

# SQL projekt: dostupnost potravin v Česku

Kurz: Engeto – Datová Akademie, 18. 6. 2024 – 3. 9. 2024

Zpracovala: Adéla Prystaszová

## Zadání

Úkolem bylo připravit pomocí SQL z datových sad v Engeto databázi robustní datové podklady, ve kterých bude možné porovnat v Česku dostupnost potravin na základě průměrných příjmů za vhodné časové období (**primární tabulka**). Jako dodatečný materiál měla být za stejné období vytvořena tabulka s HDP, Giniho koeficientem a populací evropských států (**sekundární tabulka**).

Zároveň bylo třeba pomocí dat v primární tabulce zodpovědět **pět výzkumných otázek**:

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?

## Data

### Primární tabulka

Tato tabulka obsahuje všechna data potřebná pro zodpovězení výzkumných otázek a byla vytvořena z dat v následujících tabulkách z Engeto databáze:

- czechia\_payroll (data o mzdách), czechia\_payroll\_industry\_branch (číselník)
- czechia\_price (data o cenách potravin), czechia\_price\_category (číselník)
- economies (data o HDP)

Celkem má vzniklá tabulka 6840 řádků a 8 sloupců:

- Sloupec **odvetvi** obsahuje 19 pracovních odvětví. Je jediným sloupcem v tabulce obsahujícím NULL hodnoty. Ty v něm označují data za všechna odvětví dohromady.
- Co se týče časového období, zdrojová tabulka czechia\_payroll obsahuje průměrné mzdy za jednotlivá čtvrtletí let 2000–2021, zatímco tabulka czechia\_price obsahuje ceny potravin v několikadenních časových úsecích, přičemž tyto úseky spadají vždy jen do jednoho roku, ale ne vždy jen do jednoho čtvrtletí. Tato tabulka pokrývá roky 2006–2018. Tabulka economies pak obsahuje údaje pouze za celé roky, a to 1960–2020. Při vytváření primární tabulky tedy byla využita jen data z období, které je obsaženo ve všech těchto tabulkách a data jsou zároveň rozlišena jen podle let. V primární tabulce se tedy nachází sloupec **rok** obsahující roky 2006–2018.
- Sloupec **prumerna\_mdza** obsahuje průměrnou hrubou měsíční mzdu v daném roce v korunách. Tato mzda byla přepočítána ze čtvrtletních údajů na údaje pro celé roky.

- Sloupec **potravina** obsahuje 27 druhů potravin. Data za každou z těchto potravin jsou připojena na data o každém z pracovních odvětví v každém roce, čímž se v tabulce mnohonásobně zvyšuje počet řádků. V tabulce jsou data za každou z potravin ve všech letech, jedinou výjimkou je jakostní víno bílé, které se vyskytuje jen v letech 2015–2018.
- Sloupec **prumerna\_cena\_potraviny** obsahuje průměrnou cenu potraviny v daném roce v korunách. Tato cena byla vypočítána z cen potraviny zjištěných v různých obdobích a regionech Česka v průběhu daného roku.
- Sloupce **mnozstvi\_potraviny** a **jednotka\_potraviny** dohromady udávají, k jakému množství potraviny se vztahuje její průměrná cena (např. 10 ks, 1 kg apod.).
- Sloupec **HDP** obsahuje hrubý domácí produkt Česka v amerických dolarech.

## Sekundární tabulka

Tato tabulka je pouze dodatečným materiálem s ekonomickými daty za celou Evropu, s nímž se už v projektu dál nepracuje. Byla vytvořena z dat v tabulce economies z Engeto databáze a obsahuje 585 řádků a 5 sloupců:

- Sloupec **stat** obsahuje 45 evropských států nebo území.
- Sloupec **rok** obsahuje roky 2006–2018, stejně jako primární tabulka.
- Sloupec **populace** obsahuje počet obyvatel.
- Sloupec **HDP** obsahuje hrubý domácí produkt státu v amerických dolarech. Tento údaj chybí (a je tak nahrazen NULL) pro Gibraltar, Faerské ostrovy (kromě roku 2010) a Lichtenštejnsko (kromě roku 2010).
- Sloupec **giniho\_index** obsahuje Giniho koeficient v procentech, jde tedy o Giniho index. Vyjadřuje rozložení bohatství ve společnosti. Tento údaj chybí (a je tak nahrazen NULL) pro 124 řádků.

## Výsledky

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

Mezi roky 2006 a 2018 vzrostly v Česku mzdy ve všech odvětvích. K nejmenšímu procentuálnímu nárůstu došlo v peněžnictví a pojišťovnictví (35,5 %), kde mzdy vzrostly z 39 tisíc korun v roce 2006 na 53 tisíc korun v roce 2018. Nejvyšší procentuální nárůst naopak zaznamenala zdravotní a sociální péče, kde mzdy vzrostly z 18 tisíc na necelých 32 tisíc korun, tedy o 76,0 %.

O žádném z odvětví se tedy nedá říct, že by v něm mzdy dlouhodobě klesaly, v některých odvětvích nicméně došlo ke krátkodobým poklesům. Nejvyšší pokles byl zaznamenán v **peněžnictví a pojišťovnictví** v roce 2013, a to 9,0 %. Nejvíce meziročních poklesů pak zaznamenala **těžba a dobývání**, kde k poklesům došlo v letech 2009, 2013, 2014 a 2016.

Poklesy se nicméně objevily i v odvětvích:

- Činnosti v oblasti nemovitostí (2013)
- Informační a komunikační činnosti (2013)
- Kulturní, zábavní a rekreační činnosti (2011, 2013)
- Profesní, vědecké a technické činnosti (2010, 2013)
- Stavebnictví (2013)
- Ubytování, stravování a pohostinství (2009, 2011)
- Velkoobchod a maloobchod; opravy a údržba motorových vozidel (2009, 2013)

- Veřejná správa a obrana; povinné sociální zabezpečení (2010, 2011)
- Výroba a rozvod elektřiny, plynu, tepla a klimatizovaného vzduchu (2013, 2015)
- Vzdělávání (2010)
- Zásobování vodou; činnosti související s odpady a sanacemi (2013)
- Zemědělství, lesnictví, rybářství (2009)

Z dat je tedy patrné, že nejkrizovější byl v tomto ohledu rok 2013.

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

Mléka (mléko polotučné pasterované) bylo v roce 2006 možné za tehdejší průměrnou mzdu koupit 1309 l. V roce 2018 už to bylo 1564 l. I u chleba (chléb konzumní kmínový) došlo k nárůstu, a to z 1172 kg v roce 2006 na 1279 kg v roce 2018.

Procentuálně byl nárůst množství, které si lze za průměrnou mzdu koupit, vyšší u mléka. Tam byl nárůst 19,5 %, zatímco u chleba 9,1 %.

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

Pro zodpovězení této otázky byl vypočítán průměrný meziroční nárůst cen jednotlivých potravin mezi lety 2006–2018, a to s výjimkou vína (jakostní víno bílé), za které nejsou dostupná data za celé analyzované období a nelze ho tedy porovnat s ostatními potravinami.

Z potravin, pro něž jsou dostupná data za celé období, docházelo k nejnižšímu nárůstu ceny u cukru (cukr krystalový), který v průměru každoročně dokonce zlevnil, a to o 1,9 %. Zlevnila též rajčata (rajská jablka červená kulatá), a to o 0,7 %. K méně než jednorozhodnutí průměrnému ročnímu nárůstu ceny pak došlo i u banánů (banány žluté) a vepřové pečeně s kostí.

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

Meziroční nárůst cen potravin byl v této otázce počítán jako průměr z meziročních nárůstů cen 26 (v letech 2015–2018 27) potravin, za které jsou data v primární tabulce a které reprezentují všechny potraviny.

Růst cen potravin předčil růst mezd z 12 let celkem 6krát. Nejvíce byl ale nárůst cen potravin vyšší než růst mezd pouze o 6,1 % v roce 2013 (tehdy mzdy poklesly o 0,1 % a ceny narostly o 6,0 %), dále byl rozdíl vysoký také v roce 2012 (4,8 %), jindy ale nepřekročil 3 %.

Naopak nárůst mezd předčil nárůst cen potravin nejvíce dokonce o 9,9 %, a to v roce 2009.

## **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?**

K výraznějšímu růstu HDP (nad 5 %) došlo v letech 2007, 2015 a 2017. Co se týče nárůstu cen potravin, v roce 2007 a 2008 byl nejvyšší z celého analyzovaného období, a to 9,3 a 8,9 %. V roce 2014–2016 naopak ceny potravin klesaly. V roce 2017 byl růst cen 7,1 % a 2018 2,4 %. Z těchto dat tedy není patrné, že by v době výrazného růstu HDP (nebo

bezprostředně po něm) spolehlivě docházelo k výraznému nárůstu či poklesu cen potravin. Obecně se nedá tvrdit ani to, že by po výrazném nárůstu HDP docházelo v nárůstu cen potravin k nějaké skokové změně proti výši nárůstu v předchozím roce.

Co se týče nárůstu mezd, je situace odlišná. V letech výrazného růstu HDP (nad 5 %) a letech následujících bezprostředně po tomto nárůstu dosahoval nárůst průměrné mzdy z celého analyzovaného období nejvyšších hodnot, a to od 3,2 do 8,0 %. V ostatních letech se nárůst mezd pohyboval jen mezi -0,1 a 3,3 %. Zdá se tedy, že na nárůst mezd mohly mít změny HDP skutečně vliv.

Jednoznačné určení vlivu změn HDP na změny ve mzdách a cenách potravin nicméně není z pouhého srovnání meziročních růstů možné. Vztah mezi těmito proměnnými by nicméně šlo dál statisticky analyzovat, více odpovědí by případně mohla poskytnout i delší časová řada údajů o HDP, cenách potravin a průměrných mzdách.