# Projekt z SQL

## Zadání: data o mzdách a cenách potravin a jejich zpracování pomocí SQL

## 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?

K zodpovězení výzkumných otázek jsem připravil tabulku, ve které je možné vidět porovnání dostupnosti potravin na základě průměrných příjmů časové období. Dále jsem připravil dodatečnou tabulku, ve které jsou pro další evropské státy uvedeny informace o HDP, GINI koeficientem a populaci.

Tabulka údajů v ČR: t\_Lubos\_Polak\_project\_SQL\_primary\_final

Tabulka údajů evropských států: t\_Lubos\_Polak\_project\_SQL\_secondary\_final

**SQL dotazy pro zodpovězení otázek jsou uvedeny v souboru Engeto\_Project1\_querries.sql**

## Zdroje dat pro tvorbu tabulek:

Zdrojem dat je Portál otevřených dat ČR: https://data.gov.cz/datové-sady

* czechia\_payroll – Informace o mzdách v různých odvětvích za několikaleté období.
* czechia\_payroll\_calculation – Číselník kalkulací v tabulce mezd.
* czechia\_payroll\_industry\_branch – Číselník odvětví v tabulce mezd.
* czechia\_payroll\_unit – Číselník jednotek hodnot v tabulce mezd.
* czechia\_payroll\_value\_type – Číselník typů hodnot v tabulce mezd.
* czechia\_price – Informace o cenách vybraných potravin za několikaleté období.
* czechia\_price\_category – Číselník kategorií potravin.
* countries - Informace o zemích na světě.
* economies - HDP, GINI, daňová zátěž, atd. pro daný stát a rok.

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

Ze získaných dat je možné vyčíst průměrné mzdy v období od roku 2006 do roku 2018.

Z porovnání dat lze vyčíst:

* Mzda v roce 2018 je vyšší než v roce 2006, toto platí pro všechny činnosti.
* Nejvyšší nominální růst mezi rokem 2006 a 2008 vykazují obory „Informační a komunikační činnosti“ (+ 20553 Kč), Výroba a rozvod elektřiny, plynu, tepla a klimatiz. Vzduchu (+ 17506 Kč).
* Obory s nejnižším nominálním růstem mezi lety 2006 a 2018 jsou „Ostatní činnosti“ (+ 6295 Kč), Administrativní a podpůrné činnosti (+ 6752 Kč).
* Kritickým obdobím byl rok 2013, kdy došlo k propadu mezd napříč obory, i když se to nedotklo všech. Nejvíce zasažen byl obor „Peněžnictví a pojišťovnictví“, ve zřejmé souvislosti s hypoteční krizí u USA. Mezi obory, kterých se tento propad nedotkl patří například „Ubytování, stravování a pohostinství“, „Zemědělství, lesnictví, rybářství“, „Vzdělávání“.

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

Tuto otázku jsem vyhodnotil pro jednotlivé obory činností:

MLÉKO

* Nejmenší množství mléka je možné koupit za průměrnou mzdu v oboru „Ubytování, stravování a pohostinství“, 791 litrů v roce 2006 a 954 litrů v roce 2018.
* Největší množství mléka je možné koupit za průměrnou mzdu v oborech „Informační a komunikační činnosti“, 2489 litrů v roce 2006 a 2897 litrů v roce 2018, dále v  „Peněžnictví a pojišťovnictví“ 2803 litrů v roce 2006 a 2785 litrů v roce 2018.

CHLÉB

* Nejmenší množství chleba je možné koupit za průměrnou mzdu v oboru „Ubytování, stravování a pohostinství“, 656 kg v roce 2006 a 752 kg v roce 2018.
* Největší množství mléka je možné koupit za průměrnou mzdu v oborech „Informační a komunikační činnosti“, 2063 kg v roce 2006 a 2282 kg v roce 2018, dále v  „Peněžnictví a pojišťovnictví“ 2323 kg v roce 2006 a 2194 kg v roce 2018.

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

Pro zodpovězení této otázky jsem sestavil dva dotazy:

* výsledkem prvního dotazu je zobrazení procentuálních meziročních změn cen jednotlivých kategorií potravin
* výsledkem druhého dotazu je zobrazení maximálních a minimálních hodnot, které dosahují procentuální meziroční změny dle jednotlivých kategorií potravin

První výsledná tabulka je tedy detailní, zatímco druhá zobrazuje celkový trend přes jednotlivé kategorie potravin.

Z druhé tabulky lze odvodit, že nejnižší procentuální meziroční růst je možné hledat u kategorie potravin „Hovězí maso zadní bez kosti“, což potvrzuje také detailní tabulka. Tato kategorie potravin zaznamenala největší nárůst ceny v roce 2011 – o 8,8%.

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

Tuto otázku jsem hodnotil ve smyslu meziročních změn, tedy letech, kdy došlo k poklesu mzdy nebo ceny, je výsledkem absolutní rozdíl, výsledky tedy zahrnují nejen růst, ale i pokles.

Pro nalezení odpovědi na tuto otázku jsem sestavil tři dotazy:

* první dotaz agregovaně zobrazí počet odvětví v daném roce, u kterých bylo dosaženo podmínky, že rozdíl mezi procentuálními meziročními změnami v platech a cenách je větší než 10%.
* výsledkem druhého dotazu je tabulka, ze které lze vyčíst rozdíl mezi procentuální meziroční změnou platů v jednotlivých odvětvích a procentuální meziroční změnou cen jednotlivých kategorií potravin
* protože mě tato otázka dále zajímala i z jiného pohledu sestavil jsem dotaz, kde jsem zjišťoval rozdíly mezi platy a cenami, vyjádřenými spotřebitelským košem (hodnotou průměrné ceny produktu přes všechny kategorie potravin)

Pro zodpovězení otázky jsem nejprve prostudoval výsledek dotazu 3, zjistil jsem, **že neexistuje rok, kde by meziroční rozdíl** mezi změnou platu v některém odvětví a změnou ceny některé kategorie potravin **byl menší než 10%,** tedy v každém roce jsou nějaké záznamy, kdy je podmínka „> 10%“ splněna. Nejmenší množství takových záznamů přes odvětví jsem zaznamenal v roce 2012.

Následně jsem sestavil dva dotazy, jejichž výsledkem je zobrazení kombinací odvětví a cenových kategorií v roce 2012, zjistil jsem, že k překročení zvolené úrovně rozdílu > 10% došlo u 118 záznamů, zatímco rozdíl do 10% byl zaznamenán v 376 případech.

Pokud bych měl hodnotit tuto otázku optikou spotřebitelského koše, tedy idealizované ceny průměrné cenové kategorie, pak z výsledku dotazu lze vyčíst, že definovaná podmínka růstu byla splněna pouze v letech 2008, 2009, 2013 a 2018.

## Otázka 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?

MZDY:

Obecně lze říci, že se změny HDP do změn ve mzdách promítají, u některých oborů je korelace vyšší, u některých nižší.

Ve sledovaném období byly ale i roky, kdy se změna výše HDP ve mzdách nepromítla, nejmarkantněji je to vidět v roce 2009, kdy výše HDP klesla o 4,66 %, zatímco mzdy ve většině oborů rostly. Oborem, kterého se pokles HDP v roce 2009 dotkl nejvíce, byla „Těžba a dobývání“, kde ale v následném roce průměrné mzdy prudce rostly. Dalším obdobím poklesu HDP byl rok 2013, kdy zaznamenalo pokles o 0,79%. Největší propad v průměrných mzdách v roce 2013 zaznamenal obor „Peněžnictví a pojišťovnictví“ (-9,0%).

CENY:

Trend změny potravin (ceny průměrné potraviny v nákupním koši) ve sledovaném období ve větší míře kopíruje trend vývoje HDP.

V letech 2012 a 2013, kdy došlo k poklesu a následné stagnaci HDP ceny potravin rostly, tento trend se otočil v následujících letech 2014 a 2015.