

# Průvodní listina

## SQL projektu kurzu Engeto

### Zadání 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?
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?

## 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).

## Postup vypracování SQL tabulek a SQL dotazů

---

Projekt byl vypracován v aplikaci DBeaver v 24.1.2 z dat lokální databáze. Tato databáze byla stažena z kurzu datové analýzy Engeto.

Jako hlavní datový podklad většiny výzkumných otázek byly dle zadání připraveny 2 tabulky:

- Primární tabulka „`t_arnost_krizan_project_sql_primary_final`“ obsahuje data potravin, jejich ceny a průměrné mzdy v letech.
- Sekundární tabulka „`t_arnost_krizan_project_sql_secondary_final`“ obsahuje data názvů jednotlivých světových zemí a hodnotu jejich HDP v letech.

Pro sestavení tabulek bylo nutné najít společný klíč, pomocí kterého bylo možné seskupit data z různých tabulek. Některá data buď chyběla, nebo bylo nutné určit, kdy jsou hodnoty správné, či nikoliv. Například pro dosažení správných hodnot mzdy bylo nutné opravit dotaz k tabulce `czechia_payroll`. Hodnota mzdy se nachází ve sloupci „`value`“. Část neplatných řádků mzdy byly navíc prázdné. Hodnota mzdy platí pouze v případě, pokud parametr `value_type_code` nabýval hodnoty 5958 (Průměrná hrubá mzda na zaměstnance) oproti druhé nežádané možnosti 316 (Průměrný počet zaměstnaných osob) a zároveň parametr „`unit_code`“ je rovno hodnotě 100 (fyzický) proti nežádoucí druhé možnosti 200 (přepočtený).

## Odpovědi na výzkumné otázky

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

**V průběhu let rostou mzdy ve všech odvětvích s výjimečnými mírnými poklesy proti předchozím rokům, především v roce 2013.**

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

**Odpověď byla vypracována do následující tabulky. Hodnoty množství byly zaokrouhleny na celá čísla.**

Potravina	Rok 2006	Rok 2018
Mléko [l]	1 444	1 667
Chléb konzumní kmínový [kg]	1 197	1 314

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

**Z kategorie potravin zdražuje v průběhu let nejméně „Cukr krystalový“, který má průměrnou hodnotu meziročního změny v letech 2006 až 2017 dokonce zápornou -0,21 %. Následují „Banány žluté“, které meziročně zdražují ve stejném časovém rozmezí s průměrnou hodnotou 0,71 %.**

**V případě kategorie vína neexistují data od roku 2006 do 2014 včetně. Data od 2015 – 2018 existují.**

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

**V roce 2013 byl meziroční nárůst cen potravin vyšší než 10 % oproti růstu mezd.**

**Potraviny zdražily o 31,7 % a průměrná mzda o pouhých 2,6 %.**

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?

Výsledky byly zpracovány SQL dotazem. Hodnoty byly přeneseny do následujících tabulek.

Z hodnot není zcela patrný jednoznačný vztah HDP (GDP) se společně kladnou nebo zápornou změnou mezd (Payroll) a potravin (Food) v prvním [1-y(%)] nebo druhém [2-y(%)] následujícím roce.

Po aplikaci barevného formátování v následujících tabulkách (tabulka pouze kladných nebo záporných hodnot) lze vzájemný vliv pozorovat mnohem lépe. Z dat vyplývá, častější souvislost růstu HDP s růstem mezd i cen potravin. V případě záporné změny HDP na záporný růst platů a cen potravin nelze souvislost potvrdit z důvodu ojedinělosti hodnot.

Tabulka: Hodnoty získané SQL dotazem

Year	GDP_1-y(%)	Food_1-y(%)	Food_2-y(%)	Payroll_1-y(%)	Payroll_2-y(%)
2006	-	-	-	-	-
2007	5,6	12,7	-	6,8	-
2008	2,7	0,1	12,8	8,1	15,4
2009	-4,7	-8,7	-8,6	3,3	11,6
2010	2,4	7,8	-1,5	2	5,3
2011	1,8	4,4	12,6	2,3	4,3
2012	-0,8	2,7	7,3	3,1	5,5
2013	0	8,5	11,5	-1,5	1,6
2014	2,3	-1,4	7	2,6	1
2015	5,4	-0,5	-1,9	2,5	5,1
2016	2,5	5,3	4,8	3,7	6,2
2017	5,2	8,2	14	6,4	10,4
2018	3,2	-0,1	8,2	7,6	14,5

**Tabulka: Pouze kladné hodnoty**

Year	GDP_1-y(%)	Food_1-y(%)	Food_2-y(%)	Payroll_1-y(%)	Payroll_2-y(%)
2006	5,6	12,7		6,8	
2007	2,7	0,1	12,8	8,1	15,4
2008				3,3	11,6
2009	2,4	7,8		2	5,3
2010	1,8	4,4	12,6	2,3	4,3
2011		2,7	7,3	3,1	5,5
2012		8,5	11,5		1,6
2013	2,3		7	2,6	1
2014	5,4			2,5	5,1
2015	2,5	5,3	4,8	3,7	6,2
2016	5,2	8,2	14	6,4	10,4
2017	3,2		8,2	7,6	14,5
2018					

**Tabulka: Pouze záporné hodnoty**

Year	GDP_1-y(%)	Food_1-y(%)	Food_2-y(%)	Payroll_1-y(%)	Payroll_2-y(%)
2006	-4,7	-8,7	-8,6		
2007			-1,5		
2008					
2009	-0,8				
2010				-1,5	
2011		-1,4			
2012		-0,5	-1,9		
2013					
2014					
2015		-0,1			
2016					
2017					
2018					