

# Review projektu týmu E týmem D

## 1 Shrnutí projektu

Předložený report se zabývá vztahem mezi přítomností cizinců v populaci a kriminalitou na úrovni okresů v České republice. Cílem analýzy je zodpovědět otázku, zda podíl cizinců v okrese zvyšuje kriminalitu, případně jak tento jev ovlivňují vybrané socioekonomické faktory (nezaměstnanost, exekuce, sociální dávky a další). Autoři se současně snaží rozlišit, zda hraje roli pohlaví (muži vs. ženy) a specifická národnost cizinců, a také se zaměřují na různé druhy trestné činnosti (násilná, majetková, drogová atd.).

Tým udělal datovou analýzu, vizualizaci pro lepší pochopení rozložení jednotlivých proměnných a jejich distribuci. Sestavili korelační matice napříč všemi relevantními proměnnými, aby odhalily vysoké korelace. Identifikovány byly silné vazby mezi proměnnými téhož druhu (např. různé typy majetkové trestné činnosti) a proměnnými popisujícími stejnou veličinu (např. cizinci – muži vs. cizinci – ženy). Zároveň vyšla najevo korelace socioekonomických proměnných (nezaměstnanost, exekuce, různé druhy dávek). Vytvořené lineární modely byly postupně rozšiřovány o nezaměstnanost, exekuce a sociální dávky. GLM s logaritmickou linkovací funkcí a gamma distribucí byly použity pro lepší zachycení případné nelinearity vztahů. Autoři došli k závěrům, že

- **Podíl cizinců** v populaci má kladnou souvislost s mírou kriminality, ale nejde o jediný určující faktor.
- **Socioekonomické faktory** (nezaměstnanost, exekuce, sociální dávky) významně zvyšují vysvětlitelnou variabilitu, pokud jsou do modelu přidány.
- **Pohlaví cizinců** (konkrétně mužů, u žen nikoliv) je statistický významný.
- **Národnost**: některé konkrétní národnosti jsou v pozitivním vztahu ke kriminalitě, zatímco např. Ukrajinci v datech vyšli spíše neutrálně, určité národnosti pak mají vliv na konkrétní typ kriminality.
- **Druh kriminality**: větší podíl cizinců zvyšuje zejména majetkovou trestnou činnost.

## 2 Naše hodnocení a poznámky

- **Data z více zdrojů**: Oceňujeme širokou škálu dat: kriminalita, demografie, geografie i socioekonomické proměnné.
- **Dobře zvolený postup**: Postup EDA → korelační analýza → lineární regrese → GLM nám přijde zcela relevantní, stejně tak jako kontrola normality reziduí, heteroskedasticity a užití AIC při výběru modelů.
- **Závěrečné hodnocení** je srozumitelné a výstižně napsané.
- **Korelační matice** je hezky vizualizovaná a interpretovaná.
- **Kauzalita vs. korelace**: Z reportu není zřejmé, zda si růst počtu cizinců nevybírá regiony, které již před tím vykazovaly vyšší kriminalitu. Bez podrobné časové analýzy zůstává otázka kauzality částečně otevřená.
- **Vizualizace**: Grafy v sekci Vizualizace obsahují pouze část okresů (po Přerov, resp. Rakovník, zbytek je oříznut). Dále jsou hodnoty v absolutních číslech, kvůli čemuž jsou méně přehledné. Bylo by vhodné graf normalizovat počtem obyvatel. Tím se „snaha o zjištění, zda existuje nějaká vizuální korelace“ výrazně komplikuje.
- **Normalizace**: V sekci explorační analýza dat by bylo vhodnější zabarvení map, nebo alespoň histogramů, normalizovat napříč všemi. Takto například žlutá barva v jedné mapě neznamená to samé jako žlutá barva ve vedlejší mapě. V případě, že by v některých mapách poté nebyly rozdíly mezi regiony zřetelné, by bylo vhodné toto znormalizování použít alespoň v histogramech.
- **Úhlednost dokumentu**: Vizuální nedostatky jako zarovnání textu, chybějící popisky grafů a os. Hodnoty u histogramů začínající větším množstvím nul by mohly být zapsány například  $123 \times 10^{-5}$ . Takto jsou těžko pro čtenáře interpretovatelné a porovnatelné. Lehká úprava textu by zlepšila zážitek ze čtení.
- **Další drobné nedokonalosti**, jako jsou překlepy, stylistické nedostatky či překročení limitu počtu stránek reportu (myslíme si, že by bylo lepší vybrat ty nejdůležitější grafy a vizualizace a zbytek dát do přílohy).

## 3 Otázky

- Uvažovali jste nad použitím dat z většího časového rozsahu (než jen rok 2023) pro ověření, že cizinci míru kriminality skutečně ovlivňují, a nejedná se pouze o to, že se cizinci z různých důvodů usídlují v oblastech, kde již dříve byla vyšší kriminalita?
- Kdybyste měli více času, jak byste tuto studii rozšířili?
- Co vás během vaší statistické analýzy nejvíce překvapilo?
- Změnila tato práce váš pohled na celkovou populaci a cizince?