

# 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.

Potřebuji 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` – Všechné 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 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?

### 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.

	year	industry_code	industry_name	average_pay	product_code	product_name	price_value	price_unit	average_price
1	2006	N	Administrativní a podpůrné činnosti	13,929	116,103	Banány žluté	1	kg	27.305
2	2006	N	Administrativní a podpůrné činnosti	13,929	118,101	Čukr krystalový	1	kg	21.727
3	2006	N	Administrativní a podpůrné činnosti	13,929	114,501	Eidamská cihla	1	kg	110.947
4	2006	N	Administrativní a podpůrné činnosti	13,929	112,101	Hovězí maso zadní bez kosti	1	kg	166.337
5	2006	N	Administrativní a podpůrné činnosti	13,929	111,301	Chléb konzumní krmínový	1	kg	16.124
6	2006	N	Administrativní a podpůrné činnosti	13,929	116,104	Jablka konzumní	1	kg	30.715
7	2006	N	Administrativní a podpůrné činnosti	13,929	114,401	Jogurt bílý netučný	150	g	5.829
8	2006	N	Administrativní a podpůrné činnosti	13,929	2,000,001	Kapři živý	1	kg	69.35
9	2006	N	Administrativní a podpůrné činnosti	13,929	117,401	Konzumní brambory	1	kg	12.075
10	2006	N	Administrativní a podpůrné činnosti	13,929	112,401	Kuřata kuchařská celá	1	kg	47.474
11	2006	N	Administrativní a podpůrné činnosti	13,929	115,101	Máslo	1	kg	104.388
12	2006	N	Administrativní a podpůrné činnosti	13,929	114,201	Mléko polotučné pasteurované	1	l	14.438
13	2006	N	Administrativní a podpůrné činnosti	13,929	117,106	Mříkev	1	kg	14.409
14	2006	N	Administrativní a podpůrné činnosti	13,929	117,103	Papriky	1	kg	35.31
15	2006	N	Administrativní a podpůrné činnosti	13,929	111,303	Pecivo pšeničné bílé	1	kg	38.596
16	2006	N	Administrativní a podpůrné činnosti	13,929	213,201	Pivo výčepní, světlé, lahvé	0.5	l	8.45
17	2006	N	Administrativní a podpůrné činnosti	13,929	116,101	Pomeranče	1	kg	24.733
18	2006	N	Administrativní a podpůrné činnosti	13,929	122,102	Přírodní minerální voda uhlíkatá	1	l	7.688

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

economies 1 ✕					
SELECT e.country, e.year, e.GDP, e.population, e.gini FROM   <sup>1</sup> Enter a SQL expression to filter res					
	country	year	GDP	population	gini
100	Croatia	2,014	57,909,810,141.88	4,238,389	32.1
101	Croatia	2,015	59,318,112,197.572	4,203,604	31.1
102	Croatia	2,016	61,393,139,836.765	4,174,349	30.9
103	Croatia	2,017	63,504,583,355.145	4,124,531	30.4
104	Croatia	2,018	65,287,544,197.869	4,087,843	29.7
105	Czech Republic	2,006	197,470,142,753.551	10,238,905	26.7
106	Czech Republic	2,007	208,469,898,850.69	10,298,828	26
107	Czech Republic	2,008	214,070,259,127.502	10,384,603	26.3
108	Czech Republic	2,009	204,100,298,391.036	10,443,936	26.2
109	Czech Republic	2,010	209,069,940,963.177	10,474,410	26.6
110	Czech Republic	2,011	212,750,323,790.759	10,496,088	26.4
111	Czech Republic	2,012	211,080,224,602.858	10,510,785	26.1
112	Czech Republic	2,013	210,983,331,025.624	10,514,272	26.5
113	Czech Republic	2,014	215,755,991,069.112	10,525,347	25.9
114	Czech Republic	2,015	227,381,745,549.048	10,546,059	25.9
115	Czech Republic	2,016	233,151,067,380.883	10,566,332	25.4
116	Czech Republic	2,017	245,202,002,265.920	10,594,428	24.9

# 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.

123 payroll_year	nsc_code	nsc_name	123 avarage_pay	123 payroll_year_2	123 avarage_pay_2
2,015	N	Administrativní a podpůrné činnosti	17,065	2,016	18,058
2,016	N	Administrativní a podpůrné činnosti	18,058	2,017	19,215
2,017	N	Administrativní a podpůrné činnosti	19,215	2,018	20,560
2,018	N	Administrativní a podpůrné činnosti	20,560	2,019	22,407
2,019	N	Administrativní a podpůrné činnosti	22,407	2,020	23,559
2,020	N	Administrativní a podpůrné činnosti	23,559	2,021	24,685
2,000	L	Činnosti v oblasti nemovitostí	11,955	2,001	13,080
2,001	L	Činnosti v oblasti nemovitostí	13,080	2,002	14,848
2,002	L	Činnosti v oblasti nemovitostí	14,848	2,003	15,618
2,003	L	Činnosti v oblasti nemovitostí	15,618	2,004	16,647
2,004	L	Činnosti v oblasti nemovitostí	16,647	2,005	17,120
2,005	L	Činnosti v oblasti nemovitostí	17,120	2,006	18,396
2,006	L	Činnosti v oblasti nemovitostí	18,396	2,007	19,851
2,007	L	Činnosti v oblasti nemovitostí	19,851	2,008	20,223
2,008	L	Činnosti v oblasti nemovitostí	20,223	2,009	20,313
2,009	L	Činnosti v oblasti nemovitostí	20,313	2,010	20,803

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

123 payroll_year	nsc_code	nsc_name	123 avarage_pay	123 payroll_year_2	123 avarage_pay_2	123 compare
2,000	N	Administrativní a podpůrné činnosti	10,101	2,001	10,833	1
2,001	N	Administrativní a podpůrné činnosti	10,833	2,002	11,582	1
2,002	N	Administrativní a podpůrné činnosti	11,582	2,003	11,697	1
2,003	N	Administrativní a podpůrné činnosti	11,697	2,004	12,462	1
2,004	N	Administrativní a podpůrné činnosti	12,462	2,005	12,940	1
2,005	N	Administrativní a podpůrné činnosti	12,940	2,006	13,929	1
2,006	N	Administrativní a podpůrné činnosti	13,929	2,007	14,676	1
2,007	N	Administrativní a podpůrné činnosti	14,676	2,008	14,989	1
2,008	N	Administrativní a podpůrné činnosti	14,989	2,009	15,481	1
2,009	N	Administrativní a podpůrné činnosti	15,481	2,010	15,582	1
2,010	N	Administrativní a podpůrné činnosti	15,582	2,011	16,028	1
2,011	N	Administrativní a podpůrné činnosti	16,028	2,012	16,488	1
2,012	N	Administrativní a podpůrné činnosti	16,488	2,013	16,428	0
2,013	N	Administrativní a podpůrné činnosti	16,428	2,014	16,696	1
2,014	N	Administrativní a podpůrné činnosti	16,696	2,015	17,065	1
2,015	N	Administrativní a podpůrné činnosti	17,065	2,016	18,058	1
2,016	N	Administrativní a podpůrné činnosti	18,058	2,017	19,215	1
2,017	N	Administrativní a podpůrné činnosti	19,215	2,018	20,560	1
2,018	N	Administrativní a podpůrné činnosti	20,560	2,019	22,407	1
2,019	N	Administrativní a podpůrné činnosti	22,407	2,020	23,559	1
2,020	N	Administrativní a podpůrné činnosti	23,559	2,021	24,685	1

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

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 mléka.

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

year	avg_pay_in_year	product_name	avg_product_price_in_year	quantity_to_buy
2,006	20,754	Chléb konzumní kmínový	16.1	1,289
2,006	20,754	Mléko polotučné pasterované	14.4	1,441
2,018	32,536	Chléb konzumní kmínový	24.2	1,344
2,018	32,536	Mléko polotučné pasterované	19.8	1,643



### 3. Která kategorie potravin zdražuje nejpomaleji (je u ní nejnižší procentuálně 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`.

yearly_price	code	name	avarage_price	yearly_price_2	avarage_price_2	percentual_change
2,006	116,103	Banány žluté	27.305	2,007	30.848	13
2,011	116,103	Banány žluté	26.966	2,012	31.003	15
2,012	116,103	Banány žluté	31.003	2,013	31.789	2.5
2,014	116,103	Banány žluté	31.648	2,015	31.914	0.8
2,006	118,101	Cukr krystalový	21.727	2,007	21.921	0.9
2,010	118,101	Cukr krystalový	18.269	2,011	22.45	22.9
2,011	118,101	Cukr krystalový	22.45	2,012	24.267	8.1
2,016	118,101	Cukr krystalový	17.587	2,017	19.961	13.5
2,006	114,501	Eidamská cihla	110.947	2,007	122.68	10.6
2,007	114,501	Eidamská cihla	122.68	2,008	135.279	10.3
2,009	114,501	Eidamská cihla	111.666	2,010	119.952	7.4
2,010	114,501	Eidamská cihla	119.952	2,011	122.383	2
2,011	114,501	Eidamská cihla	122.383	2,012	128.08	4.7
2,012	114,501	Eidamská cihla	128.08	2,013	141.969	10.8
2,013	114,501	Eidamská cihla	141.969	2,014	150.609	6.1
2,016	114,501	Eidamská cihla	110.895	2,017	142.581	28.6
2,006	112,101	Hovězí maso zadní bez kosti	166.337	2,007	169.459	1.9
2,007	112,101	Hovězí maso zadní bez kosti	169.459	2,008	175.549	3.6
2,008	112,101	Hovězí maso zadní bez kosti	175.549	2,009	179.006	2
2,010	112,101	Hovězí maso zadní bez kosti	176.153	2,011	180.729	2.6
2,011	112,101	Hovězí maso zadní bez kosti	180.729	2,012	200.806	11.1

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

name	avg_raise
Jakostní víno bílé	2.7
Pivo výčepní, světlé, lahvé	2.85833
Přírodní minerální voda uhlíčitá	3.26667
Šunkový salám	3.64286
Hovězí maso zadní bez kosti	3.77778
Vepřová pečeně s kostí	4
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 kuchařská celá	7.47143
Banány žluté	7.825
Rýže loupaná dlouhozrná	8.5875
Jablka konzumní	9.84286
Eidamská cihla	10.0625
Chléb konzumní kmínový	10.57143
Cukr krystalový	11.35
Pecivo 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, lahvého** – 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.

payroll_year	avarage_payment	avarage_product_price
2,006	20,754	45.52
2,007	22,173	48.59
2,008	23,918	51.6
2,009	24,674	48.29
2,010	25,156	49.23
2,011	25,736	50.88
2,012	26,516	54.3
2,013	26,103	57.07
2,014	26,771	57.49
2,015	27,443	57.18
2,016	28,444	56.49
2,017	30,231	61.93
2,018	32,536	63.27

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

payroll_year	avarage_payment	avarage_product_price	payroll_year	avarage_payment	avarage_product_price
2,006	20,754	45.52	2,007	22,173	48.59
2,007	22,173	48.59	2,008	23,918	51.6
2,008	23,918	51.6	2,009	24,674	48.29
2,009	24,674	48.29	2,010	25,156	49.23
2,010	25,156	49.23	2,011	25,736	50.88
2,011	25,736	50.88	2,012	26,516	54.3
2,012	26,516	54.3	2,013	26,103	57.07
2,013	26,103	57.07	2,014	26,771	57.49
2,014	26,771	57.49	2,015	27,443	57.18
2,015	27,443	57.18	2,016	28,444	56.49
2,016	28,444	56.49	2,017	30,231	61.93
2,017	30,231	61.93	2,018	32,536	63.27

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.



yearly_pay_change	yearly_price_change
6.8	6.7
7.9	6.2
3.2	-6.4
2	1.9
2.3	3.4
3	6.7
-1.6	5.1
2.6	0.7
2.5	-0.5
3.6	-1.2
6.3	9.6
7.6	2.2

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?

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.

year	GDP	average_payment	average_product_price	year	GDP	average_payment	average_product_price	gdp_change	pay_change	product_change
2006	197,470,142,753.551	20,754	45.52	2007	208,469,898,850.69	22,173	48.59	5.57	6.84	6.74
2007	208,469,898,850.69	22,173	48.59	2008	214,070,259,127.502	23,918	51.6	2.69	7.87	6.19
2008	214,070,259,127.502	23,918	51.6	2009	204,100,298,391.036	24,674	48.29	-4.66	3.16	-6.41
2009	204,100,298,391.036	24,674	48.29	2010	209,069,940,963.177	25,156	49.23	2.43	1.95	1.95
2010	209,069,940,963.177	25,156	49.23	2011	212,750,323,790.759	25,736	50.88	1.76	2.31	3.35
2011	212,750,323,790.759	25,736	50.88	2012	211,080,224,602.858	26,516	54.3	-0.79	3.03	6.72
2012	211,080,224,602.858	26,516	54.3	2013	210,983,331,025.624	26,103	57.07	-0.05	-1.56	5.1
2013	210,983,331,025.624	26,103	57.07	2014	215,755,991,069.112	26,771	57.49	2.26	2.56	0.74
2014	215,755,991,069.112	26,771	57.49	2015	227,381,745,549.048	27,443	57.18	5.39	2.51	-0.54
2015	227,381,745,549.048	27,443	57.18	2016	233,151,067,380.883	28,444	56.49	2.54	3.65	-1.21
2016	233,151,067,380.883	28,444	56.49	2017	245,202,003,265.939	30,231	61.93	5.17	6.28	9.63
2017	245,202,003,265.939	30,231	61.93	2018	253,045,172,103.95	32,536	63.27	3.2	7.62	2.16

gdp_change	pay_change	product_change
5.57	6.84	6.74
2.69	7.87	6.19
-4.66	3.16	-6.41
2.43	1.95	1.95
1.76	2.31	3.35
-0.79	3.03	6.72
-0.05	-1.56	5.1
2.26	2.56	0.74
5.39	2.51	-0.54
2.54	3.65	-1.21
5.17	6.28	9.63
3.2	7.62	2.16

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.