

SQL Projekt – Stoulil Patrik. Discord jméno – patrik09899.

Dle zadání se v následujícím projektu snažím pomocí SQL zodpovědět několik výzkumných otázek, které se týkají dostupnosti základních potravin široké veřejnosti.

Na začátek bylo potřeba připravit podklady, ve kterých je možné vidět porovnání dostupnosti potravin na základě průměrných příjmů za určité časové období.

Bylo tedy nejprve třeba vytvořit dvě tabulky, jejich popis je k nalezení níže. Následně bylo pomocí těchto tabulek možné zodpovědět definované otázky.

Primary table

Primární tabulku bylo dle zadání potřeba sjednotit na stejných letech pro potraviny a mzdy. Aby tedy v tabulce nebyly přebytné hodnoty, byla tabulka sjednocena (pomocí JOINu) na letech 2006-2018. (V hodnotách pro potraviny jsou pouze roky 2006-2018, v tabulce pro mzdy jsou hodnoty od roku 2000 do roku 2021).

Do tabulky bylo zvoleno 5 sloupců - payroll_year, industry_name, salary, name_of_commodity, price_of_commodity.

Tyto sloupce byly zvoleny z následujících důvodů:

payroll_year - pro označení jednotlivých roků.

industry_name - tento sloupec obsahuje názvy jednotlivých odvětví v každém roce. Tento sloupec je zde primárně kvůli první otázce, ve které bylo třeba zjistit, jak se vyvíjí mzdy v jednotlivých odvětvích. Pro existenci tohoto sloupce bylo třeba pomocí JOINu připojit názvy odvětví z tabulky czechia_payroll_industry_branch.

salary - následuje sloupec salary, který vyjadřuje průměrnou mzdu pro každé odvětví v jednotlivých letech. Tato hodnota byla průměrována pomocí funkce AVG, jelikož v tabulce mezd (czechia_payroll) jsou vždy mzdy uváděny za každý rok 4x - jednou za každé čtvrtletí. Tento sloupec bylo do tabulky třeba zahrnout kvůli otázce č. 2, taktéž kvůli otázce č.4.

name_of_commodity - udává jména jednotlivých potravin v každém roce. Potřeba pro výstupy otázek 2 a 3. Pro existenci tohoto sloupce bylo třeba pomocí JOINu připojit jména komodit z tabulky czechia_price_category.

price_of_commodity - navazuje na předchozí sloupec. Udává nám průměrnou cenu komodity v daném roce. Opět použita funkce AVG pro průměr ceny potravin za rok, aby bylo možné porovnávat se mzdami, které byly taktéž průměrovány jako hodnota za celý rok. Tento sloupec byl taktéž potřeba pro výstupy ze 2. a 3. a 4. otázky, určitě tedy nesmí v tabulce chybět.

Kvůli agregačním funkcím využitým v selectu bylo třeba před koncem tabulku patřičně zgrupovat (viz kód). Následně je tabulka jen dle preferencí seřazena pomocí ORDER BY.

Secondary Table

Dle pokynů ze zadání projektu do tabulky zahrnuji z tabulky economies zemi, HDP, gini a velikost populace. Pro přehlednost přidávám také dané období (roky).

Vzhledem k tomu, že v zadání je, že tabulka má být připravena pouze pro evropské státy, bylo třeba JOINnout tabulku countries, aby bylo následně pomocí WHERE možné omezit výčet zemí pouze na evropské státy.

Časové je tabulka omezena na období 2006 - 2018, dle zadání projektu.

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

Z dlouhodobého hlediska rostou mzdy ve všech odvětvích, což lze vidět při porovnání v prvním selectu - porovnání prvního a posledního roku v dostupných datech mezd - tedy porovnání mzdy v každém odvětví v roce 2006 a 2018.

Druhý select je podrobnější. V něm zjistíme, ve kterých letech a ve kterém konkrétním odvětví došlo k poklesu mzdy. V tomto selectu je použita funkce lag k tomu, aby bylo možné porovnat mzdu z aktuálního roku se mzdou z roku předchozího.

Z dlouhodobého hlediska tedy mzdy ve všech odvětvích bezesporu stoupají (první select), v několika případech však lze u některých odvětví zpozorovat mírný pokles (druhý select).

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

Tabulka je sjednocena na datech od roku 2006 do roku 2018. Prvním a posledním srovnatelným obdobím jsou tedy roky 2006 a 2018.

V otázce není specifikováno, jestli je dotaz mířen na jednotlivá profesní odvětví. Z toho důvodu tedy mzdu za roky 2006 a 2018 počítám průměrnou, tedy za všechna pracovní odvětví v ČR.

Z tohoto selectu lze po vydělení průměrné mzdy cenou komodity zjistit, že v roce 2006 bylo možné si průměrně koupit 1 287,46 kg chleba a 1 437,24 l mléka. V roce 2018 se dalo z průměrné mzdy pořídit 1 342,24 kg chleba a 1 641,57 l mléka.

Otázka 3. – Která kategorie potravin zdražuje nejpomaleji?

V kódu pro třetí otázku byly nejprve vybrány ceny jednotlivých potravin v konkrétních letech, následně je primární tabulka pomocí JOINu propojena sama se sebou na předešlém roce a jménu komodity, aby bylo možné vypočítat meziroční změnu jednotlivých potravin.

Po spočítání meziroční změny je pak pro každou potravinu pomocí celkového selectu stanovena průměrná meziroční změna, aby bylo možné odpovědět na otázku.

Výsledky jsou následně seřazeny od nejmenšího výsledku po nejvyšší. První dvě komodity nebereme v potaz (cukr a rajská jablka), jelikož u nich průměrně docházelo k poklesu ceny.

První komoditou, pro kterou vychází výsledek kladně (tedy jedná se o průměrný meziroční růst), jsou Banány žluté. Nejnižší zdražování potravin se tedy projevuje u Banánů.

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

První (vnitřní) select vytvořen pro zjištění meziroční změny ve mzdách za sledované období (průměrně za veškerá odvětví).

Druhý select vytvořen pro zjištění meziroční změny v cenách potravin (taktéž průměrně za všechny potraviny). Následně jsou tyto selecty spojeny pomocí JOINU, aby bylo možné zjistit, zdali v některém roce meziroční nárůst potravin dosáhl o 10 % většího růstu než růst mezd.

Výsledek celkového selectu nám říká, že v žádném roce nedošlo k tomu, že by potraviny zdražili o 10 nebo více %, než růst mezd.

Otázka 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?

Za významný růst byla zvolena hladina 5%.

Pro zjednodušení práce byla velká část kódu pro tuto otázku (celá WITH klauzule) zkopírována z předchozí otázky (otázka č.4), kde byla zjišťována meziroční změna cen potravin a mezd.

Následně je k této WITH klauzili 2x pomocí JOINu napojena Secondary_table pro možnost vytvoření meziročního růstu HDP.

Data jsou samozřejmě omezeny pouze na Českou republiku, jelikož informace o cenách potravin a mezd pro ostatní státy v databázi nemáme.

Finální (celkový) Select nám ukazuje data za roky, kde došlo k významnému růstu HDP (alespoň 5 %) oproti roku minulému - stalo se tak ve třech rocích. Připojení CASE Expression nám dále řekne (1 - ano, 0 - ne), zdali je v některém z těchto roků splněna požadovaná podmínka, tedy jestli v tom samém nebo následujícím roce došlo k významnému růstu potravin nebo mezd. Za významný růst jsem zvolil totožnou hodnotu jako u HDP, tedy 5%. Case expression nám říká, že ve dvou ze tří případů je tato podmínka splněna, tedy ve dvou ze tří období došlo současně s růstem HDP k navýšení cen potravin nebo mezd ve stejném, nebo následujícím roce (nebo obojí).

Ve dvou ze tří případů tedy zvýšení HDP má vliv na navýšení cen potravin nebo mezd. Tedy ve většině případů existuje souvislost mezi navýšením HDP a navýšením ostatních dvou veličin v totožném směru (navýšení).