

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](#) - 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

Zadáním projektu bylo zodpovězení na 5 výzkumných otázek, týkajících se dat ohledně cen, mezd a HDP v průběhu pozorovaného období.

Na začátku bylo nutné vytvořit základní tabulky pro následnou práci s daty, na základě již předem připravených dat o cenách a mzdách v České republice, společně s daty o, mimo jiném, HDP ve všech státech světa.

Tyto tabulky byly vytvořeny a pojmenovány `t_Zana_Masarova_project_SQL_primary_final` (pro data mezd a cen potravin za Českou republiku sjednocených na totožné porovnatelné období a `t_Zana_Masarova_project_SQL_secondary_final` (pro dodatečná data o dalších evropských státech). Soubor s query pro vytvoření těchto tabulek je pojmenován `project_tables`.

Na základě těchto dat bylo poté možné zpracovat jednotlivé otázky a zjistit případně souvislosti mezi daty.

Pro jednotlivé otázky jsou vytvořeny samostatné soubory s query, pod názvy `project_task1` – `project_task5`.

Odpovědi na otázky:

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

Na základě dat pro mzdy v jednotlivých odvětvích v průběhu let lze pozorovat, že mzdy ve většině období v průběhu let meziročně rostou.

Lze však na datech pozorovat i období, kdy mzda v následujícím roce poklesla.

Pro rozlišení těchto případů je v tabulce vytvořen sloupec `salary_change`, ve kterém lze vyčíst zda mzda rostla (*salary is higher compared to previous year*) či klesala (*salary is lower compared to previous year*).

Na základě tohoto sloupce vidíme, že ve 23 z 228 případů, pro některá odvětví v některých letech pozorujeme pokles mzdy. Nicméně ve většině pozorovaného období jde o meziroční růst.

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

Prvním pozorovaným obdobím, které bylo společné pro daná data, byl rok 2006. Posledním pak rok 2018.

Pro zodpovězení výzkumné otázky jsme v přidaném sloupci *quantity_bought* vypočítali následující hodnoty:

- V roce 2006 bylo možné v průměru nakoupit 1 192 kg chleba a 1 331 l mléka za danou mzdu
- V roce 2018 bylo možné v průměru nakoupit 1 300 kg chleba a 1 590 l mléka za danou mzdu

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

Na základě dat cen jednotlivých kategorií potravin v daných letech jsme zkoumali procentuální růst v průběhu sledovaného období pro jednotlivé kategorie.

Pro toto pozorování jsme vytvořili dodatečný sloupec *avg_price_change_percentage*, ve kterém jsme spočítali průměrný nárůst/pokles cen pro celé pozorované období podle jednotlivých kategorií.

Pomocí dat v tomto sloupci můžeme pozorovat, že nejpomaleji rostla cena žlutých banánů, která za celé obdob vzrostla pouze o 0,81%, dále pak cena vepřové pečeně vzrostla o 0,99% a cena minerální vody pak o 1,03%.

Zároveň můžeme pozorovat, že za sledované období u dvou kategorií cena poklesla, a to u cukru krystal o 1,92% a u rajských jablek pak o 0,74%.

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

Pro zodpovězení této otázky jsme srovnali data pro změny mezd a cen v jednotlivých letech, sjednocených pro všechna odvětví a kategorie potravin. Pracovali jsme tedy s průměry mezd i potravin pro jednotlivé roky.

Na základě těchto dat jsme vytvořili sloupce pro pozorování meziročních změn mezd a cen a následně pro jejich rozdíl (*price_salary_change_diff*). Na základě dat z tohoto sloupce můžeme pozorovat, že nárůst cen potravin v pozorovaném období významně nepřevyšuje nárůst mezd (více než 10%*).

* V tomto případě považuji za cíl otázky zjistit rozdíl v procentních bodech, jelikož pozorované změny jsou v procentech.

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?

Na základě dat cen, mezd a HDP jsme pozorovali jednotlivé změny pro srovnání vztahu mezi nimi.

Pro těchto změn byly vytvořeny sloupce s hodnotami procentuálních meziročních změn – *avg_salary_perc_change* pro změny průměrných mezd, *avg_price_perc_change* pro změny průměrných cen a *GDP_perc_change* pro změny HDP.

Na základě těchto dat lze pozorovat, že v případě výraznějšího růstu HDP (pro dané data byla zvolena hodnota 5% jako významnější) lze pozorovat jistý vztah mezi růstem mzdy a cen v následujícím roce. Významnější růst můžeme pozorovat v roce 2007, 2015 a 2017.

V roce 2007 došlo k růstu HDP o 5,57%, mzdy vzrostly v témže roce o 7,23% a ceny o 9,26%.

V roce 2008 poté rostlo HDP pouze polovičním tempem, o 2,69%. Meziroční nárůst mezd byl vyšší, 7,91% a nárůst cen naopak poklesl na 8,91%. Tudíž i přes nižší HDP v daném roce došlo stále k růstu mezd i cen, nicméně mzdy rostly rychlejším tempem, zatímco ceny pomalejším.

Stejná změny můžeme pozorovat mezi lety 2015 a 2016.

V roce 2015 meziročně vzrostlo HDP o 5,39%, nicméně mzdy vzrostly pouze o 3,18%, zatímco ceny naopak klesaly o 0,7%.

V roce 2016 následně HDP opět rostlo polovičním tempem, mzdy opět vzrostly o 4,43% (tedy rychlejším tempem než předchozí rok), zatímco ceny ještě více klesaly, o 1,4%.

Stejně tak mezi lety 2017 a 2018 lze pozorovat podobný trend.

V roce 2017, při růstu HDP i 5,17%, došlo k růstu mezd o 6,84% a cen o 7,06%.

V následujícím roce pak, při růstu HDP 3,2% ještě více rostly mzdy, o 8,1% a naopak ceny rostly méně, o 2,41%.

Na základě těchto informací můžeme pozorovat, že významnější růst HDP může ovlivňovat míru růstu mezd a cen. Nicméně pro odvození míry korelace je třeba delší pozorování.