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.
- 2. czechia district Číselník okresů České republiky dle normy LAU.

Dodatečné tabulky:

- 1. countries Všemožné 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 percentuální 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 násdujícím 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).

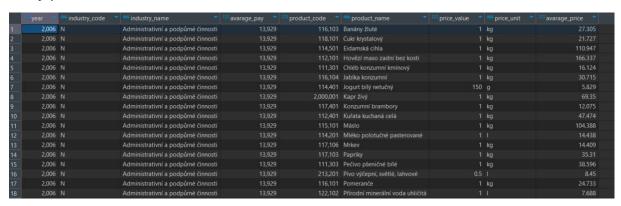
Dále připravte sadu SQL, které z vámi připravených tabulek získají datový podklad k odpovězení na vytyčené výzkumné otázky. Pozor, otázky/hypotézy mohou vaše výstupy podporovat i vyvracet! Záleží na tom, co říkají data.

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.).

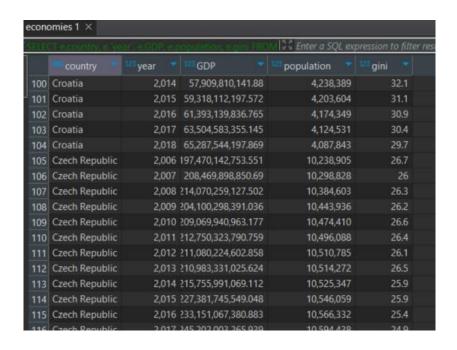
Zpracování finálních tabulek

Primární tabulka **t_petr_melicharik_project_SQL_primary_final** byla vytvořena spojením dvou tabulek: **t_avg_pay** (průměrné mzdy v různých odvětvích za určité období) a **t_avg_price** (průměrné ceny různých potravin v určitém období).

Výsledkem je tabulka, která zobrazuje průměrné mzdy dle odvětví a průměrné ceny potravin v období let 2006 – 2018.



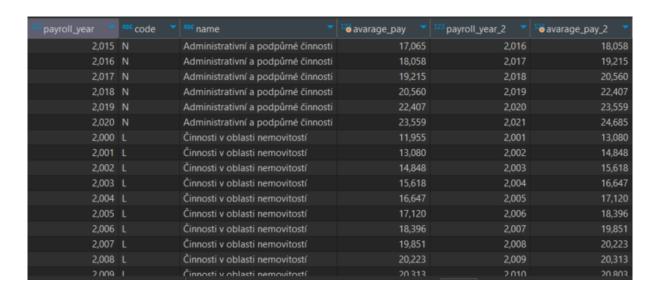
Sekundární tabulka zobrazuje výši HDP, populace a gini koeficient v letech 2006 – 2018 v evropských státech.



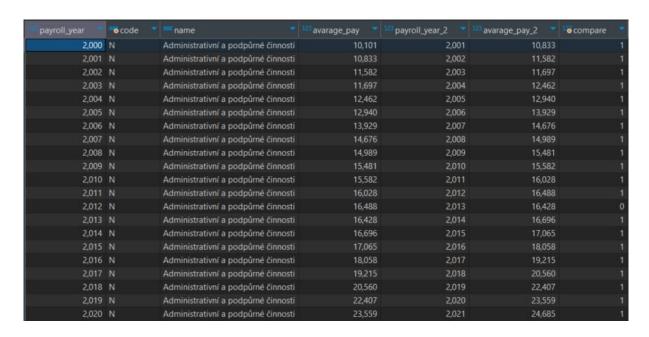
Výzkumné otázky

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

Vytvořením porovnávací tabulky t_avg_pay_yearly získáme porovnání průměrných mezd v jednom roce a roce následném u všech sledovaných odvětví.



Pomocí funkce CASE si tyto roky porovnáme a v případě růstu se v novém sloupci compare vypíše 1, v případě poklesu 0.



Pokud si projdu každé odvětví postupně, dojdu ke zjištění, že jediná odvětví, ve kterých v každém roce došlo k růstu mzdy, jsou: **Doprava a skladování**, **Ostatní činnosti**, **Zdravotní a sociální péče** a **Zpracovatelský průmysl**.

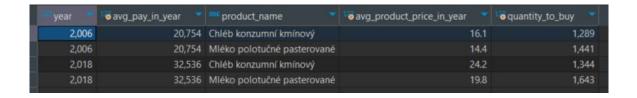
payroll_year	ocode •	name	avarage_pay	payroll_year_2	avarage_pay_2	ocompare 🔻
2,000	Н	Doprava a skladování	13,279	2,001	14,201	1
2,001		Doprava a skladování	14,201	2,002	15,295	
2,002		Doprava a skladování	15,295	2,003	16,124	
2,003		Doprava a skladování	16,124	2,004	17,223	
2,004		Doprava a skladování	17,223	2,005	18,066	
2,005		Doprava a skladování	18,066	2,006	19,125	
2,006		Doprava a skladování	19,125	2,007	20,524	
2,007		Doprava a skladování	20,524	2,008	22,191	
2,008		Doprava a skladování	22,191	2,009	22,849	
2,009		Doprava a skladování	22,849	2,010	22,901	
2,010		Doprava a skladování	22,901	2,011	22,909	
2,011		Doprava a skladování	22,909	2,012	23,146	
2,012		Doprava a skladování	23,146	2,013	23,271	
2,013		Doprava a skladování	23,271	2,014	23,737	
2,014		Doprava a skladování	23,737	2,015	24,507	
2,015		Doprava a skladování	24,507	2,016	25,660	
2,016		Doprava a skladování	25,660	2,017	27,289	
2,017		Doprava a skladování	27,289	2,018	29,298	
2,018		Doprava a skladování	29,298	2,019	31,393	
2,019		Doprava a skladování	31,393	2,020	31,644	
2,020		Doprava a skladování	31,644	2,021	31,938	

V ostatních odvětvích se vždy našel minimálně 1 rok, kdy došlo ke snížení mzdy oproti roku předcházejícímu.

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

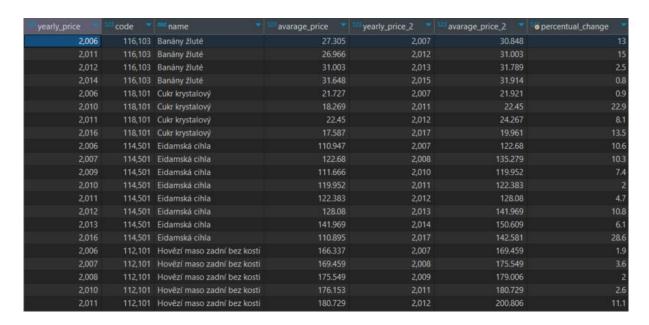
V roce 2006 si za průměrnou mzdu je možné koupit 1289 ks chleba a 1441 litrů mláka.

V roce 2018 si za průměrnou mzdu je možné koupit 1344 ks chleba a 1643 litrů mléka.



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

Nejprve jsem si vytvořil pomocnou tabulku t_avg_price_raise, která mi zobrazuje meziroční změny průměrných cen potravin. Zároveň jsem si vyfiltroval pouze ty roky, ve kterých došlo k růstu cen (ke zdražení). Procentuální nárůst zobrazuje sloupec percentual_change.



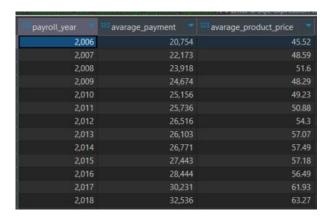
Z této tabulky jsem si vypočítal průměrný nárůst v těchto letech pro každou potravinu a seřadil.

name 🔻	¹ã avg_raise ▼
Jakostní víno bílé	2.7
Pivo výčepní, světlé, lahvové	2.85833
Přírodní minerální voda uhličitá	3.26667
Šunkový salám	3.64286
Hovězí maso zadní bez kosti	3.77778
Vepřová pečeně s kostí	
Kapr živý	4.18889
Těstoviny vaječné	5.25833
Jogurt bílý netučný	5.53
Rostlinný roztíratelný tuk	5.675
Pomeranče	6.85556
Mléko polotučné pasterované	6.93333
Kuřata kuchaná celá	7.47143
Banány žluté	7.825
Rýže loupaná dlouhozrnná	8.5875
Jablka konzumní	9.84286
Eidamská cihla	10.0625
Chléb konzumní kmínový	10.57143
Cukr krystalový	11.35
Pečivo pšeničné bílé	13.61667
Rajská jablka červená kulatá	14.16
Máslo	15.42857
Vejce slepičí čerstvá	15.48571
Pšeničná mouka hladká	16.58571
Mrkev	18.95
Konzumní brambory	21.28333
Papriky	22.61667

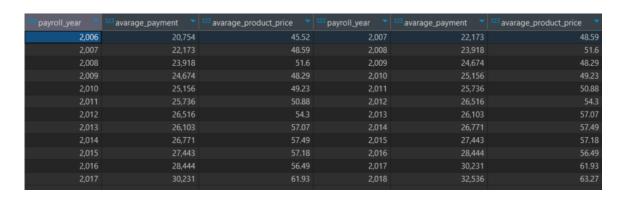
Nejpomalejší nárůst můžeme tedy zaznamenat u **Jakostního vína bílého** a **Piva výčepního, světlého, lahvového** – průměrný nárůst cen < 3 %.

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

Nejdříve jsme si připravil pomocnou tabulku, ve které se nachází průměrné mzdy a průměrné ceny potravin v ČR za období 2006 – 2018.



Tuto tabulku jsem napojil na sebe samou, abych měl v řádku srovnání dvou let.

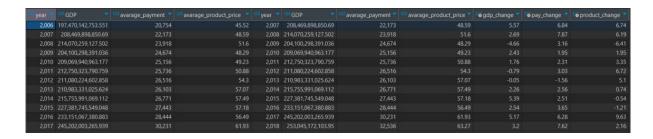


Po potřebných výpočet lze dojít k závěru, že v žádném roce nebyl meziroční nárůst cen potravin větší o více jak 10 procent než meziroční nárůst mezd.



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?

Potřebná tabulka vznikla spojením tabulky s průměry mezd a cen a sekundární tabulky s informacemi o výši HDP za potřebné období pro Českou republiku. Poslední 3 sloupce jsou potřebné výpočty o meziročních změnách HPD, mezd, cenách produktů.



400	
¹6 pay_change ▼	¹ product_change ▼
6.84	6.74
7.87	6.19
3.16	-6.41
1.95	1.95
2.31	3.35
3.03	6.72
-1.56	5.1
2.56	0.74
2.51	-0.54
3.65	-1.21
6.28	9.63
7.62	2.16
	6.84 7.87 3.16 1.95 2.31 3.03 -1.56 2.56 2.51 3.65 6.28

Pro zodpovězení výzkumné otázky jsem vybral roky 2007, 2015 a 2017, kdy došlo k procentuálnímu nejvyššímu růstu HDP (> 5 %).

V roce 2007 došlo k růstu HDP o 5,57 %. Ve stejném roce se mzdy navýšily o 6,84% a ceny produktů o 6,74 %. V následujícím roce 2018 mzdy stouply o 7,87 % a ceny produktů o 6,19%.

V roce 2015 došlo k růstu HDP o 5,39%. Ve stejném roce se výše mezd zvýšila o 2,51%, ceny produktů se snížily o 0,54 procent. V následujícím roce 2016 se mzdy zvýšily o 3,65%, ceny produktů opět klesly a to o 1,21%.

V roce 2017 došlo k růstu HDP o 5,17 %. Mzdy se v tomto roce zvýšily o 6,28 %, ceny potravin o 9,63%. V následujícím roce 2018 výše mezd nadále rostla, konkrétně o 7,62 %, ceny potravin vzrostly také, avšak nešlo již o tak razantní růst (2,16%).

Závěr:

Při výraznějším růstu HDP docházelo v témže nebo následujícím roce také k vyššímu růstu průměrné mzdy, avšak průměrné ceny potravin dokonce v některých letech při růstu HDP klesaly.

Vyšší růst HDP tedy ovlivnil i růst průměrné mzdy, na průměrnou cenu potravin však vliv neměl.