**PROJEKT SQL\_FARADAY MARTIN**

**Úvod**

V rámci zadání projektu mám za úkol zpracovat / odpovědět na pár definovaných výzkumných otázek, které adresují dostupnost základních potravin široké veřejnosti. V zadání mám již definované základní otázky, na které se pokusím odpovědět. Výsledky prezentuji v následujících přílohách.

1. Připravil jsem robustní datové podklady, se kterých je možné vidět porovnání dostupnosti potravin na základě průměrných příjmů za určité časové období.
2. Jako dodatečný materiál jsem připravil i tabulku s HDP, GINI koeficientem a populací dalších evropských států ve stejném období, jako primární přehled pro ČR.

Otázky pro získání informací vyplývající ze zadání projektu:

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 percentuálně 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?

**POSTUP**

1. Vytvoření 1 robustní tabulky > t\_Martin\_Faraday\_project\_SQL\_primary\_final
2. Nejprve jsem si musel vytvořit dva dotazy pro spojení tabulek.

Dotaz 1 –

**SELECT** \*

**FROM** czechia\_payroll *cp*

**LEFT** **JOIN** czechia\_payroll\_calculation *cpc*

**ON** *cp*.calculation\_code = *cpc*.code

**LEFT** **JOIN** czechia\_payroll\_industry\_branch *cpib*

**ON** *cp*.industry\_branch\_code = *cpib*.code

**LEFT** **JOIN** czechia\_payroll\_unit *cpu*

**ON** *cp*.unit\_code = *cpu*.code

**LEFT** **JOIN** czechia\_payroll\_value\_type *cpvt*

**ON** *cp*.value\_type\_code = *cpvt*.code

**JOIN** czechia\_price *cp2*

**ON** **YEAR** (*cp2*.date\_from) = *cp*.payroll\_year

;

Dotaz 2 –

**SELECT**

*cp*.id,

*cp*.value ,

*cp*.category\_code,

*cp*.date\_from,

*cp*.date\_to,

*cp*.region\_code,

*cr*.name

**FROM** czechia\_price *cp*

**JOIN** czechia\_region *cr*

**ON** *cp*.region\_code = *cr*.code

;

Jen, že nedařilo se mi spojit obě množiny do sebe.

Snažil jsem se využít dotazu :

**WITH** regions **AS** (

**SELECT**

cp1.id **AS** cp1\_ID,

cp1.value ,

cp1.category\_code,

cp1.date\_from,

cp1.date\_to,

cp1.region\_code,

cr.name

**FROM** czechia\_price cp1

**JOIN** czechia\_region cr

**ON** cp1.region\_code = cr.code

)

**SELECT** \*

**FROM** ***regions***

.

Po neúspěších, a několika hodinách zkoušení a procházení podkladů jsem se rozhodl na tvrdo část dotazu vložit přímo do první množiny.

Ve výsledku pro zpracování hlavní tabulky jsem použil tento dotaz:

**SELECT** \*

**FROM** czechia\_payroll cp

**LEFT** **JOIN** czechia\_payroll\_calculation cpc

**ON** cp.calculation\_code = cpc.code

**LEFT** **JOIN** czechia\_payroll\_industry\_branch cpib

**ON** cp.industry\_branch\_code = cpib.code

**LEFT** **JOIN** czechia\_payroll\_unit cpu

**ON** cp.unit\_code = cpu.code

**LEFT** **JOIN** czechia\_payroll\_value\_type cpvt

**ON** cp.value\_type\_code = cpvt.code

**LEFT** **JOIN** czechia\_price cp2

**ON** **YEAR** (cp2.date\_from) = cp.payroll\_year

**LEFT** **JOIN** czechia\_region cr

**ON** cp2.region\_code = cr.code

;

A následně tento dotaz jsem přizpůsobil tak, aby duplikované sloupce, se v hlavní tabulce nezobrazovali. Dále jsem duplikované názvy sloupců, jsem ještě přejmenoval, tak aby šla vytvořit zcela nová tabulka.

Stejně jsem musel vložit WHERE klauzule abych optimalizoval čas vytvoření tabulky.

Viz. Podrobnosti v SQL příkazech.

Zároveň tabulka czechia\_district není v hlavní tabulce. Jelikož podle mého názoru nemá logiku aby v ní byla. Není na co ji napojit.

1. Vytvoření 2 robustní tabulky > t\_Martin\_Faraday\_ project\_SQL\_secondary\_final

Při spojování obou tabulek economies a countries, jsem šel stejnou cestou jako v bodu č. 1.

**CREATE** **TABLE** t\_Martin\_Faraday\_project\_SQL\_secondary\_final **AS**

**SELECT**

c.country, c.abbreviation, c.avg\_height, c.calling\_code, c.capital\_city,

c.continent, c.currency\_name, c.religion, c.currency\_code, c.domain\_tld,

c.elevation, c.north, c.south, c.west, c.east, c.government\_type, c.independence\_date,

c.iso\_numeric, c.landlocked, c.life\_expectancy, c.national\_symbol, c.national\_dish,

c.population\_density, c.population, c.region\_in\_world, c.surface\_area,

c.yearly\_average\_temperature, c.median\_age\_2018, c.iso2, c.iso3, e.**year**,

e.GDP, e.population **AS** population\_economist, e.gini, e.taxes, e.fertility, e.mortaliy\_under5

**FROM** countries c

**JOIN** economies e

**ON** c.country = e.country

;

Viz. SQL t\_Martin\_Faraday\_ project\_SQL\_secondary\_final

V jedniná věc, kterou jsem ošetřil byl zdablovaný sloupec polupace.

**ZPRACOVÁNÍ ÚKOLŮ**

task\_1

Otázka *„Kolik je možné si koupit litrů mléka a kilogramů chleba za první a poslední srovnatelné období v dostupných datech cen a mezd?“* nebyla přesně jednoznačná. Pochopil jsem to tak, že máme zjistit kolik produktů si průměrný občan ČR mohl koupit z průměrné mzdy za období 2006 a 2018.

* Nejprve jsem si vyselektoval průměrné ceny za mléko a chleba za období roku 2006 a 2018,
* Následně jsem potřeboval zjisit průměrnou mzdu zaměstnance ve výše vedeném období,
* Dlouho jsem přemýšlel jak vydělit průměrnou cenu produktů vůči průměrné mzdě,

Výsledný postup, který směřoval k výsledku viz SQL task\_1

Task\_2

Otázka „Která kategorie potravin zdražuje nejpomaleji (je u ní nejnižší percentuální meziroční nárůst)?“ nebyla jednoznačná, co máme hledat. Pokud jsem dobře pochopil, pak jsem nejprve srovnal jednotlivé kategorie produktů s průměrnými cenami v návaznosti na roky. Následně porovnal průměrné ceny v letech oproti předchozímu roku.

Z výsledku a srovnání vychází, že nejmenší cenový nárust měl produkt Rajská jablka červená a to v roce 20007, kdy cena se snížila o cca 30,28%

Více v SQL task\_2.

Task\_3

Otázka „Rostou v průběhu let mzdy ve všech odvětvích, nebo v některých klesají?“ . Po vyselektování průměrných mezd za jednotlivé odvětví v návaznosti na roky jsem následně zkusil doupřesnit dotaz tak abych viděl jen odvětví, u kterých je procentuální nárust menší než 0.

Z výsledku vychází, že celkem 23 odvětví v návaznosti na roky, měli záporný nárust mezd než v ostatních letech. Z tohoto pohledu ne všechna odvětví v průběhu let mají nárust mezd, ba naopak.

Více v SQL task\_3.

Task\_4

Otázka „Existuje rok, ve kterém byl meziroční nárůst cen potravin výrazně vyšší než růst mezd (větší než 10 %)?“ . Po vyselektování průměrných mezd a průměrných cen potravin v návaznosti na roky jsem následně zkusil doupřesnit dotaz tak abych viděl roky, u kterých je procentuální nárust větší než 1.

Z výsledku vychází, že není žádný rok u kterého by se projevil nárust cen potravin větší než 10% oproti nejvyší průměrnému nárustu mzdy v daném roce.

Více v SQL task\_4.

Task\_5

Otázka „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 násdujícím roce výraznějším růstem?“ . Po vyselektování průměrných mezd ,průměrných cen potravin a procentní rozdíl HDP v návaznosti na roky jsem došel k názoru, že vliv růstu HDP na cenovou hladinu potravin a průměrné mzdy má. Přesto pokud HDP neroste, ale je v záporném růstu, na cenovou hladinu potravin nebo mezd nemusí mít vliv.

Pokud bych měl shrnout tuto otázku, pak ve většině případu vliv HDP na výše sledované průměrné ceny a mzdy má.

Tento dotaz zpracoval jak jsem nejlépe mohl.

Více v SQL task\_5.