

Projekt SQL

Tento projekt má porovnat dostupnost potravin na základě průměrných příjmů za společné srovnatelné období pro mzdy i ceny potravin v ČR.

Popis tvorby datových podkladů

Pro zodpovězení výzkumných dotazů bude nejdříve nutné vytvořit dvě tabulky s datovými podklady.

Nejdříve primární tabulku s údaji o průměrných mzdách a průměrných cenách potravin. Následovně sekundární tabulku s hodnotami hrubého domácího produktu. Obě tabulky za společné časové období.

Popis tvorby primární tabulky

První tabulka bude obsahovat následující údaje:

- a) Průměrná výše mezd v uvedeném průmyslovém odvětví za sledovaný rok,
- b) průmyslové odvětví,
- c) rok, pro který je vykazována průměrná mzda v odvětví,
- d) průměrná cena potravin v dané kategorii potravin za sledovaný,
- e) kategorie potravin,
- f) rok, pro který je vykazována cena potravin,

Výše uvedené údaje získáme z tabulek `czechia_payroll` a `czechia_price`. Abychom mohli vytvořit požadovanou tabulku, musíme tyto dvě tabulky spojit.

Ještě před spojením tabulek si vytvořím dvě pomocné tabulky. První pomocnou tabulku nazvu jako `t_marek_lestina_primary_part1` a druhou jako `t_marek_lestina_primary_part2`. Pomocné tabulky jsou upravené pomocí agregační funkce `AVG()`, kterou jsem použil, abych získal průměrné mzdy a ceny potravin.

V dalším kroku jsem seskupil řádky, které mají stejné hodnoty pomocí příkazu `GROUP BY`. Operace jsem použil za účelem redukce počtu řádků. U pomocné tabulky č. 1 jsem vybral pouze průměrné hrubé mzdy zaměstnanců zadáním příkazu „`WHERE = 5958`“. U pomocné tabulky č. 2 jsem převedl formát datumu sloupce `Date_from`, pojmenovaný jako `price_year`, na roky pomocí funkce `YEAR`. Sloupec `Price_year` použiji jako atribut pro spojení těchto pomocných tabulek pomocí příkazu `JOIN` se sloupcem `payroll_year`. Příkaz `JOIN` mi spojí pouze ty řádky tabulek, které se rovnají pro zadané sloupce. Tím získám data za společné sledované období.

Pro jistotu jsem si ověřil sledované období společné pro všechny podkladové tabulky. Vyhledal jsem si lokální maximum a minimum pro sloupce tabulek, ve kterých je uvedený rok, ke kterému se data vztahují pomocí agregačních funkcí `MIN()` a `MAX()`. Ze získaných hodnot jsem určil průnik množin jako období mezi lety 2006 až 2018.

Výslednou tabulku `t_Marek_Lestina_project_SQL_primary_final` vytvoříme pomocí příkazu `CREATE TABLE`.

Popis tvorby sekundární tabulky

Druhá tabulka bude obsahovat následující údaje:

- a) Název státu,
- b) populace,
- c) hrubý domácí produkt (GDP),
- d) Gini,
- e) rok, pro který se vykazuje GDP a gini.

Výše uvedené údaje získáme spojením tabulek countries a economies pomocí příkazu JOIN. Pro redukci počtu řádků jsem vybral pouze evropské státy a sledované období mezi lety 2006 až 2018 zadáním příkazu “**WHERE** c.continent = 'Europe' **AND** e.`year` **BETWEEN** 2006 **and** 2018. Výslednou tabulku jsem vytvořil pomocí příkazu CREATE TABLE.

Výzkumné dotazy

Dotaz č. 1

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

Odpověď

Připravil jsem datový podklad, který zobrazuje průměrné mzdy v daném odvětví za rok 2006 a 2018. Ve všech odvětvích došlo oproti roku 2006 k nárůstu výše mezd v roce 2018.

Dotaz č. 2

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