

Project2023

Team:

- *) Ellar Seidelberg
- *) Free place
- *) Free place

Project link:

(<https://github.com/EllarSeidelberg/Project2023>)

Raport:

1. Eellugu

Kasutame <https://avaandmed.eesti.ee/> andmebaasi. Tunduvad head andmebaasid selles mõttes, et on CSV formaadis ja kergesti kättesaadavad – seega pole andmete kogumisega probleeme.

Tallinna linna tänavanimedes tähtede kasutussagedus. Ja kas selle põhjal oleks võimalik ennustada, et kui antakse nimi mõnele uuele tänavale, siis mis tähti see võiks/peaks sisaldama.

2. Äriline arusaam

- Eks alati on küsimus, milline on üks hea ja ilus ning meeldejääv tänavanimi. Eesmärk oleks leida, kas on mingid tähed, mis peaksid alati tänavanimesse kuuluma.
- Kasutame <https://avaandmed.eesti.ee/datasets/tallinna-aadresside-tabelid-asumite-kaupa> andmebaasi ja tundub, et peame natukene andmeid korrastama, enne kui uurida saame.
- Positiivne tulem oleks see, kui suudame mõne uue tänavanime puhul hinnata, kas seal on niinimetatud vajalikud tähed olemas.
- Negatiivne tulem võiks olla see, kui just enim esinevaid tähti tulevikus tänavanimedes vältida (kuna neid on näiteks seni liiga palju juba kasutatud).
- Õnnestumise tõenäosus oleks see, kui umbes 20-30% tähestiku tähtedest tõuseb esile.
- Ebaõnn oleks see, kui kõiki tähti oleks kasutatud sama sagedalt.

3. Andmete mõtestamine

Tundub, et andmekogu on meil olemas.

<https://avaandmed.eesti.ee/datasets/tallinna-aadresside-tabelid-asumite-kaupa>

Seega tänavaküsitlust ei pea tegema, ka ise ei pea tänavanimesi üles kirjutamas käima ega ka mitte google mapi või Maa-ameti kaardirakendust kasutama.

Peame tänavanimed sorteerima. Meil on sisuliselt iga maja number, seega tuleb jätta ainult tänav ja korduvad tänavanimed välja visata.

Oht on, et valim võib väikseks jääda peale korduste eemaldamist. Samas kordusi eemaldamata ka ei saa, sest meid huvitavad unikaalsed tänavanimed. Muidugi, võib juhtuda, et erinevates linnaosades kordub mõni tänavanimi, siis tuleks kordus alles jätta.

Siis on kindlasti küsimus, kas uurida sõnu tervikuna või jagada need eraldi tähtedeks ja siis uurida. Eks see selgub töö käigus ja otsustame jooksvalt.

4. Projekti planeerimine

- *) Tänavanimede korduste välja viskamine. (2h)
- *) Andmed tulpadesse ja otsustada, mis jääb ja mis välja visata. (4h)
- *) Analüüs 1 (4h)
- *) Analüüs 2 (4h)
- *) Kas annab midagi ennustada nii positiivse kui negatiivses võtmes. (2h)
- *) Graafiline esitlus (4h)