Szia,

A cikket csatoltam.

A pdf linkeket itt találod: <https://www.kozut.hu/kozerdeku-adatok/orszagos-kozuti-adatbank/forgalomszamlalas/>.

A 2022-es adathoz geojson formátumban itt férsz hozzá egy regisztráció után: <https://napportal.kozut.hu/#/>

Az adatigényléshez itt találsz információkat: <https://www.kozut.hu/download/tajekoztato-kozerdeku-adatigenyleshez/>

Én ezt a két feladatot tartom megoldhatónak egy 10 hetes internship alatt:

1. A 2022-es ODM becslése: Ez a feladat felfogható úgy mint egy y=Px lináris regresszió, ahol y tartalmazza a forgalomadatokat, P mátrix mondja meg hogy ha valaki A-ból B-be megy, akkor melyik útszakaszokon milyen valószínűséggel megy át, és x vektor tartalmazza az OD mátrix elemeit. A legegyszerűbb esetben feltételezzük, hogy P-t és y-t ismerjük és x-et keressük. Szerencsére y benne van a geojson file-ban, ez egyszerű. Ami nehezebb, hogy P-t meg kell konstruálni. Manuálisan nyilván egyszerű, a google maps nagyjából megmondja, hogy merre kell menni A-ból B-be, hogy hogyan kell ezt automatizálni, az a te feladatod lesz. Tipp: [ez hasznos lehet](https://openrouteservice.org/services/), de lehet találsz jobb megoldást. A másik nehézség, hogy attól függően, hogy hány OD mátrix elemet akarunk becsülni, a rendszer könnyen megoldhatatlanná vagy aluldetermináltá tud válni. Hosszú irodalma van annak, hogy ezt hogyan lehet megoldani (pl, alternatív célfüggvényekkel, konvex optimizálással), a csatolt cikkben erről is van szó. Ha csak a legnagyobb városokra koncentrálunk, akkor valószínűleg könnyű a feladat, kérdés, hogy milyen felbontásig tudunk elmenni. Nyilván vannak modernebb machine learning cikkek és github libraryk is (utóbbiak nagyrésze szintén student projekt), ezeket is érdemes lehet az elején átnézni.

2. A 2016-os ODM becslése és összehasonlítása a KSH mikrocenzus adatokkal: Itt a fő nehézség, hogy a 2016-os közúti adatok csak pdf-ben vannak meg. Érdemes a projektet adatigényléssel kezdeni a közútnál (2016os forgalmi adatokat), vagy akár a KSHnál (2022-es kommutálási adatokat). Ha nagyon nem megy az első feladat, akkor a KSH adatokat akár training adatnak is fel lehet használni, nem csak validálásnak.

Remélem ez elég információ a projekt elkezdéséhez.

Üdv,

Gergő

ui: hasonló adatok vannak más országokban is, pl. az [Egyesült Királyságban](https://roadtraffic.dft.gov.uk/downloads). Ez nekünk rövidtávon nem annyira érdekes, de érdemes észben tartani, hátha ezeken is lehet tesztelni. Az osztrák adatot még nem sikerült megtalálnom, ha te megtalálod kérlek szólj.