

# Projekt - porovnání dostupnosti potravin v ČR

---

*Výstupní projekt kurzu Datová akademie od firmy Engeto*

## 1. Úvod do projektu

Cílem projektu je pomoci kolegům z (fiktivního) analytického oddělení připravit datovou sadu pro vyhodnocení dostupnosti potravin na základě průměrných příjmů za určité časové období. Pro tento účel máme za úkol vytvořit dvě tabulky s datovými podklady – jednu pro analýzu vývoje cen potravin a mezd a druhou, doplňkovou, s makroekonomickými údaji dalších zemí.

V rámci analýzy budeme hledat odpovědi na připravené výzkumné otázky, které budou sloužit jako podklad pro prezentaci na konferenci.

**Výstup projektu obsahuje:**

- SQL pro vytvoření primární tabulky - `t_veronika_abonyiova_project_sql_primary_final` – obsahující podklady pro analýzu cen a mezd
- SQL pro vytvoření sekundární tabulky - `t_veronika_abonyiova_project_sql_secondary_final` – obsahující makroekonomické údaje zemí světa
- 5 SQL – pro každou výzkumnou otázku jeden – připravující data pro jednotlivé výzkumné otázky

## 2. Datové podklady

Pro přípravu datové sady byly k dispozici tabulky s údaji o vývoji cen v České republice, vývoji mezd v České republice a makroekonomických údajích zemí světa. Aby bylo možné použít data pro porovnání mezi sebou, bylo potřeba data omezit pouze na srovnatelné (společné) roky a dále provést jejich mírné čištění.

### 2.1. Úpravy dat

**Primární tabulka:**

- Omezení dat na srovnatelné roky – dostupné tabulky měly společné roky pouze 2006-2018. Z tabulek, kde bylo k dispozici datum od-do byl použit rok z datumu „od“.
- Agregace dat - do primární tabulky byly údaje o mzdách a cenách agregovány tak, aby zachovaly potřebnou strukturu, ale neobsahovaly každý jediný řádek původní tabulky. Obsahuje tedy 4 typy údajů, rozlišené hodnotou ve sloupci „value\_type“ a hodnotami industry a product:
  - Průměrné mzdy za jednotlivé roky a jednotlivá odvětví
  - Průměrné mzdy za jednotlivé roky celkově (bez detailu za odvětví – industry = null)

- Průměrné ceny jednotlivých druhů potravin za jednotlivé roky
- Průměrné ceny za jednotlivé roky celkově (bez detailu za odvětví – product = null)
- Vyloučení dat – po omezení roků analýzy bylo ještě potřeba prozkoumat data z hlediska jejich úplnosti v daném časovém období.
  - Ceny – identifikovala jsem dva produkty s nižším počtem údajů o ceně, než bylo očekávané vzhledem k časovému období analýzy, a to
    - Jakostní víno bílé – obsahovalo údaje pouze za 4 roky z dostupných 13 let. Aby byl dotaz obecněji použitelný pro případ budoucí analýzy, použila jsem v SQL dotazu pravidlo, aby byly vyřazeny ty produkty, které obsahují údaje za méně než polovinu srovnávaných roků.
    - Kapři – bylo k dispozici celkově méně údajů, protože byly k dispozici pro každý rok údaje pouze v prosinci. Jelikož to ale odpovídá povaze prodáváného produktu a údaje existovaly pro všechny srovnávané roky, ponechala jsem kapry ve zkoumaném vzorku dat.
  - Mzdy – v údajích o mzdách žádné nedostatky v chybějících údajích nebyly nalezeny.

#### Sekundární tabulka:

- Omezení dat na srovnatelné roky – tj. období roku 2006-2018
- Omezení států – byly zahrnuty pouze evropské státy.

### 3. Výzkumné otázky

Úkolem bylo připravit data pro následující výzkumné otázky:

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

Připravená datová sada nabízí pro každé odvětví vypočtenou průměrnou míru růstu mezd, celkovou míru růstu mezd, spolu s počtem období, ve kterých za dané časové období mzdy rostly a klesaly. Z dat je zřejmé, že ve všech odvětvích mzdy postupně rostou (průměrně 2 - 4% ročně), nicméně pouze ve čtyřech odvětvích rostly mzdy ve všech zkoumaných letech. V ostatních odvětvích mzdy v 1-4 letech klesaly.

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

Připravená datová sada nabízí porovnání dostupného množství mléka a chleba za rok 2006 a 2018. V obou případech bylo možné v posledním období koupit více produktu, než v prvním období.

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

Připravená datová sada nabízí pro každou kategorii potravin průměrnou roční a poté celkovou míru růstu cen za sledované období a následně je seřazena od nejnižšího růstu po nejvyšší. Nejméně rostla cena (respektive dokonce klesala) u krystalového cukru.

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

Připravená datová sada poskytuje pro každý rok míru růstu cen, míru růstu mezd a jejich porovnání. Z dat vyplývá, že v žádném roce nebyl růst cen vyšší než růst mezd o více než 10 %. Nejvíce převyšoval růst cen růst mezd v roce 2013, a to o 5,68%. Naopak v roce 2009 byl růst mezd o více než 10% vyšší, než růst cen, resp. mzdy stouply o 3,35% a ceny klesly o 6,81%.

**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?*

Údaje o mzdách a cenách máme jen pro Českou republiku, proto je analýza vztahu s růstem HDP omezena pouze pro Českou republiku. Obecný vztah mezd, cen a HDP je v připravené datové sadě vyjádřen korelačním koeficientem míry růstu těchto veličin (korelace absolutních hodnot nedává příliš smysl – obě veličiny se mění jen mírně s průměrnou pomalu rostoucí tendencí, proto je korelační koeficient absolutních hodnot téměř 1.00, což neodpovídá realitě).

Z korelačních koeficientů vyplývá, že korelace změny HDP a:

- Změny cen ve stejném období existuje, ale je relativně mírná
- Změny cen v následujícím období je téměř nulová
- Změny mezd ve stejném období existuje a je mírná
- Změny mezd v následujícím období existuje a je poměrně vysoká.

Nicméně zde je potřeba říci, že korelace pouze udává podobnost trendů obou veličin, nikoli přímo to, zda je to změna výše HDP, co způsobilo změnu ve mzdách a cenách.

Také je tato otázka položená poměrně volně – není definované, co znamená výrazný růst. Pokud bych vzala jako míru významnosti například 4% růstu HDP, pak tato situace nastala během sledovaného období třikrát a ve všech třech případech byl růst mezd následujícího období také vyšší než 4%. Myslím si ale, že pro odpověď na tuto otázku máme příliš krátké zkoumané období.

## **4. Závěr**

Připravené podklady pro výzkumné otázky poskytují adekvátní data pro získání odpovědí na otázky 1-4. Z dostupných dat vyplývá, že mzdy ve všech odvětvích mírně rostou, ceny základních potravin rostou také, ale na vzorku dvou potravin se ukazuje, že v posledním zkoumaném roce bylo možné koupit více potravin, než v prvním.

V páté otázce ohledně závislosti růstu mezd a cen na HDP získáváme údaje o závislosti především mezd stejného a následujícího období, ale pro jednoznačnou odpověď by bylo potřeba mít údaje za podstatně delší časové období.