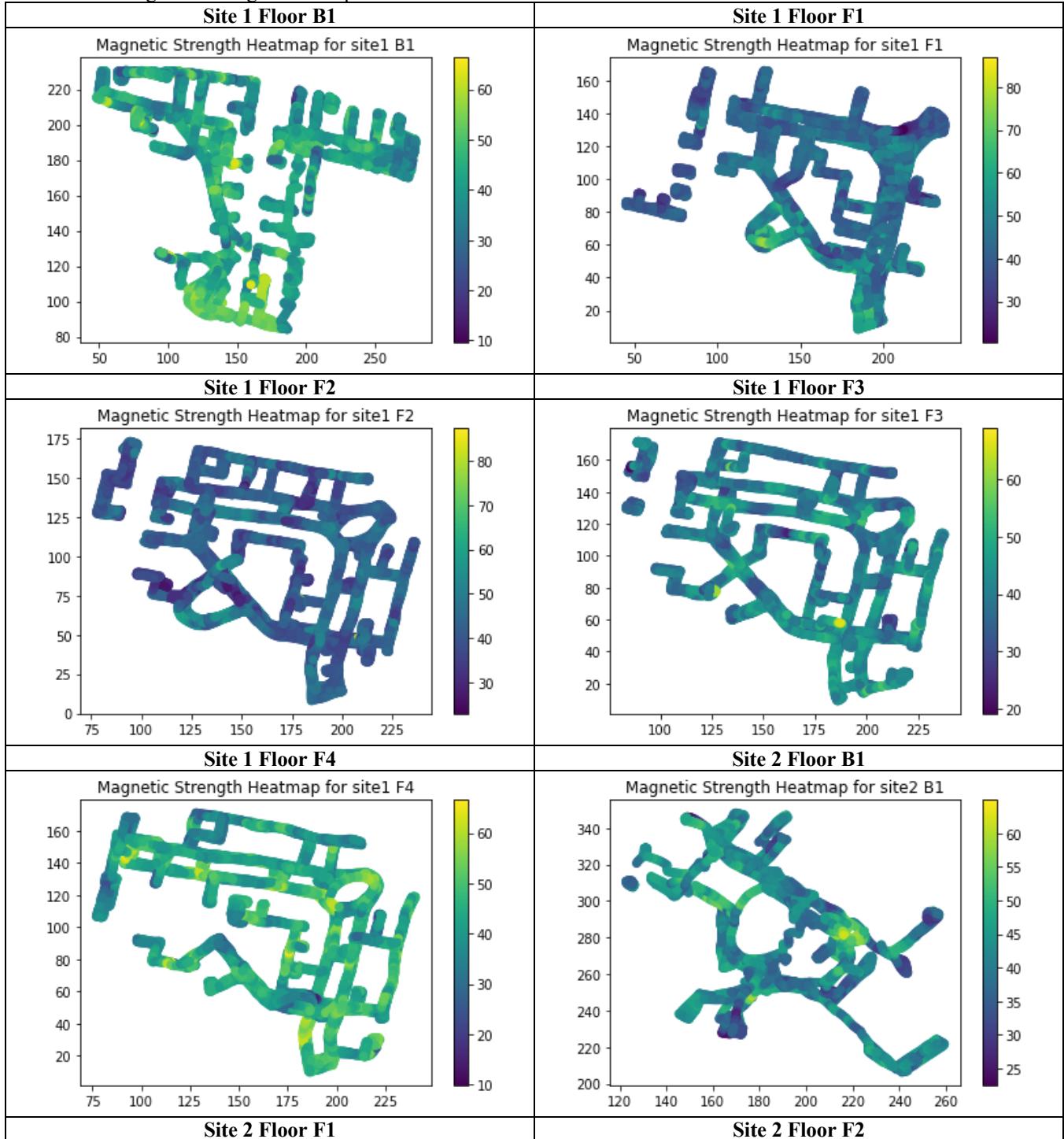
## APPENDIX (SOURCE CODE)

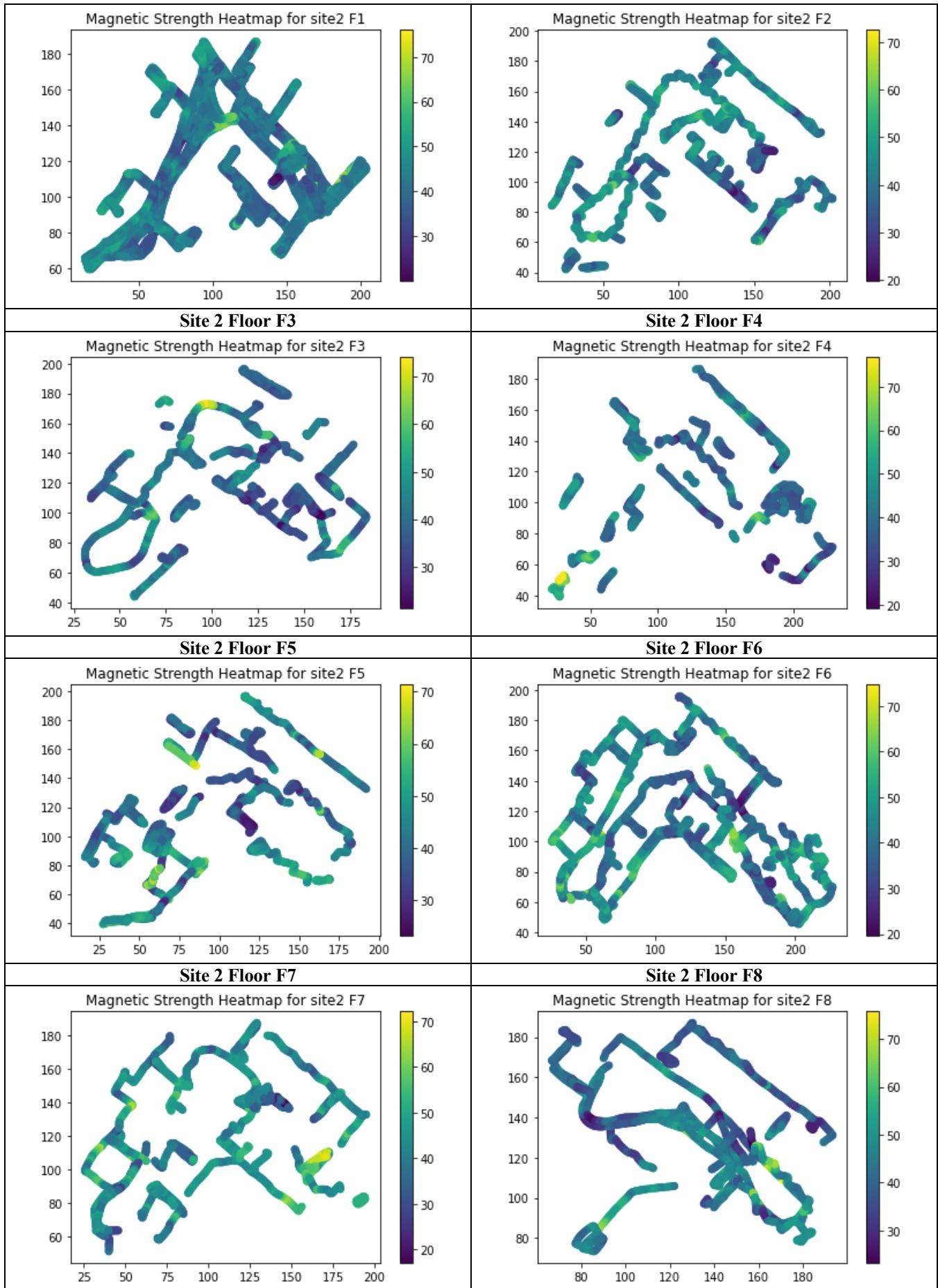
#### A. The waypoints for each floor in each site



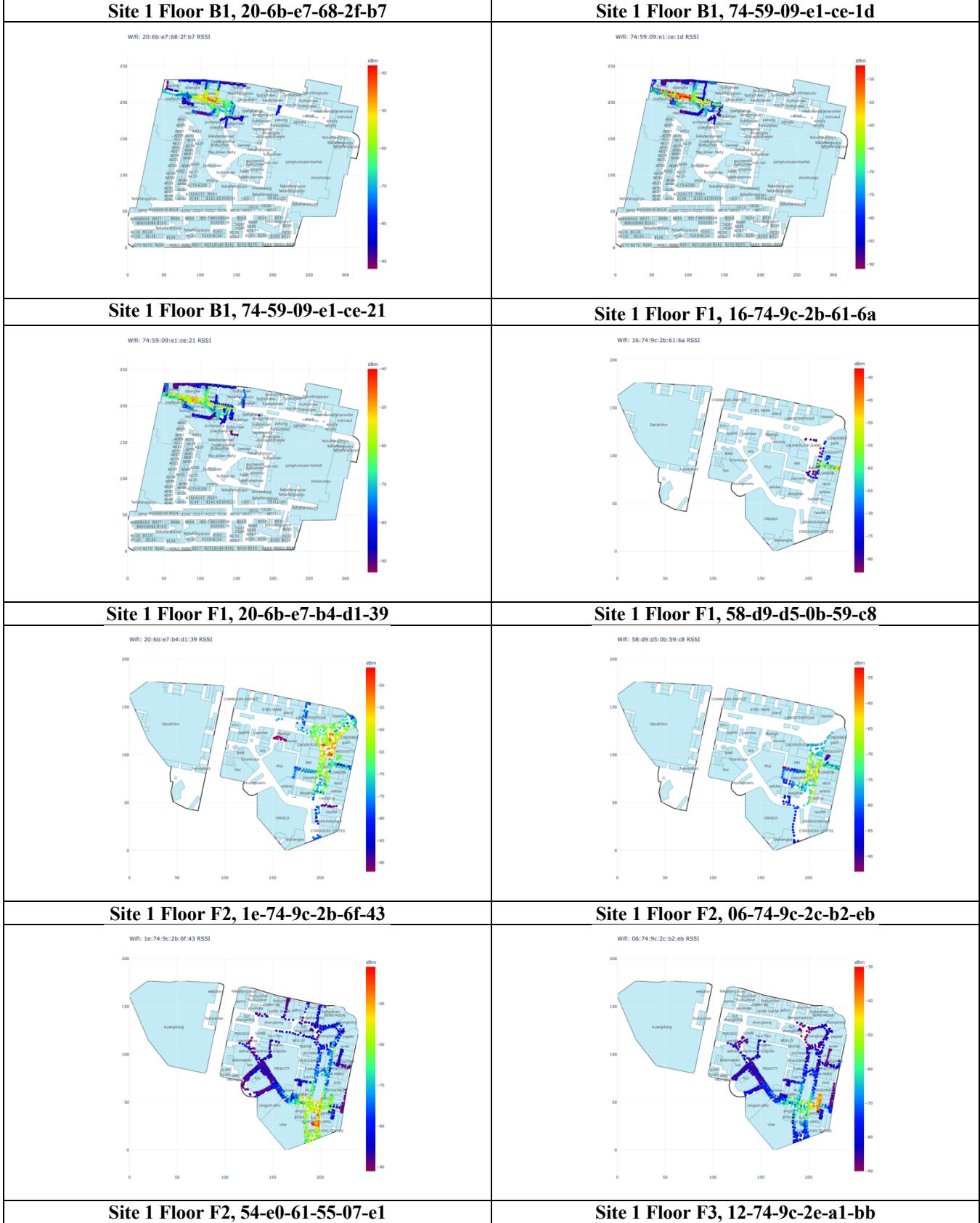


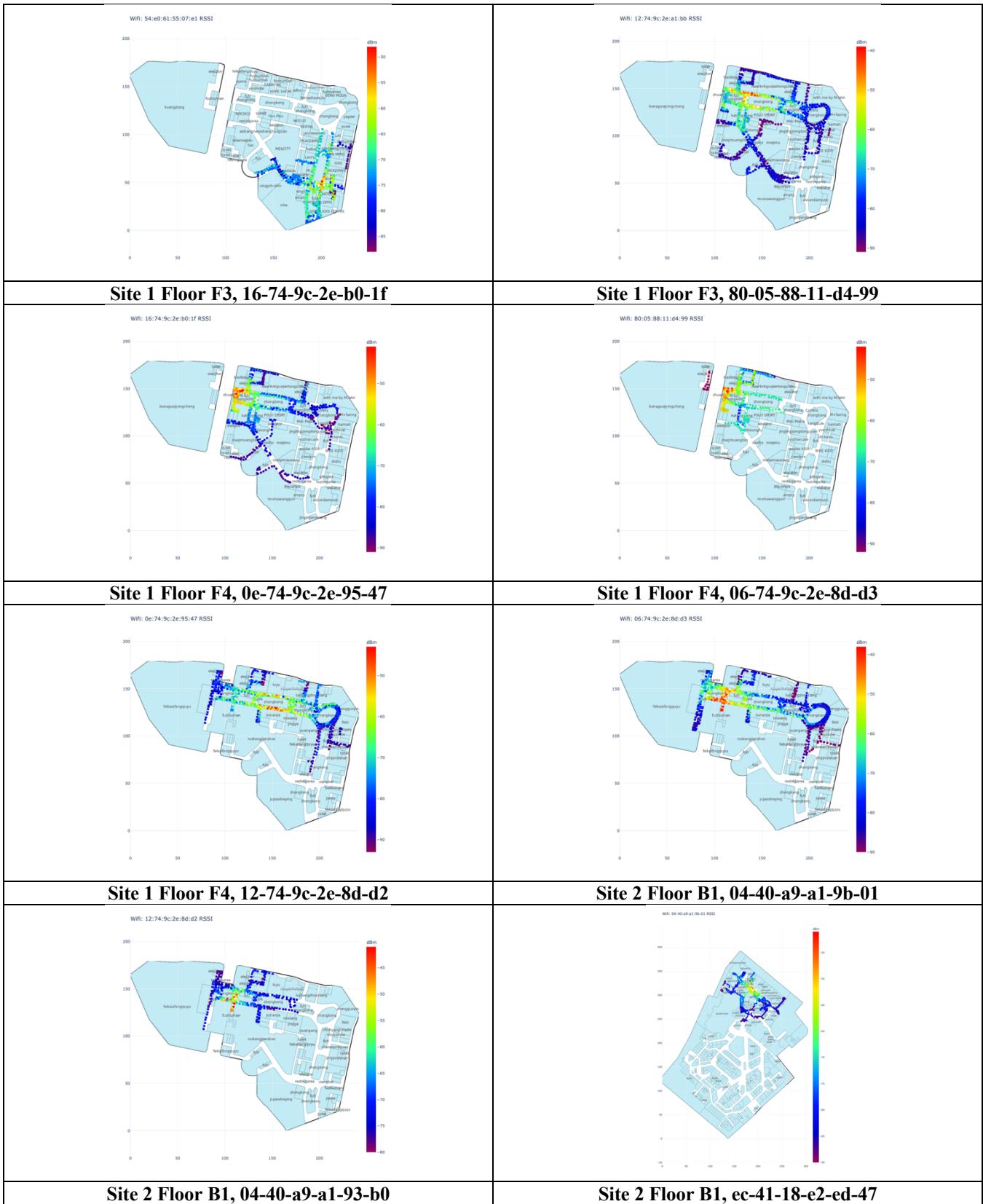
B. The Magnetic Strength Heatmap for each floor in each site

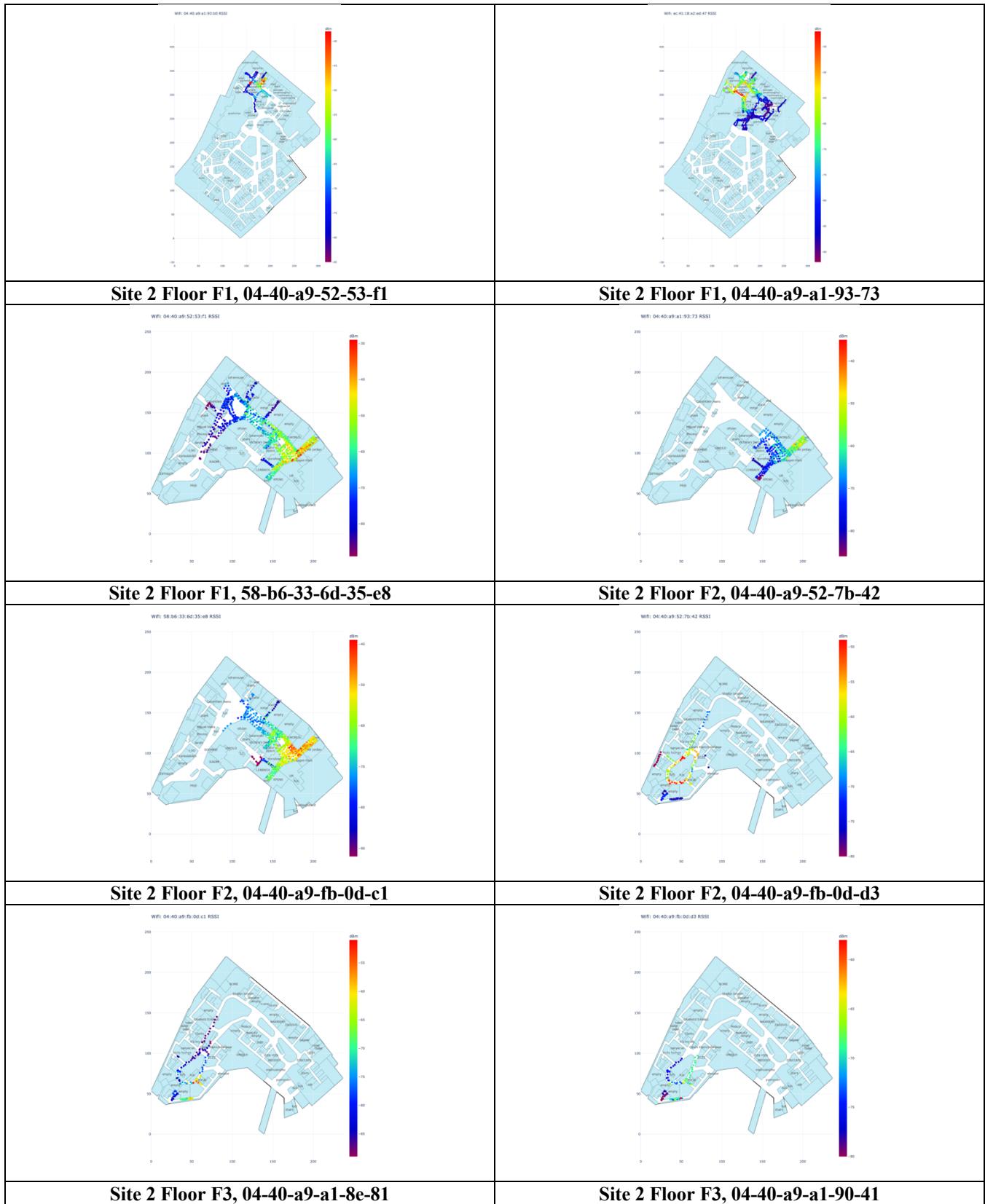


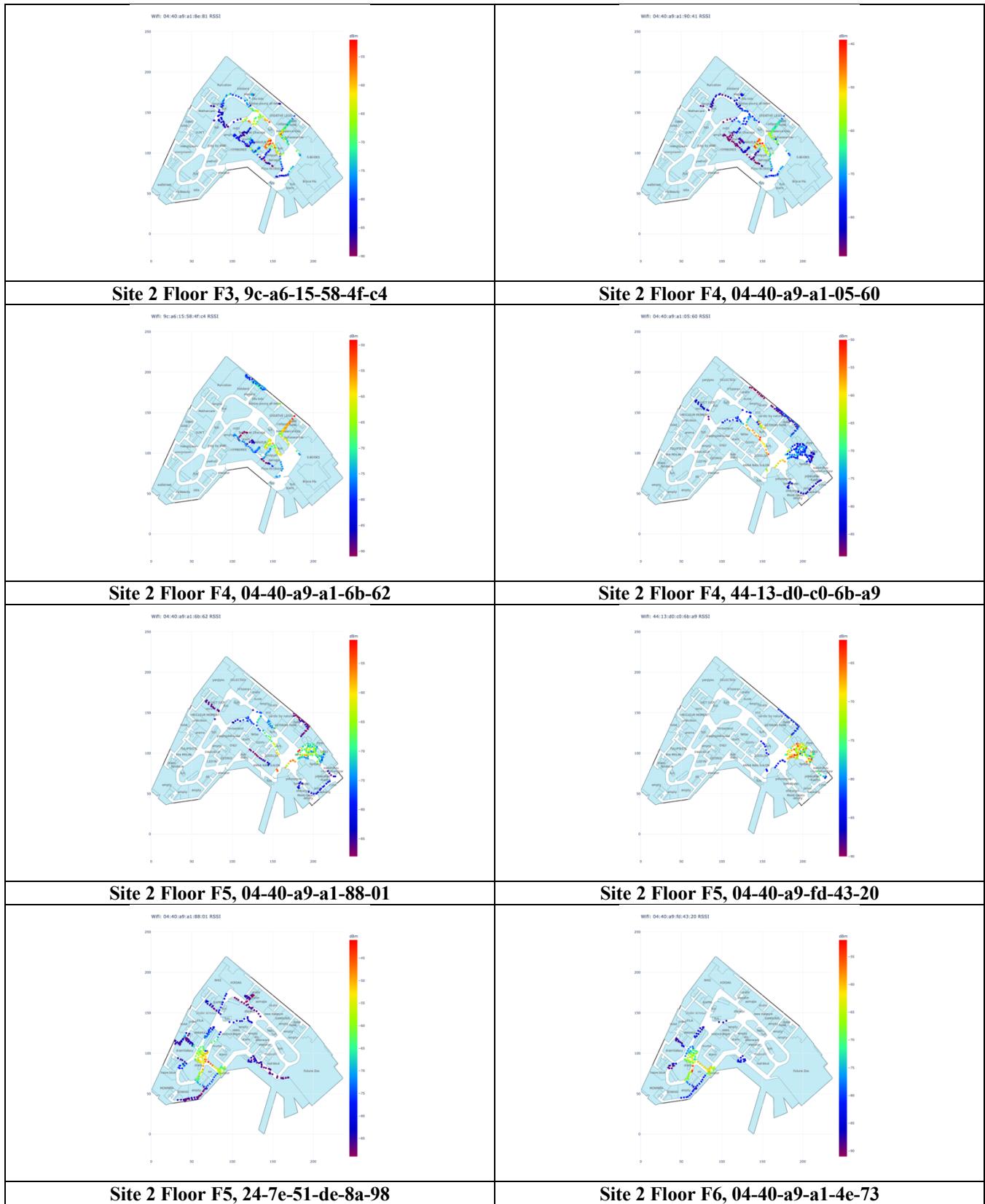


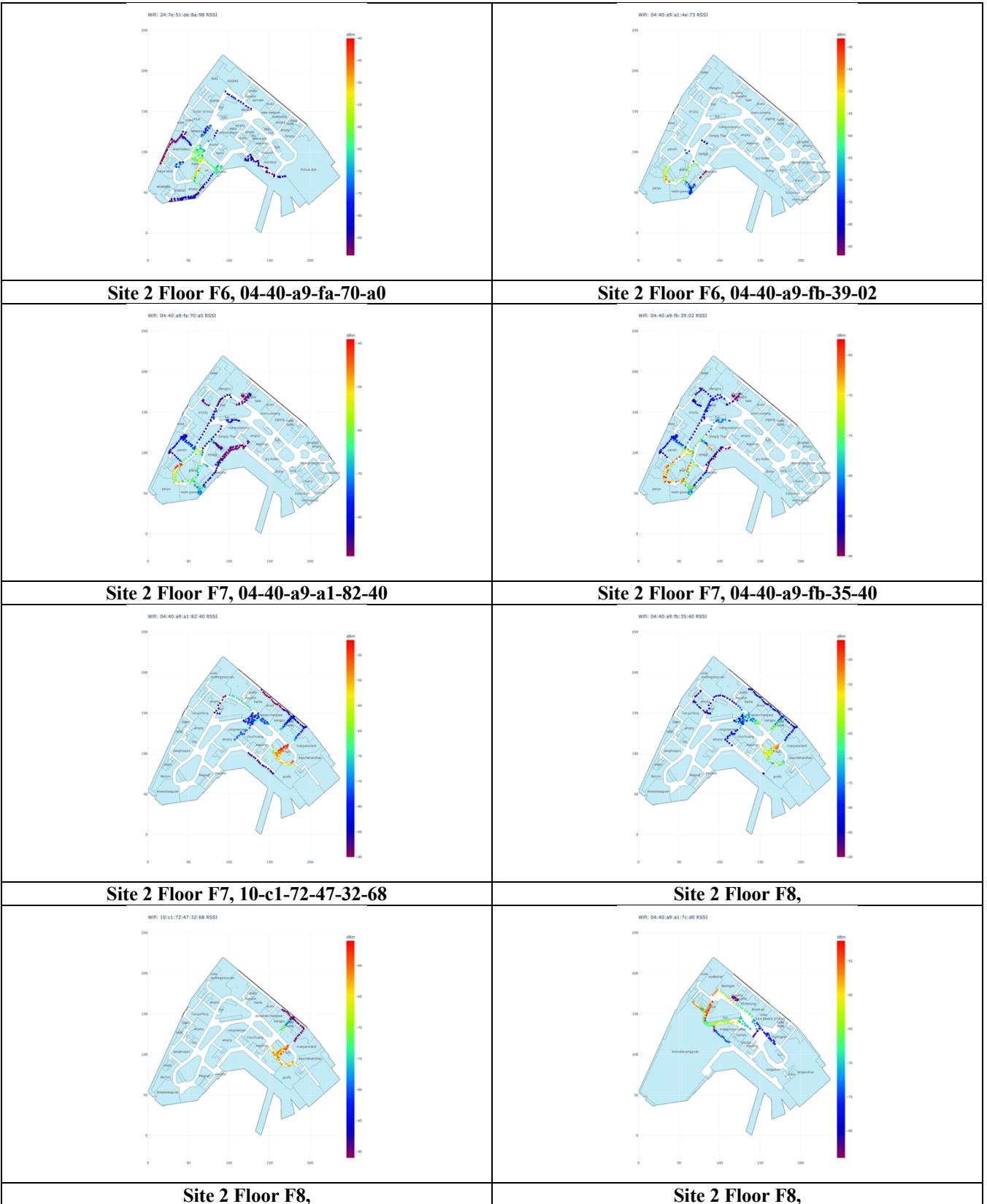
C. The Wi-Fi RSSI Heatmap for 3 APs in each floor and each site

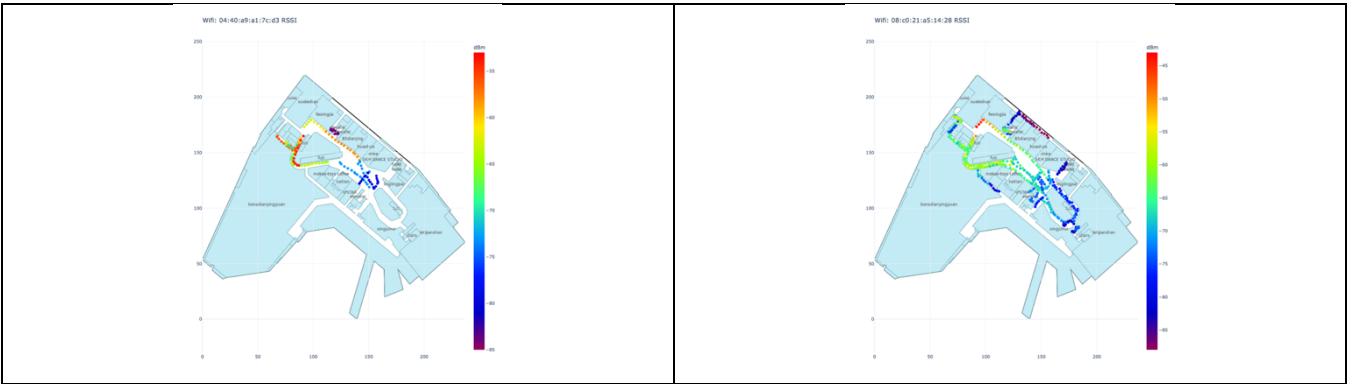




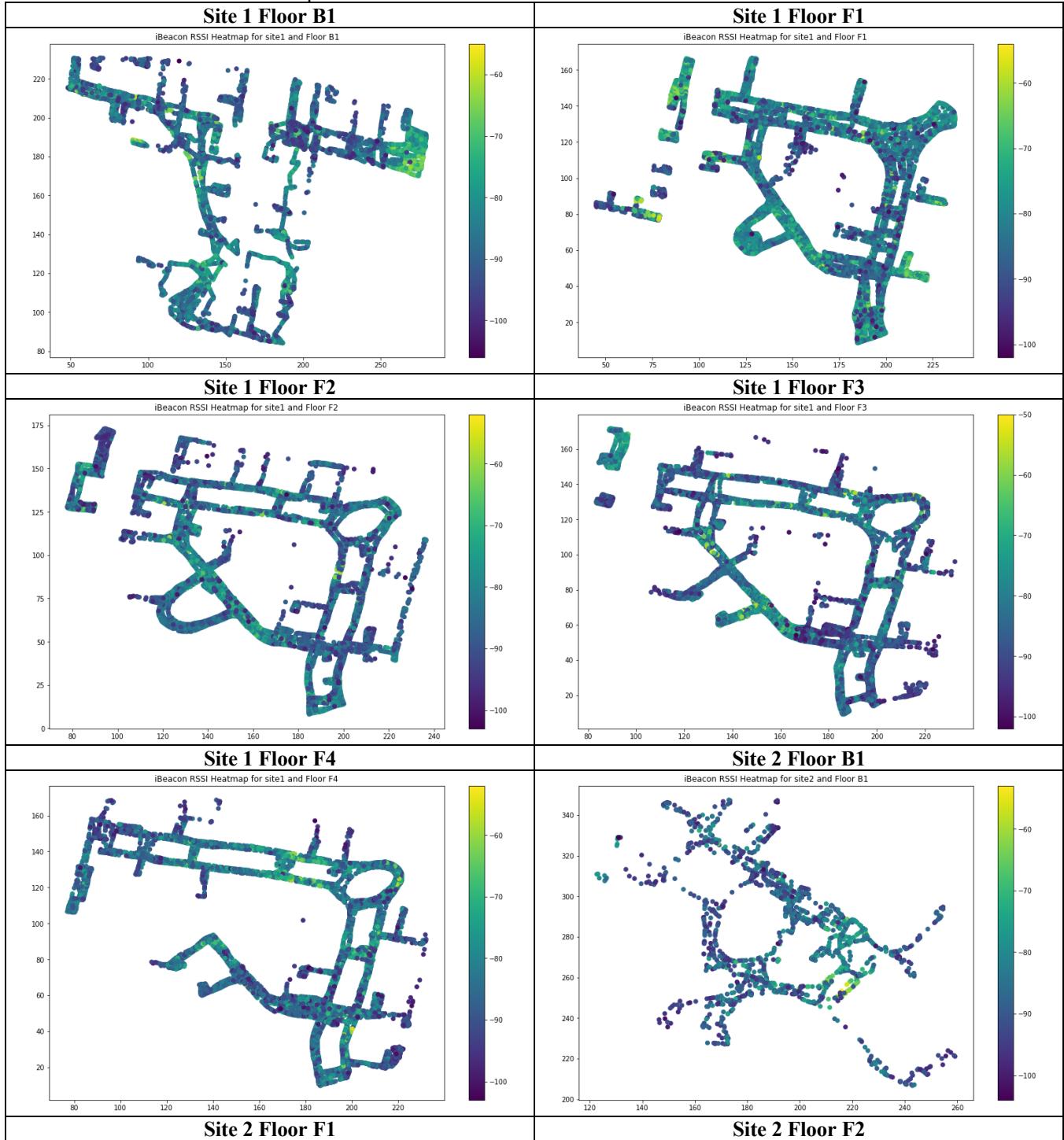


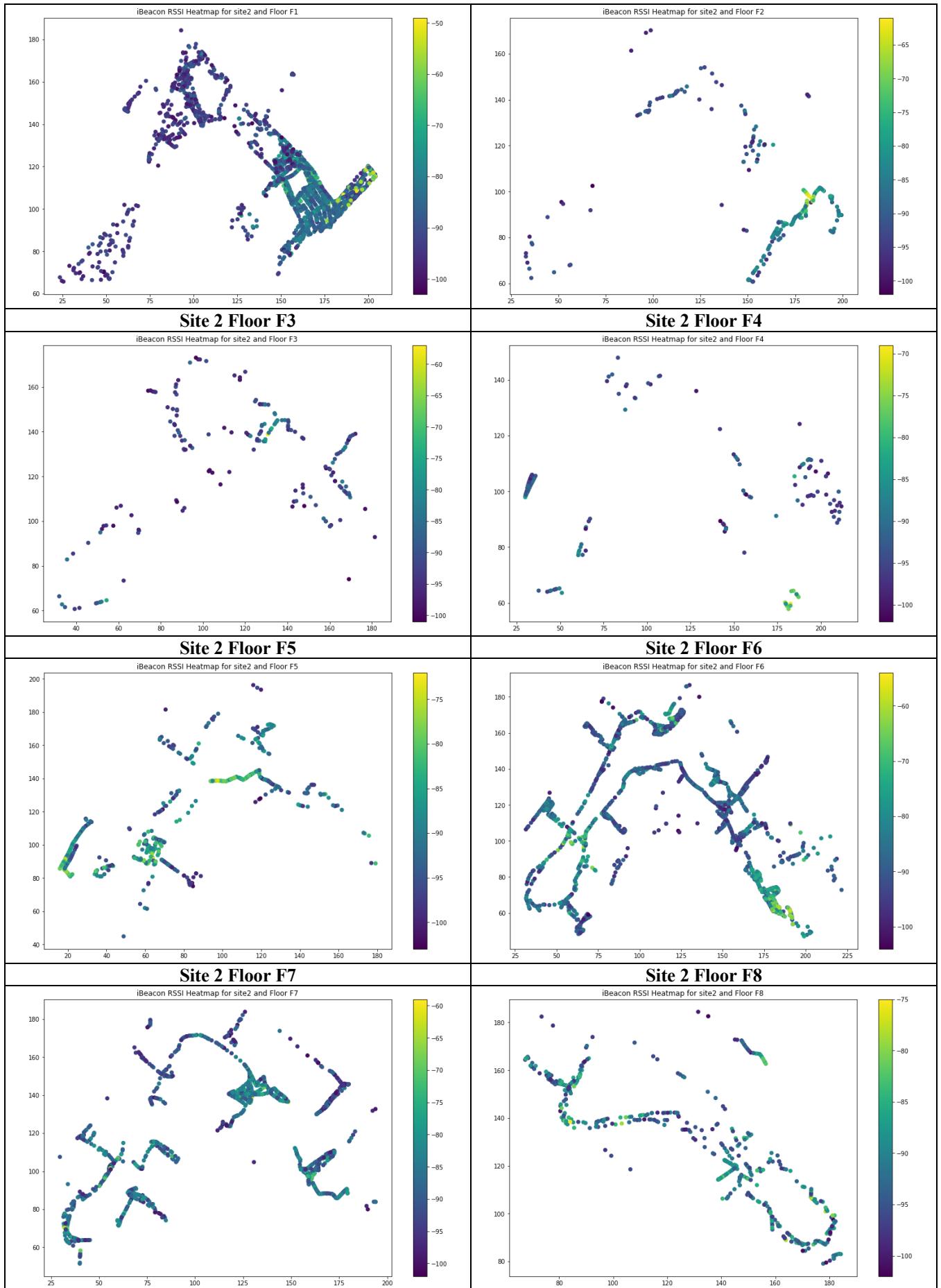






D. The iBeacon RSSI Heatmap for each floor in each site





E. Link to Google Colab for Source Code

<https://colab.research.google.com/drive/1bTll77vY9HzTA123U5MVyqaHV4P0LuW6?usp=sharing>

F. HTML files for each floor

Google Drive Link to all the html files for both sites and floors (download or open in Chrome/Safari/MicroSoft Edge):

[https://drive.google.com/drive/folders/1o8sdAQ8My6Z7zfM3JFG\\_JXkjBqVviE1Y](https://drive.google.com/drive/folders/1o8sdAQ8My6Z7zfM3JFG_JXkjBqVviE1Y)