

# PROJEKT : PROJEKT Z SQL

## ÚVOD DO PROJEKTU

Projekt se zaměřuje na analýzu dostupnosti základních potravin ve vztahu k průměrným příjmům obyvatel za určité časové období. Tento projekt vznikl v rámci analytické činnosti nezávislé organizace zabývající se hodnocením životní úrovně obyvatelstva. Cílem je odpovědět na klíčové výzkumné otázky související s ekonomickou dostupností základních potravin pro širokou veřejnost.

Výsledky těchto analýz budou sloužit jako podklad pro prezentaci na odborné konferenci, která se zaměřuje na témata spojená s ekonomickou dostupností potravin a životní úrovní.

Projekt využívá nástroje **PostgreSQL** pro správu a dotazování se do databáze, kde byla data agregována a analyzována. Pro vizualizaci výsledků byl použit **Python** s knihovnami jako *pandas*, *matplotlib* a *seaborn* což umožnilo přehledné zobrazení dat. Tím byla zajištěna jak vysoká kvalita analýzy, tak i její snadná prezentace odborné i širší veřejnosti.

## DATOVÉ SADY

### PRIMÁRNÍ TABULKA

Pro vytvoření tabulky **t\_elena\_kocumova\_project\_SQL\_primary\_final** jsem spojila informace o mzdách a cenách potravin v České republice. Data byla získána z následujících tabulek:

- **czechia\_payroll**: obsahuje údaje o mzdách podle odvětví a roku.
- **czechia\_payroll\_industry\_branch**: číselník pro názvy průmyslových odvětví.
- **czechia\_price**: údaje o cenách potravin podle kategorií a roku.
- **czechia\_price\_category**: číselník pro názvy kategorií potravin.

#### Důležité faktory:

- období bylo omezené na roky 2006 – 2018 s cílem zajistit srovnatelná data o mzdách a cenách.
- pouze záznamy s dostupnými údaji o mzdách (*value\_type\_code* = 5958) a cenách potravin (*value IS NOT NULL*).
- mezi hlavní sloupce patří rok (*payroll\_year*, *date\_from*), průmyslové odvětví (*industry*), kategorii potravin (*product*), průměrné mzdy (*avg\_salary*) a průměrné ceny potravin (*avg\_price*).

#### Postup

1. Náhled dat:
  - zobrazení obsahu tabulek pro pochopení struktury a dostupnosti dat.
  - Ověření, jaké roky obsažene v tabulkách **czechia\_payroll** a **czechia\_price**.
2. Kontrola a čištění dat:
  - Identifikace a odstranění záznamů s NULL hodnotami ve sloupcích klíčových kódů a hodnot (*value*, *industry\_branch\_code*, *category\_code*).
3. Vytvoření dílčích datových pohledů:
  - *salary\_data*: průměrné mzdy podle roku a odvětví.
  - *price\_data*: průměrné ceny potravin podle roku a kategorie.
4. Do výsledné tabulky byly zahrnuty pouze záznamy s platnými údaji o odvětvích (*industry IS NOT NULL*).

## SEKUNDÁRNÍ TABULKA

Pro vytvoření tabulky **t\_elena\_kocumova\_project\_SQL\_secondary\_final** jsem spojila údaje o ekonomických ukazatelích evropských států. Informace pocházejí z následujících tabulek:

- **economies**: obsahuje údaje o HDP, GINI koeficientu, populaci a dalších ekonomických ukazatelích států podle roku.
- **countries**: obsahuje informace o státech, včetně názvů, kontinentů a doplňujících údajů, jako je typ vlády, průměrná výška nebo očekávaná délka života.

### Důležité faktory:

- data byla omezena na roky 2006-2018, aby odpovídala časovému rozsahu primární tabulky.
- byly zahrnuty pouze evropské státy, určené pomocí sloupce *continent* v tabulce **countries**.
- do analýzy se zahrnuly pouze záznamy s dostupnými údaji o HDP (*gdp*), populaci (*population*) a GINI koeficientu (*gini*), aby byla data úplná a použitelná.
- klíčové sloupce zahrnují rok (*year*), stát (*country*), HDP (*gdp*), GINI koeficient (*gini*), populaci (*population*) a doplňující informace, například typ vlády (*government\_type*) nebo průměrnou délku života (*life\_expectancy*).

### Postup

1. Náhled dat:
  - Zobrazení obsahu tabulek **economies** a **countries**, aby bylo možné pochopit strukturu a dostupnost dat.
  - Kontrola propojení států mezi oběma tabulkami pomocí levého spojení (*LEFT JOIN*) a identifikace případných chybějících záznamů.
2. Kontrola a čištění dat:
  - Vyhledání záznamů s chybějícími klíčovými hodnotami (např. *gdp*, *population*, *gini*) a jejich odstranění.
  - Omezení na evropské státy pomocí sloupce *continent* v tabulce **countries**.
3. Vytvoření přechodné tabulky:
  - Vytvoření přechodné tabulky *filtered\_economies*, která obsahuje data z tabulek **economies** a **countries**, filtrovaná podle zadaných kritérií (*roky*, *státy*, *dostupnost hodnot*).
4. Spojení na základě názvu státu (*country*) a vytvoření sekundární tabulky.

## VÝZKUMNÉ OTÁZKY

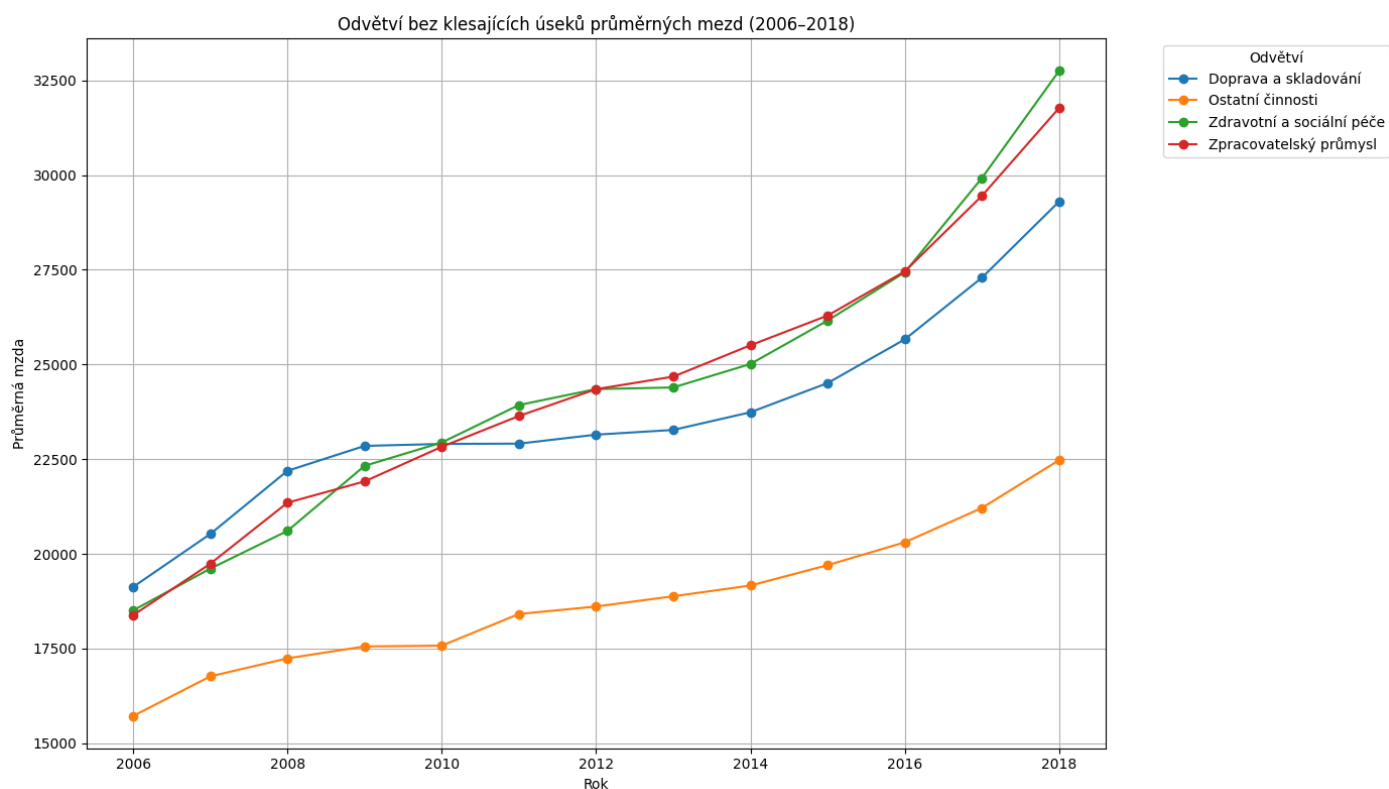
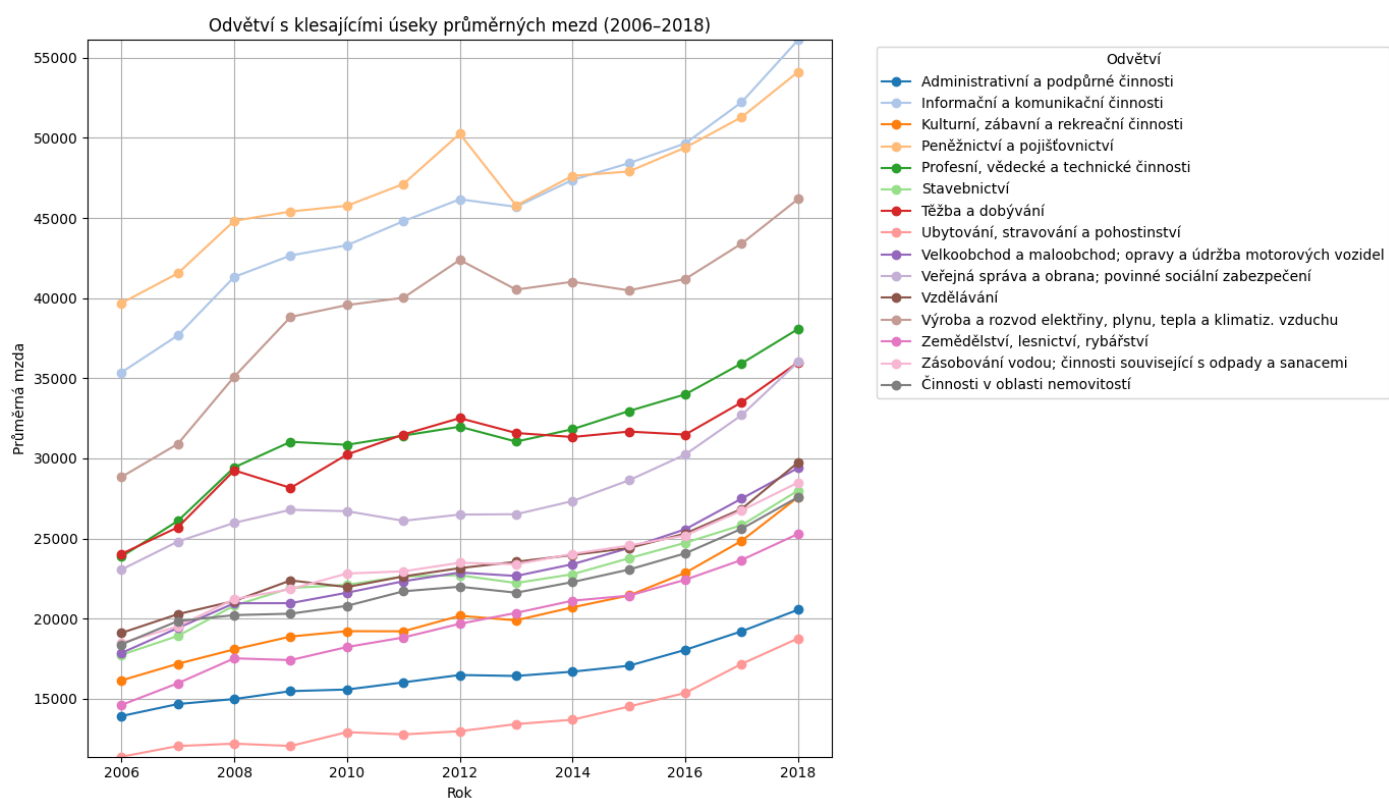
### 1. ROSTOU V PRŮBĚHU LET MZDY VE VŠECH ODVĚTVÍCH, NEBO V NĚKTERÝCH KLESAJÍ?

Pro analýzu otázky, zda mzdy v průběhu let rostou ve všech odvětvích, nebo v některých klesají, byla použita data z tabulky **t\_elena\_kocumova\_project\_SQL\_primary\_final**. Tato tabulka obsahuje informace o průměrných mzdách v jednotlivých odvětvích a letech. Nejprve byla data očištěna tak, aby obsahovala pouze záznamy s platnými hodnotami průměrných mezd, což znamená, že byly odstraněny prázdné hodnoty. Poté byl pro každé odvětví a rok spočítán trend na základě srovnání s předchozím rokem. Tento výpočet umožnil určit, zda mzdy rostly, klesaly, nebo zůstaly beze změny. Výsledná data byla seřazena podle odvětví a roku, což usnadňuje sledování vývoje.

Pro lepší porozumění a vizualizaci byly výsledky zpracovány v Pythonu a vykresleny do grafů. Díky této vizualizaci bylo zjištěno, že ve čtyřech odvětvích:

- ostatní činnosti,
- doprava a skladování,
- zdravotní a sociální péče,
- zpracovatelský průmysl

nedošlo během celého sledovaného období k žádnému poklesu mezd. U ostatních odvětví se občasné poklesy vyskytly, ale celkový trend mezd zůstal rostoucí.



## 2. KOLIK JE MOŽNÉ SI KOUPIT LITRŮ MLÉKA A KILOGRAMŮ CHLEBA ZA PRVNÍ A POSLEDNÍ SROVNATELNÉ OBDOBÍ V DOSTUPNÝCH DATECH CEN A MEZD?

Pro analýzu otázky, kolik litrů mléka a kilogramů chleba je možné koupit za průměrnou mzdu v prvním a posledním srovnatelném období, byla použita data z tabulky **t\_elena\_kocumova\_project\_sql\_primary\_final**. Tato tabulka obsahuje informace o průměrných mzdách, cenách a produktech v jednotlivých letech a odvětvích.

Následně byly určeny první a poslední roky dostupné v datech, což umožnilo provést srovnání ve dvou obdobích. V každém období byla vypočítána dostupnost produktů vydělením průměrné mzdy průměrnou cenou.

Číslo	Nazev odvětví	Chléb 2006 (kg)	Chléb 2018 (kg)	Mléko 2006 (l)	Mléko 2018 (l)
1	Administrativní a podpůrné činnosti	864	849	965	1038
2	Doprava a skladování	1187	1209	1325	1479
3	Informační a komunikační činnosti	2194	2315	2450	2831
4	Kulturní, zábavní a rekreační činnosti	1002	1138	1119	1392
5	Ostatní činnosti	975	928	1089	1134
6	Peněžnictví a pojišťovnictví	2462	2233	2750	2731
7	Profesní, vědecké a technické činnosti	1481	1571	1654	1922
8	Stavebnictví	1101	1155	1230	1413
9	Těžba a dobývání	1492	1485	1666	1816
10	Ubytování, stravování a pohostinství	707	775	789	948
11	Velkoobchod a maloobchod; opravy a údržba motorových vozidel	1109	1214	1238	1484
12	Veřejná správa a obrana; povinné sociální zabezpečení	1431	1486	1598	1818
13	Vzdělávání	1186	1228	1324	1501
14	Výroba a rozvod elektřiny, plynu, tepla a klimatiz. vzduchu	1789	1906	1998	2330
15	Zdravotní a sociální péče	1148	1352	1282	1653
16	Zemědělství, lesnictví, rybářství	907	1044	1013	1277
17	Zpracovatelský průmysl	1140	1311	1273	1603
18	Zásobování vodou; činnosti související s odpady a sanacemi	1147	1176	1281	1438
19	Činnosti v oblasti nemovitostí	1141	1139	1275	1392

### Trendy mezi lety 2006 a 2018:

- Chléb: Většina odvětví zaznamenala mírný nárůst dostupnosti chleba. Výjimku tvoří například „Administrativní a podpůrné činnosti“ (pokles z 864 na 849 kg) a „Ostatní činnosti“ (pokles z 975 na 928 kg).
- Mléko: U mléka došlo k nárůstu dostupnosti ve většině odvětví. Nejvýraznější zlepšení zaznamenaly „Informační a komunikační činnosti“ (z 2450 na 2831 litrů) a „Zdravotní a sociální péče“ (z 1320 na 1528 litrů).

### Odvětví s nižší kupní silou v roce 2018:

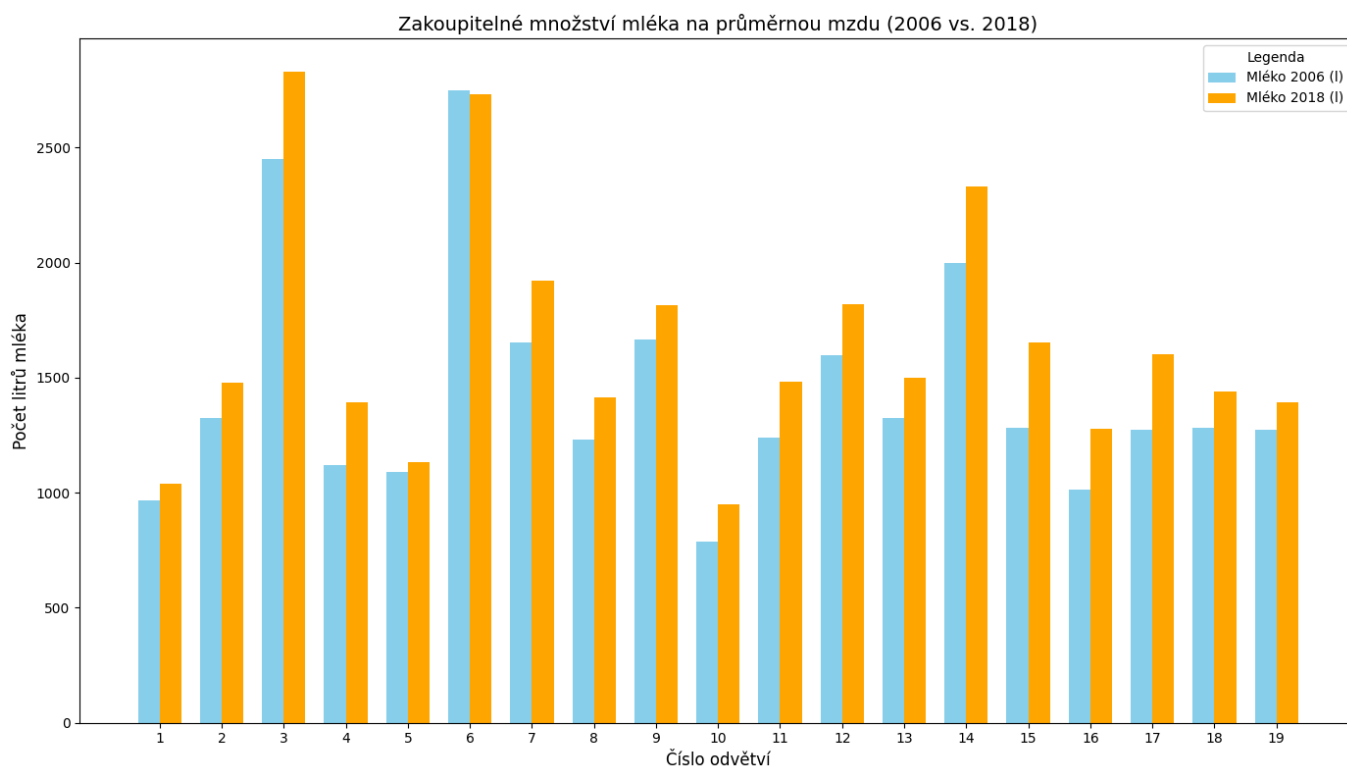
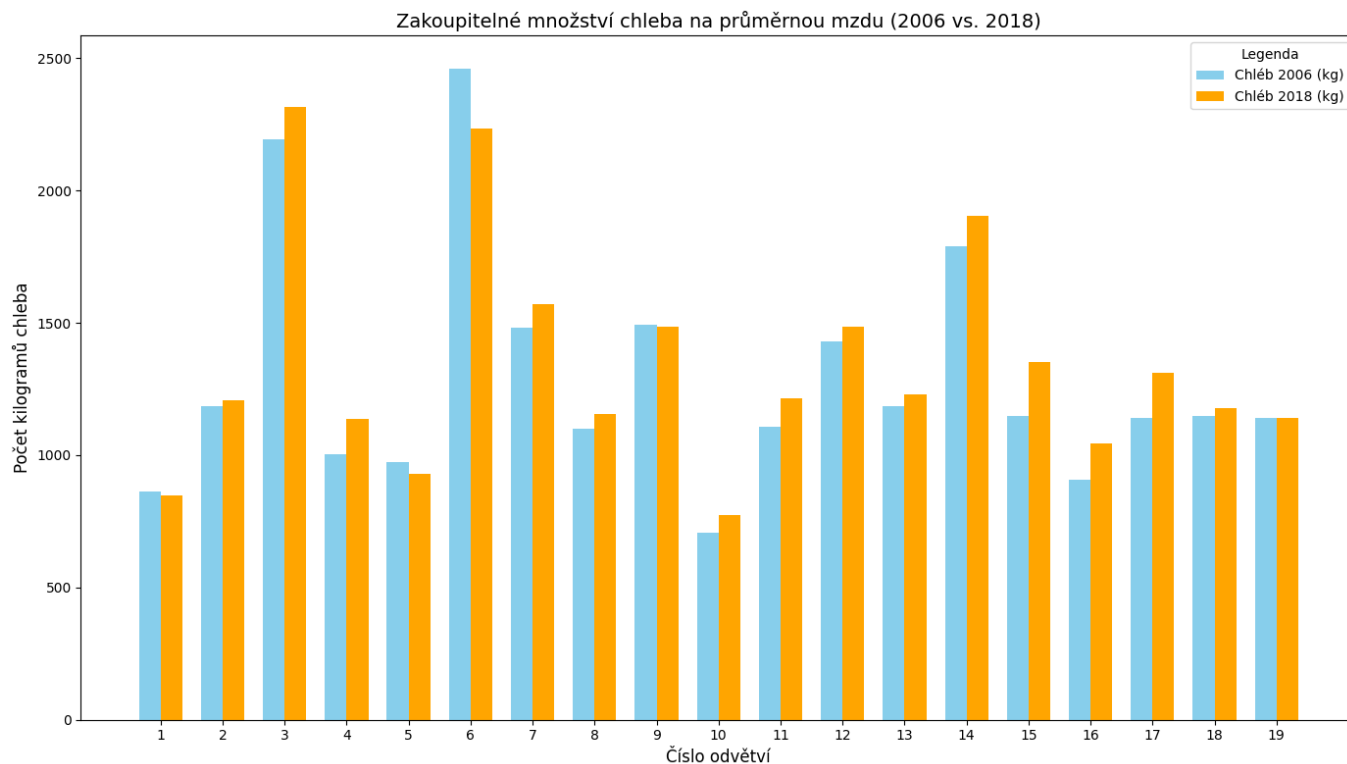
- V roce 2018 měla některá odvětví relativně nižší dostupnost obou základních potravin. Patří sem zejména:
  - Administrativní a podpůrné činnosti: Nejnižší dostupnost chleba (849 kg) i mléka (1038 litrů).
  - Ubytování, stravování a pohostinství: Druhá nejnižší dostupnost (775 kg chleba, 948 litrů mléka).
- Tato odvětví vykazují nižší kupní sílu pravděpodobně kvůli nižším průměrným mzdám.

### Zlepšení v ostatních odvětvích:

- Odvětví jako „Informační a komunikační činnosti“ nebo „Peněžnictví a pojišťovnictví“ zaznamenala významný nárůst kupní síly. Tyto sektory vykazují nejvyšší dostupnost chleba (2233 kg) i mléka (2731 litrů) v roce 2018, což ukazuje na vysoké mzdy.

**Závěr:**

Data ukazují, že většina odvětví zaznamenala nárůst kupní síly mezi lety 2006 a 2018, což odráží pozitivní ekonomický trend. Nicméně „Administrativní a podpůrné činnosti“ a „Ubytování, stravování a pohostinství“ zůstávají odvětvími s nejnižší dostupností základních potravin, což naznačuje nerovnoměrný růst mezd mezi sektory.



### 3. KTERÁ KATEGORIE POTRAVIN ZDRAŽUJE NEJPOMALEJI (JE U NÍ NEJNIŽŠÍ PERCENTUÁLNÍ MEZIROČNÍ NÁRŮST)?

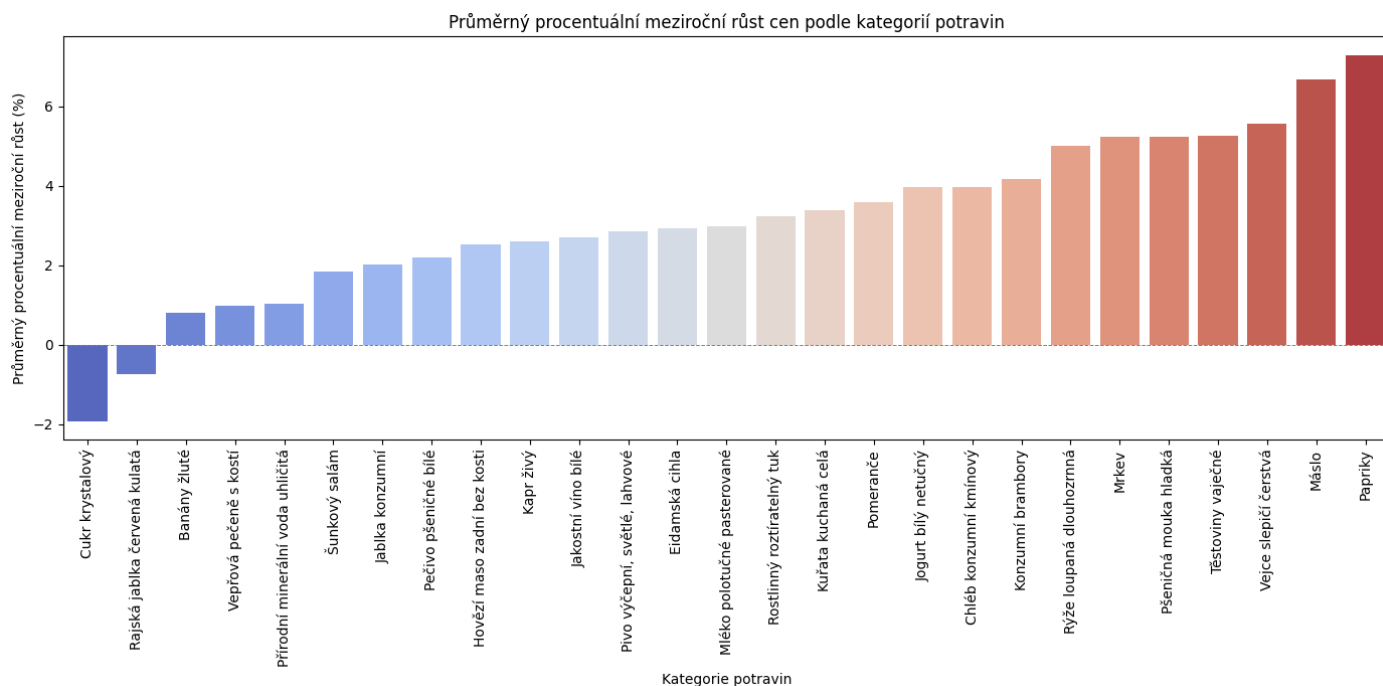
Pro analýzu byla použita tabulka `t_elena_kocumova_project_sql_primary_final`. Nejprve byla data seskupena podle produktů a roků, aby se spočítaly průměrné ceny. S pomocí funkce LAG byly získány ceny z předchozích let, což umožnilo vypočítat meziroční růst cen jako procentuální rozdíl mezi aktuální a předchozí hodnotou. Tyto meziroční změny byly poté použity pro výpočet průměrného růstu za celé sledované období.

Pro výpočet byl vytvořen SQL dotaz, který pracoval ve třech krocích. Nejprve byly vypočítány průměrné ceny produktů v jednotlivých letech a jejich meziroční změny pomocí funkce LAG. Následně byly tyto změny průměrovány pomocí funkce AVG, což umožnilo seřadit produkty podle jejich dlouhodobého tempa růstu nebo zlevňování. Nakonec byly výsledky seřazeny vzestupně, aby bylo možné snadno identifikovat produkty s nejpomalejším růstem.

Produkty byly seřazeny podle průměrného růstu:

- **Cukr krystalový** vykazuje nejnižší průměrný meziroční růst (**-1,92 %**), což znamená, že dlouhodobě zlevňuje.
- Další produkty, jako **rajská jablka červená kulatá** (-0,74 %) a **banány žluté** (0,81 %), vykazují velmi pomalý růst cen nebo zlevňování, ale v menší míře.
- Naopak nejrychleji zdražují produkty jako **papriky** (7,29 %) a **máslo** (6,67 %), které vykazují výrazné tempo růstu cen. Tyto rozdíly naznačují, že ceny některých potravinových kategorií se vyvíjejí mnohem rychleji než ostatní, což může být ovlivněno poptávkou, výrobními náklady nebo sezónními faktory.

Výsledky ukazují, že zatímco některé potraviny dlouhodobě zlevňují, jiné vykazují prudký růst cen.

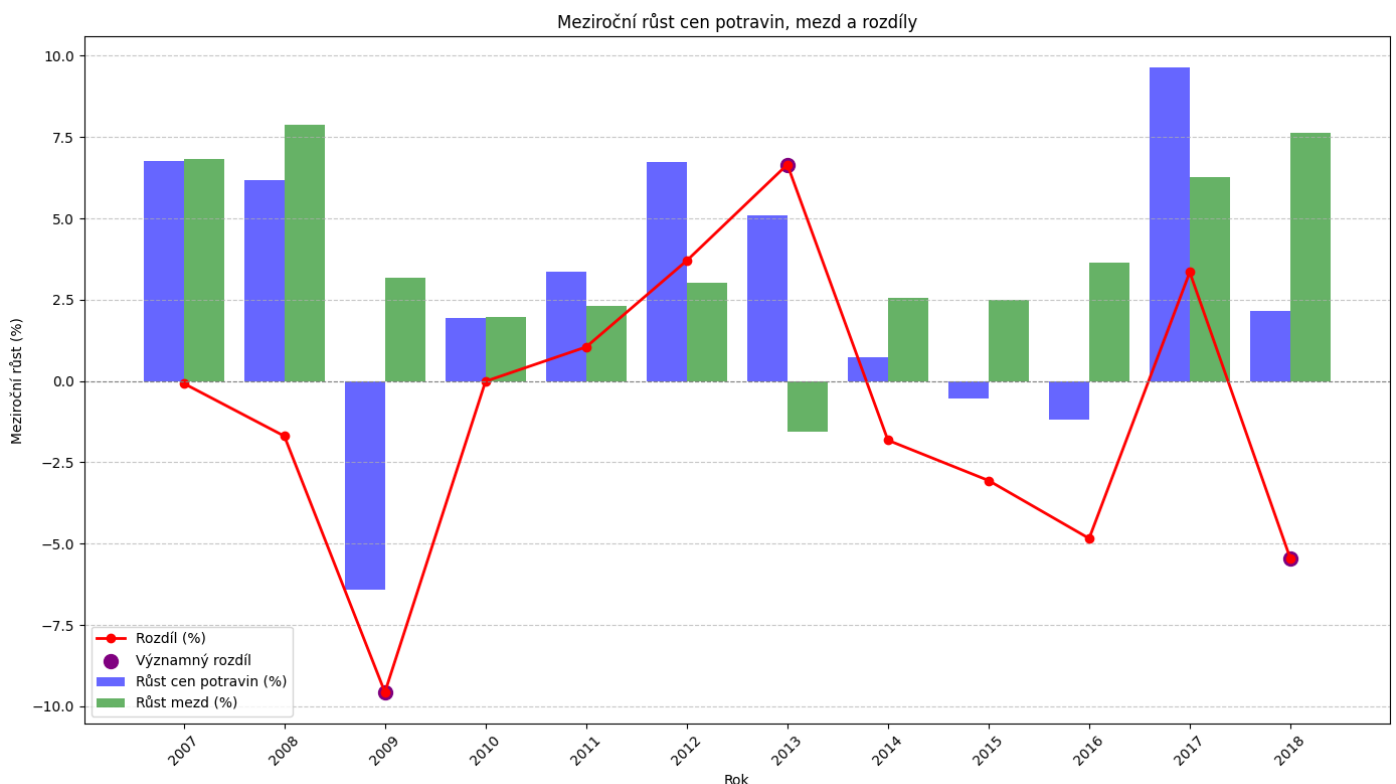


#### 4. EXISTUJE ROK, VE KTERÉM BYL MEZIROČNÍ NÁRŮST CEN POTRAVIN VÝRAZNĚ VYŠŠÍ NEŽ RŮST MEZD (VĚTŠÍ NEŽ 10 %)?

Pro analýzu otázky, zda existuje rok, byla použita tabulka `t_elena_kocumova_project_sql_primary_final`. Nejprve byla data seskupena podle roků, aby bylo možné vypočítat průměrné ceny potravin a mzdy za každý rok. Pomocí funkce LAG byly vypočítány meziroční změny cen potravin i mezd jako procentuální rozdíl mezi aktuální a předchozí hodnotou.

Analýza ukázala, že žádný rok nesplňuje podmínku růstu cen potravin výrazně vyššího než růst mezd o více než 10 %. Největší rozdíl byl zaznamenán v roce **2013**, kdy ceny potravin vzrostly o **5,1 %**, zatímco mzdy poklesly o **-1,56 %**, což vedlo k rozdílu **6,66 %**. Tento výsledek ukazuje, že růst cen potravin sice v některých letech převýšil růst mezd, ale nikdy ne v tak výrazném rozsahu.

Grafické zpracování dat ukázalo jasné rozdíly mezi růstem cen potravin a mezd v průběhu let. Křivka zobrazující rozdíly mezi těmito hodnotami odhalila, že růst cen potravin byl občas vyšší, avšak nikdy nepřekročil hranici 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?

Pro analýzu této otázky byly použity dvě tabulky: `t_elena_kocumova_project_sql_primary_final` a `t_elena_kocumova_project_sql_secondary_final`. Data byla omezena pouze na Českou republiku a seskupena podle roků, aby bylo možné vypočítat průměrné hodnoty HDP, cen potravin a mezd. Pomocí funkce **LAG** byly získány hodnoty z předchozích let, na jejichž základě byly spočítány meziroční procentuální změny. Tyto změny byly následně porovnány, aby se zjistilo, zda výrazný růst HDP ovlivňuje růst cen potravin a mezd ve stejném nebo následujícím roce.

### Výsledky analýzy:

- Rok 2007:** HDP vzrostlo o **5,57 %**, ceny potravin o **6,76 %** a mzdy o **6,84 %**.  
Tento rok ukazuje možnou souvislost mezi růstem HDP, cen potravin a mezd, kdy všechny hodnoty rostly v podobné míře.
- Rok 2009:** HDP pokleslo o **-4,66 %**, ceny potravin klesly o **-6,41 %**, ale mzdy vzrostly o **3,16 %**.  
Tento rok naznačuje, že mzdy nejsou vždy přímo ovlivněny změnami HDP, což se může lišit v závislosti na dalších faktorech.
- Rok 2013:** HDP kleslo (**-0,05 %**), ceny potravin vzrostly o **5,1 %**, ale mzdy klesly o **-1,56 %**.  
Tento rok ukazuje, že růst HDP nemusí být vždy spojen s růstem mezd nebo cen potravin.
- Rok 2015:** HDP vzrostlo o **5,39 %**, mzdy mírně vzrostly o **0,55 %**, ale ceny potravin klesly o **-0,55 %**.  
Tento rok naznačuje, že růst HDP nemusí automaticky vést k růstu cen potravin, pokud mzdy nerostou dostatečně rychle.
- Rok 2017:** HDP vzrostlo o **5,39 %**, mzdy o **6,28 %** a ceny potravin o **2,54 %**.  
Tento rok potvrzuje, že růst HDP může být často spojen s růstem mezd i cen potravin.

### Shrnutí:

Výsledky analýzy ukazují, že výška HDP má vliv na mzdy a ceny potravin, ale tento vztah není vždy přímý. V některých letech, jako v roce 2007 a 2017, byl růst HDP spojen s růstem mezd i cen potravin, zatímco v roce 2015 růst HDP nevedl k růstu cen potravin kvůli nízkému růstu mezd.

