

Datová akademie

# Projekt SQL

V Olomouci dne 28.04.2024

Zpracovala Petra Jungová

Email: [pjungova@seznam.cz](mailto:pjungova@seznam.cz)

Nick Github: Pjungova

# Obsah

Úvod .....	3
Tvorba tabulek .....	4
Výzkumné otázky .....	6
Závěr .....	10

# Úvod

Projekt se zaměřuje na zkoumání dostupnosti základních potravin široké veřejnosti a jejich vztah k růstu mezd v různých odvětvích. Dále se soustředí na změny cen potravin, životní úroveň a vliv HDP na mzdy a ceny potravin.

Pro získání dat jsou použity statistické zdroje, konkrétně tabulky 'czechia\_payroll' a 'czechia\_price', které obsahují relevantní informace o mzdách, cenách potravin a dalších ekonomických ukazatelích. Další pomocné tabulky, jako jsou 'czechia\_payroll\_calculation', 'czechia\_payroll\_industry\_branch', 'czechia\_payroll\_unit' a 'czechia\_payroll\_value\_type', poskytují informace o kalkulacích, odvětvích, jednotkách a typech hodnot v tabulce 'czechia\_payroll'. Pro kategorie potravin je k dispozici přidružená tabulka 'czechia\_price\_category'. Pro získání informací o HDP a dalších ekonomických ukazatelích je použita tabulka 'economies', která obsahuje údaje o ekonomickém vývoji, včetně HDP, GINI indexu, daňové zátěže a dalších relevantních informací.

K dosažení cílů projektu jsou využity SQL dotazy pro extrakci, transformaci a analýzu dat. Výsledky těchto analýz poskytnou odpovědi na stanovené 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?

Kromě odpovědí na výzkumné otázky bude výsledkem projektu vytvoření dvou tabulek. První z těchto tabulek bude obsahovat data o mzdách a cenách potravin v České republice, která budou sloučena do porovnatelných období. Druhá tabulka poskytne doplňující informace o dalších evropských státech. Pro vytvoření těchto tabulek budou k dispozici samostatné SQL soubory, stejně jako pro každou odpověď na jednotlivou výzkumnou otázku.

# Tvorba tabulek

Proces tvorby tabulky byl zaměřen na sběr a zpracování klíčových dat, která umožní odpovědět na stanovené výzkumné otázky. Primární tabulka obsahuje podrobnosti o mzdách, cenách potravin a jejich změnách v čase. Sekundární tabulka poskytuje informace o HDP a dalších ekonomických ukazatelích v různých zemích. Tyto tabulky slouží jako základ analytického procesu, který umožní porozumět vzájemným vztahům mezi různými ekonomickými faktory.

## Primární tabulka

Nová tabulka s názvem 't\_petra\_jungova\_project\_sql\_primary\_final' byla vytvořena pomocí příkazu CREATE TABLE a je klíčovým prvkem pro analýzu dat nezbytných k zodpovězení výzkumných otázek. Zdrojem informací o mzdách a cenách potravin v České republice jsou tabulky 'czechia\_payroll' a 'czechia\_price'.

Pro každou výzkumnou otázku byla identifikována nezbytná data. Základní informace o mzdách byly získány z tabulky 'czechia\_payroll', kde byly filtrovány hodnoty na průměrnou hrubou mzdu na zaměstnance (value\_type\_code = '5958'). Potřebná data o cenách potravin byla vybrána z tabulky 'czechia\_price'. Pro sledování změn cen v čase byl zvolen sloupec date\_from.

Pomocí klauzule WITH byly vytvořeny čtyři dočasné tabulky: 'previous\_price', 'actualy\_price', 'previous\_payroll' a 'actualy\_payroll', které obsahují průměrné ceny a mzdy za aktuální a předchozí roky. Pro zjištění průměrných hodnot byla použita funkce AVG. Data z předchozích let byla získána přičtením jedné jednotky k roku. Což umožnilo porovnání s daty aktuálních let a provedení analýzy změn v cenách potravin mezi dvěma po sobě jdoucími roky.

Pro propojení všech dočasných tabulek a získání společných dat byla využita klauzule LEFT JOIN s podmínkou společných roků z tabulek 'czechia\_price' a 'czechia\_payroll'. Výsledná tabulka obsahuje relevantní informace pro analýzu, včetně názvů kategorií potravin a odvětví.

## Výsledná tabulka obsahuje následující sloupce:

- year\_price\_payrall – rok potravin a mezd
- category – název kategorie potravin
- price\_value – ceny potravin
- price\_unit – jednotka potravin
- avg\_price – průměrná cena potravin
- avg\_price\_previous – průměrná cena potravin z předchozího roku

- `percentage_change_price` – procentuální změna cen potravin
- `industry_branch` – odvětví
- `avg_payroll` – průměrná mzda
- `avg_payroll_previous` – průměrná mzda z předchozího roku
- `percentage_change_payroll` – procentuální změna mezd

Pro výpočet procentuálních změny byl využit vzorec:

$$= \frac{(\text{hodnota v aktuálním roce} - \text{hodnota v předchozím roce})}{\text{hodnota v předchozím roce}} \times 100$$

## Sekundární tabulka

Pro vytvoření sekundární tabulky byla využita tabulka 'economies', která obsahuje informace o HDP, GINI a dalších ekonomických ukazatelích různých zemí a let.

Nejprve byla pomocí klauzule `WITH` vytvořena dočasná tabulka 'previous\_economies' s informacemi o ekonomikách z předchozích let. Poté byl proveden hlavní dotaz, ve kterém byla tato dočasná tabulka spojena s tabulkou 'economies' pomocí funkce `JOIN` a klíčů země a rok.

Pro výpočet procentuální změny HDP mezi aktuálním a předchozím rokem byly využity hodnoty z obou tabulek. Výsledná tabulka obsahuje sloupce se zemí, rokem, aktuálním HDP, HDP z předchozího roku a procentuální změnou HDP, zaokrouhlenou na dvě desetinná místa.

# Výzkumné otázky

Výzkumné otázky se zaměřují na vztahy mezi mzdami, cenami potravin a ekonomickými ukazateli. První otázka se zabývá trendem mzdového vývoje v různých odvětvích, zatímco druhá zkoumá kupní sílu mzdy v průběhu času. Třetí otázka se vztahuje ke změně cen potravin v jednotlivých kategoriích, kdežto čtvrtá zkoumá, zda existuje rok, kdy byl meziroční nárůst cen potravin vyšší než růst mezd. Poslední otázka se zabývá vlivem výšky HDP na změny mezd a cen potravin. Otázky vedou k hlubšímu porozumění ekonomických faktorů a jejich vzájemných vztahů.

## První výzkumná otázka

Pro zodpovězení otázky, zda v průběhu let rostou mzdy ve všech odvětvích, nebo v některých klesají, byl vytvořen vnitřní dotaz. Pomocí něj došlo ke seskupení dat dle roku a odvětví a vytvoření nového sloupce payroll\_change. Ten odpovídá na otázku, zda průměrná mzda v daném roce vzrostla, klesla nebo zůstala stejná ve srovnání s předchozím rokem. V externím dotazu byla použita podmínka HAVING, která vyfiltrovala řádky tak, aby zůstaly pouze ty, kde je payroll\_change "Klesající". Díky tomu bylo zjištěno, že mzdy nerostou ve všech obdobích, ale také klesají.

123 year_price_payroll	ABC industry_branch	ABC payroll_change
2,009	Ubytování, stravování a pohostinství	Klesající
2,009	Těžba a dobývání	Klesající
2,009	Zemědělství, lesnictví, rybářství	Klesající
2,010	Vzdělávání	Klesající
2,010	Veřejná správa a obrana; povinné sociální zabezpečení	Klesající
2,010	Profesní, vědecké a technické činnosti	Klesající
2,011	Veřejná správa a obrana; povinné sociální zabezpečení	Klesající
2,011	Ubytování, stravování a pohostinství	Klesající
2,011	Kulturní, zábavní a rekreační činnosti	Klesající
2,013	Peněžnictví a pojišťovnictví	Klesající
2,013	Kulturní, zábavní a rekreační činnosti	Klesající
2,013	Velkoobchod a maloobchod; opravy a údržba motorových vozidel	Klesající
2,013	Těžba a dobývání	Klesající
2,013	Administrativní a podpůrné činnosti	Klesající
2,013	Profesní, vědecké a technické činnosti	Klesající
2,013	Informační a komunikační činnosti	Klesající
2,013	Zásobování vodou; činnosti související s odpady a sanacemi	Klesající
2,013	Činnosti v oblasti nemovitostí	Klesající
2,013	Výroba a rozvod elektřiny, plynu, tepla a klimatiz. vzduchu	Klesající
2,013	Stavebnictví	Klesající
2,014	Těžba a dobývání	Klesající
2,015	Výroba a rozvod elektřiny, plynu, tepla a klimatiz. vzduchu	Klesající
2,016	Těžba a dobývání	Klesající

## Druhá výzkumná otázka

Pro zodpovězení otázky, kolik je možné si koupit litrů mléka a kilogramů chleba za první a poslední srovnatelné období v dostupných datech cen a mezd, byla nejprve vytvořena dočasná tabulka 'first\_last\_prices', která obsahovala data o cenách v prvním a posledním roce. Tato tabulka byla následně propojena s tabulkou primárních dat k získání důležitých informací o průměrných mzdách a cenách potravin pro první a poslední rok v kategoriích chléb a mléko. Poté, co byla průměrná mzda vydělena průměrnou cenou v prvním a posledním roce, byla získána potřebná odpověď na výzkumnou otázku.

Počítáním výsledků pro každé z 19 různých odvětví, které ovlivňují výši mzdy, bylo získáno celkem 38 výsledků pro chleba a 38 výsledků pro mléko, což činí celkem 76 výsledků. Například z výsledné tabulky je patrné, že v prvním období bylo možné zakoupit 864,08 kg chleba při mzdě 13,929 Kč v administrativní a podpůrné činnosti a ceně chleba 16,12 Kč. V posledním srovnatelném období, kdy měl chléb cenu 24,24 Kč a mzda v daném odvětví dosáhla 20.559,5 Kč, bylo možné zakoupit méně chleba, a to 848,16 kg. Za cenu 14,44 Kč za litr v prvním srovnatelném období a dané mzdě bylo možné zakoupit 964,61 litrů mléka. V posledním období došlo k nárůstu výše mzdy, stejně jako ceny mléka na 19,82 Kč, což umožnilo zakoupit 1.037,31 litrů mléka.

84c industry_branch	123 payroll_first_year	123 payroll_last_year	84c category	123 price_first_year	123 price_last_year	123 quantity_first_year	123 quantity_last_year
Administrativní a podpůrné činnosti	13,929	20,559.5	Chléb konzumní kmínový	16.12	24.24	864.08	848.16
Administrativní a podpůrné činnosti	13,929	20,559.5	Mléko polotučné pasterované	14.44	19.82	964.61	1,037.31
Činnosti v oblasti nemovitostí	18,396.38	27,584.25	Mléko polotučné pasterované	14.44	19.82	1,273.99	1,391.74
Činnosti v oblasti nemovitostí	18,396.38	27,584.25	Chléb konzumní kmínový	16.12	24.24	1,141.21	1,137.96
Doprava a skladování	19,125.25	29,297.63	Mléko polotučné pasterované	14.44	19.82	1,324.46	1,478.19
Doprava a skladování	19,125.25	29,297.63	Chléb konzumní kmínový	16.12	24.24	1,186.43	1,208.65
Informační a komunikační činnosti	35,367.5	56,101.25	Mléko polotučné pasterované	14.44	19.82	2,449.27	2,830.54
Informační a komunikační činnosti	35,367.5	56,101.25	Chléb konzumní kmínový	16.12	24.24	2,194.01	2,314.41
Kulturní, zábavní a rekreační činnosti	16,152.38	27,580.25	Mléko polotučné pasterované	14.44	19.82	1,118.59	1,391.54
Kulturní, zábavní a rekreační činnosti	16,152.38	27,580.25	Chléb konzumní kmínový	16.12	24.24	1,002.01	1,137.8
Ostatní činnosti	15,719.38	22,473	Mléko polotučné pasterované	14.44	19.82	1,088.6	1,133.85
Ostatní činnosti	15,719.38	22,473	Chléb konzumní kmínový	16.12	24.24	975.15	927.1
Peněžnictví a pojištnictví	39,690.13	54,111.38	Mléko polotučné pasterované	14.44	19.82	2,748.62	2,730.14
Peněžnictví a pojištnictví	39,690.13	54,111.38	Chléb konzumní kmínový	16.12	24.24	2,462.17	2,232.32
Profesní, vědecké a technické činnosti	23,875.38	38,073.88	Mléko polotučné pasterované	14.44	19.82	1,653.42	1,920.98
Profesní, vědecké a technické činnosti	23,875.38	38,073.88	Chléb konzumní kmínový	16.12	24.24	1,481.1	1,570.7
Stavebnictví	17,747	27,994.13	Mléko polotučné pasterované	14.44	19.82	1,229.02	1,412.42
Stavebnictví	17,747	27,994.13	Chléb konzumní kmínový	16.12	24.24	1,100.93	1,154.87
Těžba a dobývání	24,042.13	35,990.88	Mléko polotučné pasterované	14.44	19.82	1,664.97	1,815.89
Těžba a dobývání	24,042.13	35,990.88	Chléb konzumní kmínový	16.12	24.24	1,491.45	1,484.77
Ubytování, stravování a pohostinství	11,390	18,769.88	Mléko polotučné pasterované	14.44	19.82	788.78	947.02
Ubytování, stravování a pohostinství	11,390	18,769.88	Chléb konzumní kmínový	16.12	24.24	706.58	774.33
Velkoobchod a maloobchod; opravy	17,864.88	29,404.5	Mléko polotučné pasterované	14.44	19.82	1,237.18	1,483.58
Velkoobchod a maloobchod; opravy	17,864.88	29,404.5	Chléb konzumní kmínový	16.12	24.24	1,108.24	1,213.06
Veřejná správa a obrana; povinné soc	23,068.63	36,013.88	Mléko polotučné pasterované	14.44	19.82	1,597.55	1,817.05
Veřejná správa a obrana; povinné soc	23,068.63	36,013.88	Chléb konzumní kmínový	16.12	24.24	1,431.06	1,485.72
Výroba a rozvod elektřiny, plynu, tepla	28,844.13	46,178.5	Mléko polotučné pasterované	14.44	19.82	1,997.52	2,329.89
Výroba a rozvod elektřiny, plynu, tepla	28,844.13	46,178.5	Chléb konzumní kmínový	16.12	24.24	1,789.34	1,905.05
Vzdělávání	19,113.5	29,745	Mléko polotučné pasterované	14.44	19.82	1,323.65	1,500.76
Vzdělávání	19,113.5	29,745	Chléb konzumní kmínový	16.12	24.24	1,185.7	1,227.1

## Třetí výzkumná otázka

Pro zodpovězení otázky, která kategorie potravin zdražuje nejpomaleji, byla vypočítána průměrná hodnota procentuální změny ceny potravin pro každou kategorii. Tyto výsledné průměrné hodnoty byly seskupeny podle kategorií potravin a seřazeny vzestupně. Nakonec byla použita klauzule LIMIT 1, aby se vrátil pouze první řádek. Výsledek tedy obsahuje pouze kategorii potravin

s nejnižším průměrným procentuálním meziročním nárůstem. Výsledkem výzkumné otázky je zjištění, že cukr krystal zdražuje nejpomaleji.

### Čtvrtá výzkumná otázka

Pro zodpovězení otázky, zda existuje rok, ve kterém byl meziroční nárůst cen potravin vyšší než růst mezd, byly nejprve vytvořeny dvě dočasné tabulky, které byly následně propojeny. První dočasná tabulka 'avg\_price\_payroll' obsahovala průměrné procentuální změny cen potravin a mezd pro každý rok, seskupené podle roku. Druhá tabulka 'price\_payroll\_growth' počítala rozdíl mezi průměrnými změnami cen potravin a mezd pro každý rok. Spojením těchto tabulek bylo možné zjistit, zda v daném roce byl meziroční nárůst cen potravin vyšší než růst mezd. Z výsledků vyplývá, že v žádném roce nebyl meziroční nárůst cen potravin výrazně vyšší než růst mezd.

<sup>123</sup> year_price_payroll	<sup>123</sup> avg_price_change	<sup>123</sup> avg_payroll_change	<sup>123</sup> price_vs_payroll_growth	<sup>abc</sup> price_vs_payroll_growth_10_percent
2,007	9.26	6.91	2.35	Ne
2,008	8.92	7.24	1.68	Ne
2,009	-6.59	2.97	-9.56	Ne
2,010	1.52	2.17	-0.65	Ne
2,011	4.84	2.24	2.6	Ne
2,012	7.47	2.72	4.75	Ne
2,013	6.01	-0.78	6.79	Ne
2,014	-0.62	2.52	-3.14	Ne
2,015	-0.69	2.84	-3.53	Ne
2,016	-1.41	3.95	-5.36	Ne
2,017	7.06	6.57	0.49	Ne
2,018	2.41	7.78	-5.37	Ne

### Pátá výzkumná otázka

Pro zodpovězení otázky, zda má výška HDP vliv na změny ve mzdách a cenách potravin, byly vytvořeny dvě dočasné tabulky. Tabulka 'GDP\_change' obsahovala údaje o procentuálních změnách HDP pro Českou republiku. Druhá dočasná tabulka 'price\_payroll\_change' obsahovala průměrné procentuální změny cen potravin a mezd pro každý rok. Tyto tabulky byly propojeny na základě roku. Z výsledných dat nelze jednoznačně určit, zda má výška HDP vliv na změny ve mzdách a cenách potravin. Například pokud se podíváme na procentuální změny HDP vyšší než 5 %, kterých ve výsledné tabulce nalezneme tři, není možné jednotně určit, jaký vliv má zvýšení HDP na stejný či následující rok.



<sup>123</sup> year_price_payroll ▼	<sup>123</sup> avg_price_change ▼	<sup>123</sup> avg_payroll_change ▼	<sup>123</sup> percentage_change_GDP ▼
2,007	9.26	6.91	5.57
2,008	8.92	7.24	2.69
2,009	-6.59	2.97	-4.66
2,010	1.52	2.17	2.43
2,011	4.84	2.24	1.76
2,012	7.47	2.72	-0.79
2,013	6.01	-0.78	-0.05
2,014	-0.62	2.52	2.26
2,015	-0.69	2.84	5.39
2,016	-1.41	3.95	2.54
2,017	7.06	6.57	5.17
2,018	2.41	7.78	3.2

# Závěr

V rámci projektu byla provedena komplexní analýza dat zaměřená na zkoumání dostupnosti základních potravin pro širokou veřejnost a jejich vztah k růstu mezd v různých odvětvích. Dále došlo ke zkoumání změn cen potravin, životní úroveň a vlivu HDP.

Na základě statistických zdrojů a využití SQL dotazů byly vytvořeny dvě klíčové tabulky obsahující data o mzdách a cenách potravin v České republice a ekonomických ukazatelích ve světě. Tyto tabulky poskytly ucelený přehled, který umožnil získat odpovědi na stanovené výzkumné otázky, které se týkaly trendů mzdového vývoje, kupní síly mzdy v průběhu času, změn cen potravin, existujících rozdílů mezi růstem cen a mezd a vlivu HDP na změny mezd a cen potravin.

Z analýzy dat vyplývá, že mzdy neustále nerostou ve všech obdobích, ale také klesají. Došlo ke zjištění, že kupní síla mzdy se v průběhu času mírně zvyšuje, avšak zároveň se zvyšují i ceny potravin. Zajímavě, nejpomaleji zdražuje kategorie cukru krystal. Dále se nepodařil potvrdit výrazný rozdíl mezi meziročním růstem cen potravin a mezd. Analýza neposkytla jasný důkaz vlivu HDP na změny mezd a cen potravin.