

Zadání projektu

Zadáním projektu je připravit robustní datové podklady a na jejich základě se pokusit odpovědět na několik výzkumných otázek, které se zabývají **dostupností základních potravin široké veřejnosti**.

Otázky jsou následující:

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 procentuálně 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? Nebo-li, pokud HDP vzroste výrazněji v jednom roce, projeví se to na cenách potravin či mzdách ve stejném nebo následujícím roce výraznějším růstem?*

K zodpovězení dotazů nám slouží několik tabulek:

czechia_payroll
czechia_payroll_calculation
czechia_payroll_industry_branch
czechia_payroll_unit
czechia_payroll_value_type
czechia_price
czechia_price_category
czechia_region
czechia_district
countries
economies

Na základě výše uvedených tabulek mají být vytvořeny dvě nové **výchozí tabulky**, ze kterých bude možné získat data potřebná pro zodpovězení otázek.

První tabulka **t_kristyna_popova_project_SQL_primary_final** - pro data mezd a cen potravin za ČR sjednocených na totožné porovnatelné období.

Druhá tabulka **t_kristyna_popova_project_SQL_secondary_final** – pro dodatečná data o dalších evropských státech.

Výchozí tabulky

Primární výchozí tabulka **t_kristyna_popova_project_SQL_primary_final** byla vytvořena sjednocením (join) dvou pohledů *primary_part_one* a *primary_part_two*.

Pohled *primary_part_one* byl vytvořen v zásadě z dat, které poskytovala tabulka, *czechia_payroll*. Byly k ní však dodány data z tabulky *czechia_payroll_industry_branch*, konkrétně název odvětví. Tabulky byly sjednoceny za základě kódu.

Obdobně tomu bylo i u pohledu *primary_part_two*, ten byl vytvořen sjednocením *czechia_price*, kam byly dodány data o množství (*price_value*) a jednotky (*price_unit*) z *czechia_price_category*.

Během tvoření pohledů docházelo pomocí podmínek k filtraci dat. Snahou bylo získat převážně jen ta data, které budou relevantní k zodpovězení otázek.

Pro zúžení dat bylo učiněno několik kroků:

➔ Data byla omezena na ty, která se týkají průměrné hrubé mzdy na zaměstnance:

WHERE cp1.value_type_code = 5958

➔ Data byla omezena na průměr ceny potravin pro celou ČR: WHERE region_code IS NULL .
Tato podmínka byla zvolena proto, že zadání nevyžaduje pracovat s jednotlivými kraji.
Vyfiltrují se data uvedená pro celou ČR.

Sjednocení pohledů a tvorba finální primární tabulky.

Pohledy byly sjednoceny na základě společných let.

Payroll_year byl tedy z prvního rozmezí let 2000-2021 zúžen na 2006-2018. To odpovídalo časovému rozpětí, které bylo poskytnuto v tabulce *czechia_price*.

Sekundární tabulka *t_kristyna_popova_project_SQL_secondary_final* byla vytvořena dle zadání:

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

z tabulky *economies* a tabulky *countries*, ze které byly údaje prostřednictvím vnořeného selectu vymezeny pouze pro EVROPU. Z tabulky *economies* byly vybrány sloupce *country*, *year*, *GDP*, *population*, *gini*.

Původní výstup čítal místa, kde nebylo uvedeno GDP, proto byla zvolena podmínka *e.gdp IS NOT NULL*, pro získání dat, kde byly údaje vyplněny. Obdobně tomu bylo i se sloupcem *gini*.

Dále bylo vymezeno i období *year BETWEEN 2006 AND 2018* a to z toho důvodu, že primární tabulka *t_kristyna_popova_project_SQL_primary_final* pracuje s údaji pouze z časového období mezi lety 2006-2018.

Výzkumné otázky

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

K zodpovězení otázky posloužily sloupce *hruba_mzda*, *odvětví*, *payroll_year* a *industry_branch_code* z tabulky *t_kristyna_popova_project_SQL_primary_final*.

Byla zvolena podmínka pro vymezení konkrétního odvětví *industry_branch_code = 'kód příslušného odvětví'*.

Dále byla použita klauzule GROUP BY, čímž jsme dostali příslušné roky do jednoho řádku.

Seřazení vzestupně ASC dle hrubé mzdy, mělo poukázat na to, zda vzestupně stoupající mzda odpovídá vzestupně stoupajícím letům (payroll_year).

Dotaz byl kladen pro každé odvětví zvlášť.

Výsledky:

Žádné odvětví nevykazovalo s každým rokem postupný nárůst hrubé mzdy. V každém odvětví docházelo občas k meziročnímu poklesu.

Některá odvětví byla stabilnější. Jako příklad uvedu *Informační a komunikační činnosti*, kde k poklesu došlo pouze mezi rokem 2010 a 2009 nebo *Zemědělství, lesnictví, rybářství*, kde k poklesu došlo mezi roky 2009 a 2008.

Vysoké výkyvy například vykazovalo odvětví *Peněžnictví a pojišťovnictví*, kde mezi jednotlivými roky docházelo opakovaně k nárůstům a poklesům hrubé mzdy. Největší pokles je viděn mezi roky 2012 a 2013.

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

Jako srovnatelné období byl stanoven rok 2006 (první srovnatelné období) a rok 2018 (poslední srovnatelné období).

Jelikož se hrubá mzda v odvětvích liší, bude se lišit i počet litrů mléka a kilogramů chleba, které by se daly koupit za hrubou mzdu v příslušných odvětvích. Zadání bylo počítáno pro každé odvětví jednotlivě.

Data byla rozdělena na stanovená léta (2006, 2018), na odvětví a rovněž byla rozdělena na dvě vyžadované kategorie potravin – mléko a chléb.

Výsledky:

Výpis výsledků pro jednotlivá odvětví a roky je uveden v tabulce.

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

Zde bylo využita funkce LAG() díky níž jsme do nového sloupce dostali nadefinovaný záznam z předchozího řádku. To nám umožnilo následně vytvořit výpočetní vzorec pro procentuální nárůst, který byl zaznamenán do nového sloupce.

Výstup byl následně vzestupně seřazen dle *procentuálního nárůstu*. Míňsové hodnoty ve sloupci *procentuální nárůst* nevykazují, že cena v daném období vzrostla (→nedošlo tedy k nárůstu), ale že klesla. Jako „nejnižší meziroční nárůst“ byly v tabulce pozorovány hodnoty, které se nejvíce blížily 0.

Prostřednictvím klauzule WHERE category_code a zadáním „příslušného kódu“ pro danou kategorii potravin (kódy v tabulce czechia_price_category), byly generovány výsledky pro příslušnou kategorii potravin. Výsledky pak byly jednotlivě projduy a pozorovány.

Výsledky:

Z výstupů pro jednotlivé kategorie potravin bylo vypořádováno, že nejvyšší meziroční nárůst měl Chléb konzumní kmínový v roce 2018. Meziroční procentuální nárůst byl 0,04 %.

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

Zde byla jako v předchozí otázce využita funkce LAG(). Byl vytvořen meziroční procentuální nárůst průměrné hrubé mzdy a meziroční procentuální nárůst průměrné ceny zmiňovaných potravin. Nakonec od sebe byly meziroční procentuální nárůsty každé kategorie odečteny – získal se rozdíl.

Výsledky:

V žádném roce nedošlo k tomu, že by meziroční nárůst cen potravin byl výrazně vyšší než růst mezd. K rozdílu 10% jsme se nejvíce přiblížili v roce 2009 – 9,97%.

5. Má výška HDP vliv na změny ve mzdách a cenách potravin? Nebo-li, pokud HDP vzroste výrazněji v jednom roce, projeví se to na cenách potravin či mzdách ve stejném nebo následujícím roce výraznějším růstem?

Výpočet byl prováděn pro Českou republiku.

V zásadě posloužily dotazy z předchozích otázek - funkce LAG(). Funkce LAG byla nejprve samostatně provedena pro výpočet meziročního nárůstu HDP (vycházelo se z tabulky [t_kristyna_popova_project_SQL_secondary_final](#)).

Následně došlo ke spojení obou tabulek přes roky. (fin1. payroll_year = fin2. year), přičemž byla použita funkce LAG() i pro meziroční procentuální nárůst mezd a potravin.

V poslední fázi byl dotaz upraven tak, že vygeneroval pouze roky a sloupce pro meziroční procentuální nárůsty pro jednotlivé kategorie – za účelem přehlednějšího porovnání.

Výsledky:

Z výsledků je patrné, že se výška HDP promítá i do cen potravin a mezd. Nelze však říci, že by bylo možné striktně paušálně aplikovat tvrzení, že vzroste-li HDP, vzrostou zákonitě ceny potravin a výška mezd. A naopak, klesne-li HDP projeví se to na cenách i mzdách.

Např.:

V roce 2012 HDP kleslo, ale nárůst v cenách potravin byl celkem značný.

V roce 2008 nedošlo k tak velkému nárůstu HDP, přičemž u potravin a mezd došlo, oproti HDP, k poměrně značnému nárůstu.