

## Primární projekt SQL

### Odpovědi na výzkumné otázky 1 - 5

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

**Odpověď:** Z dlouhodobého hlediska rostou mzdy ve všech odvětvích, ale z krátkodobého hlediska v některých odvětvích v průběhu let mzdy klesají. Nejčastěji v oboru Těžby a dobývání.

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

**Odpověď:** Za průměrnou mzdu v roce 2006 lze koupit 1 213,4 Kg Chleba nebo 1 356,7 litrů Mléka polotučného pasteurovaného. Za průměrnou mzdu v roce 2018 lze zakoupit 1 324,1 Kg chleba nebo 1618,3 litrů Mléka polotučného pasteurovaného.

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

**Odpověď:** Nejpomaleji mezi rokem 2006 a 2018 zdražuje Cukr krystalový, který za tuto dobu zlevnil o 37,3 %.

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

**Odpověď:** NE, nejvyšší meziroční nárůst cen potravin oproti nárůstu mezd byl v roce 2013 o 5,4 %.

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 následujícím roce výraznějším růstem?

**Odpověď:** Není tomu tak vždy. Chybí data z roku 2019 pro tabulku cen potravin. Kdybychom měli tato data a ukázalo by se, že mezi rokem 2018 - 2019 byl nárůst ceny vyšší než 4 % tak dle vypočtených dat vychází: Pokud HDP vzrostlo o více jak 4 %, tak nárůst cen potravin i mezd narostl ve stejném a zároveň v následujícím roce o více jak 4 % přesně ve dvou ze tří případů. S daty, které máme, můžeme říci, že pokud HDP vzrostlo o více jak 4 %, tak o více jak 4 % vzrostly mzdy i ceny potravin ve dvou ze tří případů. V následujícím roce mzdy ve dvou ze tří případů a ceny potravin v jednom ze tří případů.

### Průvodní listina

#### Vytvoření výsledné tabulky „t\_petr\_luka\_project\_SQL\_primary\_final“

Výslednou tabulku „t\_petr\_luka\_project\_SQL\_primary\_final“ ze které jsem získával data pro otázky č.1 – 5 jsem vytvořil pomocí dvou VIEW.

a) První VIEW „joining\_payroll\_tables“ vzniklo spojením tabulek:

- czechia\_payroll
- czechia\_payroll\_calculation
- czechia\_payroll\_industry\_branch
- czechia\_payroll\_unit

Rovnou v „SELECTu“ jsem určil Aliasy pro dané sloupce, aby se s nimi lépe pracovalo.

Taktéž jsem udělal i pro následující VIEW, Tabulku, CTE, vnořené SELECTY a ostatní skupiny příkazů. Jelikož zde byly některé názvy sloupců mezi tabulkami stejné, musel jsem za názvy sloupců přidávat názvy tabulek. Z kategorie mezd, které nebyly pojmenované neboli znamenaly průměr, jsem přejmenoval pomocí „CASE“ na „Průměrná hodnota“.

Zároveň jsem zprůměroval mzdu (kterou jsem zaokrouhlil na jedno desetinné místo), protože mzdy byly měřeny kvartálně a já potřeboval průměrnou mzdu za celý rok. To jsem následně provedl na konci tabulky pomocí funkce „GROUP BY“, která mi seskupila dané řádky do sebe dle roku a kategorie.

Rovnou jsem si také přejmenoval zprůměrované hodnoty, které již v tabulce byly ve sloupci `industry_branch` jako nepojmenované na název „Průměrná hodnota“, aby se mi s hodnotami lépe pracovalo.

Poté jsem postupně spojoval tabulky do sebe, kde ve většině případech stačilo použít `INNER JOIN`, ale v tabulce „`czechia_payroll_industry_branch`“ jsem musel použít `LEFT JOIN`, protože jsem potřeboval získat i výše zmíněné `NULL` hodnoty (nepojmenované průměrné hodnoty).

Abych následně získal jen hodnoty, se kterými budu dále pracovat, tak jsem pomocí funkce `WHERE` vyfiltroval hodnoty v Kč, protože jsem potřeboval pouze mzdy a zároveň jsem vyfiltroval hodnoty, kde se jednalo o přepočítanou mzdu, protože se více hodí pro statistické výpočty.

b) Druhé VIEW „`joining_price_tables`“ vzniklo spojením tabulek:

- `czechia_price`
- `czechia_price_category`

Zde jsem si hned se začátku pomocí „`YEAR`“ transformoval datum ze sloupce „`date_from`“ na samotný rok, protože následné spojování tabulek je právě přes samotné roky.

Následně jsem zprůměroval a zaokrouhlil na jednu desetinnou čárku sloupec „`value`“ s cenou kvůli tomu, že jsem ho ke konci seskupoval pomocí „`GROUP BY`“.

Poté jsem spojil tabulku „`czechia_price`“ s „`czechia_price_category`“, abych získal názvy kategorií a vyfiltroval jsem data pomocí „`WHERE`“, kde ve sloupci „`region_code`“ byli jen hodnoty „`NULL`“ protože tyto hodnoty znamenaly celkové průměry, se kterými jsem následně počítal.

Jak jsem již zmínil, tak jsem poté data seskupil podle sloupců a podle jmen a následně jsem je vzestupně seřadil, abych s nimi mohl dále pracovat.

Tabulku „`czechia_region`“ jsem nespojoval, protože to nebylo nutné. Jinak bych na spojení použil „`LEFT JOIN`“, aby se mi ze sloupce „`region_code`“ neodstranily hodnoty „`NULL`“.

### Výsledná tabulka „`t_petr_luka_project_SQL_primary_final`“

Tuto tabulku jsem složil ze dvou CTE, které byly napojené na obě výše zmíněné VIEW.

- a) Nejdříve jsem si vytvořil první CTE s názvem „`price_group`“, kde jsem potřeboval získat z prvního „VIEW - `joining_payroll_tables`“ rozdíl ceny oproti předchozímu roku a s tím vytvořit nový sloupec „`price_value_diff`“, který byl rozdílem hodnot sloupce „`price_value`“ a hodnotami ze sloupce „`price_value`“ o jeden řádek níže. Udělal jsem to pomocí „`CASE`“, kde jsem potřeboval, aby hodnoty roku 2006 a hodnoty `NULL` z ceny byly nula, aby následné výpočty nezkreslovaly hodnoty z roku 2006 (případně jiné) a poté jsem použil „`LAG`“, který mi posunul řádek ve sloupci o jeden řádek níže pro pozdější výpočty.
  - b) Vytvoření druhého CTE s názvem „`payroll_group`“ kde jsem potřeboval udělat u druhého „VIEW“ to samé jako u prvního „VIEW“ s tím rozdílem, že zde jsem nepotřeboval, aby hodnoty z roku 2006 byly nulové, protože „`czechia_payroll`“ má hodnoty od roku 2000 a pro porovnání v daných úkolech jsem potřeboval především data od roku 2006 – 2018.
  - c) Poté jsem pomocí „`JOIN`“ spojil obě CTE („`INNER JOIN`“ proto, abych měl pouze společné roky, jak bylo požadováno v zadání. Jinak bych použil „`LEFT JOIN`“ z `payroll_group` na `price_group`, aby mi zůstala data z předchozích i následujících let, které `price_group` nemá). Zároveň jsem zde potřeboval vytvořit procentuální rozdíl mezi roky pro ceny potravin jako sloupec „`price_value_percent`“ a pro mzdy jako sloupec „`payroll_value_percent`“.
- Nakonec jsem vše seskupil podle roku, názvu kategorie mezd a názvu potravin.

## Vytvoření výsledné tabulky „t\_petr\_luka\_project\_SQL\_secondary\_final“

Výslednou tabulku „t\_petr\_luka\_project\_SQL\_secondary\_final“ jsem vytvořil pomocí „LEFT JOIN“ tabulky „countries“ s tabulkou „economies“. Spojení „LEFT JOIN“ jsem zvolil proto, aby se daly případně zjistit státy, pro které máme data v „economies“, ale nejsou v „countries“.

Následně jsem pro tabulku použil filtr pomocí „WHERE“, abych získal pouze data Evropských států a zároveň jsem použil filtr, abych získal data mezi roky 2006 a 2018 (jak bylo požadováno v zadání).

Nakonec jsem data seřadil vzestupně podle roku a státu.

## Vytvoření dotazů pro jednotlivé otázky

1. **Otázka:** Zde jsem pro dosažení odpovědi zvolil „Vnořený SELECT“.

Abych získal správnou četnost klesajících hodnot, tak jsem musel nejdříve seskupit kategorii mezd a roky, a poté jsem použil funkci „COUNT“ pro počet seskupených řádků.

Následně jsem potřeboval zobrazit hodnoty z rozdílu mezd mezi roky, které byly v záporných hodnotách neboli „pokles mezi roky“ pomocí „WHERE“, kde jsem zároveň odstranil řádky „Průměrná hodnota“ z kategorie mezd, protože jsem je potřeboval zobrazit zvlášť.

Poté jsem seskupil kategorie mezd záporných hodnot a seřadil počty sestupně a zároveň kategorie mezd vzestupně.

Níže je výsledek dotazu pro otázku č.1

A-Z payroll_industry_branch	123 payroll_count
Těžba a dobývání	4
Výroba a rozvod elektřiny, plynu, tepla a klimatiz. vzduchu	3
Činnosti v oblasti nemovitostí	2
Profesní, vědecké a technické činnosti	2
Ubytování, stravování a pohostinství	2
Veřejná správa a obrana; povinné sociální zabezpečení	2
Administrativní a podpůrné činnosti	1
Informační a komunikační činnosti	1
Kulturní, zábavní a rekreační činnosti	1
Peněžnictví a pojišťovnictví	1
Stavebnictví	1
Velkoobchod a maloobchod; opravy a údržba motorových vozidel	1
Vzdělávání	1
Zásobování vodou; činnosti související s odpady a sanacemi	1
Zemědělství, lesnictví, rybářství	1

2. **Otázka:** Pro tuto otázku jsem potřeboval nejdříve zjistit průměrné hodnoty rozdílu mezi mzdami a cenami potravin. Tím jsem získal sloupec „count\_value“, který mi zobrazoval počet, kolik si mohu koupit dané potraviny za danou mzdu.

Následně jsem použil filtr pro „Mléko polotučné pasterované“ a „Chléb konzumní kmínový“ na které se v otázce ptali a zároveň jsem také použil filtr pro průměrné hodnoty z kategorie mezd.

Nakonec jsem seskupil rok a název potraviny.

Níže je výsledek dotazu pro otázku č.2

123 common_year	A-Z price_name	123 count_value	A-Z price_unit
2,006	Chléb konzumní kmínový	1,213.4	kg
2,006	Mléko polotučné pasterované	1,356.7	l
2,018	Chléb konzumní kmínový	1,324.1	kg
2,018	Mléko polotučné pasterované	1,618.3	l

3. **Otázka:** Zde jsem si nejdříve vytvořil dvě CTE, abych získal ceny potravin z roku 2006 a ceny potravin z roku 2018.

Následně jsem obě CTE spojil a poté použil vzoreček pro výpočet procent rozdílu mezi cenou z roku 2006 a 2018 a pojmenoval sloupeček jako „price\_perc“.

Vše jsem seřadil vzestupně a nakonec jsem použil „LIMIT“, abych viděl jen hodnotu nejmenšího nárustu ceny potravin.

Níže je výsledek dotazu pro otázku č.3

A:2 name_2018	123 price_perc
Cukr krystalový	-37.3

4. **Otázka:** Nejdříve jsem si vytvořil sloupeček s názvem „diff\_value\_percent“, který vznikl rozdílem mezi procentuálním nárustem cen potravin a procentuálním nárustem mezd. Nakonec jsem seskupil roky a výsledné hodnoty seřadil sestupně a omezil na jeden řádek, abych viděl jen nejvyšší nárůst cen potravin.

Níže je výsledek dotazu pro otázku č.4

123 common_year	123 diff_value_percent
2,013	5.4

5. **Otázka:** Pro odpověď na tuto otázku jsem musel nejdříve:

- použít CTE „gdp\_join“ pro vytvoření sloupce „gdp\_value\_percent“, který zobrazoval procentuální meziroční růst HDP a zároveň v této tabulce „economies“ zobrazit pomocí WHERE pouze data pro Českou republiku.
- Poté jsem vytvořil druhé CTE „price\_and\_payroll“, kde jsem potřeboval vytvořit sloupečky pro cenu potravin a pro mzdy, kde bude řádek o jeden výše pomocí LEAD, pro pozdější výpočty. Následně jsem data seskupil podle roku a seřadil vzestupně.
- Následně jsem spojil obě CTE, pomocí CASE vytvořil sloupeček „same\_year\_differences“, kde jsem potřeboval znázornit, zda pokud je nárůst HDP vyšší než 4 %, tak jsou ve stejném roce o 4 % vyšší mzdy, ceny potravin, obě hodnoty najednou anebo ani jedna. To samé jsem potřeboval udělat i pro předchozí rok, pro sloupec „previous\_year\_differences“ kde jsem využil výše zmíněný sloupec s použitím „LEAD“. Poté jsem použil filtr pro rok mezi 2007 až 2018 a zároveň kde bylo HDP větší než 4 %. To jsem seskupil podle roku a seřadil roky vzestupně.

Níže je výsledek dotazu pro otázku č.2

123 common_year	A:2 same_year_differences	A:2 previous_year_differences
2,007	Payroll and Price are Higer	Payroll and Price are Higer
2,015	Has NO effect	Has NO effect
2,017	Payroll and Price are Higer	Payroll is Higher