

Projekt SQL

Libuše Gregorová

Discord nick: Liba G.

GitHub účet: LibuseG

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.

Potřebuji 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.

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 procentuální 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 následujícím roce výraznějším růstem?

Postup

Po prostudování zadání si projdu tabulky, které mám k úkolu k dispozici a stanovím sloupce, které budu z jednotlivých tabulek potřebovat k získání potřebných dat. U některých hodnot si dohledám jejich význam na webu Open Data (např. `calculation_code` v tabulce `czechia_payroll`).

Pro výzkumné otázky 1. až 4. si vytvořím tabulku:

[t_libuse_gregorova_project_SQL_primary_final](#)

(pro data mezd a cen potravin za Českou republiku sjednocených na totožné porovnatelné období – společné roky)

Ceny jednotlivých potravin v daných letech z `Czechia_price` doplním o měrné jednotky a názvy potravin z `tabulky Czechia_price_category`. Je potřeba vybrat jen data pro ceny potravin, které nejsou v jednotlivých krajích ale pro celou Českou republiku.

Připojím mzdy za konkrétní odvětví a HDP za jednotlivé roky. Průměrná hrubá mzda je `value` s `value_type_code` 5958 z `tabulky czechia:payroll`.

Období – pro mzdy je rok `payroll_year` a pro ceny období `date_from` a `date_to` v `tabulce Czechia_price`.

Nakonec připojím názvy odvětví.

Pro výzkumný dotaz 5. sestavím `tabulku`:

[t_libuse_gregorova_project_SQL_secondary_final](#)

(pro dodatečná data o dalších evropských státech).

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

Srovnání vývoje, mezd, cen potravin a HDP můžu udělat jen pro Českou republiku, protože pro ostatní země nemám potřebná data. V zadání je ale i požadavek pro přidání údajů o GINI koeficientu a populaci dalších evropských, takže tyto údaje do `tabulky` doplním.

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.

Řešení

1. Rostou v průběhu let mzdy ve všech odvětvích, nebo v některých klesají? [Projekt_3.sql](#)

Potřebuji meziroční srovnání mezd za jednotlivá odvětví. Protože ve výchozí `tabulce` mám odvětví definována názvem a ne kódem, je zaručeno, že ve výsledku nebudou hodnoty mezd pro řádky bez odvětví. Pro zjednodušení porovnávám růst mezi prvním a posledním srovnatelným obdobím.

V pomocných sloupcích vypíšu pro jednotlivá odvětví průměrnou mzdu pro první a pro poslední období a v dynamicky vytvořeném sloupci se porovnáním stanovím, jestli mezi daty došlo k poklesu nebo růstu. V našem případě došlo u všech odvětví k růstu mezd.

2. Kolik je možné si koupit litrů mléka a kilogramů chleba za první a poslední srovnatelné období v dostupných datech cen a mezd? [Projekt_4.sql](#)

Nejprve si pomocným dotazem zjistím, pro které roky mám údaje o mzdách a cenách potravin. Zároveň tím určím první a poslední srovnatelné období, v tomto případě roky 2006 a 2018.

Zjistím průměrnou cenu za litr mléka a za kilogram chleba v jednotlivých letech.

Zjistím průměrnou mzdu v jednotlivých letech.

Vyfiltruji data pro minimální a maximální období (rok).

Za první srovnatelné období (v roce 2006) je možné si koupit 1313 Kg chleba a 1466 litrů mléka za průměrnou mzdu. Za poslední srovnatelné období (v roce 2018) je možné si koupit 1365 Kg chleba a 1670 litrů mléka za průměrnou mzdu.

3. Která kategorie potravin zdražuje nejpomaleji (je u ní nejnižší procentuální meziroční nárůst)? Projekt_5.sql

Potřebuji přehled kategorií potravin a ceny jednotlivých let.

Vytvořím si dočasnou tabulku s hodnotami, které jsou pro výpočet potřeba. V další dočasné tabulce si zobrazím ceny a nárůst oproti předchozímu roku pro jednotlivé kategorie potravin a roky.

V následujícím dotazu určím průměrný meziroční nárůst podle kategorií potravin a seřadím vzestupně.

Nejpomaleji zdražuje Jakostní víno bílé.

Následuje Pivo výčepní, světlé, lahvové a Přírodní minerální voda uhlíčitá.

4. Existuje rok, ve kterém byl meziroční nárůst cen potravin výrazně vyšší než růst mezd (větší než 10 %)? Projekt_6.sql

Vytvořím si dočasnou tabulku s hodnotami průměrných mezd a průměrných cen pro jednotlivé roky. Potom vypočítám meziroční nárůst mezd a cen a rozdíl mezi těmito hodnotami. V dynamickém sloupci určím roky, pro které byla nebo nebyla splněna podmínka zadání.

V žádném roce není růst cen potravin výrazně vyšší než růst mezd. Nejvyšší nárůst cen potravin oproti předchozímu roku byl v roce 2017 a to 10%, mzdy v tomto roce vzrostly o 6,2%.

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? Projekt_7.sql

Nejdřív si připravím dočasnou tabulku s hodnotami HDP, mezd a cen potravin pro Českou republiku.

Pro jednotlivé roky vyčíslím meziroční změnu všech těchto sledovaných hodnot. Vytvořím dynamický sloupec, který vybere roky, kdy došlo k výraznějšímu růstu HDP. Pro každý rok ještě vytvořím dynamické sloupce, které popisují změnu mezd a cen potravin.

HDP vzrostlo výrazněji v letech 2006, 2014, 2016 a 2017. Ve většině případů došlo v následujícím roce k výraznějšímu zvýšení cen potravin a mezd, ale ne ve všech případech. Na růst cen a mezd tedy mohou mít vliv i jiné faktory a nebo se růst HDP může projevit s větším odstupem.