Practical Exercise 1

Comparison of different datasets

Sample Data

http://bit.ly/mpd-kigali

- Location events of one person
- Domestic & Outbound roaming data combined
- margus_cdr_1week.csv
  - 2014-09-16 - 2014-09-20
- margus_signalling_1week.csv
  - 2014-09-16 - 2014-09-20
- margus_cdr_3yr.csv (for your interest, 3 years of his life)
  - 2014-09-02 - 2017-08-31

margus_cdr_1week.csv

pos_usr_id	pos_time	country	cell_id	event_type	lat	lon	lau1_code	lau1_name	lau2_code	lau2_name
5433301	16.09.2014 8:48	EE	2145	6	58.3652694	26.7194472	78	Tartu maakond	795	Tartu linn
5433301	16.09.2014 8:52	EE	4274	6	58.3705527	26.7052666	78	Tartu maakond	795	Tartu linn
5433301	16.09.2014 10:26	EE	4274	5	58.3705527	26.7052666	78	Tartu maakond	795	Tartu linn
5433301	16.09.2014 14:27	EE	6695	1	58.3779694	26.7050166	78	Tartu maakond	795	Tartu linn
5433301	16.09.2014 14:33	EE	6695	1	58.3779694	26.7050166	78	Tartu maakond	795	Tartu linn
5433301	16.09.2014 16:45	EE	6695	1	58.3779694	26.7050166	78	Tartu maakond	795	Tartu linn
5433301	16.09.2014 20:49	EE	6695	5	58.3779694	26.7050166	78	Tartu maakond	795	Tartu linn
5433301	16.09.2014 21:12	EE	6695	1	58.3779694	26.7050166	78	Tartu maakond	795	Tartu linn
5433301	17.09.2014 9:55	EE	4521	1	59.4038833	24.8047138	37	Harju maakond	653	Rae vald
5433301	17.09.2014 10:10	EE	4778	2	59.43993573	24.73716217	37	Harju maakond	784	Tallinn
5433301	17.09.2014 10:45	EE	4604	1	59.4302777	24.7436083	37	Harju maakond	784	Tallinn
5433301	17.09.2014 12:45	EE	4604	1	59.4302777	24.7436083	37	Harju maakond	784	Tallinn
5433301	17.09.2014 12:52	EE	3692	1	59.4358138	24.7372833	37	Harju maakond	784	Tallinn
5433301	17.09.2014 13:10	EE	4704	2	59.4287805	24.7670472	37	Harju maakond	784	Tallinn
5433301	17.09.2014 17:53	LV		5						
5433301	17.09.2014 19:28	LV		2						
5433301	18.09.2014 10:24	LU		5						
5433301	18.09.2014 10:34	LU		2						
5433301	18.09.2014 20:54	LU		1						
5433301	18.09.2014 22:00	LU		5						
5433301	18.09.2014 22:09	LU		2						
5433301	18.09.2014 22:12	LU		1						
5433301	19.09.2014 0:26	LU		1						
5433301	19.09.2014 0:26	LU		2						
5433301	19.09.2014 0:27	LU		2						
5433301	19.09.2014 0:29	LU		2						
5433301	19.09.2014 7:57	LU		5						

57 location events over 5 days

CDR Data: margus_cdr_1week.csv

- Id – event ID
- pos_usr_id – subscriber ID (5433301=Margus)
- pos_time – time of the event
- country – country where the event happens
- cell_id – antenna ID
- event_type – event type (1- Mobile-Originating Call start; 2- Mobile-Terminating Call start; 5- Mobile-Originating SMS; 6- Mobile-Terminating SMS)
- Lat – latitude coordinate (WGS84)
- Lon – longitude coordinate (WGS84)
- lau1_code – Estonian county code where the event took place
- lau1_name – Estonian county name where the event took place
- lau2_code – Estonian municipality code where the event took place
- lau2_name – Estonian municipality name where the event took place

margus_cdr_1week.csv

- 57 location events over 5 days

margus_signalling_1week.csv

- 400 location events over 5 days

margus_cdr_3yr.csv

- 4868 location events over 3 years

Exercise 1.1

- Try to calculate the size of the data (file size / estimated number of records) for your country based on the example data from Margus
  - Per month
  - Inbound roaming data
    - CDR
    - Signalling

Exercise 1.1 – Answer from The Gambia

- Try to calculate the estimated size of data for The Gambia based on the example data from Margus
  - Per month
  - Inbound roaming data
    - CDR
    - Signalling
  - Assumptions
    - Population using mobile phones – 1 700 000
    - Subscribers of one MNO – 1 500 000
    - One week of data – 4 kb
    - 52 weeks
  - Formula = # subscribers of one MNO * one week of data * 52 weeks
    - CDR 24 GB → Fits on a laptop
    - Signalling 1.08 TB

Exercise 1.2

- What can you tell about Margus based on his data?
- What can you tell from signaling data that you cannot from CDR?

Exercise 1.2 – Answers from the group

- What can you tell about Margus based on his data?
  - He visited 4 countries in 1 week
  - His CDR is showing 4 different event types
  - Margus is calling more than using SMS
  - He spends a lot of time in Tartu
  - During 1 week he used 8/11 cell id
  - He uses the mobile more in the afternoon than in the morning
  - He did more calls on Friday, September 19 than any other day

Exercise 1.2 – Feedback on data exploration task

- What can you tell about Margus based on his data?
  - He visited 4 countries in 1 week
  - His CDR is showing 4 different event types
  - Margus is calling more than using SMS
  - He spends a lot of time in Tartu
  - During 1 week he used 8/11 cell id
  - He uses the mobile more in the afternoon than in the morning
  - He did more calls on Friday, September 19 than any other day