

Průvodní listina

Úvod – Na základě předchozí zpětné vazby a konzultace mého minulého SQL projektu, kde jsem se dopustil funkčních chyb, posílám nově vypracovaný projekt, kde jsem se snažil zmíněné chyby odladit. Všechny řešení daných úkolů jsem se snažil koncipovat tak, aby bylo dosaženo výsledku pomocí jednoho kliknutí, celkem hojně jsem používal klauzuli WITH díky níž jsem spojoval a dále třídil a upravoval data. Kromě dvou hlavních výchozích tabulek, jsem netvořil žádné další stálé tabulky, zároveň při tvorbě hlavních tabulek, jsem se snažil data upravit tak, aby se s nimi co nejlépe pracovalo. Projekt jsem umístil na GitHub do stejného adresáře jako ten předchozí, nicméně pro lepší orientaci jsem zde vytvořil složky. Složka obsahující části nového (opravného) projektu nese název „new_project_sql_second_attempt_th“ a jednotlivé části projektu jsou pojmenovány následovně:

- primární tabulka – new_t_tomas_holub_SQL_primary_final.sql
- sekundární tabulka – new_t_tomas_holub_SQL_secondary_final.sql
- první otázka – new_task_one.sql
- druhá otázka – new_task_two.sql
- třetí otázka – new_task_three.sql
- čtvrtá otázka – new_task_four.sql
- pátá otázka - new_task_five.sql

Výzkumné otázky

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

Odpověď: Z dostupných dat vyplývá, že mzdy rostou ale zároveň i mírně klesají ve všech odvětvích. Ve většině odvětví jsou poklesy v řádu desítek až stovek korun, ale v několika odvětvích se poklesy pohybují v řádech stovek až tisíců korun.

Název - new_task_one.sql

Tento úkol jsem se rozhodl vyřešit jedním příkazem SELECT kde jsem vyfiltroval sledované odvětví průměrných mezd pomocí jejich kodu. Sloupce, ve kterých jsem hledal řešení tohoto úkolu nesou názvy „payroll_year“- měřený rok, „branch_name“- název odvětví, „final_avg_payroll“- průměrná měsíční mzda v daném roce.

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

Odpověď: Z dostupných dat vyplývá, že za první měřené období (rok 2006) je při uvážení nejvyšší naměřené hodnoty průměrné mzdy možné koupit maximálně 3097 litrů mléka a 2567 kilogramů chleba. Za poslední sledované období (rok 2018) je možné při uvážení nejvyšší hodnoty naměřené průměrné mzdy koupit maximálně 3030 litrů mléka a 2387 kilogramů chleba. V prvním období byla nejvyšší hodnota průměrné mzdy naměřena v odvětví

peněžnictví a pojišťovnictví a v posledním období to bylo v odvětví informačních a komunikačních technologií.

Název - new_task_two.sql

Při vytváření řešení tohoto úkolu jsem použil řetězení dočasných tabulek pomocí funkce WITH. Hledaná data jsem vyřídil podle posuzovaných let a podle kodu kategorie potravin. Výsledná dočasná tabulka obsahuje sloupce „measured_year“ – měřený rok, „food_category“ – kategorie jídla, „final_avg_price“ – průměrná cena položky v daném roce, „food_amount“ – měřené množství vzhledem k ceně, „measure_unit“ – jednotka měření, „branch_name“ – nejlépe finančně ohodnocené odvětví, „first_max_payroll“ – nejvyšší hodnota průměrné mzdy v prvním odvětví, „second_max_payroll“ – nejvyšší hodnota průměrné mzdy v posledním odvětví, „2006_max_bread_amount“ – maximální množství chleba v prvním období, „2006_max_milk_amount“ – maximální množství mléka v prvním období, „2018_max_bread_amount“ – maximální množství chleba v posledním období, „2018_max_milk_amount“ – maximální množství mléka v posledním období.

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

Odpověď: Z dostupných dat vyplývá, že kategorií potravin, která zdražuje nejpomaleji je cukr krystalový, jehož celkový procentuální meziroční nárůst (pokles) ceny je – 2,48 %.

Název - new_task_three.sql

Pro řešení tohoto úkolu jsem pomocí příkazu WITH vytvořil dočasnou tabulku s výpočty jednotlivých meziročních nárůstů cen potravin. Z této tabulky jsem vybral potřebné sloupce s daty a také jsem vytvořil sloupec součtu meziročních nárůstů ve sledovaném období, který jsem seřadil k získání konečných výsledků. Použil jsem i funkce LAG – pro získání předcházejících hodnot a určení meziročního růstu, AVG – pro získání průměrných hodnot a ROUND pro zaokrouhlení hodnot. Sloupce tabulky nesou názvy: „price_year“ – rok, „food_category“ – kategorie potravin, „final_avg_price“ – průměrná cena dané kategorie potravin v daném roce, „food_code“ – kod kategorie potravin (slouží pro lepší orientaci v datech), „percentual_difference“ – meziroční procentuální rozdíl, „percentage_growth“ – meziroční procentuální růst (výsledný sloupec).

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

Odpověď: Z dostupných dat vyplývá, že k významnému nárůstu cen potravin (většímu než 10%) došlo, ale pouze u některých kategorií potravin a to v letech: 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018.

Název - new_task_four.sql

Pro získání odpovědi na tuto otázku jsem použil 4 výběry a dvě dočasné tabulky sdružené pomocí klauzule WITH do jedné konečné tabulky. Také jsem použil funkce LAG – pro získání předcházejících hodnot a určení meziročního růstu, AVG – pro získání průměrných hodnot a ROUND pro zaokrouhlení hodnot. Jelikož byla data mezd i cen potravin spojena do

jedné tabulky, použil jsem k jejich rozlišení funkce CASE. Hledaná data jsem získával z primární tabulky. Hlavní sloupce tabulky jsou „measured_year“ – měřený rok, „category“ – kategorie odvětví mezd a potravin, „avg_value“ – průměrná výše mezd a cen potravin v daném roce, „perc_growth“ – meziroční nárůst cen potravin a výše mezd, „category_sign“ – slouží pro rozlišení, zda se jedná o mzdy nebo potraviny a „significant_growth“ – označuje, zda byl nárůst významný nebo ne.

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?

Odpověď: Z dostupných dat vyplývá že, růst výše HDP má vliv především na růst mezd v sektoru zpracovatelského průmyslu.

Název - new_task_five.sql

K vyhledání řešení úkolu jsem použil podobné konstrukce klauzule WITH jako v úkole č. 4. Pomocí WITH jsem vytvořil 2 dočasné tabulky které jsem následně spojil pomocí sloupce měřeného roku. Současně jsem implementoval také několik různých výběrů k získání hledaných dat a dále jsem použil funkce LAG – pro získání předcházejících hodnot a určení meziročního růstu, AVG – pro získání průměrných hodnot a ROUND pro zaokrouhlení hodnot. Hledaná data jsem získával jak z primární tabulky, tak ze sekundární. Jelikož byla data mezd i cen potravin spojena do jedné tabulky, použil jsem k jejich rozlišení funkce CASE. Nejdůležitější sloupce tabulky nesou názvy „measured_year“ – měřený rok, „category“ – kategorie odvětví mezd a potravin, „avg_value“ – průměrná výše mezd a cen potravin v daném roce, „perc_growth“ – meziroční nárůst cen potravin a výše mezd, „gdp“ – výše HDP pro daný rok a „gdp_growth“ naměřený vliv růstu HDP na mzdy. Zbytek sloupců jsem používal jako pomocné sloupce pro lepší orientaci v datech.

Doufám, že se mi povedlo předchozí chyby napravit a velmi děkuji za opravení a zpětnou vazbu

S pozdravem

Tomáš Holub