

Projekt z SQL

Popis projektu:

Týmto projektom som sa pokúsila odpovedať na päť definovaných výskumných otázok, ktoré sa týkajú dostupnosti základných potravín širokej verejnosti. Najprv som si pripravila dátové podklady, v ktorých je možno vidieť porovnanie dostupných potravín na základe priemerných príjmov za určité časové obdobie. V poslednej časti som vytvorila dodatočnú tabuľku, ktorá rozšírila už známe informácie o výšku HDP v Českej republike za určité časové obdobie.

Tvorba tabuliek:

Na zistenie toho, ktoré atributy (stĺpce) sú potrebné na vytvorenie tabuľky, som najprv pripravila skripty pre každú výskumnú otázku z originálnych dátových sád a následne retrospektívne zistila, ktoré informácie sú potrebné na vytvorenie najprv prvej a následne druhej tabuľky. Na získanie odpovedí na prvé štyri výskumné otázky mi postačila tabuľka s piatimi atributami: category – SALARY alebo ITEM, year, avg_value, industry_name a item_name. Z tabuľky czechia_payroll a k tomu príslušných číselníkov som za pomoci inner joinu získala informácie o Průměrné hrubé mzde na zaměstnance, pomenovaní jednotlivých industry, aby sa mi hodnoty neduplikovali vybrala som si v časti calculation iba dáta s fyzickým kódom a zároveň som odfiltrovaná nezistené hodnoty pomocou IS NOT NULL. Tieto informácie som prepojila s informáciami z tabuľky czechia_price a k nej príslušnému číselníku. Keďže hodnota v category_code bola pre každý záznam vyplnená, použila som inner join. Informácie som spracovala na úrovni rokov a priemerných hodnôt.

Na zistenie odpovede na piatu výskumnú otázku mi postačila tabuľka s tromi atribútmi: category, year a value. Atribút category pozostáva z hodnôt SALARY, ITEM a GDP. Zjednotila som vybrané dáta z tabuľky czechia_payroll a jej číselníkov, czechia_price a jej číselníka a tabuľky economies.

Výskumné otázky:

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

Na základe dostupných dát vyplynulo, že v štyroch odvetviach rastú mzdy kontinuálne v priebehu rokov a to Ostatní činnosti, Doprava a skladování, Administrativní a podpůrné činnosti a Zdravotní a sociální péče. Vo všetkých ostatných sledovaných odvetviach t.j. 15 odvetví bol zaznamenaný aspoň jeden pokles.

Vyfiltrovala som si údaje ako year, avg_value, industry_name a z atribútu category ma zaujímali iba dáta s názvom SALARY. Vytvorila som si dočasnú tabuľku TEMP_BASE, nasledne som si túto tabuľku najoinovala v rámci sledovaných rokov 2000 až 2021. Dôležité bolo porovnať po sebe každý rok v poradí. Platy som transformovala na úroveň záznamu pre

každé odvetvie. Na konci za pomoci case expression som vyhodnotila trend, či v jednotlivých odvetviach rástli platy v každom roku alebo aj klesali.

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

Za prvé sledované obdobie, t.j. rok 2006 je možné kúpiť 1 262 kg chleba konzumného kmínového a 1 409 l mléka polotučného pasterovaného. Za posledné sledované obdobie t.j. rok 2018 je možno kúpiť 1 319kg chleba konzumného kmínového a 1 614 l mléka polotučného pasterovaného.

V podkladovej tabuľke mám priemerné mzdy za jednotlivé odvetvia po rokoch. Použila som AVG funkciu, ktorou som získala priemerný plat v danom roku. Vytvorila som si dočasné tabuľky TEMP_BASE priemerných miezd a priemerných cien a prepojila ich cez join podmienkou na úrovni roku. Pre lepšiu čitateľnosť a jednoduchšie vyhodnotenie som pri položkách nezvolila kód položiek 'Chléb konzumní kmínový' a 'Mléko polotučné pasterované', ale ich názov. V praxi by som zvolila kód položiek. Použila som funkciu round, ktorá mi zabezpečila zaokrúhlenie hodnoty a lepšiu čitateľnosť množstva chleba a mlieka. Keďže ma zaujíma len prvé a posledné zrovnateľné obdobie, využila som IN max a min obdobia z tabuľky cien. UNION ALL mi spravilo zjednotenie minimálneho roka 2006 a maximalného rok 2018.

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

Z kategórie potravín zdražuje najpomalejšie Cukr krystalový s percentuálnym nárastom - 2.115902000467536%.

V rámci atribútu category ma tentokrát zaujímajú len hodnoty ITEM. Vytvorila som dočasné tabuľky TEMP_FIRST_PERIOD a TEMP_LAST_PERIOD so zoznamom potravín s priemernými cenami v prvom a poslednom sledovanom období. Vypočítala som percentuálny nárast medzi prvým a posledným rokom. Vydělila som to cenou za prvý rok, aby som zisťovala percentuálny nárast za celé obdobie. Priemer som získala delením počtom rokov. Dáta som si zoradila od najmenšieho nárastu, keďže ma zaujíma iba najpomalšie zdražujúca kategória potravín použila som funkciu LIMIT.

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

Ani v jednom zo sledovaných rokov nerástli ceny potravín viac ako o 10% oproti mzdám.

Spravila som medzivýpočet medziročného nárastu cien potravín za pomoci dočasných tabuliek TEMP_PRICE_PCT a TEMP_SALARY_PCT, aby sa mi jednotlivé hodnoty zobrazili vo výsledku v % jednotkách. Za podmienky v case expression, podľa ktorej ak percentuálny nárast cien je väčší ako percentuálny nárast miezd plus 10 som si vytvorila stĺpec vo výsledku pod názvom result a zároveň prepojila sledované roky v dočasných tabuľkách a zoradila dáta zostupne.

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?

Z dostupných dát je HDP pre Českú republiku sledované v rokoch 1991 až 2020. Spoločný prienik, kde boli vyplnené hodnoty pre všetky sledované kategórie ako HDP, ceny potravín a mzdy nastal v rokoch 2007 až 2018. V tomto sledovanom období nemá výška HDP vplyv na zmeny v mzdách a cenách potravín. Napríklad v roku 2010 HDP vzrástlo oproti roku 2009, vzrástli aj ceny potravín, ale mzdy klesli. V roku 2014 v porovnaní s rokom 2013 vzrástlo HDP, ceny poklesli, ale mzdy vrástli.

Z tabuľky t_dana_lancaricova_project_SQL_secondary_final som si pripravila medziročnú percentuálnu zmenu HDP, platov a miezd. Spravila som to cez inner join sám na seba s tým, že rok pri joine sa bude navyšovať o 1. Tak som transformovala výstup z TEMP_TABLE tabuľky po rokoch, čo rok to je jeden záznam. Pre každý rok v dočasnej tabuľke maximálne 3 záznamy. Keďže chcem výstup len pre roky, ku ktorým mám všetky 3 hodnoty, použila som inner join. Výstup zo skriptu mi dal percentuálnu zmenu sledovaných ukazovateľov po rokoch. Do agregáčnej funkcie max som dala case expression, ktoré mi vyhodnotilo či daný riadok sa týka mzdy, platu, HDP.