

# SQL project Engeto - průvodní listina

Michaela Segers

## Tvorba první tabulky t\_michaela\_segers\_SQL\_primary\_final

Do tabulky budu chtít vložit následující sloupce, které potřebuji znát pro zodpovězení zadaných otázek.

Z tabulky payroll:

- rok
- value (průměrná mzda)
- industry branch (odvětví)

Z tabulky prices:

- rok
- value (průměrná cena)
- category code (produkt)

Všechna data budu zjišťovat za společné období obou tabulek, tedy 2006-2018, toto ale nijak v této fázi nefiltruji, roky se samy takto definují, až později spojím data z obou tabulek pomocí inner join právě přes hodnotu roků.

Prvním skriptem tedy vyberu požadovaná data z tabulky payroll.

[https://github.com/MichaelaSegers/SQL\\_project\\_engeto/blob/14b4feaa868a7139fd07b0c625ad8ae4fe76e4ac/SQL\\_project.sql#L15](https://github.com/MichaelaSegers/SQL_project_engeto/blob/14b4feaa868a7139fd07b0c625ad8ae4fe76e4ac/SQL_project.sql#L15)

Tabulka payroll obsahuje data po kvartálech, potřebuji tedy získat průměrnou mzdu za roky. Počítám s calculation\_code 200 pro mzdy přepočtené na celé úvazky.

Druhým skriptem vyberu požadovaná data z tabulky prices.

[https://github.com/MichaelaSegers/SQL\\_project\\_engeto/blob/14b4feaa868a7139fd07b0c625ad8ae4fe76e4ac/SQL\\_project.sql#L25](https://github.com/MichaelaSegers/SQL_project_engeto/blob/14b4feaa868a7139fd07b0c625ad8ae4fe76e4ac/SQL_project.sql#L25)

Protože v celém projektu pracuji pouze s roky, data seskupím podle roku počátku měření. Budu zodpovídat dotazy za celou republiku bez ohledu na regiony, region\_code tedy volím NULL.

Na základě předchozích skriptů jsem si vytvořila na localhost dočasné tabulky, ze kterých pak vytvořím tabulku finální.

[https://github.com/MichaelaSegers/SQL\\_project\\_engeto/blob/14b4feaa868a7139fd07b0c625ad8ae4fe76e4ac/SQL\\_project.sql#L34](https://github.com/MichaelaSegers/SQL_project_engeto/blob/14b4feaa868a7139fd07b0c625ad8ae4fe76e4ac/SQL_project.sql#L34)

Oproti předchozím selectům jsem ještě přidala jmenné názvy odvětví a produktů.

Tvořím finální tabulku pomocí inner join, abych vyfiltrovala pouze společné roky.

[https://github.com/MichaelaSegers/SQL\\_project\\_engeto/blob/14b4feaa868a7139fd07b0c625ad8ae4fe76e4ac/SQL\\_project.sql#L66](https://github.com/MichaelaSegers/SQL_project_engeto/blob/14b4feaa868a7139fd07b0c625ad8ae4fe76e4ac/SQL_project.sql#L66)

Na Engeto serveru nemám povolení tvořit dočasné tabulky, finální tabulku tedy můžu vytvořit dvěma alternativními způsoby.

Buď přes common table expression, kde 3 kroky z původního způsobu spojím do jednoho.

[https://github.com/MichaelaSegers/SQL\\_project\\_engeto/blob/14b4feaa868a7139fd07b0c625ad8ae4fe76e4ac/SQL\\_project.sql#L74](https://github.com/MichaelaSegers/SQL_project_engeto/blob/14b4feaa868a7139fd07b0c625ad8ae4fe76e4ac/SQL_project.sql#L74)

Anebo pomocné tabulky payroll a prices vytvořím jako standardní, ne dočasné. A po vytvoření finální tabulky je smažu.

[https://github.com/MichaelaSegers/SQL\\_project\\_engeto/blob/14b4feaa868a7139fd07b0c625ad8ae4fe76e4ac/SQL\\_project.sql#L107](https://github.com/MichaelaSegers/SQL_project_engeto/blob/14b4feaa868a7139fd07b0c625ad8ae4fe76e4ac/SQL_project.sql#L107)

## Tvorba druhé tabulky (další země)

V tabulce economies chybí poměrně hodně dat pro GDP a GINI za různé země v různých letech.

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

Potřebuji porovnat mzdy vždy s předchozím rokem. Protože v tabulce mám jako nejstarší rok v tabulce 2006, za tento rok mi bude růst vycházet NULL, protože nemám s čím porovnávat (pouze pokud bych importovala data z původní tabulky payroll v databázi za rok 2005).

U roku 2006 v této otázce proto nehodnotím růst, pouze využiji data o mzdách.

[https://github.com/MichaelaSegers/SQL\\_project\\_engeto/blob/4e04224488debf88f68293e56ada5054467d71b5/SQL\\_project.sql#L207](https://github.com/MichaelaSegers/SQL_project_engeto/blob/4e04224488debf88f68293e56ada5054467d71b5/SQL_project.sql#L207)

V tomto selectu vidím konkrétní meziroční poklesy v jednotlivých odvětvích od největších poklesů mezd.

Můžeme také zhodnotit průměrné meziroční pohyby mezd, tedy pro jednotlivá odvětví za celé období.

[https://github.com/MichaelaSegers/SQL\\_project\\_engeto/blob/4e04224488debf88f68293e56ada5054467d71b5/SQL\\_project.sql#L243](https://github.com/MichaelaSegers/SQL_project_engeto/blob/4e04224488debf88f68293e56ada5054467d71b5/SQL_project.sql#L243)

#### **Odpověď na otázku č.1:**

V průběhu let 2006-2018 došlo k největšímu meziročnímu poklesu mezd v:

- peněžnictví a pojišťovnictví v roce 2013 (-8,83%)

- výrobě a rozvodu elektřiny a plynu, tepla a klimatiz. vzduchu v roce 2013 (-4,44%)
- těžbě a dobývání v roce 2013 (-3,24%)

Naopak největší meziroční růst mezd proběhl v:

- výrobě a rozvodu elektřiny a plynu, tepla a klimatiz. vzduchu v roce 2008 (13,76%)
- těžbě a dobývání v roce 2008 (13,75%)
- profesní, vědecké a technické činnosti v roce 2008 (12,41%)

Za celkové období 2006-2018 je ve všech odvětvích průměrný růst mezd pozitivní. Nejnižší průměrný meziroční procentuální růst zaznamenalo peněžnictví a pojišťovnictví (2,75%), nejvyšší naopak zdravotní a sociální péče (4,95%).

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

Za průměrné roční mzdy napříč odvětvími bylo možné pořídit (zaokrouhleno dolů na celá čísla):

- 1353 l mléka v roce 2006
- 1211 kg chleba v roce 2006
- 1616 l mléka v roce 2018
- 1321 kg chleba v roce 2018

Pro zajímavost je ještě možné se podívat, jak se toto liší v různých odvětvích.

[https://github.com/MichaelaSegers/SQL\\_project\\_engeto/blob/377a89c4c2a42d352c3aa719c3801b465206cecd/SQL\\_project.sql#L285](https://github.com/MichaelaSegers/SQL_project_engeto/blob/377a89c4c2a42d352c3aa719c3801b465206cecd/SQL_project.sql#L285)

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

Data za bílé víno jsou dostupná pouze od roku 2015.

Nejméně zdražila rajská jablka mezi lety 2006-2007 (zlevnění o 30,28%), průměrně nejpomaleji za celé období měření (tedy 2006-2018, ale v případě bílého vína měřeno pouze 2015-2018) zdražil cukr krystalový (dokonce meziročně průměrně zlevňoval o 1,92%).

Pro kontrolu jsem si zobrazila ceny cukru v letech 2006 a 2018.

[https://github.com/MichaelaSegers/SQL\\_project\\_engeto/blob/377a89c4c2a42d352c3aa719c3801b465206cecd/SQL\\_project.sql#L339](https://github.com/MichaelaSegers/SQL_project_engeto/blob/377a89c4c2a42d352c3aa719c3801b465206cecd/SQL_project.sql#L339)

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

Jednotlivé produkty měly meziroční průměrný růst cen často výrazně vyšší než průměrný meziroční růst mezd.

[https://github.com/MichaelaSegers/SQL\\_project\\_engeto/blob/377a89c4c2a42d352c3aa719c3801b465206cecd/SQL\\_project.sql#L373](https://github.com/MichaelaSegers/SQL_project_engeto/blob/377a89c4c2a42d352c3aa719c3801b465206cecd/SQL_project.sql#L373)

Průměrný růst cen všech produktů nebyl nikdy o 10% vyšší než průměrný růst mezd - nejvyšší rozdíl byl v roce 2013 o 6,14%.

[https://github.com/MichaelaSegers/SQL\\_project\\_engeto/blob/377a89c4c2a42d352c3aa719c3801b465206cecd/SQL\\_project.sql#L411](https://github.com/MichaelaSegers/SQL_project_engeto/blob/377a89c4c2a42d352c3aa719c3801b465206cecd/SQL_project.sql#L411)

#### 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?

U této otázky hodně záleží na tom, jak si definujeme výraznější růst mezd, cen i HDP. Vzhledem k tomu, že v rámci projektu toto definováno není, stanovila jsem výraznější růst nad 3% a výraznější pokles pod -3%, hodnoty mezi nimi ponechávám jako neutrální.

[https://github.com/MichaelaSegers/SQL\\_project\\_engeto/blob/377a89c4c2a42d352c3aa719c3801b465206cecd/SQL\\_project.sql#L458](https://github.com/MichaelaSegers/SQL_project_engeto/blob/377a89c4c2a42d352c3aa719c3801b465206cecd/SQL_project.sql#L458)

Významně by pomohla vizualizace dat spíše než je číst z tabulky a podobnosti v pohybech odhadovat.

V mé klasifikaci stouplo HDP výrazně v letech 2007, 2015, 2017 a 2018. HDP výrazně pokleslo v roce 2009.

V roce 2007 se projevil výrazný růst HDP i v růstu mezd a cen, které potom výrazně narostly i v následujícím roce.

Vyššímu nárůstu HDP v letech 2015, 2017 a 2018 odpovídá i signifikantní růst mezd v celém období 2015-2018. Ceny však výrazně vzrostly z tohoto období pouze v roce 2017.

V roce 2009 spolu s HDP významně poklesly i ceny. Ačkoliv v tomto roce mzdy vyrostly dle mého kritéria stále výrazně (o 3,37%), oproti růstu v předchozím roce (o 7,85%) je to značný pokles. Proto bych zhodnotila, že se pokles HDP ukázal i ve mzdách.

---

## FEEDBACK:

- vylepšit commit messages
- rozdělit skript do více souborů dle jednotlivých otázek
- upravit formulaci např. pro případ přidání nových položek:

[https://github.com/MichaelaSegers/SQL\\_project\\_engeto/blob/9e5aae5353aaaf8c74c0f3eb3514afc1c1c1a5e8/SQL\\_project.sql#L40](https://github.com/MichaelaSegers/SQL_project_engeto/blob/9e5aae5353aaaf8c74c0f3eb3514afc1c1c1a5e8/SQL_project.sql#L40)