

SQL Projekt – sprievodný list

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 - Všechné 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 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?

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).

Primárna tabuľka

(SQL – Primárna tabuľka)

Pre tvorbu primárnej tabuľky boli zvolené dáta s ohľadom na stanovené výskumné otázky. Dáta o mzdách a cenách potravín, ktoré sú prepojené na základe časových období umožňujú detailnú analýzu vzťahu medzi nimi. Filtrácia dát pomocou klauzule “SELECT DISTINCT“ a množinového operátora “INTERSECT“ zaisťuje, že sú zahrnuté len tie roky, ktoré sú spoločné ako pre dáta o mzdách tak pre dáta o cenách potravín.

Podmienka “cpay.value_type_code = 5958“ bola zvolená kvôli jednoznačnej identifikácii záznamov v tabuľke “czechia_payroll“, ktoré sa vzťahujú k priemernej hrubej mzde zamestnanca. Táto hodnota poskytuje ucelený obraz o tom, aké sú priemerné príjmy zamestnancov v rôznych rokoch a odvetviach.

Sekundárna tabuľka

(SQL - Sekundárna tabuľka)

Dotaz je štruktúrovaný tak, aby získal presne tie dáta, ktoré sú relevantné pre piatu výskumnú otázku. Týmto prístupom sa zabezpečilo, že výsledná tabuľka bude obsahovať presné a usporiadané dáta pre ďalšiu analýzu. Ročníky 2006 až 2008 boli zvolené pomocou operátora “BETWEEN“, aby sa preskúmali trendy pri zmenách HDP so zameraním sa na Európsky región.

Výskumné otázky

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

(SQL – Výskumná_otázka_1)

Na základě analyzovaných údajů mzdy ve všech odvětvích mezi lety 2006 a 2018 vzrůstaly. Tento trend naznačuje, že v průběhu zvoleného období se průměrné mzdy ve všech uvedených obdobích nezměnily, ale naopak vzrůstaly. Pozitivní trend ve všech sektorech může naznačovat obecný ekonomický růst, nebo zlepšení pracovních podmínek v těchto oblastech.

Najvýraznější růst byl zaznamenán v oblasti Zemědělství, lesnictví, rybářství a nejnižší růst byl zaznamenán u Peněžnictví a pojišťovnictví.

Hoci Peněžnictví a pojišťovnictví vykázalo největší absolutní nárůst miezd, Zemědělství, lesnictví, rybářství, málo nejvyšší procentuální nárůst, kde mzda vzrástla z 12,543 CZK na 28,153 CZK, což představuje nárůst o 124%. Nejnižší procentuální růst byl zaznamenán u Peněžnictví a pojišťovnictví, kde mzda vzrástla z 35,838 CZK na 58,840 CZK, což představuje nárůst o 64%.

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

(SQL – Výskumná_otázka_2)

Analýza ukazuje, že v roce 2018 si můžeme za průměrný měsíční plat koupit více litrů mléka a kilogramů chleba než v roce 2006. Může to být v důsledku zvýšení průměrného měsíčního platu v porovnání se zvýšením cen těchto potravin.

V roce 2006 si za průměrný měsíční plat můžeme koupit přibližně 1,437 litrů mléka nebo 1,287 kilogramů chleba.

V roce 2018 si můžeme za průměrný měsíční plat koupit přibližně 1,641 litrů mléka nebo 1,345 kilogramů chleba.

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

(SQL – Výskumná_otázka_3)

Najnižší procentuální nárůst cen byl zaznamenán u kategorie potravin Cukr krystalový. Výsledky ukazují, že se cena meziročně snižovala a to v průměru o -1,92%. Tento pokles cen mohl být způsoben různými faktory, jako snížením poptávky,

nadprodukcia alebo zmeny v distribúcii.

Naopak potravinou s najvyšším percentuálnym nárastom sú papriky, kde cena vzrástla o 7,29%. Tento nárast môže byť ovplyvnený faktormi ako je sezónnosť, alebo zvýšený dopyt.

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

(SQL – Výskumná_otázka_4)

Áno, existuje rok, kde bol medziročný nárast cien potravín výrazne vyšší než rast miezd o viac než 10%. Na základe skúmaných údajov bol rok s najvyšším medziročným nárastom cien potravín v porovnaní s rastom miezd rok 2008.

Najvyšší medziročný nárast cien potravín v tomto roku dosiahol 44,04%, zatiaľ čo rast miezd bol 13,34%, čo predstavuje rozdiel 30,7%.

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?

(SQL – Výskumná_otázka_5)

Na základe analýzy obdobia 2006 – 2008 došlo v rokoch 2007 a 2008 k nárastu priemerných cien potravín a miezd. Hypotézu o vplyve HDP na tieto faktory nie je možné s úplnou istotou potvrdiť ani vyvrátiť. Na presnejšie určenie vplyvu HDP je potrebné zohľadniť aj ďalšie ekonomické a tržné faktory okrem základných ukazovateľov.