



# PROJEKT 1

Dostupnost základních potravin v ČR

ENGETO - KURZ DATOVÁ ANALÝZA

Mgr. Veronika Čížková  
*veronika.cizkova21@gmail.com*

## Obsah

|   |    |
|---|----|
| Výzkumné otázky a stanovení srovnatelného období..... | 2  |
| Výzkumné otázky.....                                  | 2  |
| Srovnatelné období .....                              | 2  |
| Otázka 1.....   | 2  |
| Otázka 2.....   | 3  |
| Otázka 3.....   | 3  |
| Otázka 4.....   | 3  |
| Otázka 5.....   | 4  |
| SQL dokumentace .....                                 | 5  |
| Tabulka 1 (SQL).....                                  | 5  |
| Tabulka 2 (SQL).....                                  | 6  |
| SQL dotazy .....                                      | 7  |
| Otázka 1 .....  | 7  |
| Otázka 2 .....  | 8  |
| Otázka 3 .....  | 8  |
| Otázka 4 .....  | 9  |
| Otázka 5 .....  | 10 |
| Zadání projektu .....                                 | 12 |

# Výzkumné otázky a stanovení srovnatelného období

## Výzkumné otázky

Tento projekt má za cíl odpovědět následující 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?

## Srovnatelné období

Dle dostupných dat je možno zkoumat vývoj cen potravin a mezd v letech 2006 až 2018 (viz primární tabulka). Pro toto období jsou dostupná také data týkající se evropských států (viz sekundární tabulka).

## Otázka 1

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

Mzdy během sledovaného období vzrostly ve všech odvětvích, ačkoliv některá odvětví zaznamenala během dotčené doby pokles. Všechny poklesy nastaly mezi roky 2009 až 2016. Největší pokles nastal v roce 2013 u Peněžnictví a pojišťovnictví, a to 9 %. Další významnější poklesy mezd nastaly u Těžby a dobývání v roce 2009 a u Výroby a rozvodu elektřiny, plynu, tepla a klimatiz. vzduchu v roce 2013, kdy poklesy činily pře 4 %.

Největší procentní změnu ve sledovaném období zaznamenalo odvětví Zdravotní a sociální péče, a to nárůst o téměř 76 %. Naopak nejméně vzrostly mzdy v oboru Peněžnictví a pojišťovnictví, kdy vzrostly o 35,54 %.

**Závěr:** Mzdy ve všech odvětvích vzrostly. Poklesy mezd byly pouze ojedinělé a dočasné a proběhly u některých odvětvích v letech 2009 – 2016. Oproti roku 2006 byly mzdy v roce 2018 ve všech odvětvích vyšší minimálně o 35,54 %.

## Otázka 2

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

V roce 2006 byla průměrná mzda 20 342,- Kč. V posledním srovnatelném roce, v roce 2018, byla průměrná mzda 31 980,- Kč.

V roce 2006 tak bylo možné koupit za průměrnou mzdu celkem 1 262 kg chleba při ceně 16,12,- Kč/kg nebo 1 409 l mléka při ceně 14,44,- Kč/l. V roce 2018, bylo možné koupit za průměrnou mzdu 1 319 kg chleba při ceně 24,24,- Kč nebo 1 614 l mléka při ceně 19,82,- Kč/l.

**Závěr:** Z výše uvedeného je patrné, že se kupní síla obyvatel u chleba a mléka zvýšila.

## Otázka 3

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

Nejpomaleji zdražuje kategorie Banány žluté. Jejich průměrný meziroční nárůst ceny je o 0,59 %. O necelé procento také průměrně meziročně zdražuje Vepřová pečeně s kostí a Přírodní minerální voda uhlíčitá.

Ovšem, ne všechny kategorie potravin od roku 2006 zdražily. Během sledovaného období klesla cena u Cukru krystalového (o 2,65 % ročně) a u kategorie Rajská jablka červená kulatá (o 2,16 % ročně).

Během sledovaného období v některém z let zaznamenaly (ojedinělý) pokles ceny všechny kategorie potravin kromě kategorií: Jakostní víno bílé, Těstoviny vaječné a Pivo výčepní, světlé, lahvé.

**Závěr:** Nejnižší procentuální meziroční nárůst je vzhledem k zápornému geometrickému koeficientu u Cukru krystalového, tento typ potravin je v čase průměrně meziročně levnější o 2,65 %. Banány žluté jsou potom kategorií potravin, která zdražuje ve sledovaném období nejpomaleji, a to o 0,59 % ročně.

## Otázka 4

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

Žádný meziroční nárůst cen potravin nebyl výrazně vyšší než růst mezd. Nejvíce oproti mzdám vzrostly ceny potravin meziročně v roce 2013, kdy rozdíl byl 6,79 %. Tento trend rychleji rostoucích cen potravin nežli mezd začal již v roce 2011.

Naopak nejvíce vzrostly mzdy oproti cenám potravin v roce 2009, kdy rozdíl meziročních růstů činil 9,69 %.

**Závěr:** Ve většině případů rostly meziročně mzdy více nežli ceny potravin. Žádný meziroční nárůst cen potravin oproti růstu mezd nebyl vyšší než 10 %.

## 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?**

Nejdříve budeme definovat výraznější růst HDP ve výši 5 % a jako výraznější růst u mezd nebo cen potravin definujeme na 3 %.

Nejvýraznější meziroční růst HDP nastal v letech 2007, 2015 a 2017, kdy HDP meziročně vzrostlo o více než 5 %. Můžeme pozorovat, že mzdy poté vždy v témže roce vzrostly minimálně průměrně o 2,78 % a v dalším roce vzrostly minimálně o 3,97 %.

Co se týče cen potravin ve stejném roce, kdy HDP vzrostlo o více než 5 %, jednou potraviny zlevnily o 0,7 % a dvakrát ceny potravin vzrostly. V dalším roce byla situace stejná (jednou zlevnily, dvakrát zdražily).

**Závěr:** Z dat můžeme usuzovat, že pokud v jednom roce HDP vzroste minimálně o 5 %, zvýší se mzdy v roce následujícím. Nicméně pro potvrzení nebo vyvrácení tohoto tvrzení, by bylo potřeba mít delší časovou řadu.

## SQL dokumentace

### Tabulka 1 (SQL)

```
CREATE OR REPLACE TABLE t_veronika_cizkova_project_SQL_primary_final
AS
SELECT
    a.code_food,
    a.category_food,
    a.average_price,
    a.price_value,
    a.price_unit,
    b.average_payroll,
    b.payroll_year AS 'year',
    b.industry_branch_code,
    b.name AS industry_branch_name
FROM
    (SELECT
        cp.category_code AS code_food,
        cpc.name AS category_food,
        ROUND(AVG(value),2) AS average_price,
        cpc.price_value,
        cpc.price_unit,
        YEAR(cp.date_to) AS year_food
    FROM czechia_price cp
    LEFT JOIN czechia_price_category cpc
        ON cp.category_code = cpc.code
    WHERE region_code IS NULL
        AND YEAR(cp.date_to) BETWEEN 2006 AND 2018
    GROUP BY cp.category_code, YEAR(cp.date_to), cpc.price_unit,
    cpc.price_value
    ) a
JOIN (SELECT
    ROUND(AVG(cp2.value),2) AS average_payroll,
    cp2.payroll_year,
    cp2.industry_branch_code,
    cpib.name
FROM czechia_payroll cp2
LEFT JOIN czechia_payroll_industry_branch cpib
    ON cp2.industry_branch_code = cpib.code
WHERE cp2.payroll_year BETWEEN 2006 AND 2018
    AND cp2.value_type_code = 5958
    AND cp2.calculation_code = 100
    AND cp2.industry_branch_code IS NOT NULL
GROUP BY cp2.payroll_year, cp2.industry_branch_code, cpib.name) b
    ON a.year_food = b.payroll_year
;
```

## Tabulka 2 (SQL)

```
CREATE OR REPLACE TABLE
t_veronika_cizkova_project_SQL_secondary_final AS
SELECT
    c.country,
    e.population,
    e.gini,
    e.gdp,
    e.year
FROM economies e
JOIN countries c
    ON e.country = c.country
WHERE c.continent LIKE 'Europe'
    AND (e.year BETWEEN 2006 AND 2018)
ORDER BY e.year, c.country
;
```

## SQL dotazy

### Otázka 1

```
CREATE OR REPLACE VIEW v_payroll_change AS
SELECT
    tc1.year,
    tc1.industry_branch_name,
    ROUND(tc1.average_payroll,0) AS average_payroll,
    LAG(tc1.average_payroll, 1) OVER
        (PARTITION BY tc1.industry_branch_name ORDER BY tc1.year)
AS last_payroll,
    (ROUND(tc1.average_payroll/LAG(tc1.average_payroll, 1) OVER
        (PARTITION BY tc1.industry_branch_name ORDER BY
tc1.year), 4) -1) * 100 AS percent_changes
FROM t_veronika_cizkova_project_SQL_primary_final tc1
GROUP BY year, tc1.industry_branch_name, tc1.average_payroll
ORDER BY percent_changes;
```

```
SELECT
    *
FROM v_payroll_change vpc
WHERE vpc.percent_changes IS NOT NULL
    AND vpc.percent_changes < 0
ORDER BY vpc.percent_changes, vpc.year, vpc.industry_branch_name
;
```

```
SELECT
    *
FROM v_payroll_change vpc
WHERE vpc.percent_changes IS NOT NULL
    AND vpc.percent_changes < 0
ORDER BY vpc.year
;
```

```
SELECT
    vpc.year,
    vpc.industry_branch_name,
    vpc.average_payroll,
    LAG(average_payroll, 12) OVER
        (PARTITION BY industry_branch_name ORDER BY year) AS
total_difference_payroll,
    ROUND(average_payroll/LAG(average_payroll, 12) OVER
        (PARTITION BY industry_branch_name ORDER BY year),4)*100-
100 AS total_percent_change
FROM v_payroll_change vpc
GROUP BY vpc.year, vpc.industry_branch_name, vpc.average_payroll
ORDER BY total_percent_change DESC
;
```



## Otázka 2

```
SELECT
    tc1.code_food,
    tc1.category_food,
    tc1.year,
    ROUND(AVG(tc1.average_payroll),2) AS avg_payroll,
    ROUND(AVG(tc1.average_price),2) AS avg_price,
    ROUND(ROUND(AVG(tc1.average_payroll),2)/ROUND(AVG(tc1.average_price),2)) AS purchasing_power,
    tc1.price_unit
FROM t_veronika_cizkova_project_SQL_primary_final tc1
WHERE tc1.code_food IN (111301, 114201)
    AND year IN (2006, 2018)
GROUP BY tc1.code_food, tc1.category_food, tc1.year, tc1.price_unit
ORDER BY tc1.code_food
;
```

## Otázka 3

```
SELECT
    category_food,
    MIN(a.percent_changes) AS min_percent_change,
    MAX(a.percent_changes) AS max_percent_change,
    ROUND((EXP(AVG(LOG(1 + (percent_changes / 100)))) - 1)*100 ,2)
AS geometric_mean,
    MIN(average_price) AS min_average_price,
    MAX(average_price) AS max_average_price
FROM (
    SELECT
        year,
        tc1.category_food,
        tc1.code_food,
        tc1.average_price,
        LAG(average_price, 1) OVER
            (PARTITION BY code_food ORDER BY year) AS last_price,
        (ROUND(average_price/LAG(average_price, 1) OVER
            (PARTITION BY code_food ORDER BY year), 4) -1) * 100 AS
percent_changes
    FROM t_veronika_cizkova_project_SQL_primary_final tc1
    GROUP BY year, tc1.category_food, tc1.code_food, average_price
    ORDER BY percent_changes
) a
GROUP BY category_food
HAVING geometric_mean > 0
ORDER BY geometric_mean
;
```

#### Otázka 4

```
CREATE OR REPLACE VIEW v_food AS
SELECT year,
       ROUND(AVG(food_percent_changes),2) AS average_change_food
FROM (
SELECT
    year,
    tc1.category_food,
    tc1. average_price,
    LAG(average_price, 1) OVER
        (PARTITION BY category_food ORDER BY year) AS last_price,
    (ROUND(average_price/LAG(average_price, 1) OVER
        (PARTITION BY category_food ORDER BY year), 4) -1) * 100
AS food_percent_changes
FROM t_veronika_cizkova_project_SQL_primary_final tc1
GROUP BY year, tc1.category_food, tc1.code_food, average_price
) a
GROUP BY a.year
ORDER BY a.year;
```

```
CREATE OR REPLACE VIEW v_payroll AS
SELECT
    a.year,
    ROUND(AVG(a.percent_changes),2) AS average_change_payroll
FROM (
SELECT
    tc1.year,
    tc1.industry_branch_name,
    ROUND(AVG(tc1.average_payroll),2) AS average_payroll,
    LAG(tc1.average_payroll, 1) OVER
        (PARTITION BY tc1.industry_branch_name ORDER BY tc1.year)
AS last_payroll,
    (ROUND(tc1.average_payroll/LAG(tc1.average_payroll, 1) OVER
        (PARTITION BY tc1.industry_branch_name ORDER BY
tc1.year), 4) -1) * 100 AS percent_changes
FROM t_veronika_cizkova_project_SQL_primary_final tc1
GROUP BY year, tc1.industry_branch_name
) a
GROUP BY a.year
ORDER BY a.year;
```

```
SELECT
    v_food.year,
    v_food.average_change_food,
    v_payroll.average_change_payroll,
    average_change_food - average_change_payroll AS difference
FROM v_food
LEFT JOIN v_payroll
    ON v_food.year = v_payroll.year
-- WHERE average_change_food - average_change_payroll > 10
ORDER BY difference DESC
;
```

## Otázka 5

```
CREATE OR REPLACE VIEW v_cze_gdp AS
SELECT
    tc2.year,
    tc2.gdp,
    LAG(tc2.gdp,1) OVER
        (PARTITION BY country ORDER BY year) AS last_gdp,
    (ROUND(gdp/LAG(gdp, 1) OVER
        (PARTITION BY country ORDER BY year),4) -1)*100 AS percentual_change_gdp
FROM t_veronika_cizkova_project_SQL_secondary_final tc2
WHERE country LIKE 'czech_republic'
;
```

```
CREATE OR REPLACE VIEW v_gdp_food_payroll AS
SELECT
    v_cze_gdp.*,
    v_food.average_change_food,
    v_payroll.average_change_payroll,
    LEAD(v_food.average_change_food, 1) OVER (ORDER BY year) AS
next_average_change_food,
    LEAD(v_payroll.average_change_payroll, 1) OVER (ORDER BY year) AS
next_average_change_payroll
FROM v_cze_gdp
LEFT JOIN v_food
    ON v_food.year = v_cze_gdp.year
LEFT JOIN v_payroll
    ON v_payroll.year = v_cze_gdp.year
;

SELECT *,
    CASE
        WHEN (average_change_food > 3 OR next_average_change_food > 3) AND
(average_change_payroll > 3 OR next_average_change_payroll > 3) THEN 'payroll and
food'
        WHEN (average_change_food < 3 AND next_average_change_food < 3) AND
(average_change_payroll < 3 AND next_average_change_payroll < 3) THEN 'not food
not payroll'
        WHEN (average_change_payroll < 3 AND next_average_change_payroll < 3)
AND (average_change_food > 3 OR next_average_change_food > 3) THEN 'food not
payroll'
        WHEN (average_change_food < 3 AND next_average_change_food < 3) AND
(average_change_payroll > 3 OR next_average_change_payroll > 3) THEN 'payroll not
food'
        ELSE 0
    END AS influence_gdp
FROM v_gdp_food_payroll
-- WHERE percentual_change_gdp > 5
;
```

```
SELECT a.influence_gdp,
    COUNT(a.influence_gdp) AS count_of_influence
FROM (SELECT *,
    CASE
        WHEN (average_change_food > 3 OR next_average_change_food > 3) AND
(average_change_payroll > 3 OR next_average_change_payroll > 3) THEN 'payroll and
food'
```

```

        WHEN (average_change_food < 3 AND next_average_change_food < 3) AND
(average_change_payroll < 3 AND next_average_change_payroll < 3) THEN 'not food
not payroll'
        WHEN (average_change_payroll < 3 AND next_average_change_payroll < 3)
AND (average_change_food > 3 OR next_average_change_food > 3) THEN 'food not
payroll'
        WHEN (average_change_food < 3 AND next_average_change_food < 3) AND
(average_change_payroll > 3 OR next_average_change_payroll > 3) THEN 'payroll not
food'
        ELSE 0
    END AS influence_gdp
FROM v_gdp_food_payroll
WHERE percentual_change_gdp > 5
) a
GROUP BY a.influence_gdp
;

```

## Zadání projektu

### Úvod do 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. `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).

Dále připravte sadu SQL, které z vámi připravených tabulek získají datový podklad k odpovězení na vytyčené výzkumné otázky. Pozor, otázky/hypotézy mohou vaše výstupy podporovat i vyvracet! Záleží na tom, co říkají data.

Na svém GitHub účtu vytvořte repozitář (může být soukromý), kam uložíte všechny informace k projektu – hlavně SQL skript generující výslednou tabulku, popis mezivýsledků (průvodní listinu) a informace o výstupních datech (například kde chybí hodnoty apod.).

**Neupravujte data v primárních tabulkách! Pokud bude potřeba transformovat hodnoty, dělejte tak až v tabulkách nebo pohledech, které si nově vytváříte.**