

Průvodní listina k SQL projektu Datové akademie Engeto

Tato průvodní listina k SQL projektu Datové akademie se skládá ze dvou částí. V první popisují, z jakých dat jsem tvořil primární a sekundární tabulku potřebnou pro zodpovězení výzkumných otázek projektu. Ve druhé části se zabývám konstrukcí SQL skriptů a odpověďmi na jednotlivé výzkumné otázky.

Primární tabulka

Pro zodpovězení první otázky týkající se růstu mezd v odvětvích jsem použil rok, kód odvětví a průměrnou mzdu pro jednotlivá odvětví z tabulky *czechia_payroll* a název odvětví z tabulky *czechia_payroll_industry_branch*. Tabulky jsem spojil pomocí kódu odvětví a omezil na *value_type* = 5958, který odpovídá hrubé mzdě.

```
SELECT
    cp.payroll_year,
    cp.industry_branch_code,
    cpib.name AS industry_branch_name,
    round(avg(cp.value), 0) AS average_industry_wage
FROM czechia_payroll cp
LEFT JOIN czechia_payroll_industry_branch cpib -- připojení názvu odvětví
    ON cp.industry_branch_code = cpib.code
WHERE value_type_code = "5958"
GROUP BY payroll_year, industry_branch_code
ORDER BY payroll_year
```

Pro získání dat k dalším výzkumným otázkám jsem tabulku následně doplnil o ceny potravin z *czechia_price* a názvy potravin z *czechia_price_category*, spojeno

```
SELECT
    year(cp.date_from) AS `year`,
    round(avg(cp.value), 2) AS average_price,
    cpc.code AS food_code,
    cpc.name AS food_name
FROM czechia_price cp
LEFT JOIN czechia_price_category cpc
    ON cp.category_code = cpc.code
GROUP BY `year`, cpc.code, cpc.name
```

Konečně pro pátou otázku jsem k tabulce připojil údaje o hrubém domácím produktu v České republice z tabulky *economies*.

```
SELECT
    `year`,
```

```
gdp
FROM economies e
WHERE country = 'czech republic'
ORDER BY `year`
```

Sekundární tabulka

Tabulku jsem vytvořil s daty o evropských státech, konkrétně hrubým domácím produktem, GINI koeficientem a údaji o daních za totožné období, tj. za roky 2006 až 2018. Data z tabulky *economies* jsem dále propojil s tabulkou *countries*, abych mohl odfiltrovat státy podle evropského kontinentu.

```
SELECT
    e.country,
    e.`year`,
    e.GDP,
    e.gini,
    e.taxes
FROM economies e
JOIN countries c
    ON c.country = e.country
WHERE c.continent = "europe"
AND e.`year` BETWEEN 2006 AND 2018
ORDER BY e.country, e.`year`
```

Výzkumná otázka 1

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

Příprava SQL skriptu

Pomocí Common table Expression *WITH year_comparision* jsem spojil primární tabulku s primární tabulkou, ve které jsem payroll_year upravil o + 1. Následně jsem pomocí sloupce s rokem + 1 zobrazil rozdíl ve mzdách mezi požadovaným rokem a rokem + 1, vyjádřil ho v procentech a klauzulí *WHERE wage_change < 0* zobrazil pouze ta odvětví, ve kterých meziročně mzdy klesaly.

Odpověď

Meziročně mzdy klesaly celkem ve 29 případech, z nichž se desetkrát jednalo o pokles do 1 %. Nejvýraznější propad nastal v roce 2013 v peněžnictví a pojišťovnictví o 8,9 % z 50254 Kč na 45775 Kč, v roce 2020 v činnosti v oblasti nemovitostí o 7 % z 30607 Kč na 28466 Kč a taktéž v roce 2020 o 5,9 % v odvětví ubytování, stravování a pohostinství z 20367 Kč na 19170 Kč. Další meziroční poklesy mezd byly nižší než 5 %.

K alespoň minimálnímu meziročnímu poklesu mezd došlo v 15 z 19 sledovaných odvětvích, pouze ve 4 odvětvích nedošlo ve sledovaném období, tj. v letech 2000 až 2021, k meziročnímu poklesu mezd.

Výzkumná otázka 2

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

Příprava SQL skriptu

Data pro ceny potravin pocházela z let 2006 až 2018, což jsem odfiltroval pomocí WHERE klauzule. Zaokrouhleným rozdílem průměrné mzdy a ceny potravin jsem zjistil množství potravin, jaké je možné za mzdu nakoupit. Data jsem omezil pro kód potravin 114201 a 111301, což odpovídá 1 l mléka, resp. 1 kg chleba.

Odpověď

V prvním srovnatelném období – v roce 2006, bylo možné za průměrnou mzdu zakoupit 1192 kilogramů chleba a 1331 litrů mléka. V posledním srovnatelném období – v roce 2018, bylo možné zakoupit 1300 kilogramů chleba a 1590 litrů mléka. V obou případech došlo k nárůstu potravin, které je možné za mzdy zakoupit. V porovnání s rokem 2006 bylo v roce 2018 možná si za mzdu zakoupit o cca 9 % více chleba a o cca 19,5 % více mléka.

Výzkumná otázka 3

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

Příprava SQL skriptu

Za pomoci Window Function LAG jsem vyjádřil meziroční změny cen pro každou ze sledovaných druhů potravin.

Odpověď

Nejnižší meziroční procentuálně nárůst, čili nejvyšší meziroční procentuálně pokles, byl u rajských jablek, a to o cca 30 % v roce 2007 a o cca 28 % v roce 2011. K meziročnímu procentuálnímu poklesu o více než 25 % došlo i u bílého pšeničného pečiva v roce 2009, konkrétně o 28,5 %.

Výzkumná otázka 4

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

Příprava SQL skriptu

Pro přípravu podkladu k této otázce jsem opět použil Window Function LAG, se kterou jsem získal procentuální změny ceny potravin, mezd a nakonec i rozdílu mezi meziročními změnami cen potravin a mezd.

Odpověď

Ze získaných dat je zřejmé, že ceny potravin i výše mezd mají dlouhodobě tendenci růst, avšak ani jedna kategorie se ve sledovaném období nevyhnula meziročnímu poklesu. Největší rozdíl mezi nárůstem

cen potravin oproti růstu mezd byl v roce 2013, kdy se cena potravin meziročně zvýšila o 5,1 %, zatímco mzdy poklesly o 1,5 %, rozdíl byl 6,6 %. K většímu rozdílu nárůstu cen potravin oproti mzdám než v roce 2013 ve sledovaném období nedošlo, naopak častěji rostou mzdy více než ceny potravin.

Výzkumná 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?“

Příprava SQL skriptu

Skript z předchozí otázky jsem doplnil o procentuální meziroční změnu HDP v České republice.

Odpověď

Hrubý domácí produkt vzrostl v České republice ve sledovaném období nejvýrazněji v roce 2007 o 5,6 %, v roce 2015 o 5,4 % a v roce 2017 o 5,2 %. V ostatních letech byl růst HDP nižší než 5 %, ve dvou případech dokonce záporný.

V roce 2007, ve kterém byl nárůst HDP o 5,6 % došlo k nárůstu cen potravin o 6,7 % a nárůstu mezd o 6,9 %, v následujícím roce o 6,2 %, resp. o 7,9 %.

Vlivem hospodářské krize a ekonomické recese došlo v roce 2009 k poklesu HDP o 4,7 %, meziročně klesly i ceny potravin, konkrétně o 6,4 %, ale u mezd nedošlo k meziročnímu poklesu, pouze ke zpomalení růstu na 3,2 %.

K dalšímu výraznému nárůstu HDP o 5,4 % došlo v roce 2015, ve kterém však ceny potravin poklesly o 0,5 %, mzdy zaznamenaly nárůst 2,5 %. V roce následujícím ceny potravin opět poklesly, konkrétně o 1,2 % a mzdy rostly o 3,7 %.

Rok 2017 znamenal růst HDP o 5,2 % a cen potravin o rekordních 9,6 %. Mzdy se v tomto roce zvýšily o 6,3 %. V roce 2018 se ceny potravin zvýšily o 2,2 % a mzdy zaznamenaly nadstandardní nárůst o 7,6 %.

Z výše uvedeného a níže přiložené tabulky je zřejmé, že meziroční změna HDP přímo nekoreluje s cenami potravin a mzdami, přestože z dlouhodobého pohledu dochází k růstu HDP, potravin i mezd.

payroll_year	YoY_food_price_change	YoY_payroll_change	YoY_gdp_change
2006			
2007	6.7	6.9	5.6
2008	6.2	7.9	2.7
2009	-6.4	3.2	-4.7
2010	1.9	2.0	2.4
2011	3.4	2.3	1.8
2012	6.7	3.0	-0.8
2013	5.1	-1.5	0.0
2014	0.7	2.6	2.3
2015	-0.5	2.5	5.4
2016	-1.2	3.7	2.5
2017	9.6	6.3	5.2
2018	2.2	7.6	3.2