

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 *czechia_payroll_edit* a *czechia_price_edit*.

Pohled *czechia_payroll_edit* 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 *czechia_price_edit*. Ten byl vytvořen sjednocením *czechia_price*, kam byly dodány data o množství (*price_value* → *quantity*) a jednotky (*price_unit* → *units*) z *czechia_price_category*.

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

Dále byla pomocí klauzule GROUP BY seskupena dle *payroll_year* a *industry*.

➔ 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.

Dále byla pomocí klauzule GROUP BY seskupena dle *food_category* a *date*.

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 *gross_wage*, *industry* a *payroll_year* z tabulky [t_kristyna_popova_project_SQL_primary_final](#).

Byla použita klauzule GROUP BY, čímž bylo docíleno řazení dle odvětví (*industry*) a let (*payroll_year*).

Výsledky:

Odvětví, ve kterých mzdy v žádném roce neklesly, byly:

- Doprava a skladování
- Ostatní činnosti
- Zdravotní a sociální péče
- Zpracovatelský průmysl

Odvětví, která vykazovala snížení mzdy pouze mezi roky 2012 a 2013, jinak u nich byl patrný nárůst v každém roce, byly:

- Administrativní a podpůrné činnosti
- Činnosti v oblasti nemovitostí
- Informační a komunikační činnosti
- Peněžnictví a pojišťovnictví
- Stavebnictví
- Velkoobchod a maloobchod
- Zásobování vodou

Dále zmíněná odvětví vykazovaly poklesy v nárůstu mezd mezi více lety než jen 2012-2013

- Kulturní, zábavní a rekreační činnosti
- Profesní a vědecké, technické činnosti
- Těžba a dobývání
- Ubytování, stravování, pohostinství
- Veřejná správa a obrana
- Výroba a rozvod elektřiny
- Vzdělávání – 2009 -2010
- Zemědělství, rybářství – 2008 - 2009

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).

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

Výsledky:

- V roce 2006 si bylo možné koupit: **20 440 kilogramů chleba** a **24 627 litrů mléka**
- V roce 2018 si bylo možné koupit: **24 967 kilogramů chleba** a **30 422 litrů mléka**

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*. Mínusové 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 projduty a pozorovány.

Výsledky:

Z výstupů pro jednotlivé kategorie potravin bylo vypořezováno, že nejnižší 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:

Ano, jedná se o rok 2009 - 12.48%.

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 značnějšímu nárůstu.