# **Limited amount of usable data pairs**

From observations subscriber Id and location is not updated at regular intervals due to the nature of smartphone and end user devices. At certain timing where the next location report / request is made phones may had lost their signals, unable to respond timely, not turned on.

Within the given data set there are occurrence of one S\_ID reporting twice within 7 seconds, then the next report is at a random number of minutes or perhaps seconds (Reporting not done in fixed intervals. May be very hard to obtain mean intervals.)

The above will contribute to the following.

Between two selected timings (eg 0900 to 1000) there may contain only a small percentage of union data because the rest of the user is either not online, or did not report, at given time (accuracy to seconds).

Example 1:

Subscriber A reported at 0901 hrs

Selecting 0900 will miss out Subscriber A

Example 2:

Subscriber A reported at 0900 hrs

Reported again at different location at 1001 hrs

No movement will be recorded as Subscriber A did not qualify for point B 1000hrs.

Valid Example 3:

Subscriber A reported at 0900hrs

Reported again at different location at 1000hrs

Both time will need to have their accuracy to the seconds.

## **Mitigating solution:**

We can expand the accuracy to 5 minutes to have as much chance of recapturing Subscriber A again at point B.

However:

· Duplicate entries of Subscriber A (1000, 1001, 1004hrs) can creep into point B.

o We can use SELECT DISTINCT to cope.

· Shortest time interval user able to select to view movement will be shortest 6 minutes, depending on accuracy span adopted.

# **Data cleansing**

Using geopy certain locations are not returning lat lon pairs. We need to go through each of these location and manually key in their lat lon in order for a complete set of translation database.

The rows (locations) that need to be cleansed, and their respective coordinates are in ‘data cleansing.sql’

# **Timezone of unix time**

UNIX time when converted to readable time is in timezone of UTC. Singapore is UTC + 8.

To convert unix time to local time add 8 hours (in seconds) to UNIX time.

To convert local time to UNIX time minus 8 hours (in seconds).

Time period in master data starts from 2015-08-20 00:00:00 UTC to 2015-08-21 00:00:00 UTC.