

# Průvodní zpráva

## 1. Zadání 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 požit 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?

### **Výstupy z 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 veřejný repozitář, kam uložíte všechny informace k projektu – hlavně SQL skript generující výslednou tabulku, popis mezivýsledků (průvodní listinu) ve formátu markdown (.md) a informace o výstupních datech (například kde chybí hodnoty apod.).

**Neupravujte data v primárních tabulkách! Pokud bude potřeba transformovat hodnoty, dělejte tak až v tabulkách nebo pohledech, které si nově vytváříte.**

## 2. Popis tvorby primární a sekundární tabulky

Primární tabulku jsem vytvořil ve třech krocích. Nejdřív jsem spojil všechny potřebné informace z tabulek *payroll\_xxx* do tabulky *t\_jaroslav\_demek\_payroll*. Potom jsem spojil potřebné informace z *czechia\_price* a *czechia\_price\_category* do tabulky *jaroslav\_demek\_price*. A jako poslední krok jsem obě tabulky spojil do požadované tabulky *t\_jaroslav\_demek\_project\_SQL\_primary\_final*.

Při tom jsem zjistil nesrovnalosti v tabulkách *payroll\_xx* mezi *value* + *value\_type\_code* na jedné straně s *unit\_code* + *unit\_name*.

Pokud by jsme považovali hodnotu 11332 za tis osob, pak by jsme se dostali ca 11,3 mil osob. To v ČR nemůže být pravda. Hodnota 48 určitě také není průměrná mzda. .

Proto vycházím z toho, že je správně vztah *value*, *value\_type\_code* a *value\_type\_name*. Hodnoty *unit\_code* + *unit\_name* k výpočtům stejně nebudu potřebovat.

123 id	123 value	123 payroll_year	123 value_type_code	A-Z value_type_name	123 unit_code	A-Z unit_name	123 calculate
741,373,810	40	2,006	316	Průměrný počet zaměstnaných osob	80,403	Kč	100
741,373,811	41	2,006	316	Průměrný počet zaměstnaných osob	80,403	Kč	100
741,373,860	47	2,013	316	Průměrný počet zaměstnaných osob	80,403	Kč	100
741,373,862	48	2,013	316	Průměrný počet zaměstnaných osob	80,403	Kč	100
741,373,878	8,793	2,000	5,958	Průměrná hrubá mzda na zaměstnanc	200	tis. osob (tis. os.)	100
741,373,879	9,942	2,000	5,958	Průměrná hrubá mzda na zaměstnanc	200	tis. osob (tis. os.)	100
741,373,880	10,601	2,000	5,958	Průměrná hrubá mzda na zaměstnanc	200	tis. osob (tis. os.)	100
741,373,881	11,332	2,000	5,958	Průměrná hrubá mzda na zaměstnanc	200	tis. osob (tis. os.)	100

Sekundární tabulku jsem vytvořil spojením tabulek *countries* a *economies*. Potom jsem vyseletoval Evropské státy a následně roky 2006-2018. Pak jsem ale zjistil, že v tabulkách nejsou žádné informace o mzdách a cenách, pouze o HDP a GINI.

To mi neumožňuje porovnávat vliv vývoje HDP na vývoj cen a mezd v Evropě. Jedinou možností bylo vyselektovat vývoj HPD pro ČR a porovnat to s vývojem průměrných mezd a cen z předchozích selektů z tabulky *primary*.

### 3. Otázky a odpovědi

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

Většinou mzdy z roku na rok rostly ale neplatí to stoprocentně.

Některé meziroční pohyby průměrných mezd v některých oborech byly nepatrně klesající. Týkalo se to především krizových let 2009-2013 ale nejen jich.

123 payroll_year	A-Z industry_b	Ctrl+click to open SQL console	g_salary	123 prev_year_salary	123 increase_salary
2,009	Těžba a dobývání		28,360.5	29,272.5	0
2,009	Ubytování, stravování a pohostinství		12,333.5	12,471.5	0
2,009	Zemědělství, lesnictví, rybářství		17,644.5	17,763.75	0
2,009	Činnosti v oblasti nemovitostí		20,706	20,790.25	0
2,010	Profesní, vědecké a technické činnosti		31,601.75	31,791.25	0
2,010	Veřejná správa a obrana; povinné sociá		26,943.5	27,034.5	0
2,010	Vzdělávání		23,023	23,416	0
2,011	Doprava a skladování		23,062	23,062.5	0
2,011	Ubytování, stravování a pohostinství		13,131.25	13,205.25	0
2,011	Veřejná správa a obrana; povinné sociá		26,330.75	26,943.5	0
2,011	Výroba a rozvod elektřiny, plynu, tepla		40,201.5	40,295.75	0
2,013	Administrativní a podpůrné činnosti		16,829.25	17,040.5	0
2,013	Informační a komunikační činnosti		46,155	46,641	0

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

Prvním společně sledovaným rokem je rok 2006. V tomto roce bylo možné za průměrnou mzdu v celé ekonomice koupit 1211 kg chleba a 1353 L mléka. V posledním sledovaném roce 2018 to bylo 1321 kg chleba a 1616 L mléka.

123 payroll_year	A-Z category_name	123 avg_salary	123 avg_price	A-Z price_unit	123 milk_bread
2,006	Chléb konzumní kmínový	19,536	16.1234	kg	1,211
2,006	Mléko polotučné pasterované	19,536	14.438	l	1,353
2,018	Chléb konzumní kmínový	32,043	24.2383333333	kg	1,321
2,018	Mléko polotučné pasterované	32,043	19.8191666667	l	1,616

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

Postupnými kroky jsem vyselekoval průměr meziročního pohybu cen ve sledovaných letech pro jednotlivé kategorie. Pokud bych měl otázku vzít formalisticky doslova „nejnižší nárůst“ pak by tou kategorií byly „banány žluté“ s průměrným nárůstem 0,8 %. Odpovědí podle selského rozumu by ale mělo být „Cukr krystalový“ Ten má průměrnou hodnotu meziročního pohybu cen dokonce minus 1.9 %

WITH data\_with\_lag AS ( SELECT year, category\_name, p

	A-Z category_name	123 avg_ratio_percent
1	Cukr krystalový	98.1
2	Rajská jablka červená kulatá	99.3
3	Banány žluté	100.8
4	Přírodní minerální voda uhličitá	101
5	Vepřová pečeně s kostí	101
6	Šunkový salám	101.9

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

Rok ve kterém by rozdíl mezi růstem cen a růstem platů byl vyšší než 10 % ve sledovaném období není. Max. hodnotu tohoto rozdílu najdeme v roce 2007 a to 8,98 %. Pokud si můžu dovolit osobní poznámku, já tento rozdíl za výrazný považuji.

123 year	123 avg_price_change_percent	123 avg_salary_change_percent	123 price_vs_salary_diff
2,007	9.26	0.28	8.98
2,008	8.91	0.3	8.61
2,012	7.47	0.1	7.38
2,017	7.06	0.25	6.81
2,013	6.01	-0.01	6.01
2,011	4.84	0.1	4.75
2,018	2.41	0.3	2.11

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

Přiznávám, že mě výsledek velmi překvapil. Korelace mezi HDP, cenami potravin a pohybem mezd není skoro žádná. Ještě tak mezi HPD a cenami potravin v některých letech jsou hodnoty alespoň trochu u sebe. Ale spojení s mzdami není prostě žádné. Přestože HDP i ceny ve sledovaných letech létaly nahoru dolů v řádu jednotek procent, mzdy se měnily v řádu desetin procent. Srovnání platí pouze pro ČR a sledované roky 2006-2018. Pro ostatní státy EU nemáme data o pohybu cen a mezd.

123 year	123 gdp_per_capita_growth_pct	123 year	123 avg_price_change_percent	123 avg_salary_change_percent
2,006	[NULL]	2,007	9.26	0.28
2,007	4.96	2,008	8.91	0.3
2,008	1.84	2,009	-6.57	0.13
2,009	-5.2	2,010	1.51	0.08
2,010	2.14	2,011	4.84	0.1
2,011	1.55	2,012	7.47	0.1
2,012	-0.92	2,013	6.01	-0.01
2,013	-0.08	2,014	-0.63	0.11
2,014	2.15	2,015	-0.69	0.12
2,015	5.18	2,016	-1.41	0.16
2,016	2.34	2,017	7.06	0.25
2,017	4.89	2,018	2.41	0.3
2,018	2.85			