

Projekt SQL – Průvodní zpráva

Úvod

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. V porovnání s rokem 2006 byly průměrné mzdy ve všech odvětvích na konci sledovaného období v roce 2018 vyšší.

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?

Odpověď

Připravil jsem datový podklad, který zobrazuje podíl průměrné mzdy a ceny sledovaných potravin za určité odvětví v roce 2006 a 2018. Tento podíl se zobrazuje pouze pro kategorii potravin s čísly '111301' a '114201'. První zmiňované představuje chléb, druhé pak mléko.

Chleba (kg)		Mléko (l)	
2006	2018	2006	2018
871	782	995	938

Dotaz č. 3

Která kategorie potravin zdražuje nejpomaleji (je u ní nejnižší procentuální meziroční nárůst)?

Odpověď

Zde jsem napojil tabulku samu na sebe pomocí příkazu JOIN. Jako klíč jsem použil kategorii potravin. Pomocí vnořeného selektu jsem vybral rok 2006 u „první“ tabulky a rok 2018 u „druhé“ tabulky. Vznikla tak jedna tabulka se sloupci cen potravin v jednotlivých kategoriích za oba uvedené roky. V dalším kroku vznikl nový sloupec `Lowest_price_increase_percent` s procentuálním vyjádřením změny cen potravin za sledované roky, který jsem si seřadil vzestupně od nejnižší procentuální změny cen.

- a) Nejvíce zdražilo máslo o +99,04%;
- b) Pokles ceny zaznamenán pro Cukr o -27,27% a Rajská jablka -24,14%;
- c) Nejméně zdražily banány +7,41%.

Dotaz č. 4

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

Odpověď

Datové podklady jsem tvořil podobně jako u předchozího dotazu. Tentokrát jsem neprovedl rozpad na kategorie potravin a průmyslové odvětví. Tabulku jsem na samu sebe napojil pomocí klíčů `price_year` a `food_category`. Ke klíči pravé strany `price_category` jsem přičetl 1. Jako datový podklad jsem získal tabulku s celkovým meziročním růstem průměrných mezd a cen potravin.

Pomocí klauzule CASE jsem nastavil nejdříve dvě podmínky.

- a) meziroční změna mezd bude větší než nula
 - $\text{round}((t.\text{avg_wages}-t2.\text{avg_wages})/ t2.\text{avg_wages} * 100) > 0$
- b) meziroční změna cen potravin bude větší než nula
 - $\text{round}((t.\text{avg_price}-t2.\text{avg_price})/ t2.\text{avg_price} * 100) > 0$

Pokud tyto podmínky nebudou splněny současně, vypíše se hodnota „0“, rok jsme nenalezli.

- c) Pokud jsou podmínky současně splněny, musí být splněna poslední podmínka. Procentuální rozdíl cen potravin a výše mezd těchto hodnot musí být větší jak „10“, vypíše se hodnota „1“. Hodnota „0“ nebo „1“ se vypíše do sloupce *Difference_percent_more_than_10*.
 - $(\text{round}(((t.\text{avg_price}-t2.\text{avg_price})/ t2.\text{avg_price} * 100)-((t.\text{avg_wages}-t2.\text{avg_wages})/ t2.\text{avg_wages} * 100)) > 10$

V připraveném datovém podkladu jsem takový rok nenalezl.

Dotaz č. 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 následujícím roce výraznějším růstem?

Odpověď

V tomto úkolu budu potřebovat vyjádřit meziroční změny HDP, platů a cen potravin v ČR. Tyto změny vyjádřit v procentuální hodnotě. Následně zhodnotit pozorované výsledky.

Postup: Abych získal meziroční změny hodnot, bude třeba napojit agregovanou sekundární tabulku příkazem JOIN na samu sebe pomocí klíče `a.`year` = b.`year` + 1`. Následně vytvořím sloupec `GDP_growth_percent` představující meziroční změny HDP v ČR. Data za ČR ve sloupci `country` zobrazí příkaz `WHERE country = „Czech republic“`. Toto zadám vnořeným příkazem do klauzule `FROM`. Na tuto vnořenou tabulku „gdp“ napojím další vnořenou tabulku „pp“, kterou vytvořím napojením agregované primární tabulky příkazem JOIN na samu sebe pomocí klíče `c.payroll_year = d.payroll_year + 1`. Následně vytvořím sloupce `prices_increase` a `payrol_increase` vyjadřující meziroční změny cen potravin a mezd v procentech. Tyto vnořené tabulky `gdp` a `pp` spojím pomocí klíče `gdp.`year` = pp.payroll_year`.

V průběhu let lze zaznamenat při změnách HDP růst i pokles cen potravin a mezd.

Z připraveného datového podkladu lze pozorovat, že mezi lety 2006-2008 došlo k růstu HDP a současně vzrostly i ceny potravin a výše mezd. Zde bychom mohli tvrdit, že pokud se hospodářství daří, porostou ceny potravin i mzdy a nebo naopak. Ovšem přelom let 2008-2009 propadem HDP o -5% a současným 3% nárůstem mezd nebo meziroční 5% nárůst HDP mezi lety 2014-2015 s poklesem cen potravin o -2% tento trend vyvrací.

Na změny ve mzdách a cenách potravin může mít vliv více faktorů. Datový podklad nezohledňuje například míru nezaměstnanosti, spotřebitelské ceny jiných odvětví, celkovou meziroční míru inflace.

- a) Inflace - Prudkým nárůstem cen celého spotřebního koše dochází k poklesu reálných mezd;
- b) Nezaměstnanost – Klesající spotřeba domácností;
- c) Ceny jiných odvětví – Zvýšení nákladů vstupů.