

Okruh C – vytvoření interaktivního dashboardu

Cílem úlohy je vytvoření **interaktivního dashboardu**, který se zaměřuje na problematiku gender pay gap ve vybraných zemích světa a na jeho souvislost s různými demografickými charakteristikami států. Preferovaným programem pro zpracování úlohy je Power BI, v případě zájmu je možné použít i jiný software, např. Tableau, Qlik, Shiny v R, Streamlit nebo Dash v Pythonu.

Gender pay gap index (rozdíl ve mzdách mužů a žen) je pro účely této úlohy definován jako podíl mediánu výdělku žen a mediánu výdělku mužů za dané časové období. Základní informace o gender pay gap: https://en.wikipedia.org/wiki/Gender_pay_gap (stránka je k dispozici i v češtině).

Datasetsy

- Gender pay gap index (GPGI) v letech 2016–2023: `Global_gender_gap_index_2023.xlsx`
- HDP na hlavu: `GDP_per_capita.csv`
- ISO kódy států světa:
https://en.wikipedia.org/wiki/List_of_ISO_3166_country_codes
- Naděje dožití: `Life_expectancy.csv`
- Počet let, který stráví ve škole děvčata: `School_life_expectancy_female.csv`
- Počet let, který stráví ve škole chlapci: `School_life_expectancy_male.csv`
- Počet obyvatel: `Total_population.csv`
- Nezaměstnanost dle pohlaví a věku: `Unemployment_rate_by_sex_age.csv`
- Úmrtnost v prvním roce života: `Mortality_rate.csv`
- Venkovské obyvatelstvo: `Rural_population.csv`

Pokyny

- 1) Načtěte data z jednotlivých souborů a odkazů do Power BI. Každý z těchto datasetů obsahuje data různých typů. Zkontrolujte kvalitu dat a proveďte úpravy, jako je odstranění chybějících hodnot a duplicit, data převed'te do vhodných datových typů. Pokud uznáte za vhodné, můžete zadané datasety doplnit i jinými, které se vztahují ke zkoumanému tématu (v tomto případě zdůvodněte jejich použití).
- 2) Jednotlivé datové soubory propojte do datového modelu.
- 3) V tabulce s ISO kódy států přidejte sloupec, který bude určovat, zda leží stát v Evropě, nebo ne (kritérium pro evropský stát si určete sami – přístupů je více).

- 4) V tabulce **Gender pay gap index** vytvořte dva nové sloupce:
- procentuální změna v hodnotě GPGI mezi lety 2016 a 2023 (jako 2023/2016),
 - rozdíl v pořadí státu mezi roky 2023 a 2016 (jako pořadí 2023 – pořadí 2016).
- 5) V případě potřeby vytvořte další počítané sloupce.
- 6) Vytvořte několikastránkovou sestavu, na jejíchž dashboardech bude k dispozici:
- přehledová tabulka všech dostupných informací,
 - vývoj GPGI v letech 2016 až 2023 (státy na začátku a konci žebříčku, hodnoty GPGI; můžete se zaměřit i na významné posuny mezi prvními 20 státy světa v tomto období),
 - demografické charakteristiky 10 nejlepších a 10 nejhorších států v GPGI (využijte ostatní datasety),
 - vyhodnocení GPGI v České republice a porovnání jejích demografických charakteristik se skupinou 10 nejlepších států světa (viz předchozí bod), dále vyhodnocení České republiky v rámci Evropy,
 - vizualizace nebo tabulkové znázornění souvislostí GPGI s demografickými charakteristikami z ostatních datasetů; inspirovat se můžete např. na <https://ourworldindata.org/grapher/gender-wage-gap-vs-gdp-per-capita>; ne vždy jsou demografické charakteristiky dostupné pro všechna období – v tomto případě vyberte vždy nejbližší rok k roku, v němž určujete GPGI,
 - další tabulky a vizualizace týkající se tématu.
- Předpokládá se, že stránky sestavy budou vhodně pojmenované, grafy a tabulky budou mít vhodné názvy a popisky os, v případě potřeby budou jednotlivé stránky doplněny stručnými komentáři k použitým datům/tabulkám/grafům. Stránky sestavy budou obsahovat různé typy vizualizací, které odpovídají znázorněným datům.
- 7) Umožněte uživateli v rámci jednotlivých stránek filtrovat minimálně dle příslušnosti do Evropy, státu a roku, kdy byl počítán GPGI (např. pomocí průřezu).
- 8) Vhodně nastavte interakce mezi jednotlivými tabulkami/grafy.
- 9) Interpretujte získané výsledky (zpracujte jednostránkové shrnutí Vašich výsledků).