

Duomenų analitikų bandomosios užduotys

Valstybės duomenų agentūra

2023-09-27

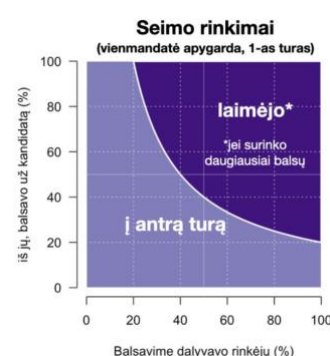
Būsimieji kolegos. Jūsų pasirinktų užduočių atsakymuose labai tikimės pamatyti ne vien rezultatus ir išvadas, bet ir jūsų komentarais papildytą kompiuterinį kodą, sugeneravusį tuos rezultatus. Pasirinkite jums tinkamą užduočių sudėtingumo lygį ir pamėginkite jas atlikti kuo geriau. Galite išmėginti savo jėgas ir kelių lygių užduotyse. Linkime kūrybiškumo, drąsos ir sėkmės!

Junior level

Naudodamiesi R arba Python programavimo kalbomis bei aktualiais Lietuvos Respublikos įstatymais septyniose iliustracijose pavaizduokite sąlygas, kurios yra būtinos, kad:

1. kandidatas į Lietuvos Respublikos Prezidentus būtų išrinktas pirmajame rinkimų ture;
2. kandidatas į mero postą būtų išrinktas pirmajame rinkimų ture;
3. kandidatas į Seimo nario postą būtų išrinktas vienmandatėje apygardoje pirmajame rinkimų ture;
4. konsultaciniame referendume būtų sėkmingai pritarta referendumui pateikto sprendimo priėmimui;
5. privalomajame referendume būtų sėkmingai pritarta referendumui pateikto sprendimo priėmimui;
6. privalomajame referendume būtų sėkmingai pritarta dėl Konstitucijos XIV skirsnio nuostatų pakeitimo;
7. privalomajame referendume būtų sėkmingai pritarta dėl Konstitucijos 1 straipsnio nuostatos pakeitimo.

pavyzdys:



Mid level A

Naudodamiesi R arba Python programavimo kalbomis bei atvirais pandemijos rodiklių duomenimis:

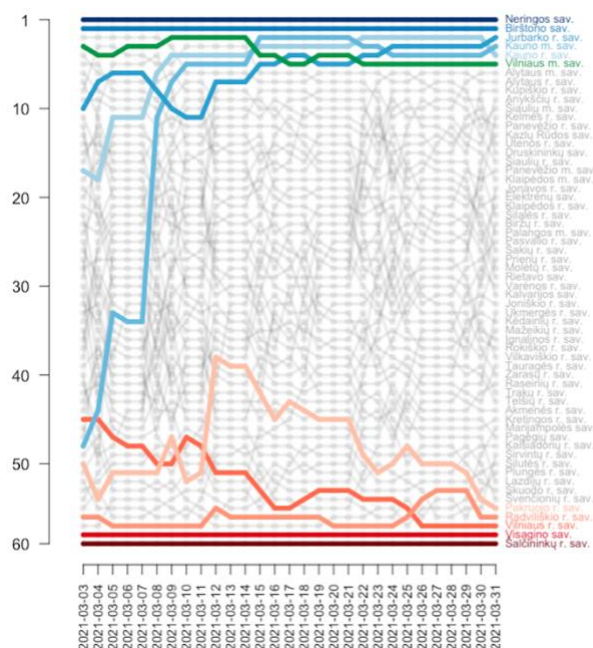
<https://osp.stat.gov.lt/covid-dashboards> → „Atviri duomenys“ → „Agreguoti COVID19 atvejai ir mirtys“ ir oficialiosios statistikos demografijos duomenimis:
<https://osp.stat.gov.lt/> → „Visi rodikliai“ → „Rodiklių duomenų bazė“ → ...

vizualizuokite Lietuvos savivaldybes pagal jų **antrinio sergamumo** (*prevalence*) rangą padieniu didžiausio sergamumo 30 dienų laikotarpiu. Pateikiame panašų pavyzdį iš kitur (dešinėje).

- Pirminis sergamumas (*incidence*) yra užsikrėtimų skaičius per dieną.
- Antrinis sergamumas (*prevalence*) yra užsikrėtimų skaičius per dieną, tenkantis 100 000 gyventojų

Savivaldybės turi būti demografiškai palyginamos tarpusavyje, tad skaičiuodami sergamumo rodiklį, jį normalizuoti ne tik gyventojų skaičiui savivaldybėje, tačiau ir tos savivaldybės gyventojų skaičiui visose amžiaus bei lyties grupėse.

Rango linijas nuspalvinkite gradiento spalva (pvz., kad spalva indikuotų didžiausią teigiamą ar neigiamą rango pokytį arba didžiausius ir mažiausius rangus paskutinę laikotarpio dieną).



Mid level B

Naudodamiesi R arba Python programavimo kalbomis:

1. pavaizduokite 2010-aisiais ir 2020-aisiais metais gimusių Lietuvos piliečių asmens kodų 7-to skaitmens skirstinius (2 lentelės arba grafikai), paaiškinkite išskirtis arba jų nebuvimą;
2. pavaizduokite Lietuvos piliečių asmens kodų 10-to ir 11-to skaitmens skirstinius (2 lentelės arba grafikai), paaiškinkite išskirtis arba jų nebuvimą;
3. paaiškinkite 11-to skaitmens skirstinyje matomą asmens kodų generatoriaus defektą ir jo pasekmes, pasiūlykite defekto sprendimą.

Senior level

Naudodamiesi R arba Python programavimo kalbomis bei Vyriausiosios rinkimų komisijos valdomu API gaukite 2023 m. savivaldybių tarybų ir merų rinkimų II turo duomenis bei:

1. raskite kandidatus, kurių gauto palaikymo dalis, skirta balsais paštu, buvo statistiškai reikšmingai didesnė, lyginant su jų konkurentais toje pačioje rinkimų apygardoje tame pačiame rinkimų ture;
2. žemėlapyje pavaizduokite rinkimų apylinkes ir partijas, kurių kandidatų tose apylinkėse gauto palaikymo dalis, skirta balsais paštu, buvo statistiškai reikšmingai didesnė, lyginant su jų konkurentais toje pačioje rinkimų apylinkėje tame pačiame rinkimų ture.