PRŮVODNÍ LISTINA PRO SQL PROJEKT

Tento dokument popisuje jednotlivé kroky a mezivýsledky při přípravě datových podkladů pro výzkumné otázky týkající se dostupnosti základních potravin a průměrných příjmů v České republice. Projekt byl realizován pomocí SQL dotazů na databázi MariaDB.

**Krok 1: Příprava primární tabulky pro Českou republiku**

1. **Vytvoření tabulky *t\_NgocKhanhVy\_Tranova\_project\_SQL\_primary\_final***

**CREATE** **TABLE** t\_NgocKhanhVy\_Tranova\_project\_SQL\_primary\_final (

**year** **INT**,

industry\_branch\_id **CHAR**(1),

industry\_branch **VARCHAR**(255),

avg\_payroll **INT**,

food\_category **VARCHAR**(255),

food\_price **INT**

);

1. **Vytvoření dočasné tabulky pro data průměrných mezd v ČR**

**CREATE** **TEMPORARY** **TABLE** temp\_payroll **AS**

**SELECT**

cpay.payroll\_year **AS** **year**,

cpay.industry\_branch\_code **AS** industry\_branch\_id,

cpib.name **AS** industry\_branch,

**AVG**(cpay.value) **AS** avg\_payroll

**FROM** czechia\_payroll cpay

**JOIN** czechia\_payroll\_industry\_branch cpib

**ON** cpay.industry\_branch\_code = cpib.code

**JOIN** czechia\_payroll\_value\_type cpvt

**ON** cpay.value\_type\_code = cpvt.code

**JOIN** czechia\_payroll\_calculation cpcal

**ON** cpay.calculation\_code = cpcal.code

**WHERE**

cpay.value\_type\_code = 5958 **AND** cpay.calculation\_code = 200

**GROUP** **BY**

cpay.payroll\_year, cpay.industry\_branch\_code, cpib.name;

1. **Vytvoření dočasné tabulky pro data o cenách potravin v ČR**

**CREATE** **TEMPORARY** **TABLE** temp\_food **AS**

**SELECT**

**YEAR**(cp.date\_from) **AS** **year**,

cpc.name **AS** food\_category,

**AVG**(cp.value) **AS** food\_price

**FROM** czechia\_price cp

**JOIN** czechia\_price\_category cpc

**ON** cp.category\_code = cpc.code

**GROUP** **BY**

**YEAR**(cp.date\_from), cpc.name;

1. **Sjednocení dat o mzdách a cenách z dočasných tabulek do primární tabulky**

**INSERT** **INTO** t\_NgocKhanhVy\_Tranova\_project\_SQL\_primary\_final (**year**, industry\_branch\_id, industry\_branch, avg\_payroll, food\_category, food\_price)

**SELECT**

tp.**year**,

*tp*.industry\_branch\_id,

*tp*.industry\_branch,

*tp*.avg\_payroll,

*tf*.food\_category,

*tf*.food\_price

**FROM** ***temp\_payroll*** *tp*

**LEFT** **JOIN** ***temp\_food*** *tf*

**ON** tp.**year** = tf.**year**

**UNION**

**SELECT**

tf.**year**,

*tp*.industry\_branch\_id,

*tp*.industry\_branch,

*tp*.avg\_payroll,

*tf*.food\_category,

*tf*.food\_price

**FROM** ***temp\_food*** *tf*

**LEFT** **JOIN** ***temp\_payroll*** *tp*

**ON** tf.**year** = tp.**year**;

**Krok 2: Výpočty a analýza pro výzkumné otázky**

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

**SELECT**

**`year` ,**

**industry\_branch\_id,**

**industry\_branch,**

**avg\_payroll**

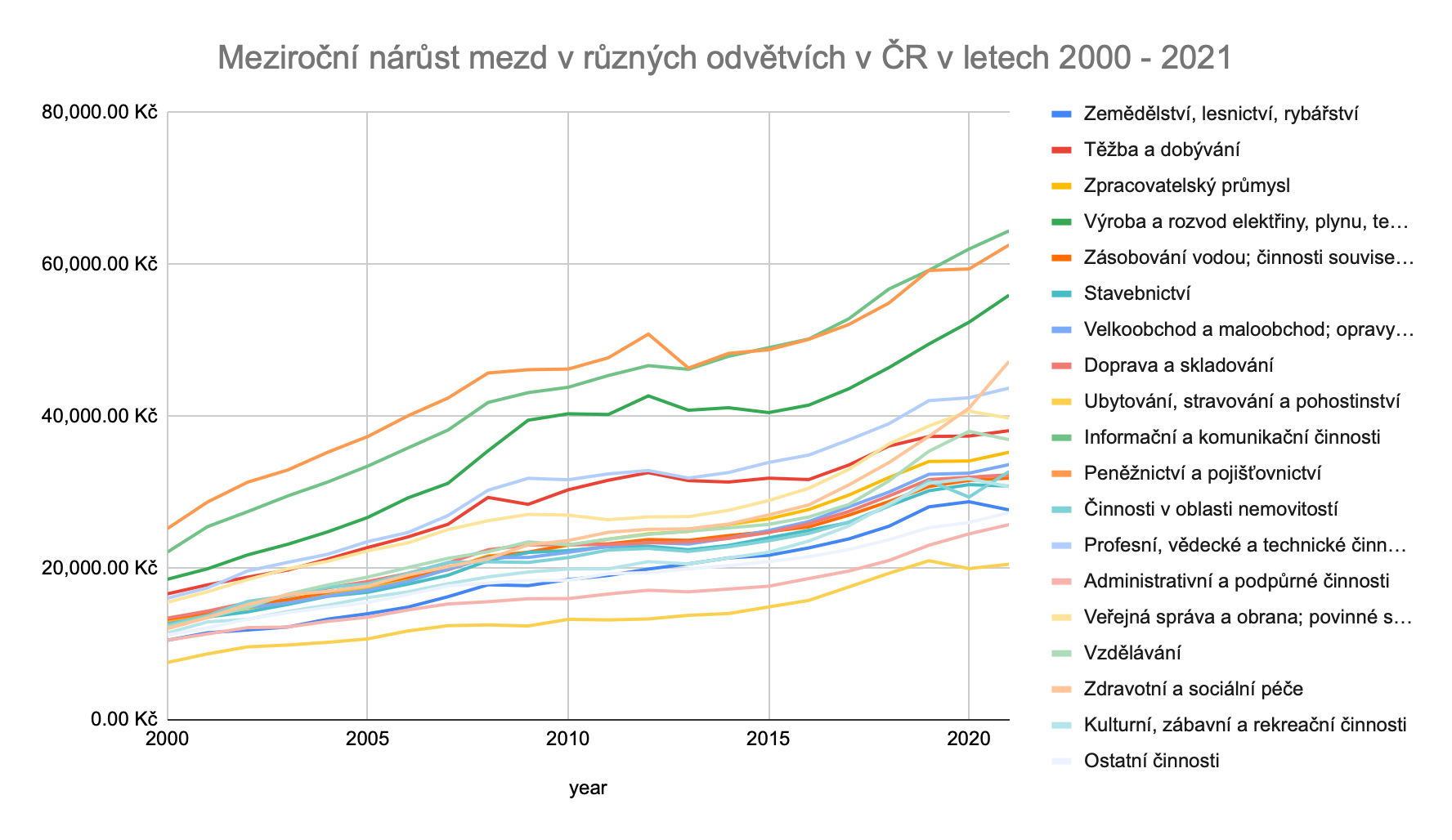
**FROM t\_NgocKhanhVy\_Tranova\_project\_SQL\_primary\_final *t***

**GROUP BY industry\_branch\_id , `year`**

**ORDER BY industry\_branch\_id , `year`;**

**Odpověď:** Data ukazují, že mzdy rostly ve všech uvedených odvětvích od roku 2000 do roku 2021. Nicméně v některých letech došlo k poklesu nebo stagnaci.

* V roce 2021 zažila poklesy tyto odvětví:
  + Zemědělství, lesnictví, rybářství
  + Stavebnictví
  + Veřejná správa a obrana; povinné sociální zabezpečení
  + Kulturní, zábavní a rekreační činnosti

****

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

* **Zjištění prvního a posledního srovnatelného období:**

**SELECT**

**MIN**(`year`) **AS** *first\_year*, **MAX**(`year`) **AS** *last\_year*

**FROM** t\_NgocKhanhVy\_Tranova\_project\_SQL\_primary\_final *t*

**WHERE** avg\_payroll **IS** **NOT** **NULL** **AND** food\_price **IS** **NOT** **NULL**;

* **První srovnatelné období 2006:**

**SELECT**

(**SELECT** **AVG**(avg\_payroll)

**FROM** t\_NgocKhanhVy\_Tranova\_project\_SQL\_primary\_final *t*

**WHERE** `year` = 2006)/

(**SELECT** **AVG**(food\_price)

**FROM** t\_NgocKhanhVy\_Tranova\_project\_SQL\_primary\_final *t*

**WHERE** food\_category = **'Mléko polotučné pasterované'** **AND** `year` = 2006)

**SELECT**

(**SELECT** **AVG**(avg\_payroll)

**FROM** t\_NgocKhanhVy\_Tranova\_project\_SQL\_primary\_final *t*

**WHERE** `year` = 2006)/

(**SELECT** **AVG**(food\_price)

**FROM** t\_NgocKhanhVy\_Tranova\_project\_SQL\_primary\_final *t*

**WHERE** food\_category = **'Chléb konzumní kmínový'** **AND** `year` = 2006)

* **Poslední srovnatelné období 2018:**

**SELECT**

(**SELECT** **AVG**(avg\_payroll)

**FROM** t\_NgocKhanhVy\_Tranova\_project\_SQL\_primary\_final *t*

**WHERE** `year` = 2018)/

(**SELECT** **AVG**(food\_price)

**FROM** t\_NgocKhanhVy\_Tranova\_project\_SQL\_primary\_final *t*

**WHERE** food\_category = **'Mléko polotučné pasterované'** **AND** `year` = 2018)

**SELECT**

(**SELECT** **AVG**(avg\_payroll)

**FROM** t\_NgocKhanhVy\_Tranova\_project\_SQL\_primary\_final *t*

**WHERE** `year` = 2018)/

(**SELECT** **AVG**(food\_price)

**FROM** t\_NgocKhanhVy\_Tranova\_project\_SQL\_primary\_final *t*

**WHERE** food\_category = **'Chléb konzumní kmínový'** **AND** `year` = 2018)

**Odpověď:** Za průměrnou mzdu v roce 2006 bylo možné koupit 1 511 litrů mléka a 1 322 kg chleba a v roce 2018 1 654 litrů mléka a 1 378 kg chleba

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

* **Vytvoření dočasné tabulky pro průměrné ceny potravin podle roku a kategorie:**

**CREATE TEMPORARY TABLE temp\_avg\_price AS**

**SELECT**

**year,**

**food\_category,**

**AVG(food\_price) AS avg\_price**

**FROM t\_NgocKhanhVy\_Tranova\_project\_SQL\_primary\_final t**

**WHERE food\_category IS NOT NULL**

**GROUP BY food\_category, year;**

* **Vytvoření dočasné tabulky pro meziroční změny cen v procentech:**

**CREATE TEMPORARY TABLE temp\_yearly\_changes AS**

**SELECT**

**cur.year,**

**cur.food\_category,**

**cur.avg\_price,**

**prev.avg\_price AS prev\_avg\_price,**

**((cur.avg\_price - prev.avg\_price) / prev.avg\_price) \* 100 AS yearly\_change**

**FROM temp\_avg\_price cur**

**JOIN temp\_avg\_price prev**

**ON cur.food\_category = prev.food\_category AND cur.year = prev.year + 1;**

* **Výběr kategorie potravin s nejnižším průměrným meziročním nárůstem:**

**SELECT**

***food\_category*,**

**AVG(*yearly\_change*) AS *avg\_yearly\_increase***

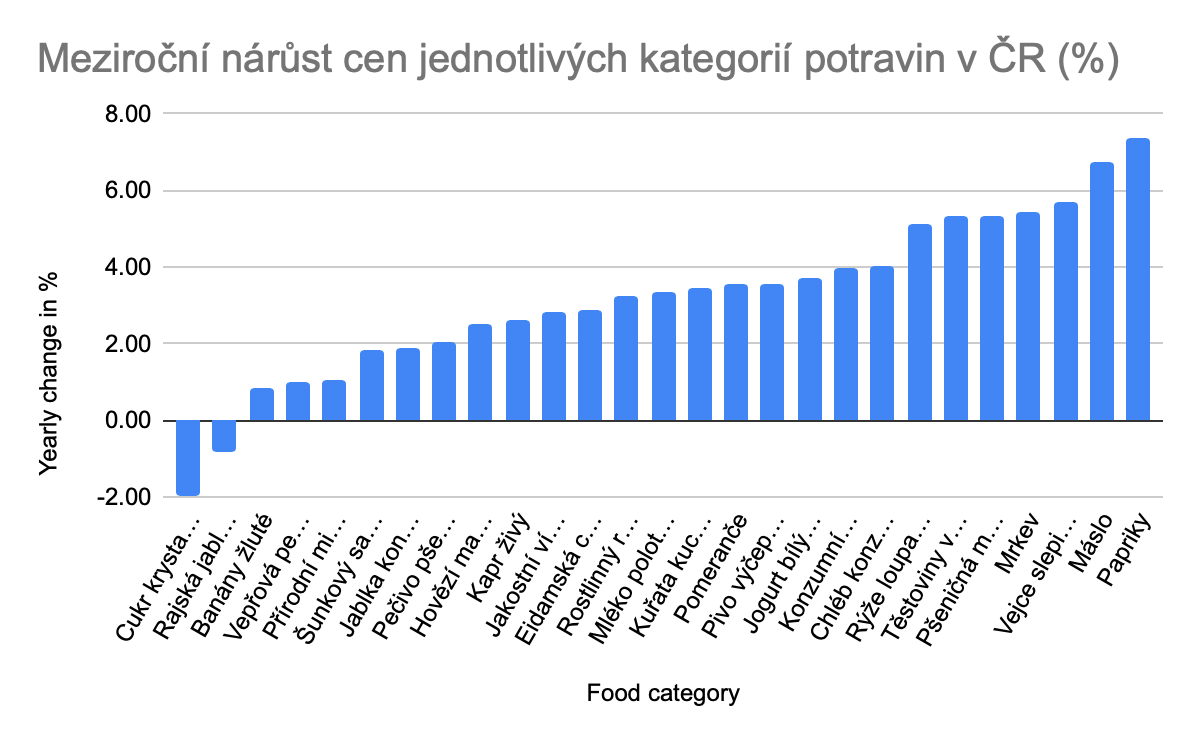
**FROM *temp\_yearly\_changes***

**GROUP BY food\_category**

**ORDER BY *avg\_yearly\_increase* ASC**

**LIMIT 1;**

**Odpověď:** Nejpomaleji zdražující kategorie potravin je “Cukr krystalový” s meziročním nárůstem -1,95%.



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

* **Vytvoření dočasné tabulky pro výpočet meziročního nárůstu mezd:**

**CREATE TEMPORARY TABLE temp\_payroll\_changes AS**

**SELECT**

**cur.year,**

**((cur.avg\_payroll - prev.avg\_payroll)/prev.avg\_payroll) \* 100 AS payroll\_change**

**FROM (**

**SELECT**

**year,**

**AVG(avg\_payroll) AS avg\_payroll**

**FROM t\_NgocKhanhVy\_Tranova\_project\_SQL\_primary\_final t**

**GROUP BY year) cur**

**JOIN (**

**SELECT**

**year,**

**AVG(avg\_payroll) AS avg\_payroll**

**FROM t\_NgocKhanhVy\_Tranova\_project\_SQL\_primary\_final t**

**GROUP BY year) prev**

**ON cur.year = prev.year + 1;**

* **Vytvoření dočasné tabulky pro výpočet meziročního nárůstu cen potravin:**

**CREATE TEMPORARY TABLE temp\_price\_changes AS**

**SELECT**

**cur.year,**

**((cur.avg\_price - prev.avg\_price)/prev.avg\_price) \* 100 AS price\_change**

**FROM (**

**SELECT**

**year,**

**AVG(food\_price) AS avg\_price**

**FROM t\_NgocKhanhVy\_Tranova\_project\_SQL\_primary\_final t**

**GROUP BY year) cur**

**JOIN (**

**SELECT**

**year,**

**AVG(food\_price) AS avg\_price**

**FROM t\_NgocKhanhVy\_Tranova\_project\_SQL\_primary\_final t**

**GROUP BY year) prev**

**ON cur.year = prev.year + 1;**

* **Kontrola výsledků:**

**SELECT**

pr.**year**,

*pr*.***price\_change***,

*pay*.***payroll\_change***

**FROM** ***temp\_price\_changes*** *pr*

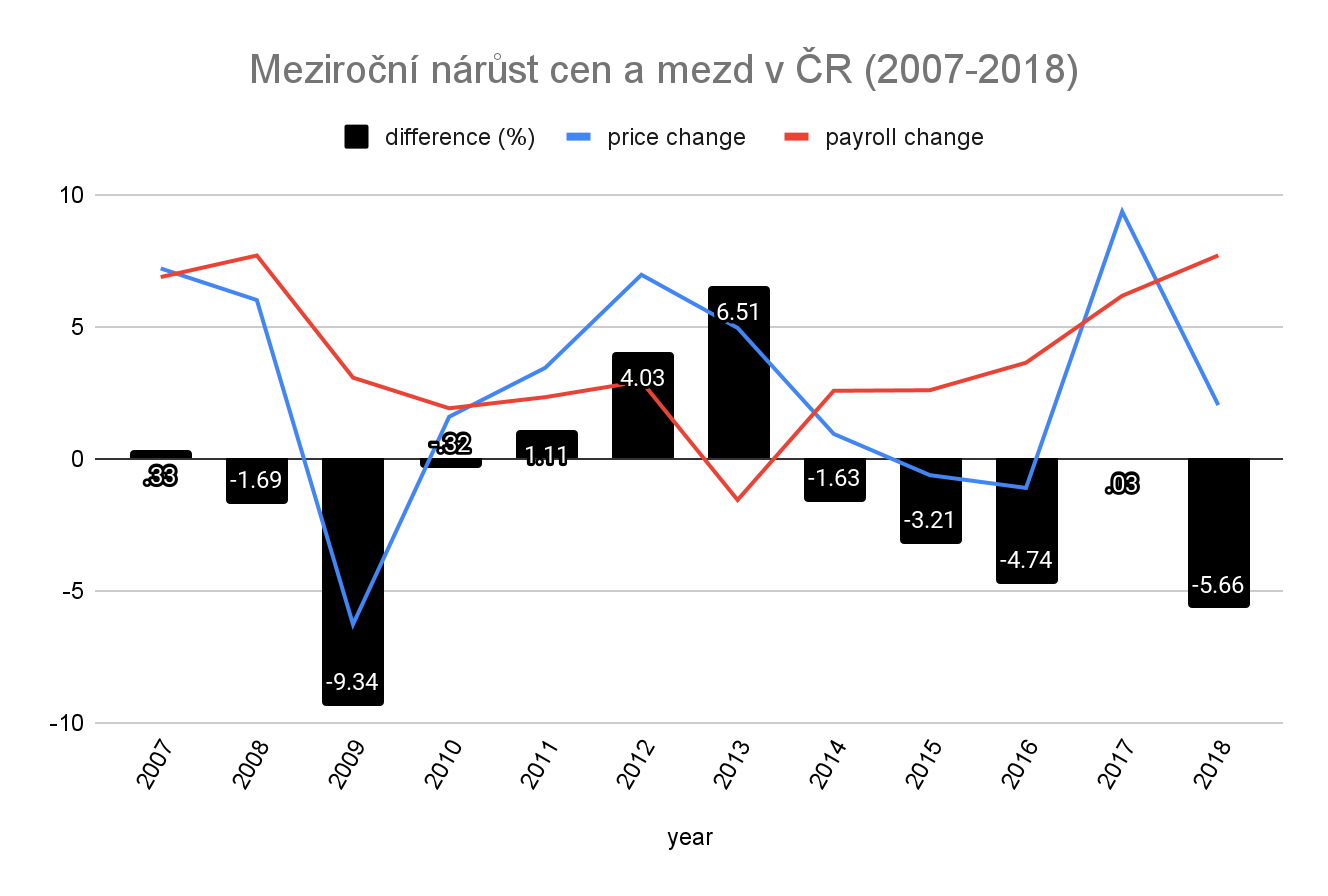
**JOIN** ***temp\_payroll\_changes*** *pay*

**ON** pr.**year** = pay.**year**

**WHERE** *pr*.***price\_change*** - *pay*.***payroll\_change*** > 10;

* V letech 2001 až 2005 a od 2019 nejsou dostupná data ohledně průměrných cen potravin.

**Odpověď:** V žádném roce nebyl nárůst cen potravin o více než 10% vyšší než meziroční nárůst mezd. Nejvyšší rozdíl, který byl zaznamenán je 6,51% v roce 2013.



1. **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?**

* **Přidání sloupce pro HDP:**

**ALTER TABLE t\_NgocKhanhVy\_Tranova\_project\_SQL\_primary\_final**

**ADD COLUMN gdp DOUBLE;**

* **Naplnění sloupce HDP:**

Z důvodu dlouhého trvání dotazů bylo nutné aktualizovat data pro každý rok zvlášť

**UPDATE** t\_NgocKhanhVy\_Tranova\_project\_SQL\_primary\_final *t*

**JOIN** economies *e*

**ON** *t*.`year` = *e*.`year`

**SET** *t*.gdp = *e*.GDP

**WHERE** *e*.country = **'Czech republic'** **AND** *e*.GDP **IS** **NOT** **NULL** **AND** *t*.`year` = 2000 ;

* **Vytvoření dočasné tabulky pro meziroční změny HDP:**

**CREATE TEMPORARY TABLE temp\_gdp\_changes AS**

**SELECT**

**cur.year,**

**((cur.gdp - prev.gdp) / prev.gdp) \* 100 AS gdp\_change**

**FROM (**

**SELECT**

**year,**

**AVG(gdp) AS gdp**

**FROM t\_NgocKhanhVy\_Tranova\_project\_SQL\_primary\_final t**

**GROUP BY year) cur**

**JOIN (**

**SELECT**

**year,**

**AVG(gdp) AS gdp**

**FROM t\_NgocKhanhVy\_Tranova\_project\_SQL\_primary\_final t**

**GROUP BY year) prev**

**ON cur.year = prev.year + 1;**

* **Spojení dočasných tabulek a zjištění vlivu HDP na změny ve mzdách a cenách potravin**

**SELECT**

**gdp.year,**

***gdp*.*gdp\_change*,**

***pay*.*payroll\_change*,**

***pr*.*price\_change***

**FROM *temp\_gdp\_changes* *gdp***

**JOIN *temp\_payroll\_changes* *pay***

**ON gdp.year = pay.year**

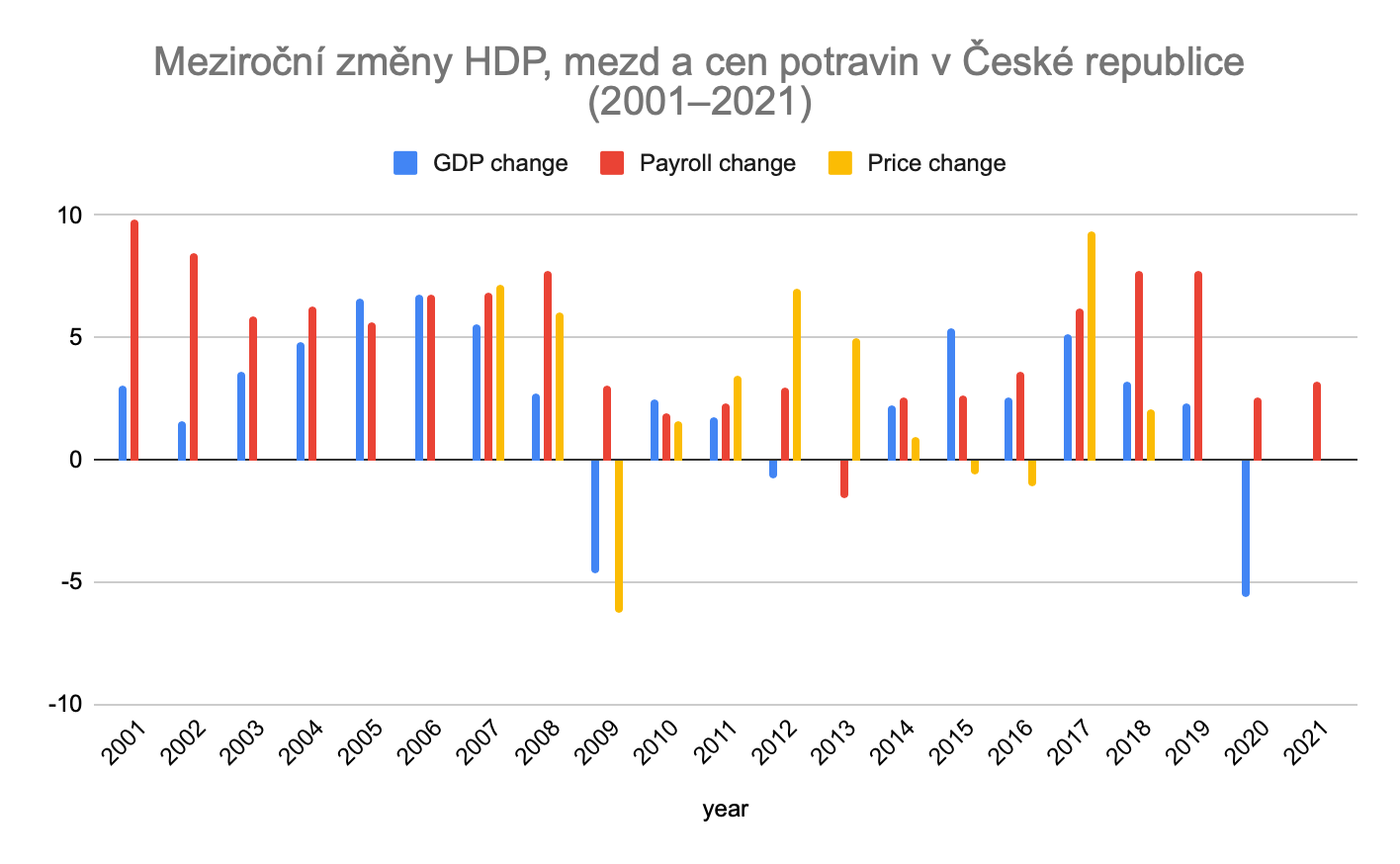
**JOIN *temp\_price\_changes* *pr***

**ON gdp.year = pr.year**

**ORDER BY gdp.year;**

* Jak už bylo zmíněno, v letech 2001 až 2005 a od 2019 nejsou dostupná data ohledně průměrných cen potravin.

**Odpověď:** Ano, obecně existuje pozitivní korelace mezi růstem HDP a růstem mezd a cen potravin. Nicméně, vztah není absolutní a v některých letech tento vztah neplatí.



**Krok 3: Příprava dat pro ostatní evropské státy**

1. **Vytvoření tabulky *t\_NgocKhanhVy\_Tranová\_project\_SQL\_secondary\_final***

***CREATE******TABLE*** *t\_NgocKhanhVy\_Tranova\_project\_SQL\_secondary\_final (*

*country* ***VARCHAR****(255),*

***year******INT****,*

*gdp* ***DOUBLE****,*

*gini* ***DOUBLE****,*

*population* ***DOUBLE***

*)*

1. **Vytvoření dočasné tabulky pro evropské státy**

**CREATE** **TEMPORARY** **TABLE** temp\_eu\_countries **AS**

**SELECT**

country

**FROM** countries c

**WHERE** c.region\_in\_world **LIKE** **'%Europe%'**

1. **Vložení dat do sekundární finální tabulky**

**INSERT** **INTO** t\_NgocKhanhVy\_Tranova\_project\_SQL\_secondary\_final (country, **year**, gdp, gini, population)

**SELECT**

*e*.country,

*e*.`year` ,

*e*.GDP ,

*e*.gini ,

*e*.population

**FROM** economies *e*

**JOIN** ***temp\_eu\_countries*** *eu*

**ON** *e*.country = *eu*.country

**WHERE** *e*.`year` **BETWEEN** (**SELECT** **MIN**(**year**) **FROM** t\_NgocKhanhVy\_Tranova\_project\_SQL\_primary\_final) **AND** (**SELECT** **MAX**(**year**) **FROM** t\_NgocKhanhVy\_Tranova\_project\_SQL\_primary\_final);