

## Power Query

(nebyla jsem si jistá, zda mám popsat postup i zde, tak raději uvádím stručně)

- změna kódování (852 Střední Evropa DOS)
- přejmenování sloupců – snake\_case plus otázka, zda nechat vše v češtině nebo angličtině
- odebrání duplicit – sloupec RUIAN ID (identifikátor objektu)
- rfs date a timestamp – ověření, že je to datum
- hp – celé číslo, fbb\_customers – celé číslo, fbb\_penetration – desetinné číslo (nový sloupec – s procenty se někdy špatně pracuje)
- x, y – longitude, latitude
- odstranění řádků, kde chybí latitude, longitude, hp – popřípadě je možný převod souřadnic v systému JTSK na WGS84 (GPS) – v Pythonu Pyproj a Transformer – případně ve výsledném souboru odebrat (prázdné), já jsem si to upravila až v Pythonu
- uložení souboru
- přeuložení původního .csv souboru do .xlsx souboru (shoda formátů)
- obohacení dat
  - nový ukazatel *neuspokojena\_poptavka* ( $hp - fbb\_customers$ ) = kolik bytových jednotek ještě nemá pevný internet – kolik zákazníků lze získat – indikátor obchodního potenciálu
  - nový ukazatel *potencialni\_skore* ( $neuspokojena\_poptavka * 1000 / hp$ ) = lepší viditelnost (škála od 1 do 1000) – nedosažená poptávka podle velikosti objektu (jaký podíl na trhu je nezískaný, převedený na hodnotu mezi 0 a 1000, větší číslo = větší potenciál) – možná vizualizace, podmíněné formátování
  - nový ukazatel *sjv\_flag* – jestli je bytová jednotka součástí SJV (0 – ne, 1 – ano) – urychlí to pravděpodobně jednání o zavedení pevného internetu (=KDYŽ(Y2<>""; 1; 0))