

Datový analytik s Pythonem

Projekt z SQL

Lenka Krčmáriková

2025

1. Zadání projektu

1.1. 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.

1.2. 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 České republice:

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.

1.3. 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?

1.4. Výstupy z projektu

Výstupem projektu bylo tvoření datové sady pro vytvoření vhodného datového podkladu:

- t_lenka_krcmarikova_projekt_sql_primary_final – pro data mezd a cen potravin za Českou republiku sjednocených totožné porovnatelné období,
- t_lenka_krcmarikova_projekt_SQL_secondary_final – pro dodatečná data o dalších evropských státech

2. Analýza dat

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

Obecně lze předpokládat, že mzdy mají rostoucí tendenci bez ohledu na odvětví. Přesto na základě analýzy dat mezd v České republice za období 2007 až 2018 vyplývá, že v některých letech a v odvětvích došlo k jejich poklesům.

Před analýzou bychom si mohli stanovit hypotézu, že nejvýznamnější pokles byl v roce 2008/2009, kdy byla globální krize a pak sledovat, zda data naši hypotézu potvrdí či vyvrátí.

Data ukazují, že nejvýraznější pokles byl zaznamenán v roce 2013, kdy došlo ke snížení průměrné mzdy v celkem 11 odvětvích z celkového počtu 19.

Závěrem bychom mohli říci, že zatímco globální krize v roce 2008/2009 neměla okamžitý a výrazný dopad na průměrné mzdy ve většině odvětví, v roce 2013 dochází v České republice k největšímu počtu poklesů. Může jít o důsledek domácí ekonomické stagnace a vládních úspor. To naznačuje, že nejen globální události, ale i lokální ekonomická situace může ovlivnit vývoj mezd v jednotlivých odvětvích.

Když se na data dále podíváme, zjistíme, že:

- největší pokles průměrných mezd ve sledovaném období byl v Peněžnictví a pojišťovnictví, a to zhruba o Kč 4 480,-,
- v několika odvětvích došlo k poklesu opakovaně, a to:
 - Těžba a dobývání v letech 2009, 2013, 2014, 2016,
 - Veřejná správa v letech 2010, 2011,
 - Kulturní a zábavní činnosti v letech 2011, 2013,
 - Výroba a rozvod elektřiny v letech 2013, 2015.

Rok	Počet odvětví	Počet rostoucích odvětví	Počet klesajících odvětví	Počet odvětví beze změny
2007	19	19	0	0
2008	19	19	0	0
2009	19	16	3	0
2010	19	16	3	0
2011	19	16	3	0
2012	19	19	0	0
2013	19	8	11	0
2014	19	18	1	0
2015	19	18	1	0
2016	19	18	1	0
2017	19	19	0	0
2018	19	19	0	0

Tabulka 1: Počet rostoucích / klesajících odvětví

Rok	Název odvětví	Průměrná mzda	Mzda za předchozí rok	Rozdíl mezd	Procentuální změna mezd
2009	Těžba a dobývání	28160.50	29254.25	-1093.75	-3.74
2009	Ubytování, stravování a pohostinství	12058.13	12204.50	-146.37	-1.20
2009	Zemědělství, lesnictví, rybářství	17418.75	17528.13	-109.38	-0.62
2010	Profesní, vědecké a technické činnosti	30848.25	31037.63	-189.38	-0.61
2010	Veřejná správa a obrana; povinné sociální zabezpečení	26707.00	26794.88	-87.88	-0.33
2010	Vzdělávání	21975.50	22388.38	-412.88	-1.84
2011	Kulturní, zábavní a rekreační činnosti	19213.63	19222.50	-8.87	-0.05
2011	Ubytování, stravování a pohostinství	12781.25	12924.88	-143.63	-1.11
2011	Veřejná správa a obrana; povinné sociální zabezpečení	26109.00	26707.00	-598.00	-2.24
2013	Administrativní a podpůrné činnosti	16427.63	16487.75	-60.12	-0.36
2013	Informační a komunikační činnosti	45697.50	46165.25	-467.75	-1.01
2013	Kulturní, zábavní a rekreační činnosti	19901.75	20179.38	-277.63	-1.38
2013	Peněžnictví a pojišťovnictví	45775.00	50253.50	-4478.50	-8.91
2013	Profesní, vědecké a technické činnosti	31049.50	31979.00	-929.50	-2.91
2013	Stavebnictví	22226.50	22710.75	-484.25	-2.13
2013	Těžba a dobývání	31586.13	32513.63	-927.50	-2.85
2013	Velkoobchod a maloobchod; opravy a údržba motorových vozidel	22666.75	22882.38	-215.63	-0.94
2013	Výroba a rozvod elektřiny, plynu, tepla a klimatiz. vzduchu	40530.75	42382.25	-1851.50	-4.37
2013	Zásobování vodou; činnosti související s odpady a sanacemi	23400.00	23489.63	-89.63	-0.38
2013	Činnosti v oblasti nemovitostí	21622.75	21995.50	-372.75	-1.69
2014	Těžba a dobývání	31336.00	31586.13	-250.13	-0.79
2015	Výroba a rozvod elektřiny, plynu, tepla a klimatiz. vzduchu	40483.50	41022.50	-539.00	-1.31
2016	Těžba a dobývání	31487.13	31674.00	-186.87	-0.59

Tabulka 2: Přehled klesajících odvětví ve vybraných letech

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

Na základě analýzy dat mezd a cen základních potravin (v tomto případě chleba a mléka) v České republice lze říci, že:

- v roce 2026 bylo možné si za průměrnou mzdu koupit:
 - 1287 kilogramů chleba,
 - 1437 litrů mléka a
- v roce 2018:
 - 1342 kilogramů chleba, čímž došlo k nárůstu o 57 kg,
 - 1642 litrů mléka, čímž došlo k nárůstu o 205 litrů.

Závěrem lze konstatovat, že kupní síla se za sledované období u obou sledovaných potravin zvýšila. Z tohoto zjednodušeného pohledu to znamená, že si člověk za svou průměrnou mzdu mohl v roce 2018 koupit více chleba a mléka než v roce 2016, což by mohlo znamenat, že došlo k zvýšení životní úrovně obyvatel v této oblasti.

Rok	Potravina	Průměrná cena potravin	Jednotka	Roční průměrná mzda	Kupní síla
2006	Chléb konzumní kmínový	16.12	kg	20753.79	1287.46
2006	Mléko polotučné pasterované	14.44	l	20753.79	1437.24
2018	Chléb konzumní kmínový	24.24	kg	32535.86	1342.24
2018	Mléko polotučné pasterované	19.82	l	32535.86	1641.57

Tabulka 3: Kupní síla vybraných potravin v prvním a posledním srovnatelném období

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

Z analýzy dat cen základních potravin v České republice ve sledovaném období 2006-2018 bylo zjištěno:

- potraviny s nejnižším nárůstem (případným poklesem) cen jsou:
 - Cukr krystalový,
 - Rajská jablka červená kulatá,
 - Banány žluté a
- potraviny s nejvyšším nárůstem cen jsou:
 - Papriky,
 - Máslo,
 - Vejce slepičí čerstvá.

Název potraviny	Průměrná meziroční změna	Sledované období (roky)
Cukr krystalový	-1.92	12
Rajská jablka červená kulatá	-0.74	12
Banány žluté	0.81	12
Vepřová pečeně s kostí	0.99	12
Přírodní minerální voda uhličitá	1.02	12
Šunkový salám	1.86	12
Jablka konzumní	2.01	12
Pečivo pšeničné bílé	2.20	12
Hovězí maso zadní bez kosti	2.54	12
Kapr živý	2.60	12
Pivo výčepní, světlé, lahvové	2.85	12
Eidamská cihla	2.92	12
Mléko polotučné pasterované	2.98	12
Rostlinný roztíratelný tuk	3.23	12
Kuřata kuchaná celá	3.38	12
Pomeranče	3.60	12
Jogurt bílý netučný	3.95	12
Chléb konzumní kmínový	3.97	12
Konzumní brambory	4.18	12
Rýže loupaná dlouhozrná	5.00	12
Mrkev	5.23	12
Pšeničná mouka hladká	5.24	12
Těstoviny vaječné	5.26	12
Vejce slepičí čerstvá	5.56	12
Máslo	6.68	12
Papriky	7.29	12

Tabulka 4: Přehled meziročních cenových změn potravin

Do analýzy byly zařazeny pouze potraviny, které mají alespoň pět let sledování, z tohoto důvodu bylo vyřazeno Jakostní víno bílé (aby se vyhnulo zkresleným výsledkům potravin).

Pro zjištění, jestli Jakostní víno bílé zdražuje pomaleji nebo rychleji, bylo třeba omezit sledované období na roky 2016 až 2018. Z těchto dat pak vyplývá, že:

- potraviny s nejpomalejším nárůstem cen jsou:
 - Cukr krystalový,
 - Pšeničná mouka hladká,
 - Banány žluté a

- potraviny s nejvyšším nárůstem cen:
 - Máslo,
 - Vejce slepičí čerstvá,
 - Eidamská cihla.

Název potraviny	Průměrná meziroční změna	Sledované období (roky)
Cukr krystalový	-2.93	3
Pšeničná mouka hladká	-2.89	3
Banány žluté	-2.76	3
Papriky	-1.6	3
Přírodní minerální voda uhlíčitá	-0.26	3
Kuřata kuchaná celá	-0.12	3
Rýže loupaná dlouhozrná	0.24	3
Mléko polotučné pasterované	0.73	3
Rajská jablka červená kulatá	1.52	3
Kapr živý	2.42	3
Vepřová pečeně s kostí	2.53	3
Těstoviny vaječné	2.53	3
Jogurt bílý netučný	2.62	3
Jakostní víno bílé	2.7	3
Chléb konzumní kmínový	2.77	3
Rostlinný roztíratelný tuk	2.81	3
Hovězí maso zadní bez kosti	3.23	3
Pivo výčepní, světlé, lahvové	3.43	3
Pečivo pšeničné bílé	3.73	3
Šunkový salám	4.05	3
Pomeranče	4.07	3
Konzumní brambory	4.47	3
Mrkev	6.47	3
Jablka konzumní	6.69	3
Eidamská cihla	7.12	3
Vejce slepičí čerstvá	8.15	3
Máslo	10.83	3

Tabulka 5: Přehled meziročních cenových změn potravin za rok 2016 až 2018

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

Z dat bylo zjištěno, že neexistuje rok, ve kterém by byl meziroční nárůst cen potravin výrazně vyšší než růst mezd, tedy vyšší než 10 %. Nejblíže tomu byl rok 2013, kdy rostly ceny o 5,1 % a mzdy rostly o -1,56 %.

Rok	Procentuální změna cen	Procentuální změna mezd	Rozdíl procentuální změny cena a mezd
2007	6.74	6.84	-0.09
2008	6.19	7.87	-1.68
2009	-6.41	3.16	-9.57
2010	1.95	1.95	-0.01
2011	3.35	2.3	1.05
2012	6.72	3.03	3.69
2013	5.1	-1.56	6.66
2014	0.74	2.56	-1.82
2015	-0.54	2.51	-3.05
2016	-1.21	3.65	-4.86
2017	9.63	6.28	3.35
2018	2.16	7.62	-5.46

Tabulka 6: Meziroční vývoj cen a mezd za sledované období

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?

Když se podíváme na roky s výrazným růstem HDP:

- 2010 (+9,86 %),
- 2007 (+4,75 %),
- 2018 (+2,58 %) – menší růst, ale ještě můžeme sledovat,
- 2014, 2015, 2016, 2017 – byly spíše menší.

Zkontrolujeme, jestli ve stejném roce rostou mzdy / ceny:

- 2010: HDP prudce roste (+9,86 %), ale mzdy jen +1,95 % a ceny +1,95 % => okamžitý vliv není silný,
- 2007: HDP roste +4,75 %, mzdy +6,84 %, ceny +6,74 % => rostou souběžně,
- 2018: HDP +2,58 %, mzdy +7,62 %, ceny +2,16 % => mzdy rostou více než HDP, ceny jen lehce.

Když se podíváme na následující rok (opozděný vliv):

- po 2010 (2011): mzdy +2,30 %, ceny +3,35 % => žádný skokový růst, spíše mírný,
- po 2007 (2008): mzdy +7,87 %, ceny 16,19 % => růst pokračuje, zde opožděný růst může být,
- po 2018 (2019): data nemáme.

Závěrem lze říci, že neexistuje jednoznačný pravidelný vzorec, že by silný růst HDP vždy znamenal okamžitý růst mezd nebo cen potravin.

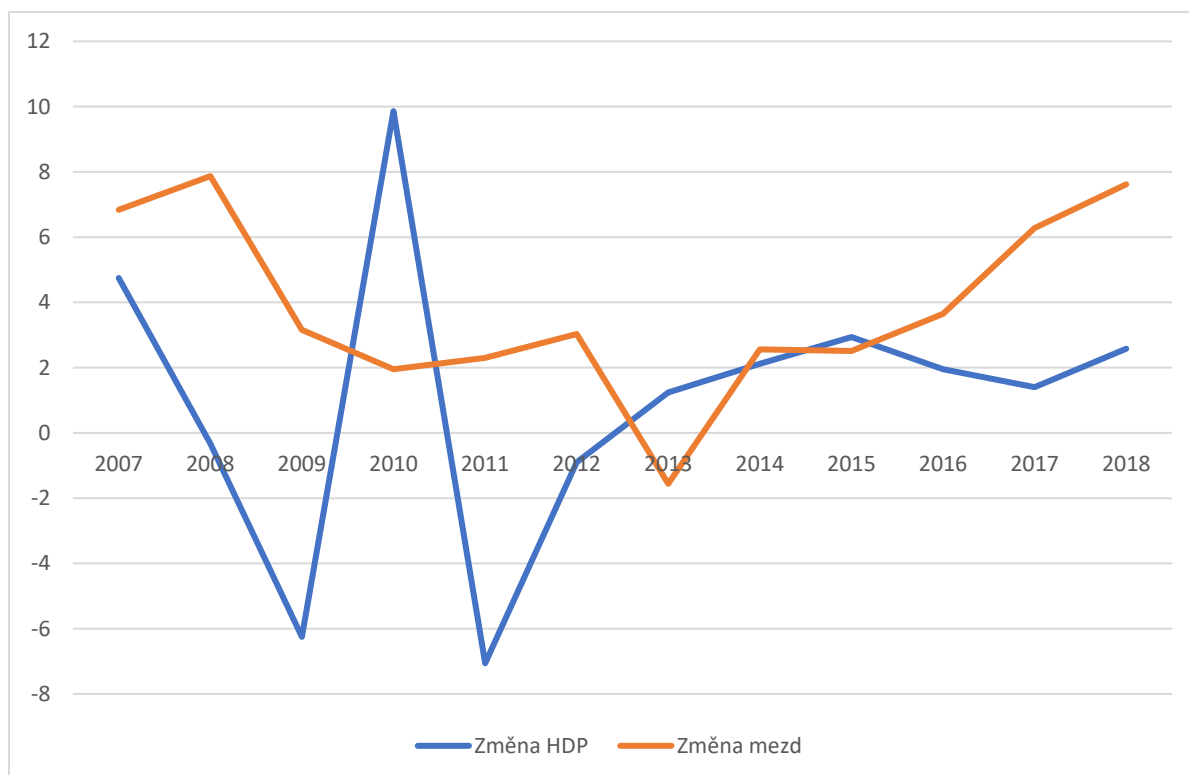
Někdy je vliv okamžitý (2007), někdy opožděný (2008, po růstu v roce 2007), a někdy žádný (2010).

U mzdového růstu je vidět, že spíše reaguje na dlouhodobější trend HDP než na jeden konkrétní rok.

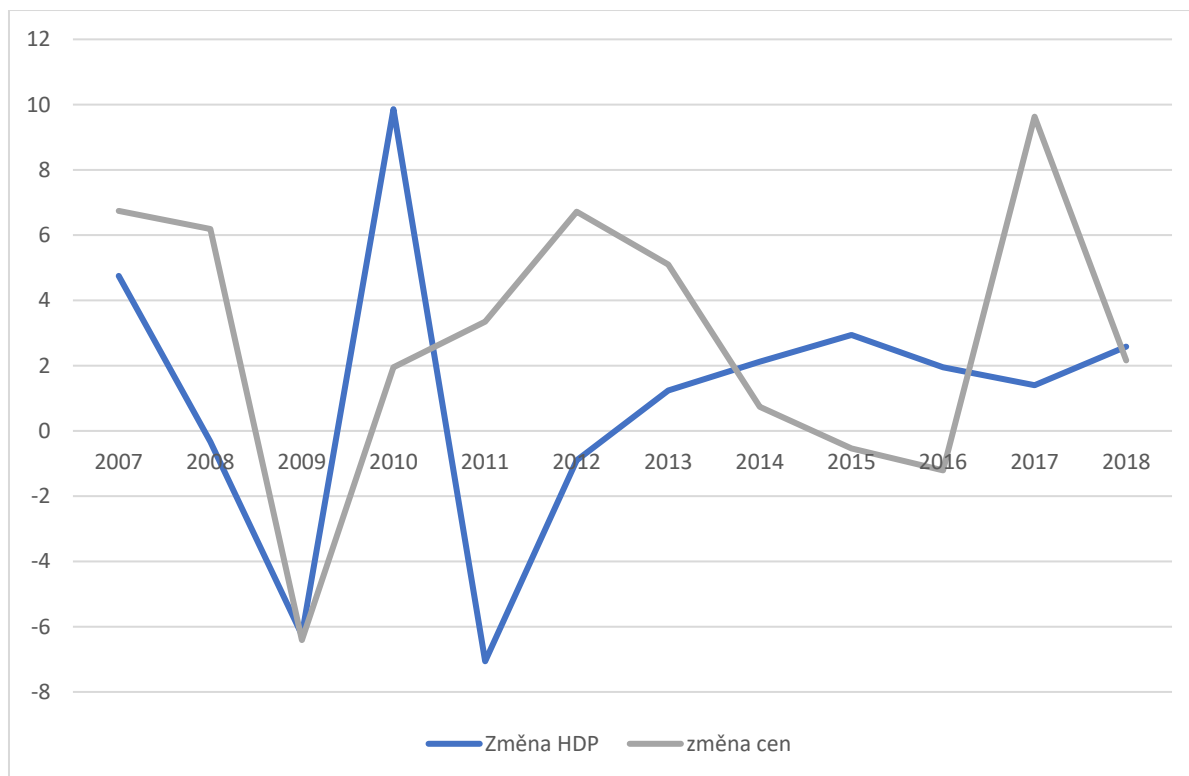
Ceny potravin se v některých letech hýbou nezávisle na HDP. Například 2009 – HDP jde prudce dolů, ceny také jdou prudce dolů, ale 2013 – HDP lehce nahoru, ceny výrazně rostou).

Rok	Změna HDP	Změna mezd	Změna cen	Rozdíl mezd a HDP	Rozdíl cen a HDP
2007	4.75	6.84	6.74	2.09	1.99
2008	-0.33	7.87	6.19	8.20	6.53
2009	-6.25	3.16	-6.41	9.41	-0.16
2010	9.86	1.95	1.95	-7.91	-7.92
2011	-7.06	2.30	3.35	9.37	10.42
2012	-0.90	3.03	6.72	3.93	7.62
2013	1.24	-1.56	5.10	-2.80	3.86
2014	2.12	2.56	0.74	0.43	-1.39
2015	2.94	2.51	-0.54	-0.44	-3.48
2016	1.95	3.65	-1.21	1.70	-3.16
2017	1.40	6.28	9.63	4.89	8.23
2018	2.58	7.62	2.16	5.05	-0.41

Tabulka 7: Vývoj HDP, mezd a cen za sledované období



Graf 1: Vývoj HDP a mezd za sledované období



Graf 2: Vývoj HDP a cen za sledované období