

Collecting mobility data

Data collection and quality

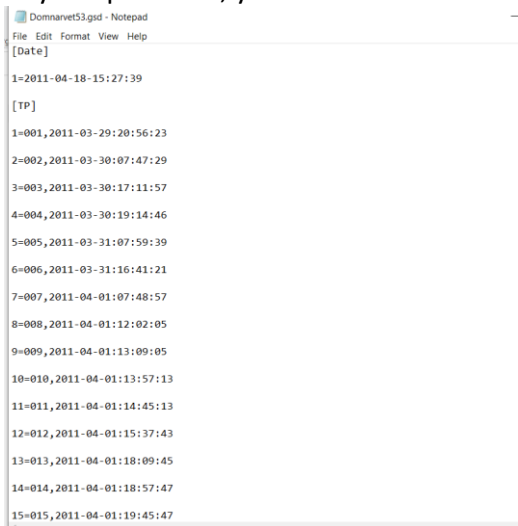
Description:

The purpose of this assignment is to perform the quality and processing of mobility data. Mobility data is one of the data collection methods used to find human mobility patterns, urban mobility models, transportations studies, etc. In this lab, you will learn to process the raw GPS data. The quality of data depends on the source of data, which needs to be evaluated. You will learn to clean and process GPS data. Task 1 is changing the format file to the proper version, while task 2 aims to clean and process the data.

Task1:

1. Download the dataset from learn.
2. Import the dataset in Excel (Import as a text).

When you open a file, you have the:



```

Dommarvet53.gsd - Notepad
File Edit Format View Help
[Date]
1=2011-04-18-15:27:39
[TP]
1=001,2011-03-29:20:56:23
2=002,2011-03-30:07:47:29
3=003,2011-03-30:17:11:57
4=004,2011-03-30:19:14:46
5=005,2011-03-31:07:59:39
6=006,2011-03-31:16:41:21
7=007,2011-04-01:07:48:57
8=008,2011-04-01:12:02:05
9=009,2011-04-01:13:09:05
10=010,2011-04-01:13:57:13
11=011,2011-04-01:14:45:13
12=012,2011-04-01:15:37:43
13=013,2011-04-01:18:09:45
14=014,2011-04-01:18:57:47
15=015,2011-04-01:19:45:47
  
```

[Date] (This is the Date when the data was transferred from the GPS logger to the computer.)

[TP] (This means how many tips are done during the period, but this information is not very useful, because it just had the Date and the starting time)

[001, 2011-04-15:07:51:03]

Below this line is the information about the recording points.

Domnarvet53.gsd - Notepad

File Edit Format View Help

```

57=057,2011-04-13:07:43:03

[001,2011-03-29:20:56:23]

1=60302958,15278668,185623,290311,5050,-1
2=60301294,15278364,185654,290311,90,-1
3=60299943,15278751,185724,290311,4560,-1
4=60298907,15277971,185754,290311,4510,-1
5=60299291,15274659,185824,290311,1860,-1
6=60298225,15276086,185854,290311,4420,-1
7=60295974,15277885,185924,290311,5230,-1
8=60293903,15276339,185954,290311,4890,-1
9=60291817,15277937,190024,290311,5560,-1
10=60289783,15280275,190054,290311,5460,-1

```

For example: 1=60267285,15286898,55103,150411,100,-1

From left to right after the = sign, it is: Latitude, Longitude, time, Date, speed, and altitude

However, they are not in the right format.

The format of the original recording for the coordinates is not in decimal degree, but in decimal degree minutes.

For example, in the original .gsd file, the original one is 60267285, which the true format is 60 (degree)26.7285(decimal minutes)

If we want to change it to decimal degree, we need to divide the decimal minutes part by 60 to get the decimal degree part

Therefore it should be:

$60 + (26.7285/60)$ Which means, to get the **WGS84 decimal degree**, you need to conduct calculation:

60267285 -> 60+ (26.7285/60)

15286898 -> 15+ (28.6898/60)

55103 -> 05:51:03

150411 -> 15-04-2011

100 -> $100/100 = 1$ km/h (this speed column should all divide 100)

-1 -> no recording for altitude, note by -1

You can decide the format for time and Date, but Y, X and Speed, you should follow exactly the calculation. Each recording should have its own id. Find your own way to preprocess the data and save the data in a CSV or XLSX file.

3. Create a Trip ID and Point ID based on the data.
4. Calculate the speed based on km/h.
5. Change the Time and Date in a proper version.
6. Calculate the distance between two points.
7. Calculate the time difference between two points.
8. Calculate the average speed on the trip.
9. Calculate the acceleration at each point.
10. Calculate average acceleration on the trip.
11. Calculate the trip duration and distance.

The output for task 1 needs to be like this:

A	B	C	D	E	F	G	H	I	J	K	L
TP_ID	TRAIL_ID	USER_ID	Y_COORDINA	X_COORDINA	TIME	DATE	SPEED	HEIGHT	SPEED(KM/H)	Y_WGS84	X_WGS84
1	1	Domnarvet53	60302958	15278668	18-56-23	29-03-11	5050	-1	50.5	60.50493	15.46444667
1	1	Domnarvet53	60301294	15278364	18-56-54	29-03-11	90	-1	0.9	60.502157	15.46394
1	1	Domnarvet53	60299943	15278751	18-57-24	29-03-11	4560	-1	45.6	60.499905	15.464585
1	1	Domnarvet53	60298907	15277971	18-57-54	29-03-11	4510	-1	45.1	60.498178	15.463285
1	1	Domnarvet53	60299291	15274659	18-58-24	29-03-11	1860	-1	18.6	60.498818	15.457765
1	1	Domnarvet53	60298225	15276086	18-58-54	29-03-11	4420	-1	44.2	60.497042	15.46014333
1	1	Domnarvet53	60295974	15277885	18-59-24	29-03-11	5230	-1	52.3	60.49329	15.46314167
1	1	Domnarvet53	60293903	15279339	18-59-54	29-03-11	4890	-1	48.9	60.489838	15.465565
1	1	Domnarvet53	60291817	15277937	19-00-24	29-03-11	5560	-1	55.6	60.486362	15.46322833
1	1	Domnarvet53	60289783	15280275	19-00-54	29-03-11	5460	-1	54.6	60.482972	15.467125
2	1	Domnarvet53	60289622	15286066	19-01-22	29-03-11	6900	-1	69	60.482703	15.47677667
3	1	Domnarvet53	60288635	15291960	19-01-52	29-03-11	6650	-1	66.5	60.481058	15.4866
4	1	Domnarvet53	60287423	15296882	19-02-22	29-03-11	2930	-1	29.3	60.479038	15.49480333
5	1	Domnarvet53	60286067	15300435	19-02-52	29-03-11	6680	-1	66.8	60.476778	15.500725
5	1	Domnarvet53	60283440	15303862	19-03-22	29-03-11	7150	-1	71.5	60.4724	15.50643667
7	1	Domnarvet53	60281111	15307665	19-03-52	29-03-11	6840	-1	68.4	60.468518	15.512775
3	1	Domnarvet53	60278621	15311063	19-04-22	29-03-11	6300	-1	63	60.464368	15.51843833
3	1	Domnarvet53	60277601	15316334	19-04-52	29-03-11	5900	-1	59	60.462668	15.52722333
1	1	Domnarvet53	60277321	15321720	19-05-22	29-03-11	6590	-1	65.9	60.462202	15.5362
1	1	Domnarvet53	60277321	15321720	19-05-22	29-03-11	6590	-1	65.9	60.462202	15.5362

Task2:

Visualization the data

- 1- Visual the data in python, R or any relevant software (longitude vs latitude).
- 2- Generate a heatmap to show the density of GPS data points in a specific area.