**Zadání projektu**

**Úvod do projektu**

Na vašem analytickém oddělení nezávislé společnosti, která se zabývá životní úrovní občanů, jste se dohodli, že se pokusíte odpovědět na pár definovaných výzkumných otázek, které adresují **dostupnost základních potravin široké veřejnosti**. Kolegové již vydefinovali základní otázky, na které se pokusí odpovědět a poskytnout tuto informaci tiskovému oddělení. Toto oddělení bude výsledky prezentovat na následující konferenci zaměřené na tuto oblast.

Potřebují k tomu **od vás připravit robustní datové podklady**, ve kterých bude možné vidět **porovnání dostupnosti potravin na základě průměrných příjmů za určité časové období**.

Jako dodatečný materiál připravte i tabulku s HDP, GINI koeficientem a populací **dalších evropských států** ve stejném období, jako primární přehled pro ČR.

**Datové sady, které je možné použít pro získání vhodného datového podkladu**

**Primární tabulky:**

1. czechia\_payroll – Informace o mzdách v různých odvětvích za několikaleté období. Datová sada pochází z Portálu otevřených dat ČR.
2. czechia\_payroll\_calculation – Číselník kalkulací v tabulce mezd.
3. czechia\_payroll\_industry\_branch – Číselník odvětví v tabulce mezd.
4. czechia\_payroll\_unit – Číselník jednotek hodnot v tabulce mezd.
5. czechia\_payroll\_value\_type – Číselník typů hodnot v tabulce mezd.
6. czechia\_price – Informace o cenách vybraných potravin za několikaleté období. Datová sada pochází z Portálu otevřených dat ČR.
7. czechia\_price\_category – Číselník kategorií potravin, které se vyskytují v našem přehledu.

**Číselníky sdílených informací o ČR:**

1. czechia\_region – Číselník krajů České republiky dle normy CZ-NUTS 2.
2. czechia\_district – Číselník okresů České republiky dle normy LAU.

**Dodatečné tabulky:**

1. countries - Všemožné informace o zemích na světě, například hlavní město, měna, národní jídlo nebo průměrná výška populace.
2. economies - HDP, GINI, daňová zátěž, atd. pro daný stát a rok.

**Výzkumné otázky**

1. Rostou v průběhu let mzdy ve všech odvětvích, nebo v některých klesají?
2. Kolik je možné si koupit litrů mléka a kilogramů chleba za první a poslední srovnatelné období v dostupných datech cen a mezd?
3. Která kategorie potravin zdražuje nejpomaleji (je u ní nejnižší percentuální meziroční nárůst)?
4. Existuje rok, ve kterém byl meziroční nárůst cen potravin výrazně vyšší než růst mezd (větší než 10 %)?
5. Má výška HDP vliv na změny ve mzdách a cenách potravin? Neboli, pokud HDP vzroste výrazněji v jednom roce, projeví se to na cenách potravin či mzdách ve stejném nebo násdujícím roce výraznějším růstem?

**Výstup projektu**

Pomozte kolegům s daným úkolem. Výstupem by měly být dvě tabulky v databázi, ze kterých se požadovaná data dají získat. Tabulky pojmenujte t\_{jmeno}\_{prijmeni}\_project\_SQL\_primary\_final (pro data mezd a cen potravin za Českou republiku sjednocených na totožné porovnatelné období – společné roky) a t\_{jmeno}\_{prijmeni}\_project\_SQL\_secondary\_final (pro dodatečná data o dalších evropských státech).

Dále připravte sadu SQL, které z vámi připravených tabulek získají datový podklad k odpovězení na vytyčené výzkumné otázky. Pozor, otázky/hypotézy mohou vaše výstupy podporovat i vyvracet! Záleží na tom, co říkají data.

Na svém GitHub účtu vytvořte repozitář (může být soukromý), kam uložíte všechny informace k projektu – hlavně SQL skript generující výslednou tabulku, popis mezivýsledků (průvodní listinu) a informace o výstupních datech (například kde chybí hodnoty apod.).

**Neupravujte data v primárních tabulkách! Pokud bude potřeba transformovat hodnoty, dělejte tak až v tabulkách nebo pohledech, které si nově vytváříte.**

**Informace o mně**

*Celé jméno*: Lucie Navrátilová, tel: +420739012698, e-mail: lnavratilova13@gmail.com

*Discord*: student Lucie N.

*GitHub*: uživatel: NavLuc, <https://github.com/NavLuc>,

*SQLProjekt*: https://github.com/NavLuc/ProjektSQL-ZivotniUrovenObcanu

**Řešení projektu**

## Datové vstupy

Použité datové struktury, popis toho, jaká data obsahují a význam jejich atributů jsou popsány viz Dokumentace\Specifikace dat

## Popis scriptů řešících projektové výzkumné otázky

### Kontrola dat

Provedeny kontroly dat viz VyzkumneOtazky-scriptySQL\ KontrolaDat.sql

### Vlastní datové struktury

Vytvořeny scripty pro vytvoření vlastních datových struktur pro získání datových sestav z výzkumných otázek viz VyzkumneOtazky-scriptySQL\ t\_lucie\_navratilova\_project\_SQL\_primary\_final-script.sql, t\_lucie\_navratilova\_project\_SQL\_secondary\_final-script.sql.

1. Rostou v průběhu let mzdy ve všech odvětvích, nebo v některých klesají?

Řešeno prostřednictvím scriptu viz: VyzkumneOtazky-scriptySQL\Otazka1-RustMezdVeVsechOdvetvich\

* Vytvořeny dvě datové sady s růsty mezd; jedna pro kalkulaci průměrných mezd typu fyzická a druhá typu přepočtená, které se omezuji pouze na klesající mzdy v příslušném odvětví.
* Pro vyjádření meziročního období je zvolen vždy rok předešlý a následující pro všechny roky, za které jsou mzdy evidovány v obdržených datových sadách, a to vždy v prvním kvartálu v příslušném roce.
* Pro zvolení způsobu meziročního období bylo postupováno dle metodiky českého statistického úřadu: https://www.statistikaamy.cz/o-slozitem-jednoduse/klice-k-porovnani-mezirocni-mezictvrtletni-a-bazicke-indexy/

**Odpověď: V některých odvětvích mzdy klesají a výsledek SQL scriptu zobrazuje ve kterých všech odvětvích (industry) mzdy klesají (average\_wages\_growth) a v jakém meziročním období(year\_prev a year) a to v případě přepočteného nebo fyzického typu (calculation\_type) výpočtu mezd.**

1. Kolik je možné si koupit litrů mléka a kilogramů chleba za první a poslední srovnatelné období v dostupných datech cen a mezd?

Řešeno prostřednictvím scriptu viz: VyzkumneOtazky-scriptySQL\Otazka2-KolikJeMozneSiKoupitMlekaaChlebaZaPrvniAPosledniObdobi\

* Určeny společné roky pro ceny potravin a mezd, z nich zvoleno první dostupné období – nejnižší rok a první kvartál a zvoleno poslední období jako poslední rok a první kvartál.
* Vytvořená datová sada kupní síly obyvatel, co se týká nákupu chleba či mléka dle mezd v jednotlivých odvětvích, vztažena k výše určeným obdobím.
* Pro vyjádření meziročního období je zvolen vždy rok předešlý a následující pro všechny roky, za které jsou mzdy evidovány v obdržených datových sadách, a to vždy v prvním kvartálu v příslušném roce.
* Pro zvolení způsobu meziročního období bylo postupováno dle metodiky českého statistického úřadu: <https://www.statistikaamy.cz/o-slozitem-jednoduse/klice-k-porovnani-mezirocni-mezictvrtletni-a-bazicke-indexy/>

**Odpověď: Výsledek SQL scriptu zobrazuje ve kterých odvětvích (industry) jaké množství potraviny (foof\_category je Mléko plnotučné … nebo chléb konzumní …) lze zakoupit za první období (amount\_of\_food\_per\_av\_wages\_min\_period) a jaké množství za poslední srovnatelné období (amount\_of\_food\_per\_av\_wages\_max\_period). Je to seřazeno sestupně podle největšího množství po nejmenší v příslušném odvětví.**

1. Která kategorie potravin zdražuje nejpomaleji (je u ní nejnižší percentuální meziroční nárůst)?

Řešeno prostřednictvím scriptu viz: VyzkumneOtazky-scriptySQL\Otazka3-KteraKategoriePotravinZdrazujeNejpomaleji\

* Vytvořena sada průměrných meziročních růstů jednotlivých potravin dle jejich kategorie a zvolena potravina s nejmenším průměrným růstem.
* Pro vyjádření meziročního období je zvolen vždy rok předešlý a následující pro všechny roky, za které jsou mzdy evidovány v obdržených datových sadách, a to vždy v prvním kvartálu v příslušném roce.
* Pro zvolení způsobu meziročního období bylo postupováno dle metodiky českého statistického úřadu: https://www.statistikaamy.cz/o-slozitem-jednoduse/klice-k-porovnani-mezirocni-mezictvrtletni-a-bazicke-indexy/

**Odpověď: Šunkový salám.**

1. Existuje rok, ve kterém byl meziroční nárůst cen potravin výrazně vyšší než růst mezd (větší než 10 %)?

Řešeno prostřednictvím scriptu viz: VyzkumneOtazky-scriptySQL\Otazka4-RokKdyBylRustCenPotravinVyrazneVyssiNezRustMezd\

* Vytvořena datová sada, která obsahuje porovnání kupní síly obyvatel v jednotlivých odvětvích vůči jednotlivým cenám potravin (=kolik potravin v příslušné kategorii a ceny ve srovnatelném období si mohl koupit obyvatel v daném odvětví dle jeho průměrné mzdy.) a zvolené jsou roky, kdy procentuální rozdíl kupních sil meziročně byl větší než 10%.
* Pro vyjádření meziročního období je zvolen vždy rok předešlý a následující pro všechny roky, za které jsou mzdy evidovány v obdržených datových sadách, a to vždy v prvním kvartálu v příslušném roce.
* Pro zvolení způsobu meziročního období bylo postupováno dle metodiky českého statistického úřadu: https://www.statistikaamy.cz/o-slozitem-jednoduse/klice-k-porovnani-mezirocni-mezictvrtletni-a-bazicke-indexy/

**Odpověď: Ano, šlo o roky: 2006, 2007, 2008, 2013,2016,2017.**

1. Má výška HDP vliv na změny ve mzdách a cenách potravin? Neboli, pokud HDP vzroste výrazněji v jednom roce, projeví se to na cenách potravin či mzdách ve stejném nebo následujícím roce výraznějším růstem?

Řešeno prostřednictví scriptu viz: VyzkumneOtazky-scriptySQL\Otazka5-MaVyskaHDPVlivNaZmenyVeMzdachaCenechPotravin\ MaVyskaHDPVlivNaZmenyVCenachPotravin-script.sql

* Určeny společné roky pro ceny potravin a mezd a rok, ve kterém byl naměřen největší nárůst HDP.
* Určeny průměrné ceny potravin v roce (shlukem přes kvartály v roce).
* Vytvořena datová sada průměrných ročních růstů cen potravin v roce viz výše s největším nárůstem HDP a v následujícím roce a jejich porovnání.
* Pro vyjádření meziročního období je zvolen vždy rok předešlý a následující pro všechny roky, za které jsou mzdy evidovány v obdržených datových sadách.
* Pro zvolení způsobu meziročního období bylo postupováno dle metodiky českého statistického úřadu: <https://www.statistikaamy.cz/o-slozitem-jednoduse/klice-k-porovnani-mezirocni-mezictvrtletni-a-bazicke-indexy/>

Řešeno prostřednictvím scriptu viz: VyzkumneOtazky-scriptySQL\Otazka5-MaVyskaHDPVlivNaZmenyVeMzdachaCenechPotravin\MaVyskaHDPVlivNaZmenyVeMzdacha-script.sql

* Určeny společné roky pro průměrné mzdy a rok, ve kterém byl naměřen největší nárůst HDP.
* Určeny průměrné mzdy v roce (shlukem přes kvartály v roce).
* Vytvořena datová sada průměrných ročních mezd v roce viz výše s největším nárůstem HDP a v následující, roce a jejich porovnání.
* Pro vyjádření meziročního období je zvolen vždy rok předešlý a následující pro všechny roky, za které jsou mzdy evidovány v obdržených datových.
* Pro zvolení způsobu meziročního období bylo postupováno dle metodiky českého statistického úřadu: https://www.statistikaamy.cz/o-slozitem-jednoduse/klice-k-porovnani-mezirocni-mezictvrtletni-a-bazicke-indexy/

**Odpověď:** Podrobné informace uvádí script ve sloupci price\_growth v případě ceny potravin a average\_wages\_growth v případě mezd.

**U cen potravin je to různé, u některých lze vysledovat v roce nárůstu HDP spíše pokles cen potravin a následný rok jejich nárůst, některé klesají v obou letech a některé v obou letech rostou.**

**U mezd jde o většinový nárůst v obou letech až na odvětví Peněžnictví a pojišťovnictví (kde pokles růstu HDP je v roce nárůstu HDP a v následném není kompenzován, nárůst je nižší než pokles.)**