# Dokumentace k projektu SQL zabývajícím se zodpovězení otázek životní úrovně obyvatel.

[1. Zadání projektu: 1](#_heading=h.jd29ouhu3by3)

[2. Vytvořené tabulky 2](#_heading=h.lovl9a1d6b5y)

[3. Výzkumné otázky: 3](#_heading=h.3ie7jd6r7ohe)

[4. Vytvořené SQL dotazy 5](#_heading=h.uo0jangzl2ar)

## 1. Zadání projektu:

Cílem projektu je vytvořit podklady pro zodpovězení výzkumných otázek týkajících se dostupnosti základních potravin široké veřejnosti. Materiály budou předány tiskovému oddělení pro prezentování široké veřejnosti.

Pro vytvoření základních datových podkladů máme k dispozici tyto tabulky:

czechia\_payroll – Informace o mzdách v různých odvětvích za několikaleté období. Datová sada pochází z Portálu otevřených dat ČR.  
czechia\_payroll\_calculation – Číselník kalkulací v tabulce mezd.  
czechia\_payroll\_industry\_branch – Číselník odvětví v tabulce mezd.  
czechia\_payroll\_unit – Číselník jednotek hodnot v tabulce mezd.  
czechia\_payroll\_value\_type – Číselník typů hodnot v tabulce mezd.  
czechia\_price – Informace o cenách vybraných potravin za několikaleté období. Datová sada pochází z Portálu otevřených dat ČR.  
czechia\_price\_category – Číselník kategorií potravin, které se vyskytují v našem přehledu.  
czechia\_region – Číselník krajů České republiky dle normy CZ-NUTS 2.  
czechia district – Číselník okresů České republiky dle normy LAU.  
countries - Všemožné informace o zemích na světě, například hlavní město, měna, národní jídlo nebo průměrná výška populace.  
economies - GDP, GINI, daňová zátěž, atd. pro daný stát a rok.

Z těchto tabulek vytvoříme 2 cílové tabulky:

**t\_ondrej spalek\_project\_SQL\_primary\_final:** tabulka bude primárně obsahovat data mezd a cen potravin za Českou republiku sjednocených na srovnatelné období - roky

**t\_ondrej spalek\_project\_SQL\_secondary\_final:** tabulka bude obsahovat data o evropských státech.

Tyto dvě základní tabulky nám budou následně sloužit pro zodpovězení výzkumných otázek:

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?

## 2. Vytvořené tabulky

Seznam a popis tabulek, které byly vytvořeny pro řešení zadání.

**t\_ondrej\_spalek\_project\_SQL\_primary\_final:** Obsahuje podklady pro zodpovězení otázek týkajících se mezd a potravin. Data se kterými lze pracovat jsou v rozmezí 200 - 2021. Nejsou však dostupná kompletně pro všechny roky. Proto jsem při odpovídání na otázky pracoval u jednotlivých otázek s rozdílnými časovými obdobími.

V této tabulce naleznete následující sloupce:  
industry\_name: popis odvětví  
industry\_branch\_code: kód odvětví  
payroll\_year: rok  
average\_value: průměrná hodnota mzdy pro dané odvětví a daný rok  
name: název potraviny sledované ve spotřebitelském koši  
category\_code: kód potraviny  
rok: rok  
average\_value\_item: průměrná cena potraviny v daném roce.

Je-li hodnota industry\_name, industry\_branch\_code NULL, tak to znamená, že se jedná o průměrnou mzdu za daný rok pro Českou republiku.

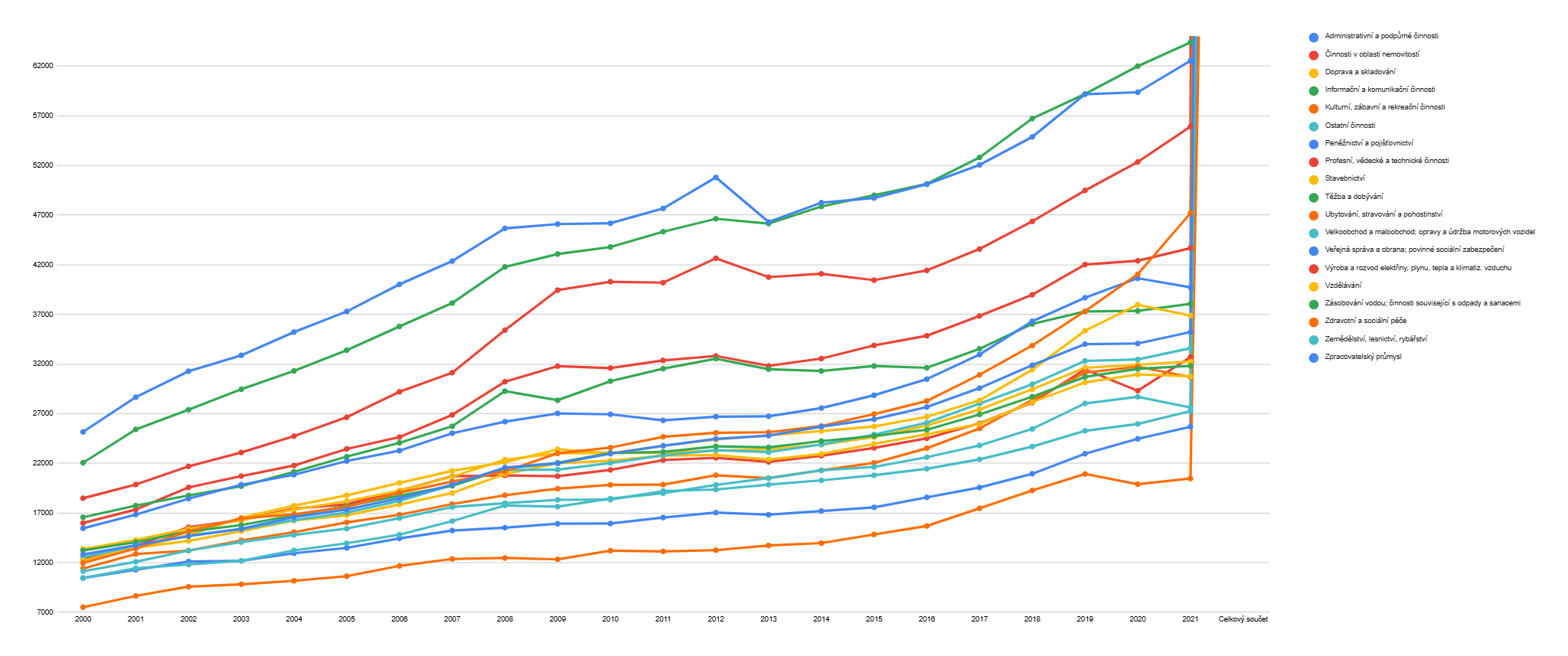
**t\_ondrej\_spalek\_project\_SQL\_secondary\_final:** Obsahuje podklady pro zodpovězení otázek týkající se porovnání evropských států.

country: název státu  
year: rok  
gdp: hodnota HDP  
population: počet obyvatel v daném roce  
gdp\_percentage\_change: změna HDP v porovnání aktuální rok oproti předchozímu, např. 2006 proti 2005 atd.

## 3. Výzkumné otázky:

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

Vývoje mezd jsem sledoval mezi lety 2000 až 2021. Obecně mzdy rostou porovnáme-li rok 2000 proti roku 2021. Při sledování mezd po jednotlivých letech jsem došel k závěru, že roční průměrná výše může kolísat a není zde trend, že ve všech odvětvích musí mzdy každý rok růst. V odvětví zemědělství, lesnictví a rybářství byl pokles mezd mezi lety 2002 a 2003. V odvětví těžby a dobývání je pokles u mezd mezi lety 2000 a 2001. Potom také mezi lety 2006 a 2007. Ve zpracovatelském průmyslu byl pokles mezd mezi lety 2005 a 2006 a potom 2008 a 2009. Takto mohu pokračovat dále napříč dalšími odvětvími. Pro lepší přehled přikládám graf, který ukazuje vývoj mezd za celé období ve všech sledovaných oblastech.

Vývoj mezd vyjádřený grafem:  


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

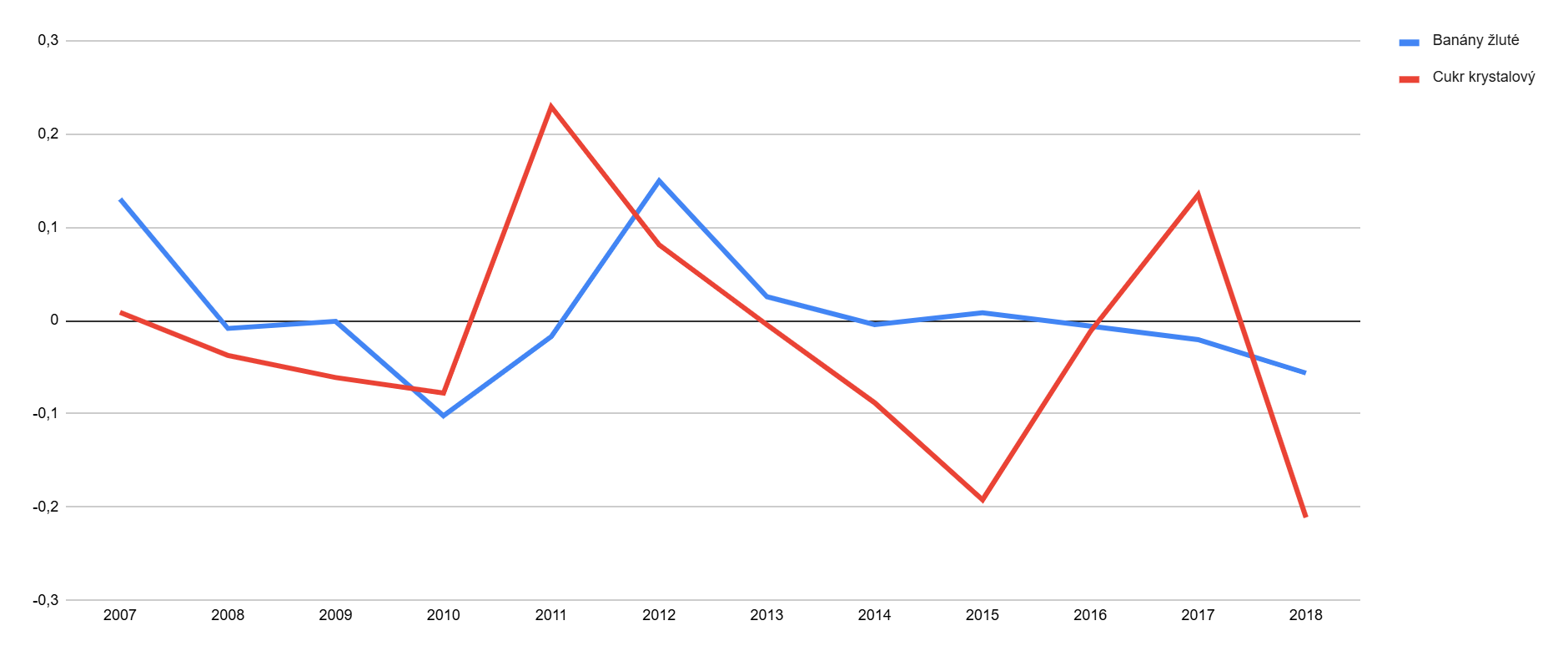
Pro zodpovězení této otázky jsem využil data z let 2006 a 2018. Jedná se o první a poslední rok, kdy byla dostupná data o cenách a mzdách.

V roce 2006 bylo možné zakoupit 1 307, 65 ks chleba konzumního kmínového a 1 460,29 l mléka polotučného pasterovaného. V roce 2018 bylo možné koupit 1 363,09 chleba konzumního kmínového a 1 667, 02 l mléka polotučného pasterovaného.

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

Pro zodpovězení této otázky jsem se podíval, jak se měnila cena dané potraviny při porovnání aktuálního roku oproti roku minulému, tzn: rok 2009 proti 2008, 2010 proti 2009 atd. Výsledkem byla tabulka s vývojem cen v čase ze které můžeme vybrat potravinu s nejmenší meziročním růstem.

Pro demonstraci výsledku přidávám zobrazení změny ceny pro Banány žluté a Cukr krystalový.



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

Ano, existuje. Ve sledovaném období 2006 až 2018 máme roky 2009, 2011, 2012, 2014 a 2016, kdy byl růst spotřebitelského koše vyšší než 10% oproti růstu mezd.

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?**

Odpovědi na tuto otázku jsem rozdělil na dvě části.   
a) Pokud HDP vzroste výrazněji v jednom roce, projeví se to na cenách potravin či mzdách ve stejném roce?   
Jako výpočet jsme použili porovnání aktuálního roku oproti roku předchozímu, tzn, pokud výsledek ukazuje, že HDP, mzdy, potraviny vzrostly v roce 2006 o X%, tak to znamená oproti roku 2005.  
Odpověď není jednoznačná. Najdeme roky, kdy růst HDP a mezd znamená i růst potravin, ale i naopak. V roce 2007, kdy HDP vzrostlo o 5,57% a mzdy vzrostly o 6,89% a ceny potravin o 6,76%. Stejně tak rok 2017, kdy HDP vzrostlo o 5,17%, mzdy o 6,19% a ceny potravin vzrostly o 9,62%. Proti tomu máme rok 2009, kdy HDP kleslo o 4,65% a mzdy vzrostly o 3,08% a ceny potravin klesly o 6,41%. V roce 2013 HDP kleslo o 0,045%, mzdy o 1,48% a cena potravin vzrostla o 5,097%

b) Pokud HDP vzroste výrazněji v jednom roce, projeví se to na cenách potravin či mzdách v následujícím roce?  
Pro odpověď na tuto otázku vychází v porovnání meziročního růstu následovně. Bereme-li rok 2008, tak zde porovnáváme změnu HDP roku 2008 proti roku 2007 a to porovnáváme ve změnách cen potravin, mzdy roku 2009 oproti roku 2008. Chceme tedy zjistit, zda trend HDP roku 2008 se propsal do vývoje mezd a cen potravin v roce 2009.  
Odpověď na tuto otázku také není jednoznačná. Máme rok 2008, kdy růst HDP znamenal růst mezd o 6,89% a růst potravin o 6,68% v roce 2009. Ovšem v roce 2009 byl pokles HDP o 4,65% avšak mzdy v roce 2010 vzrostly o 7,7% a spotřebitelský koš o 7,19%. Nebo rok 2014, kdy byl růst HDP 2,26% a mzdy v následujícím roce 2015 klesly o 1,48% a ceny potravin o 5,95%.

## 4. Vytvořené SQL dotazy pro základní tabulky

SQL skript pro vytvoření tabulky: t\_ondrej\_spalek\_project\_SQL\_primary\_final:

**CREATE** **TABLE** t\_ondrej\_spalek\_project\_SQL\_primary\_final **AS**

**WITH** avg\_payroll **AS** (

**SELECT**

cpib.**"name"** **AS** industry\_name,

cp.industry\_branch\_code ,

cp.payroll\_year,

**ROUND** (**AVG**(cp.value) :: **NUMERIC** ,2) **AS** average\_value

**FROM** czechia\_payroll cp

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

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

**WHERE** value\_type\_code = '5958'

**AND** calculation\_code = '200'

**GROUP** **BY**

cpib.**"name"** ,

cp.industry\_branch\_code ,

cp.payroll\_year

),

food\_price **AS** (

**SELECT**

cpc.**name**,

cp.category\_code,

**DATE\_PART**('year', cp.date\_from) **AS** rok,

**ROUND**(**AVG**(cp.value)::**NUMERIC**, 2) **AS** average\_value\_item

**FROM** czechia\_price cp

**LEFT** **JOIN** czechia\_price\_category cpc

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

--WHERE cp.region\_code IS NOT NULL

**GROUP** **BY**

cpc.**name**,

cp.category\_code,

**DATE\_PART**('year', cp.date\_from)

)

**SELECT** \*

**FROM** avg\_payroll

**LEFT** **JOIN** food\_price

**ON** food\_price.rok = avg\_payroll.payroll\_year;

SQL skript pro vytvoření tabulky: t\_ondrej\_spalek\_project\_SQL\_secondary\_final:

**CREATE** **TABLE** t\_ondrej\_spalek\_project\_SQL\_secondary\_final **AS**

**SELECT**

country,

**year**,

gdp,

population,

(gdp - **LAG**(gdp) **OVER** (**ORDER** **BY** **"year"**)) / **LAG**(gdp) **OVER** (**ORDER** **BY** **"year"**) \* 100 **AS** gdp\_percentage\_change

**FROM**

economies e

**WHERE**

country = 'Czech Republic'

**AND** gdp **IS** **NOT** **NULL**

**AND** **YEAR** **BETWEEN** 2006 **AND** 2018

**ORDER** **BY**

**"year"**;

## 5. Vytvořené SQL dotazy pro zodpovězení výzkumných otázek

Popis jednotlivých SQL dotazů, které byly vytvořeny pro zodpovězení položených otázek.

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

**SELECT**

payroll\_year,

industry\_name,

**ROUND**(**AVG**(average\_value)::**NUMERIC**, 2) **AS** *average\_salary*

**FROM** t\_ondrej\_spalek\_project\_SQL\_primary\_final

**WHERE**

industry\_name **IS** **NOT** **NULL**

**GROUP** **BY**

payroll\_year,

industry\_name

**ORDER** **BY**

payroll\_year,

industry\_name;

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

**SELECT**

rok,

name,

**ROUND**(**SUM**(average\_value) / **COUNT**(payroll\_year), 2) **AS** *average\_salary\_year*,

average\_value\_item,

**ROUND** (((**ROUND**(**SUM**(average\_value) / **COUNT**(payroll\_year), 0))/ average\_value\_item),2) **AS** *number\_of\_items*

**FROM** t\_ondrej\_spalek\_project\_SQL\_primary\_final

**WHERE**

rok = 2006

**AND** name **IN** ('Chléb konzumní kmínový', 'Mléko polotučné pasterované')

**GROUP** **BY**

rok,

name,

average\_value\_item

**UNION** **ALL**

**SELECT**

rok,

name,

**ROUND**(**SUM**(average\_value) / **COUNT**(payroll\_year), 2) **AS** *average\_salary\_year*,

average\_value\_item,

**ROUND** (((**ROUND**(**SUM**(average\_value) / **COUNT**(payroll\_year), 0))/ average\_value\_item),2) **AS** *number\_of\_items*

**FROM** t\_ondrej\_spalek\_project\_SQL\_primary\_final

**WHERE**

rok = 2018

**AND** name **IN** ('Chléb konzumní kmínový', 'Mléko polotučné pasterované')

**GROUP** **BY**

rok,

**name**,

average\_value\_item;

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

**SELECT**

rok,

name **AS** *item\_name*,

**ROUND** (**AVG**(average\_value\_item),2) **AS** *this\_year\_value*,

**ROUND** (**LAG**(**AVG**(average\_value\_item)) **OVER** (**PARTITION** **BY** category\_code **ORDER** **BY** rok),2) **AS** *previous\_year\_value*,

**ROUND** ((**AVG**(average\_value\_item) - **LAG**(**AVG**(average\_value\_item)) **OVER** (**PARTITION** **BY** category\_code **ORDER** **BY** rok)) / **LAG**(**AVG**(average\_value\_item)) **OVER** (**PARTITION** **BY** category\_code **ORDER** **BY** rok) \* 100,2) **AS** *percentage\_change*

**FROM** t\_ondrej\_spalek\_project\_SQL\_primary\_final

**WHERE** name **IS** **NOT** **NULL**

**GROUP** **BY**

rok,

category\_code,

name

**ORDER** **BY**

name,

rok **ASC** ,

*percentage\_change*;

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

**SELECT**

payroll\_year,

average\_value,

**ROUND** (((average\_value - **LAG**(average\_value) **OVER** (**ORDER** **BY** payroll\_year)) / **LAG**(average\_value) **OVER** (**ORDER** **BY** payroll\_year)) \* 100,2) **AS** *difference\_average\_value*,

**SUM**(average\_value\_item) **AS** *total\_value*,

**ROUND** (((**SUM**(average\_value\_item) - **LAG**(**SUM**(average\_value\_item)) **OVER** (**ORDER** **BY** payroll\_year)) / **LAG**(**SUM**(average\_value\_item)) **OVER** (**ORDER** **BY** payroll\_year)) \* 100,2) **AS** *difference\_basket\_value*,

**ROUND** (((((**SUM**(average\_value\_item) - **LAG**(**SUM**(average\_value\_item)) **OVER** (**ORDER** **BY** payroll\_year)) / **LAG**(**SUM**(average\_value\_item)) **OVER** (**ORDER** **BY** payroll\_year)) \* 100)/ (((average\_value - **LAG**(average\_value) **OVER** (**ORDER** **BY** payroll\_year)) / **LAG**(average\_value) **OVER** (**ORDER** **BY** payroll\_year)) \* 100)\*100),2) **AS** *year\_percentage\_change*

**FROM**

t\_ondrej\_spalek\_project\_SQL\_primary\_final

**WHERE**

industry\_name **IS** **NULL**

**AND** payroll\_year **BETWEEN** '2006' **AND** '2018'

**GROUP** **BY**

payroll\_year,

average\_value

**ORDER** **BY**

payroll\_year **ASC**;

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?

5A)

**WITH** *basket\_value* **AS** (

**SELECT**

rok,

**SUM**(total\_value),

(**SUM**(total\_value) - **LAG**(**SUM**(total\_value)) **OVER** (**ORDER** **BY** rok)) / **LAG**(**SUM**(total\_value)) **OVER** (**ORDER** **BY** rok) \* 100 **AS** *consumer\_basket\_percentage\_change*

**FROM** (

**SELECT**

rok,

**AVG**(average\_value\_item) **AS** total\_value,

category\_code,

**name**,

**LAG**(**SUM**(average\_value\_item)) **OVER** (**PARTITION** **BY** category\_code **ORDER** **BY** rok) **AS** previous\_year\_value,

(**SUM**(average\_value\_item) - **LAG**(**SUM**(average\_value\_item)) **OVER** (**PARTITION** **BY** category\_code **ORDER** **BY** rok)) / **LAG**(**SUM**(average\_value\_item)) **OVER** (**PARTITION** **BY** category\_code **ORDER** **BY** rok) \* 100 **AS** *percentage\_change*

**FROM** t\_ondrej\_spalek\_project\_SQL\_primary\_final

**GROUP** **BY**

rok,

category\_code,

name

**ORDER** **BY**

name,

rok,

*percentage\_change*

) *subquery*

**GROUP** **BY**

rok

**ORDER** **BY**

rok

),

*mzda\_value* **AS** (

**SELECT**

payroll\_year,

**ROUND**(**SUM**(average\_value) / **COUNT**(payroll\_year), 2) **AS** *avg\_salary*,

(**ROUND**(**SUM**(average\_value) / **COUNT**(payroll\_year), 2) - **LAG**(**ROUND**(**SUM**(average\_value) / **COUNT**(payroll\_year), 2)) **OVER** (**ORDER** **BY** payroll\_year)) / **LAG**(**ROUND**(**SUM**(average\_value) / **COUNT**(payroll\_year), 2)) **OVER** (**ORDER** **BY** payroll\_year) \* 100 **AS** *salary\_percentage\_change*

**FROM**

t\_ondrej\_spalek\_project\_SQL\_primary\_final

**WHERE**

payroll\_year **BETWEEN** '2006' **AND** '2018'

**GROUP** **BY**

payroll\_year

**ORDER** **BY**

payroll\_year

)

**SELECT**

rok,

gdp\_percentage\_change,

*salary\_percentage\_change*,

*bv*.*consumer\_basket\_percentage\_change*

**FROM** *basket\_value* *bv*

**LEFT** **JOIN** *mzda\_value* *mv* **ON** *bv*.rok = *mv*.payroll\_year

**LEFT** **JOIN** t\_ondrej\_spalek\_project\_SQL\_secondary\_final *ppp* **ON** *bv*.rok = *ppp*."year";

5B)

**WITH** *basket\_value* **AS** (

**SELECT**

rok,

**SUM**(total\_value),

(**SUM**(total\_value) - **LAG**(**SUM**(total\_value)) **OVER** (**ORDER** **BY** rok)) / **LAG**(**SUM**(total\_value)) **OVER** (**ORDER** **BY** rok) \* 100 **AS** *consumer\_basket\_percentage\_change*

**FROM** (

**SELECT**

rok,

**AVG**(average\_value\_item) **AS** total\_value,

category\_code,

**name**,

**LAG**(**SUM**(average\_value\_item)) **OVER** (**PARTITION** **BY** category\_code **ORDER** **BY** rok) **AS** previous\_year\_value,

(**SUM**(average\_value\_item) - **LAG**(**SUM**(average\_value\_item)) **OVER** (**PARTITION** **BY** category\_code **ORDER** **BY** rok)) / **LAG**(**SUM**(average\_value\_item)) **OVER** (**PARTITION** **BY** category\_code **ORDER** **BY** rok) \* 100 **AS** *percentage\_change*

**FROM** t\_ondrej\_spalek\_project\_SQL\_primary\_final

**GROUP** **BY**

rok,

category\_code,

name

**ORDER** **BY**

name,

rok,

*percentage\_change*

) *subquery*

**GROUP** **BY**

rok

**ORDER** **BY**

rok

),

*mzda\_value* **AS** (

**SELECT**

payroll\_year,

**ROUND**(**SUM**(average\_value) / **COUNT**(payroll\_year), 2) **AS** *avg\_salary*,

(**ROUND**(**SUM**(average\_value) / **COUNT**(payroll\_year), 2) - **LAG**(**ROUND**(**SUM**(average\_value) / **COUNT**(payroll\_year), 2)) **OVER** (**ORDER** **BY** payroll\_year)) / **LAG**(**ROUND**(**SUM**(average\_value) / **COUNT**(payroll\_year), 2)) **OVER** (**ORDER** **BY** payroll\_year) \* 100 **AS** salary\_percentage\_change

**FROM**

t\_ondrej\_spalek\_project\_SQL\_primary\_final

**WHERE**

payroll\_year **BETWEEN** '2006' **AND** '2018'

**GROUP** **BY**

payroll\_year

**ORDER** **BY**

payroll\_year

),

gdp\_value **AS** (

**SELECT**

**"year"**,

gdp,

(gdp - **LAG**(gdp) **OVER** (**ORDER** **BY** **"year"**)) / **LAG**(gdp) **OVER** (**ORDER** **BY** **"year"**) \* 100 **AS** gdp\_percentage\_change

**FROM**

t\_ondrej\_spalek\_project\_SQL\_secondary\_final ppp

**WHERE**

country = 'Czech Republic'

**AND** gdp **IS** **NOT** **NULL**

**AND** **"year"** **BETWEEN** 2006 **AND** 2018

)

**SELECT**

gdp\_value.**"year"** **AS** gdp\_year,

gdp\_value.gdp\_percentage\_change,

mv.salary\_percentage\_change,

bv.consumer\_basket\_percentage\_change

**FROM**

gdp\_value

**LEFT** **JOIN** basket\_value bv

**ON** gdp\_value.**"year"** = bv.rok + 1

**LEFT** **JOIN** mzda\_value mv

**ON** gdp\_value.**"year"** = mv.payroll\_year + 1

**ORDER** **BY**

gdp\_value.**"year"**;