

Porovnání dostupnosti potravin na základě průměrných příjmů za určité časové období.

Popis projektu

Zadaný projekt hledá odpovědi na pět výzkumných otázek (viz níže) o dostupnosti potravin široké veřejnosti v ČR a okrajově prokázat vztah mezi HDP a cenami potravin ve světovém měřítku v letech 2016-2018 z dat z Portálu otevřených dat ČR.

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?

Řešení

Finální tabulky

Pro zadaný projekt byly vytvořeny dvě tabulky: primární a sekundární. Primární tabulka `t_Sophia_Syrska_project_SQL_primary_final` zpracovává data mezd a cen potravin pro Českou republiku. Tato tabulka obsahuje:

- `price_category_name`,

- category_code,
- economic_sector,
- code,
- year,
- avg_price_value,
- measurement_unit,
- avg.salary

Následně vznikla sekundární tabulka t_Sophia_Syrská_project_SQL_secondary_table získaná ze zadaných tabulek countries a economies pocházející z Otevřených dat ČR Tabulka obsahuje sloupce:

- country,
- region_in_world,
- capital_city,
- year,
- population,
- GDP,
- Giny_index

Odpovědi na otázky

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

Všechna sledovaná odvětví v letech 2016-2018 zaznamenala nárůst mezd. Největšího procentuálního nárůstu mezd dosáhlo odvětví zdravotnictví a sociální péče 77,84 %. Narozdíl od odvětví peněžnictví a pojišťovnictví, které mělo nejnižší nárůst 37,12 %.

Pokud se na data podíváme podrobněji pohledem meziročního růstu, zjistíme, že poklesu ve mzdách docházelo ale jen ojediněle. Nejčastěji se tak dělo v letech 2012/2013, kdy se pokles týkal mírné nadpoloviční většiny kategorií odvětví, u nichž mzdy klesaly průměrně o -0,4 % až o -4,4 %. Výmečný pokles můžeme sledovat v roce 2013 v oboru Peněžnictví a bankovníctví a to o -8,83 %. K dalšímu poklesu mezd dále docházelo především v období let 2009–2011, propady byly ale velmi mírné, naprostá většina z nich se pohybovala kolem -1 %. Tedy meziročně mzdy nerostly u všech odvětví po celé sledované období. Naopak v některých kategoriích oborů mzdy konstantně rostly po celou dobu měřeného období, jedná se o Zpracovatelský průmysl, Dopravu a skladování (stagnovala v roce 2011), Zdravotní a sociální péči a Ostatní činnosti.

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

V roce 2006 bylo možné za průměrnou mzdu pořídit 1466 litrů mléka a 1313 kg chleba, v roce 2018 pak 1670 litrů mléka a 1365 kg chleba. Údaje poukazují na trend zvedající se kupní síly obyvatelstva V České republice.

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

Pokud neuvažujeme potraviny, které zlevnily (cukr krystalový -1,92 %, banány žluté - 0,74 %), v průběhu let 2016-2018 nejpomaleji zdražovaly vepřová pečeně (0,81 %) a přírodní minerální voda uhlíčitá (1,0 %).

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

Meziroční nárůst cen potravin v rozmezí let 2006-2018 nebyl nikdy větší než 10%. Nejvyšších hodnot dosahoval v roce 2013 a to 6,65%.

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?

Vliv HDP na průměrnou mzdu a vývoj cen v České republice se nepodařilo prokázat.

Kompletní zadání k vypracování projektu ENGETO 2023

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 v 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).

Dále připravte sadu SQL, které z vámi připravených tabulek získají datový podklad k odpovězení na vytyčené výzkumné otázky. Pozor, otázky/hypotézy mohou vaše výstupy podporovat i vyvracet! Záleží na tom, co říkají data.

Na svém GitHub účtu vytvořte repozitář (může být soukromý), kam uložíte všechny informace k projektu – hlavně SQL skript generující výslednou tabulku, popis mezivýsledků (průvodní listinu) a informace o výstupních datech (například kde chybí hodnoty apod.).

Neupravujte data v primárních tabulkách! Pokud bude potřeba transformovat hodnoty, dělejte tak až v tabulkách nebo pohledech, které si nově vytváříte.

