

Závěrečná zpráva SQL projektu

ENGETO Datová Akademie 04-2022

Radim Jedlička

** V této zprávě bude vždy napřed uvedena výzkumná otázka, případně mezivýpočet, který je potřeba udělat pro další výpočty. Poté následují SQL dotazy, díky kterým získáme z databází potřebné informace pro zodpovězení dané otázky. Na závěr bude krátký komentář, co nám výsledek ukazuje.*

```
/*
 * Datové sady, které je možné použít pro získání vhodného datového
 podkladu
 *
 ****
 */

-- Primární tabulky:

-- Informace o mzdách v různých odvětvích za několikaleté období:
SELECT * FROM czechia_payroll cp;

-- Číselník kalkulací v tabulce mezd:
SELECT * FROM czechia_payroll_calculation cpc;

-- Číselník odvětví v tabulce mezd:
SELECT * FROM czechia_payroll_industry_branch cpib;

-- Číselník jednotek hodnot v tabulce mezd:
SELECT * FROM czechia_payroll_unit cpu;

-- Číselník typů hodnot v tabulce mezd:
```

```

SELECT * FROM czechia_payroll_value_type cpvt;

-- Informace o cenách vybraných potravin za několikaleté období:
SELECT * FROM czechia_price cp ;

-- Číselník kategorií potravin, které se vyskytují v našem přehledu
SELECT * FROM czechia_price_category cpc;

-- Číselníky sdílených informací o ČR:

-- Číselník krajů České republiky dle normy CZ-NUTS 2:
SELECT * FROM czechia_region cr;

-- Číselník okresů České republiky dle normy LAU:
SELECT * FROM czechia_district cd;

-- Dodatečné tabulky:

-- Všechné informace o zemích na světě, například hlavní město, měna,
národní jídlo nebo průměrná výška populace:
SELECT * FROM countries c;

-- HDP, GINI, daňová zátěž, atd. pro daný stát a rok:
SELECT * FROM economies e;

```

SQL dotaz pro vytvoření view s HDP pro Českou republiku v porovnatelných letech:

```

-- vytvoreni pohledu s vyfiltrovaným GDP z tabulky 'economies'
CREATE OR REPLACE VIEW view_GDP AS
SELECT
    country AS stat,
    YEAR,
    GDP AS HDP
FROM economies e
WHERE country = 'Czech republic'
    AND `year` BETWEEN '2006' AND '2018'

```

```
ORDER BY YEAR ASC
```

```
;
```

```
-- nacteni pohledu GDP
```

```
SELECT * FROM view_GDP
```

```
;
```

SQL dotaz pro získání tabulky s robustním datovým podkladem, která bude potřeba pro zodpovězení následujících výzkumných otázek:

```
CREATE OR REPLACE TABLE t_radim_jedlicka_project_sql_primary AS
```

```
SELECT
```

```
    cpay.payroll_year AS rok,  
    cpib.name AS odvetvi,  
    cpc.code AS kod_potraviny,  
    cpc.name AS jmeno_potraviny,  
    cp.value AS cena_potraviny,  
    round(avg(cp.value), 2) AS prumerna_cena,  
    cpc.price_value AS mnozstvi,  
    cpc.price_unit AS jednotka,  
    cpay.value AS prumerna_mzda,  
    VIEW_gdp.HDP
```

```
FROM czechia_price AS cp
```

```
JOIN czechia_payroll AS cpay
```

```
    ON YEAR(cp.date_from) = cpay.payroll_year AND
```

```
    cpay.value_type_code = 5958
```

```
    AND cp.region_code IS NULL
```

```
JOIN czechia_price_category cpc
```

```
    ON cp.category_code = cpc.code
```

```
JOIN czechia_payroll_industry_branch cpib
```

```
    ON cpay.industry_branch_code = cpib.code
```

```
JOIN view_gdp
```

```
    ON cpay.payroll_year = view_gdp.year
```

```
GROUP BY `odvetvi`, `rok`, `jmeno_potraviny`
```

```
ORDER BY `rok`, `odvetvi`
```

```
;
```

```
SELECT * FROM t_radim_jedlicka_project_sql_primary
```

;

Výzkumná otázka č. 1: Rostou v průběhu let mzdy ve všech odvětvích, nebo v některých klesají?

```
CREATE OR REPLACE VIEW v_task1 AS
SELECT
    rok,
    odvetvi,
    prumerna_mzda,
    lag(prumerna_mzda) OVER (PARTITION BY odvetvi ORDER BY rok) AS
predchozi_rok,
    (prumerna_mzda) - lag(prumerna_mzda) OVER (PARTITION BY odvetvi ORDER
BY rok) AS rozdil,
    CASE
        WHEN (prumerna_mzda) - lag(prumerna_mzda) OVER (PARTITION BY
odvetvi ORDER BY rok) > 0
        THEN 1
        ELSE 0
    END AS rustXpokles_mzdy
FROM t_radim_jedlicka_project_sql_primary
GROUP BY odvetvi, rok
;

SELECT * FROM v_task1
WHERE rozdil IS NOT NULL
;
```

Poslední sloupec tohoto view ukazuje, že mzdy všech odvětví většinu let rostou (hodnota 1), avšak žádné odvětví se nesetkalo s trvalým nárůstem. Nejčastějšími roky, kdy mzdy klesly, jsou roky 2009 (světová hospodářská krize), dále roky 2012 a 2013.

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

```

SELECT
    rok,
    odvetvi,
    jmeno_potraviny,
    prumerna_cena,
    prumerna_mzda,
    concat(round(prumerna_mzda / prumerna_cena), ' ', jednotka) AS
'pocet_litru/kilogramu_za_prumerny_plat'
FROM t_radim_jedlicka_project_sql_primary
WHERE (jmeno_potraviny = 'Mléko polotučné pasterované' OR jmeno_potraviny
= 'Chléb konzumní kmínový')
    AND (rok = '2006'
    OR rok = '2018')
ORDER BY jmeno_potraviny, odvetvi

```

Poslední sloupec ukazuje množství kilogramů chleba nebo litrů mléka, které je možno si koupit za první a poslední rok srovnávaného období pro každé odvětví průmyslu.

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

```

CREATE OR REPLACE VIEW v_task3 AS
SELECT
    DISTINCT rok,
    jmeno_potraviny,
    prumerna_cena
FROM t_radim_jedlicka_project_sql_primary
ORDER BY jmeno_potraviny, rok

SELECT *,
    lag(prumerna_cena) OVER (PARTITION BY jmeno_potraviny ORDER BY rok)
AS 'predchozi_rok',
    concat(
        round( prumerna_cena / (lag(prumerna_cena) OVER (PARTITION BY
jmeno_potraviny ORDER BY rok) ) * 100 - 100, 2), ' %'
    )
    AS 'cenovy_rozdil'
FROM v_task3

```

```
ORDER BY jmeno_potraviny, rok
;
```

Poslední sloupec výsledné tabulky udává procentuální nárůst ceny potravin oproti předchozímu roku. Přehlednější informace o trendu růstu by nám dala grafická prezentace, nicméně letným pohledem je vidět, že Banány žluté vykazují generelně spíše pokles meziroční ceny. Také procentuální rozdíl oproti prvnímu a poslednímu roku se zdá být nejnižší.

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

```
CREATE OR REPLACE VIEW v_task4_mzdy_procenta AS
SELECT
    rok,
    odvetvi,
    prumerna_mzda,
    lag(prumerna_mzda) OVER (PARTITION BY odvetvi ORDER BY rok) AS
predchozi_rok,
    (prumerna_mzda) - lag(prumerna_mzda) OVER (PARTITION BY odvetvi ORDER
BY rok) AS rozdil,
    round(((prumerna_mzda) / lag(prumerna_mzda) OVER (PARTITION BY
odvetvi ORDER BY rok) * 100 - 100), 2) AS rozdil_mzdy_v_procentech
FROM t_radim_jedlicka_project_sql_primary
GROUP BY odvetvi, rok
;
```

```
CREATE OR REPLACE VIEW v_task4_ceny_procenta AS
SELECT *,
    lag(prumerna_cena) OVER (PARTITION BY jmeno_potraviny ORDER BY rok)
AS 'predchozi_rok',
    round((lag(prumerna_cena) OVER (PARTITION BY jmeno_potraviny ORDER BY
rok) ) / prumerna_cena * 100 - 100, 2)
    AS cenovy_rozdil_procenta
FROM v_task3
;
```

```
SELECT
    v_task4_mzdy_procenta.rok,
```

```

odvetvi,
rozdil_mzdy_v_procentech,
jmeno_potraviny,
cenovy_rozdil_procenta,
CASE
    WHEN cenovy_rozdil_procenta - rozdil_mzdy_v_procentech > 10
    THEN 1
    ELSE 0
END AS vysoky_narust_cen_potravin
FROM v_task4_mzdy_procenta
JOIN v_task4_ceny_procenta
    ON v_task4_mzdy_procenta.rok = v_task4_ceny_procenta.rok
WHERE cenovy_rozdil_procenta IS NOT NULL
    AND sign(cenovy_rozdil_procenta) != -1    # odfiltrovat potraviny, které
zlevnily
ORDER BY vysoky_narust_cen_potravin DESC, rok
;

```

Podle posledního sloupce výsledné tabulky vidíme, které potraviny měly vyšší nárůst cen (hodnota 1) v daném odvětví.

Výzkumná otázka č. 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?

```

CREATE OR REPLACE VIEW v_task5_HDP_procenta AS
SELECT
    rok,
    odvetvi,
    HDP,
    lag(HDP) OVER (PARTITION BY odvetvi ORDER BY rok) AS predchozi_rok,
    round(HDP / (lag(HDP) OVER (PARTITION BY odvetvi ORDER BY rok)) * 100
- 100, 2)
    AS HDP_rozdil_procenta
FROM t_radim_jedlicka_project_sql_primary
GROUP BY odvetvi, rok
;

```

```

CREATE OR REPLACE VIEW v_task5 AS
SELECT
    v_task5_HDP_procenta.rok,
    v_task5_HDP_procenta.odvetvi,
    HDP_rozdil_procenta,
    lag(HDP_rozdil_procenta) OVER (PARTITION BY odvetvi ORDER BY rok) AS
'predchozi_rok_HDP',
--    cenovy_rozdil_procenta,
    rozdil_mzdy_v_procentech,
    CASE
        WHEN rozdil_mzdy_v_procentech > HDP_rozdil_procenta
        THEN 1
        ELSE 0
    END rychlejsi_rust_mezd_stejny_rok
FROM v_task5_HDP_procenta
JOIN v_task4_ceny_procenta
    ON v_task5_HDP_procenta.rok = v_task4_ceny_procenta.rok
JOIN v_task4_mzdy_procenta
    ON v_task5_HDP_procenta.rok = v_task4_mzdy_procenta.rok
WHERE HDP_rozdil_procenta IS NOT NULL
GROUP BY odvetvi, rok
;

SELECT *,
    CASE
        WHEN rozdil_mzdy_v_procentech > predchozi_rok_HDP
        THEN 1
        ELSE 0
    END rychlejsi_rust_mezd_minuly_rok
FROM v_task5
;

```

Podle výsledků v posledních dvou sloupcích vidíme, že mzdy rostly generelně rychleji (hodnota 1) než HDP za stejný nebo uplynulý rok.

Dodatečný materiál: tabulka s HDP, GINI koeficientem a populací dalších evropských států ve stejném období, jako primární přehled pro ČR.


```
SELECT
    country,
    `year`,
    GDP,
    population,
    gini
FROM economies e
WHERE country
    IN (SELECT country
        FROM countries c
        WHERE continent = 'Europe'
    )
    AND YEAR BETWEEN '2006' AND '2018'
ORDER BY country, `year`
;
```

KONEC REPORTU