

Engeto SQL projekt

Úvod do projektu

Na našem analytickém oddělení nezávislé společnosti, která se zaměřuje na zkoumání životní úrovně občanů, jsme dostali za úkol pokusit se odpovědět na několik definovaných výzkumných otázek, které se týkají dostupnosti základních potravin pro širokou veřejnost. Kolegové již stanovili klíčové otázky, na něž se budu snažit nalézt odpovědi a poskytnout tuto informaci tiskovému oddělení. Výsledky naší práce bude toto oddělení prezentovat na nadcházející konferenci, věnované této problematice.

Dostala jsem za úkol připravit robustní datové podklady, které umožní porovnat dostupnost potravin v závislosti na průměrných příjmech v určitém časovém období.

Dále, jako doplňující materiál, mám připravit tabulku obsahující informace o HDP, GINI koeficientu a populaci dalších evropských států ve stejném období, která bude sloužit jako primární přehled pro Českou republiku.

Popis primární a sekundární tabulky

Před začátkem tvorby obou tabulek jsem si v editačním programu vyznačila potřebné sloupce, jak z hlavní tabulky, ze které jsem brala data, tak i z vedlejších, abych pro optimalizaci vybrala opravdu jenom ty potřebné data.

Dále jsem před začátkem tvorby svých hlavních tabulek zkoušela odpovědět na otázky s použitím původních tabulek, abych se s nimi seznámila a věděla kde začít.

Primární tabulka

Primární tabulka “**t_nikola_pincova_project_SQL_primary_final**”, je vytvořena spojením dvou pomocných tabulek “table_a” a “table_b”.

Tabulka “table_a” obsahuje průměry platů (pay), rok (payroll_year) a název odvětví (branch_name).

Rozhodla jsem se v tabulce uvádět názvy odvětví a né kódy, tak jsem tabulku spojila na kódu odvětví (branch_code), abych získala celé názvy.

Dále jsem data vyfiltrovala podle kódu 5958, kde 5958 je kód pro “Průměrnou hrubou mzdu na zaměstnance”, podle kódu 200 v řádku “calculation_code”, který znamená “Přepočtený počet zaměstnanců” a vybrala pouze data mezi roku 2006 a 2018, tohle je naše srovnatelné období.

Druhá pomocná tabulka “table_b” obsahuje cenu (price), název potravin (food_category), hmotnost a jednotka ve které je cena uvedena (kg nebo l)(price_value a price_unit) a rok (price_year).

Opět jsem se v tabulce rozhodla uvádět celé názvy potravin a né kódy, a tak je tabulka spojena na kódu kategorie, pro získání celých názvů potravin.

Nakonec jsem obě tabulky spojila v jednu pod názvem

“t_nikola_pincova_project_SQL_primary_final” pomocí JOIN na společném údaji Rok.

Sekundární tabulka

Druhá tabulka “t_nikola_pincova_project_SQL_secondary_final” byla jednodušší na vytvoření, jelikož bylo méně dat a kódů. Vybrala jsem pouze potřebná data, která byla zmíněna v zadání, a to HDP, GINI koeficientu a populaci, přidané o roky (opět pouze naše srovnatelné období, tj. 2006 až 2018) a název státu, který jsem vytáhla z tabulky countries pomocí sloupce kontinentu.

Výzkumné otázky

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

Podle dat, které nám ukazují sloupce roky, průměrné mzdy, odvětví a nárůst, můžeme vidět, že mzdy ve všech odvětvích rostou. V průměru vzrostli o 10 tisíc v každém odvětví.

Ve zdravotnictví, zpracovatelský průmysl, doprava a ostatních činnostech platy neklesly nikdy a naopak nejvíce klesaly v těžbě a dobývání, celkem 4x.

U větší části odvětví platy klesaly v roce 2013, to mohlo zavinit zavedení solidární daň z příjmů téhož roku.

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

Data nám zde ukazují, že i přes růst cen potravin si můžeme koupit víc litrů mléka a kilogramů chleba. Například v roce 2006 si můžeme z platu koupit 1472 litrů mléka a 1220 kg chleba, o 12 let později si při nárůstu cen a mezd i tak koupíme víc, a to 1696 litrů mléka, o 224 litrů více než v roce 2006, a 1336 kg chleba, což činí o 116 kg víc oproti roku 2006.

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

Za celkové období (čili 2006 - 2018) nejpomaleji zdražovaly *banány žluté*.

V tabulce ale také vidíme 2 typy potravin, které za celkové období zlevnili, konkrétně se jedná o *cukr krystalový* a *rajská jablka červená kulatá*. Cena cukru v roce 2018 byla oproti roku 2006 nižší o 27,51 %, cena rajských jablek o 23,07 %.

V tabulce můžeme vidět *jakostní víno bílé*, kde je hodnota NULL, jelikož za víno nemáme data pro naše celkové období, ale jen od roku 2015, nebudeme s ním počítat.

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

Mezi roky 2006-2018 nebyl nikde meziroční nárůst cen potravin výrazně vyšší než růst mezd. Největší nárůst cen oproti mezd byl v roce 2013, kde byl rozdíl 6.66%, ale nikdy nebyl vyšší 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?

Výška HDP nemá vliv na změnu cen a mezd.

Přestože v roce 2007 byl nárůst HDP, mezd i cen, tak naopak hned v dalším roce 2008 mělo HDP pouze 2,69% vzrůst, ale mzdy a ceny rostli výrazně víc než HDP.

Dále můžeme vidět, že v roce 2012 sice meziroční nárůst HDP byl záporný, ceny potravin ale oproti minulému roku výrazně vzrostly o 6,92%. Podobně tomu je tak i v roce 2013, kdy spolu s menším HDP klesly i mzdy, ale ceny potravin opět vzrostly o 5,55%.