

Luboš Vavruška

Email: lubos.vavruska@seznam.cz

Discord username/ displayed name: vlubos/ Luboš

Zadání projektu

Úvod do projektu

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?

```
CREATE VIEW t_Lubos_Vavruska_project_SQL_primary_final AS
SELECT
    cp.value AS mzda,
    cpib.name AS odvetvi,
    cp.payroll_year AS rok
FROM czechia_payroll cp
JOIN czechia_payroll_industry_branch cpib
    ON cp.industry_branch_code = cpib.code
WHERE
    cp.value IS NOT NULL
    AND cpib.name IS NOT NULL
    AND cp.value_type_code = '5958'
GROUP BY odvetvi, rok
ORDER BY odvetvi ASC, rok ASC;
```

Odpověď: Až na menší nepravidelnosti (menší meziroční pokles v některých odvětvích/letech) je patrným trendem růst mezd v průběhu let.

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

```
CREATE VIEW t_Lubos_Vavruska_project_SQL_primary_final AS
SELECT
    cp.payroll_year AS rok,
    AVG(cp.value) AS prumerna_mzda,
    cpc.name AS polozka,
    cp.industry_branch_code AS odvetvi,
    (cp.value/ czpr.value) AS pocet_jednotek_dostupnych_za_mzdu
FROM czechia_payroll cp
JOIN czechia_price czpr
    ON cp.payroll_year = YEAR(czpr.date_from)
JOIN czechia_price_category cpc
    ON czpr.category_code = cpc.code
WHERE
    (cpc.name = 'Mléko polotučné pasterované'
    OR cpc.name = 'Chléb konzumní kmínový')
    AND cp.value IS NOT NULL
    AND cp.payroll_year IN ('2006', '2018')
    AND cp.value_type_code = '5958'
GROUP BY rok, odvetvi;
```

Odpověď: Počty litrů mléka a kilogramů chleba, které je možné si koupit za první a poslední srovnávací období (roky 2006 a 2018), jsou uvedené ve sloupci `pocet_jednotek_dostupnych_za_mzdu`.

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

```
CREATE VIEW t_Lubos_Vavruska_project_SQL_primary_final AS
SELECT
    AVG(czpr.value) AS prumerna_cena_polozky,
    YEAR(czpr.date_from) AS rok,
    MIN(czpr.value) AS minimalni_cena,
    czpr.value/MIN(czpr.value) AS narust_2018_proti_min_cene
FROM czechia_price czpr
JOIN czechia_price_category cpc
    ON czpr.category_code = cpc.code
WHERE YEAR(czpr.date_from) = '2018'
GROUP BY cpc.name
ORDER BY narust_2018_proti_min_cene ASC
LIMIT 1;
```

Odpověď: Nejpomaleji zdražuje rýže loupaná dlouhozrnná

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

```
CREATE VIEW t_Lubos_Vavruska_project_SQL_primary_final AS
SELECT
    AVG(cp.value) AS prumerna_mzda,
    SUM(czpr.value) AS nakupni_kosik,
    cp.payroll_year AS rok,
    SUM(czpr.value)/ AVG(cp.value) AS zmena_ceny_vs_mzdy
FROM czechia_payroll cp
JOIN czechia_price czpr
    ON cp.payroll_year = YEAR(czpr.date_from)
WHERE
    cp.value IS NOT NULL
    AND cp.payroll_year BETWEEN '2006'
    AND '2018'
    AND cp.value_type_code = '5958'
GROUP BY rok
ORDER BY rok ASC;
```

Odpověď: V žádném roce ve sledovaném období nebyl růst cen výrazně vyšší než růst mezd.

1. 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?

```
SELECT
    AVG(cp.value) AS prumerna_mzda,
    SUM(czpr.value) AS nakupni_kosik,
    cp.payroll_year AS rok,
    AVG(cp.value)/SUM(czpr.value) AS kupni_sila,
    e.GDP AS HDP
FROM czechia_payroll cp
JOIN czechia_price czpr
    ON cp.payroll_year = YEAR(czpr.date_from)
JOIN economies e
    ON cp.payroll_year = e.year
WHERE
    cp.value IS NOT NULL
    AND cp.payroll_year BETWEEN '2006' AND '2018'
    AND cp.value_type_code = '5958'
    AND e.country = 'Czech Republic'
GROUP BY rok
ORDER BY rok ASC;
```

Odpověď: Korelace mezi výší HDP a výší mezd nebo cen potravin není jednoznačně patrná.