

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.

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:

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.

Číselníky sdílených informací o ČR

czechia_region – Číselník krajů České republiky dle normy CZ-NUTS 2.

czechia_district – Číselník okresů České republiky dle normy LAU.

Dodatečné tabulky:

countries - Všechné informace o zemích na světě, například hlavní město, měna, národní jídlo nebo průměrná výška populace.

economies - HDP, GINI, daňová zátěž, atd. pro daný stát a rok.

Požadované výstupy z projektu

Výstupem jsou dvě tabulky v databázi a dále odpovědi na pět dotazů, které jsou z těchto tabulek získány.

Zdrojová tabulka 1:

t_milan_krocil_SQL_primary_final (pro data mezd a cen potravin za Českou republiku sjednocených na totožné porovnatelné období – společné roky)

Tabulka je vytvořena spojením tabulek:

czechia_price,

czechia_price_category,

czechia_payroll,

czechia_payroll_calculation,

czechia_payroll_industry_branch,

czechia_payroll_unit,

czechia_payroll_value_type

Zdrojová tabulka 2:

t_{jmeno}_{prijmeni}_project_SQL_secondary_final (pro dodatečná data o dalších evropských státech).

Tabulka je vytvořena sloučením tabulky *economies* a materializovaného pohledu *mv_comparison_avg_price_vs_avg_salary*.

Tabulka *economies* : slouží k získání požadovaných dat HDP, GINI a populací pro všechny země.

Materializovaný pohled *mv_comparison_avg_price_vs_avg_salary* slouží k výběru stejného období, ve kterém byla provedena analýza v posledním dotazu 5.

Analýza dotazů:

DOTAZ 1:

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

Zdroj dat:

Data pro analýzu byla čerpána z konsolidované tabulky *t_milan_krocil_project_sql_primary_final*.

Komentář ke skriptu:

Skript zahrnuje dvě dočasné tabulky:

1. *cte_average_salary_in_branches_per_years* - obsahuje výpočet průměrných mezd v jednotlivých odvětvích pro konkrétní rok.

2. *cte_previous_year_added*, která doplňuje průměrné mzdy z předchozího roku.

Data jsou analyzována za období 2006-2018, kdy byla dostupná data pro průměrné mzdy v těchto odvětvích.

Hlavní skript potom využívá tato data k výpočtu rozdílů mezi těmito průměrnými mzdami a následně je zařazuje do jednotlivých kategorií.

Kategorie jsou definovány následovně:

Kategorie „*Increase*“ – data, kde došlo k růstu průměrné mzdy oproti roku předchozímu.

Kategorie „*Decrease*“ – data, kde došlo k poklesu průměrné mzdy oproti roku předchozímu.

Vyhodnocení:

Z dat vyplývá, že existují odvětví, kde během období došlo ke snížení průměrné mzdy.

	A# name_branch	123 payroll_year	123 average_salary	123 previous_salary	A# differences
1	Administrativní a podpůrné činnosti	2,010	15,908.82	15,923.46	Decrease
2	Administrativní a podpůrné činnosti	2,013	16,831.6	17,042.78	Decrease
3	Činnosti v oblasti nemovitostí	2,008	20,631.6	20,709.8	Decrease
4	Činnosti v oblasti nemovitostí	2,013	22,154.02	22,556.61	Decrease
5	Informační a komunikační činnosti	2,013	46,160.34	46,649.13	Decrease
6	Kulturní, zábavní a rekreační činnosti	2,013	20,515.25	20,817.1	Decrease
7	Ostatní činnosti	2,010	18,299.77	18,332.93	Decrease
8	Peněžnictví a pojišťovnictví	2,013	46,316.32	50,819.5	Decrease
9	Profesní, vědecké a technické činnosti	2,010	31,442	31,803.44	Decrease
10	Profesní, vědecké a technické činnosti	2,013	31,832.7	32,829.33	Decrease
11	Stavebnictví	2,013	22,384.34	22,857.81	Decrease
12	Těžba a dobývání	2,009	28,382.57	28,694.09	Decrease
13	Těžba a dobývání	2,013	31,505.34	32,560.9	Decrease
14	Těžba a dobývání	2,014	31,315.39	31,505.34	Decrease
15	Těžba a dobývání	2,016	31,635.88	31,823.21	Decrease
16	Ubytování, stravování a pohostinství	2,009	12,335.02	12,427.58	Decrease
17	Ubytování, stravování a pohostinství	2,011	13,132.36	13,166.85	Decrease
18	Velkoobchod a maloobchod; opravy a	2,013	23,133.4	23,330.66	Decrease
19	Veřejná správa a obrana; povinné soci	2,010	26,891.4	27,043.71	Decrease
20	Veřejná správa a obrana; povinné soci	2,011	26,337.31	26,891.4	Decrease
21	Výroba a rozvod elektřiny, plynu, tepla	2,013	40,770.25	42,681.04	Decrease
22	Výroba a rozvod elektřiny, plynu, tepla	2,015	40,461.8	41,102.15	Decrease
23	Vzdělávání	2,010	22,939.4	23,428.24	Decrease
24	Zásobování vodou; činnosti související	2,013	23,623.81	23,725.92	Decrease

Níže jsou uvedené odvětví, kde došlo minimálně v jednom ze sledovaných období k poklesu průměrné mzdy.

	A-Z branches_with DECREASING salary	A-Z differences
1	Administrativní a podpůrné činnosti	Decrease
2	Činnosti v oblasti nemovitostí	Decrease
3	Informační a komunikační činnosti	Decrease
4	Kulturní, zábavní a rekreační činnosti	Decrease
5	Ostatní činnosti	Decrease
6	Peněžnictví a pojišťovnictví	Decrease
7	Profesní, vědecké a technické činnosti	Decrease
8	Stavebnictví	Decrease
9	Těžba a dobývání	Decrease
10	Ubytování, stravování a pohostinství	Decrease
11	Velkoobchod a maloobchod; opravy a údržba motorových vozidel	Decrease
12	Veřejná správa a obrana; povinné sociální zabezpečení	Decrease
13	Výroba a rozvod elektřiny, plynu, tepla a klimatiz. vzduchu	Decrease
14	Vzdělávání	Decrease
15	Zásobování vodou; činnosti související s odpady a sanacemi	Decrease

DOTAZ 2:

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

Zdroj dat:

Data pro analýzu byla čerpána z konsolidované tabulky *t_milan_krocil_project_sql_primary_final*.

Komentář ke skriptu:

Skript zahrnuje tři dočasné tabulky. První tabulka *cte_avg_salary* slouží ke zjištění průměrných platů v jednotlivých kvartálech. Druhá tabulka *cte_price_selected_items* slouží k výběru definovaných položek mléka a chleba. Třetí tabulka *cte_selected_period* slouží k určení prvního a posledního sledovaného období.

Hlavní skript využívá zmíněné dočasné tabulky k výpočtu množství mléka a chleba, které je možné koupit v prvním (první kvartál 2006) a posledním (poslední kvartál 2018) sledovacím období.

Vyhodnocení:

V prvním kvartálu 2006 bylo možné za průměrný plat koupit 1358ks chleba a nebo 1405 litrů mléka. V posledním sledovacím období bylo možné za průměrný plat koupit 1471ks chleba nebo 1803 litrů mléka.

	A-Z f_year	A-Z f_quarter	A-Z name	123 qty_of_purchase
1	2006	1	Chléb konzumní kmínový	1,358
2	2006	1	Mléko polotučné pasterované	1,405
3	2018	4	Chléb konzumní kmínový	1,471
4	2018	4	Mléko polotučné pasterované	1,803

DOTAZ 3:

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

Zdroj dat:

Data pro analýzu byla čerpána z konsolidované tabulky *t_milan_krocil_project_sql_primary_final*.

Komentář ke skriptu:

Skript zahrnuje dvě dočasné tabulky *cte_average_category_price* a *cte_previous_year_avg_price*, ve kterých jsou vypočteny průměrné ceny jednotlivých kategorií potravin konkrétního a předchozího roku v obdobích 2006 až 2018.

Hlavní skript poté využívá tato data k výpočtu procentuálního meziročního změny. Poté jsou jednotlivé změny zobrazeny pouze pro ty kategorie, kde došlo k nárůstu. Nakonec jsou seřazeny podle tohoto nárůstu vzestupně.

Vyhodnocení:

Kategorie potravin, u které dochází k nejnižšímu meziročnímu procentuálnímu nárůstu, je Jakostní víno bílé a to o 8.1%.

	A-Z name	percentage_calculation
1	Jakostní víno bílé	8.1
2	Banány žluté	9.7
3	Vepřová pečeně s kostí	11.9
4	Přírodní minerální voda uhlíčitá	12.3
5	Šunkový salám	22.2
6	Jablka konzumní	24.2
7	Pečivo pšeničné bílé	26.4
8	Hovězí maso zadní bez kosti	30.4
9	Kapr živý	31.2
10	Pivo výčepní, světlé, lahvé	34.3
11	Eidamská cihla	35.1
12	Mléko polotučné pasterované	35.8
13	Rostlinný roztíratelný tuk	38.8
14	Kuřata kuchařská celá	40.5
15	Pomeranče	43.2
16	Jogurt bílý netučný	47.4
17	Chléb konzumní kmínový	47.6
18	Konzumní brambory	50.2
19	Rýže loupaná dlouhozrná	60
20	Mrkev	62.8
21	Pšeničná mouka hladká	62.8
22	Těstoviny vaječné	63.2

DOTAZ 4:

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

Zdroj dat:

Data pro analýzu byla čerpána z konsolidované tabulky *t_milan_krocil_project_sql_primary_final*.

Komentář ke skriptu:

Skript zahrnuje dvě dočasné tabulky *cte_average_price* a *cte_average_salary*, ve kterých jsou vypočteny průměrné hodnoty cen potravin a průměrné hodnoty mezd v obdobích 2006 až 2018. Jsou zde zároveň doplněny údaje o průměrných cenách z předchozího roku.

Hlavní skript poté využívá tato data k výpočtu procentuálního nárůstu cen potravin a mezd a porovnává jejich rozdíl v jednotlivých letech. Nakonec je ze skriptu vytvořen materializovaný pohled *mv_comparison_avg_price_vs_avg_salary*, který je použit v následujícím dotazu 5.

Vyhodnocení:

V žádném roce nebylo procentuální zvýšení cen potravin vyšší než 10% oproti procentuálním zvýšení mezd. Nejvyšší bylo o 7,11% v roce 2013.

Pro zajímavost, nejvyšší procentuální zvýšení mezd oproti cenám potravin bylo v roce 2009 a to 10,75%.

	Az f_year	123 percentage_change_price	123 percentage_change_salary	123 proc_difference_price_vs_salary
1	2006	[NULL]	[NULL]	[NULL]
2	2013	5.5471698113	-1.5662927191	7.1134625304
3	2012	6.9410815174	2.939130634	4.0019508834
4	2017	9.9837808614	6.1732250605	3.8105558009
5	2011	3.4871580706	2.6051824748	0.8819755958
6	2010	1.7637059074	1.6351613842	0.1285445232
7	2008	6.406743941	6.8451679464	-0.4384240054
8	2007	6.342447333	6.965249906	-0.622802573
9	2014	0.8938148016	2.5714602202	-1.6776454186
10	2015	-0.5669737775	2.5979068346	-3.1648806121
11	2016	-1.1225944405	3.6358409214	-4.7584353618
12	2018	1.93347534	7.6929584467	-5.7594831067
13	2009	-6.7934244405	3.9592490501	-10.7526734906

DOTAZ 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?

Zdroj dat:

Data pro analýzu byla čerpána z dodatečné tabulky *economies* a dále z materializovaného pohledu *mv_comparison_avg_price_vs_avg_salary*. Tento materializovaný pohled byl vytvořen při analýze v předchozím dotazu 4 (vytvořeno z výsledného skriptu) a obsahuje hodnoty procentuálních změn cen a platů v letech 2006 až 2018, kdy byla dostupná data pro obě kategorie.

Kategorie vyhodnocení:

Pro vyhodnocení byly nadefinovány kategorie, do kterých jsou výsledné změny zařazeny podle následujících kritérií:

-- Kategorie "low impact" = zde jsou zařazeny změny, kde procentuální změna ceny nebo mzdy byla menší než jedno procento.

-- Kategorie "medium impact" = zde jsou zařazeny změny, kde procentuální změna ceny nebo mzdy byla v rozmezí jednoho až pěti procent.

-- Kategorie "high impact" = zde jsou zařazeny změny, kde procentuální změna ceny nebo mzdy byla větší než pět procent.

-- Poznámka: rozmezí (hranice) jednotlivých kategorií se dají změnit dle požadavku zadavatele.

Komentář ke skriptu:

Skript zahrnuje dočasnou tabulku *cte_adding_gdp*, ve kterém jsou data HDP vybrána jen pro Českou republiku a doplněna o procentuální výpočet HDP. Dále pak o procentuální změnu HDP předchozího roku.

Hlavní skript spojuje zmíněné CTE a materializovaný pohled

mv_comparison_avg_price_vs_avg_salary a k datům z *cte_adding_gdp* doplňuje procentuální změny cen mezd a cen z jednotlivých let (vždy současnou a následující meziroční změnu).

Změny jsou poté zařazeny do kategorií podle kritérií zmíněných v předchozím odstavci "Kategorie vyhodnocení".

Vyhodnocení:

Výška změny HDP na ceny potravin a mzdy v tomtéž a následujícím roce je nejvíce patrná u roku 2009. Změna mzdy byla oproti změně HDP v kategorii "high impact" (větší než 5%) v daném i následujícím roce a změna cen byla větší než 5% v tom samém roce a v kategorii "medium impact" v roce následujícím.

Naopak nejmenší výška změny HDP na ceny potravin a mezd je patrná u roku 2010, kdy většina změn je v kategorii "low impact".

	year	impact_price_vs_gdp_same_year	impact_price_vs_gdp_next_year	impact_salary_vs_gdp_same_year	impact_salary_vs_gdp_next_year
1	2,006	[NULL]	low impact	[NULL]	low impact
2	2,007	low impact	low impact	medium impact	medium impact
3	2,008	medium impact	high impact	medium impact	medium impact
4	2,009	medium impact	high impact	high impact	high impact
5	2,010	low impact	medium impact	low impact	low impact
6	2,011	medium impact	high impact	low impact	medium impact
7	2,012	high impact	high impact	medium impact	low impact
8	2,013	high impact	low impact	medium impact	medium impact
9	2,014	medium impact	medium impact	low impact	low impact
10	2,015	high impact	high impact	medium impact	medium impact
11	2,016	medium impact	high impact	medium impact	medium impact
12	2,017	medium impact	medium impact	medium impact	medium impact
13	2,018	medium impact	[NULL]	medium impact	[NULL]