

PROJEKT SQL MAREK MALÝ

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

Postup:

Z vytvořené finální tabulky jsem si vytáhl pouze data, které potřebuji na zadaný úkol, v tomto případě pouze informace o mzdách. Jsou zde vidět všechna zaměstnání seřazeny podle data od roku 2006 do roku 2018 vzestupně, s průměrnou mzdou daného roku. Dále vidíme přesný meziroční nárůst, či pokles v procentech. Nakonec jsem vytvořil sloupec s odpovědí, zda mzdy rostou, či klesají.

Popis:

- V první tabulce vidíme porovnání mezd pro jednotlivé kategorie v letech 2006 až 2018.
- V druhé tabulce vidíme pouze data s rokem, kdy došlo k poklesu mzdy. Pokud je nějaká kategorie v tabulce dvakrát, znamená to, že platy rostly, ale i klesaly. Pokud je naopak nějaká kategorie pouze jednou, znamená to, že v této kategorii mzdy pouze stoupaly. A pro lepší přehled maximální procentuální rozdíl.

Odpověď:

Mzdy rostou v průběhu let pouze v pěti kategoriích - Zpracovatelský průmysl, Doprava a skladování, Administrativní a podpůrné činnosti, Zdravotní a sociální péče a ostatní činnosti.

V ostatních kategoriích mzdy v průběhu některých let klesají. Pro zajímavost, v druhé tabulce můžeme vidět přesně v jakém roce mzdy klesly.

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

Postup:

Z vytvořené finální tabulky jsem si vytáhl pouze data, které potřebuji na zadaný úkol, v tomto případě informace o mzdách, cenách potravin a datech. Musel jsem si vytvořit čtyři pohledy. Dva o množství chleba a mléka za první a poslední srovnatelné období a dva to stejné o mzdách. Spojil jsem dva pohledy o množství produktů a mzdách pro první srovnatelné období do jedné tabulky. A do druhé tabulky to stejné, akorát pro poslední srovnatelné období. Nakonec jsem si vytvořil sloupec, který vypočítá množství potravin za určitou mzdu v jednotlivých odvětvích. Poté jsem tyto dvě tabulky spojil pomocí UNION a dostal dva výstupy, které nám odpoví na otázky. Všechna data jsem čerpal z tabulky t_marek_maly_project_sql_primary_final.

Popis:

- V první tabulce jsou data o cenách a mzdách pro první i poslední srovnatelné období. Jsou zde vidět množství chleba a mléka, které je možné si koupit za každou profesi zvlášť.
- V druhé tabulce vidíme průměrné množství chleba a mléka, které je možné si koupit za první a poslední srovnatelné období za všechna odvětví.

Odpověď:

Za první srovnatelné období tj. v roce 2006, si lze koupit 1.172,48 KG chleba a 1.412,68 L mléka.

Za poslední srovnatelné období tj. v roce 2018, si lze koupit 1.291,62 KG chleba a 1.573,84 L mléka.

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

Postup:

Z vytvořené finální tabulky jsem si vytáhl pouze data, které potřebuji na zadaný úkol, v tomto případě informace o mzdách, cenách potravin a datech. Musel jsem si vytvořit čtyři pohledy. Dva o množství chleba a mléka za první a poslední srovnatelné období a dva to stejné o mzdách. Spojil jsem dva pohledy o množství produktů a mzdách pro první srovnatelné období do jedné tabulky. A do druhé tabulky to stejné, akorát pro poslední srovnatelné období. Nakonec jsem si vytvořil sloupec, který vypočítá množství potravin za určitou mzdu v jednotlivých odvětvích. Poté jsem tyto dvě tabulky spojil pomocí UNION a dostal dva výstupy, které nám odpoví na otázky. Všechna data jsem čerpal z tabulky `t_marek_maly_project_sql_primary_final`.

Popis:

- V první tabulce jsou data o cenách a mzdách pro první i poslední srovnatelné období. Jsou zde vidět množství chleba a mléka, které je možné si koupit za každou profesi zvlášť.
- V druhé tabulce vidíme průměrné množství chleba a mléka, které je možné si koupit za první a poslední srovnatelné období za všechna odvětví.

Odpověď:

Za první srovnatelné období tj. v roce 2006, si lze koupit 1.172,48 KG chleba a 1.412,68 L mléka.

Za poslední srovnatelné období tj. v roce 2018, si lze koupit 1.291,62 KG chleba a 1.573,84 L mléka.

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

Postup:

Pro tuto otázku jsem si musel vytvořit dva pohledy. Jeden na meziroční nárůst potravin v procentech. Druhý pohled na meziroční nárůst mezd. Poté jsem tyto dva pohledy spojil a porovnal meziroční nárůst cen jednotlivých potravin s každou kategorií mezd. Vytvořil jsem si nový sloupec 'more_than_10_percentage', který mi vrátil hodnotu 1, pokud byl nárůst větší než 10%, v ostatních případech pak hodnotu 0. Vytvořil jsem další sloupec 'difference', který vypočítal přesnou hodnotu rozdílu mezi cenou potravin a mzdami pouze pro hodnoty 1, tedy pouze pro potraviny, které mají větší meziroční nárůst cen, než je nárůst mezd. A nakonec jsem si vytáhl pouze údaje o počtu zvýšení z každého roku. Všechna data jsem čerpal z tabulky t_marek_maly_project_sql_primary_final.

Popis:

V této tabulce vidíme pouze roky, ve kterých je počet navýšení cen potravin oproti mzdám větší než 10%. Tudíž v roce 2013 bylo Dále si můžeme všimnout, že pouze v roce 2009 nebyl žádný záznam o navýšení cen oproti mzdám o více než deset procent.

Odpověď:

Meziroční nárůst cen potravin oproti nárůstu mezd byl ve většině případů. Pouze v roce 2009 nebyl zaznamenán průměrný meziroční nárůst cen potravin vůči platům.

5. Otázka: 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?

Postup:

U této otázky jsem musel používat často funkci LAG, abych porovnal předešlé hodnoty GDP, Mzdy a Ceny potravin. Z těchto hodnot jsem zvlášť vypočítal průměrný rozdíl, který můžu porovnávat mezi sebou a odpovídat na otázky. Všechna data jsem čerpal z tabulky t_marek_maly_project_sql_primary_final a t_Marek_Maly_project_SQL_secondary_final.

Popis:

V tabulce vidíme srovnané hodnoty GDP podle meziročního nárůstu v procentech. Tím se odvíjí i změny meziročního nárůstu mezd i cen potravin v procentech.

Odpověď:

- Například v roce 2007 byl nejvyšší nárůst GDP o 5,57 %. V tomto roce mzdy vzrostly o 6,79 a potraviny vzrostly o rekordních 13,45 %.
- Naopak při nejnižším meziročním poklesu v roce 2009 GDP kleslo o 4,66 %, mzdy stouply o 3,25 % a ceny klesly o rekordních 9,32 %.
- Ještě v roce 2017 kdy byl nárůst GDP třetí nejvyšší a to o 5,17 % se mzdy zvedly celkem o 6,4% a ceny potravin o 8,44%.
- V ostatních letech se jak mzdy i ceny potravin zvyšovaly, tak i snižovaly nezávisle na změně GDP.

Tvorba primární tabulky

Postup:

Pro tvorbu primární finální tabulky jsem použil dataset ze serveru Engeta.

Abych mohl vytvořit tabulku s cenami potravin, musel jsem spojit dvě tabulky (czechia_price a czechia_price_category). V jedné souhrn cen potravin, druhá s měnou, tj. Kč. a s názvem potravin. Spočítal jsem průměrnou cenu potravin za každý konkrétní rok a region v ČR v porovnatelném období, tj. od roku 2006 do roku 2018. Nakonec data vhodně seskupil a seřadil.

Pro tvorbu tabulky se mzdami jsem spojil tabulky tři (czechia_payroll, czechia_payroll_unit a czechia_payroll_industry_branch). Z první tabulky jsem si vytáhl pouze sloupce, které jsem potřeboval pro práci na projektu, z druhé tabulky jsem použil sloupec, který nám udává platy v Kč a třetí tabulku, která nám vyhodí názvy odvětví. Vytáhl jsem datum pouze z roku 2006 do roku 2018, jako to je v tabulce s cenami potravin. Ceny mezd jsou zprůměrované za celý rok pro každé odvětví. Po vytvoření těchto tabulek jsem obě spojil a vytvořil Finální primární tabulku.

Tvorba sekundární tabulky

Postup:

Pro tvorbu sekundární finální tabulky jsem si vytvořil novou pomocnou sekundární tabulku, do které jsem vložil sloupce s GDP, year tj. roky 2006 až 2018 a country, pouze s Českou Republikou. Tuto pomocnou tabulku jsem použil k vytvoření finální sekundární tabulky.

Spojil jsem finální primární tabulku, z které jsem si vytáhl pouze průměrnou mzdu a průměrné ceny potravin, a spojil s pomocnou tabulkou.