

Koolitus „Andmeteadus on Popp“

Excel, Python või R

November 2021

Kristjan Eljand



Euroopa Liit
Euroopa Sotsiaalfond



Eesti
tuleviku heaks

Excel, Python või R?



Euroopa Liit
Euroopa Sotsiaalfond



Eesti
tuleviku heaks

Excel, Python või R



- **Excel** – visuaalne, lihtne ja ülimalt paindlik.
- **R ja RStudio** – lihtsaim sissejuhatus andmeteadusesse Exceli kogemusega analüütikule.
- **Python** – võimas ning tipp-andmeteaduse tiimide peamine töövahend.

Kuna kasutada Excelit



Excel sobib:

- Kui andmeid on vähe ning sul on vaja sama analüüsi teha vaid 1-2 korda.
- Andmestiku mõistmiseks – Excel on väga „visuaalne“ – kõik andmed on sulle kohe näha.
- Märksõnad: andmete uurimine, prototüüpimine, testimine.

Excel ei ole hea lahendus:

- Kui andmeid on palju (enam kui miljon rida) või kui sama analüüsi on vaja teha pidevalt.

Näide Exceli kasutamisest – tutvume andmestikuga



- Kasutame Eesti avaandmete portaalist kättesaadavat ettevõtete maksulaekumiste ja käibe statistikat.
 - Leiame „INFO JA SIDE“ valdkonna kogukäibe, ettevõtete arvu ja keskmise käibe ettevõtte kohta.

Kuna kasutada R-i



R sobib:

- Kui andmeid on palju (üle miljoni rea) või sama analüüsi on vaja teha palju kordi.
- Lihtsate veebirakenduste ja automaatraportite loomiseks.
- R on lihtsaim sissejuhatus andmeteadusesse Exceli kogemusega analüütikule.

R vs Python

- R on vähem „programmeerimine“ ja rohkem analüütika.
- R'i analüütikateegid on üldiselt paremini dokumenteeritud.
- Python on aga keel, millega tehakse kõige uuemat ja ägedamat uurimistööd.

Näited R-i kasutamisest – veebirakendus ja automaatraportid



- Vaatame veebirakendust „konkurentsikaart“.
- Vaatame, milliseid automaatraporteid saab R-ga luua.

Google Colaboratory



Euroopa Liit
Euroopa Sotsiaalfond



Eesti
tuleviku heaks

Mis on Google Colaboratory



- Colaboratory või lühendatult "Colab" võimaldab programmeerida ja jooksutada seda programmi otse internetibrauseris.
- Colab on ehitatud "märkmiku" stiilis, mis tähendab, et saad kirjutada nii teksti kui ka koodi.
- [Vaatame lähemalt](#)

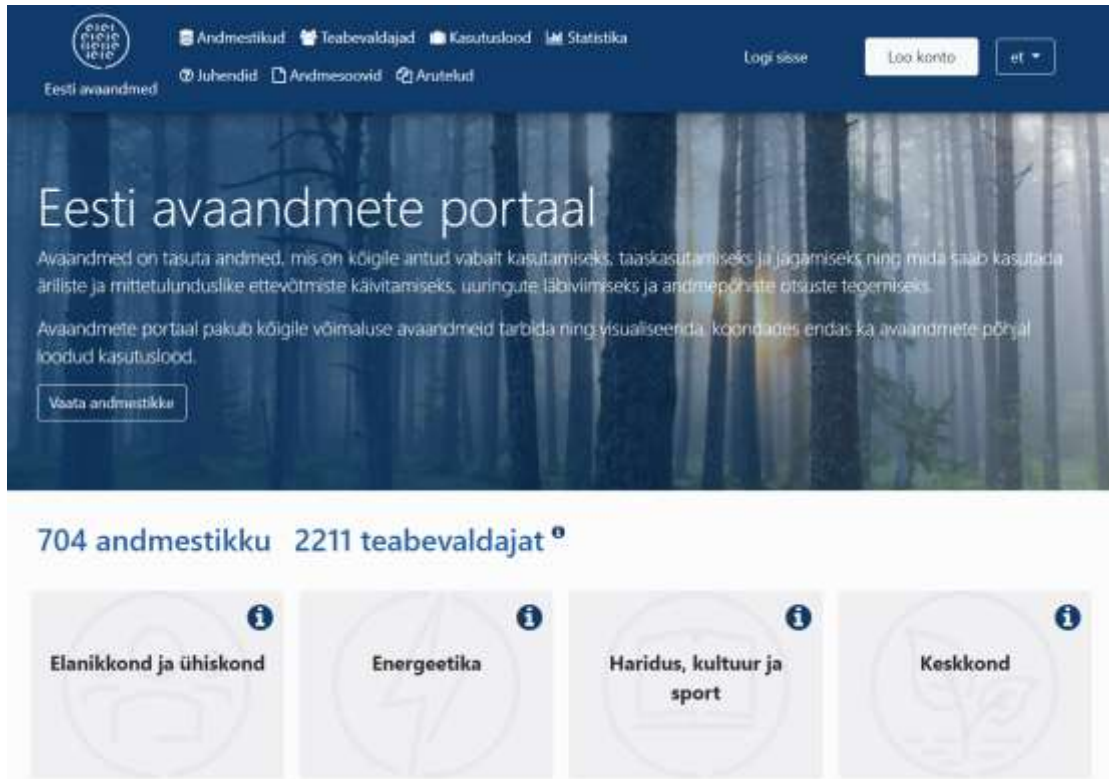
Ülesanne: Google Colaboratory



- Ava koolitusmaterjalide kaust:
https://drive.google.com/drive/folders/11WAFPW7tcO97CgteQOx3KjpdBs_wT01c3?usp=sharing
- Tee koopia failist „Ülesanne_Sissejuhatus Google Colab’i”
- Lahenda toodud ülesanded

Andmete töötlemine Pythoniga

Eesti avaandmete portaal



Allikas: avaandmed.eesti.ee

- Avaandmed on tasuta andmed, mis on kõigile antud vabalt kasutamiseks.
- Avaandmete portaal pakub kõigile võimaluse avaandmeid tarbida ning visualiseerida, koondades endas ka avaandmete põhjal loodud kasutuslood.

Avaandmete portaal – vaatame ringi



Allikas: avaandmed.eesti.ee

- Andmestikke saab [otsida](#) nii valdkondade, aasta, piirkonna kui ka otsingusõnade järgi.
- Andmete lisamise ja allalaadimiste kohta saab näha [statistikat](#).
- Avaldatud on [kasutuslood](#) ja näidisrakendused.
- [Juhendid](#) (N: API kasutamise kohta)

Ülesanne: Andmete töötlemine Pythoniga



- Ava koolitusmaterjalide kaust:
https://drive.google.com/drive/folders/11WAFPW7tcO97CgteQOx3KjpdBs_wT01c3?usp=sharing
- Tee koopia failist „Ülesanne_Andmete töötlemine Pythoniga“
- Lahenda toodud ülesanded

Koolitus „Andmeteadus on Popp“

Excel, Python või R

November 2021

Kristjan Eljand



Euroopa Liit
Euroopa Sotsiaalfond



Eesti
tuleviku heaks