

Projekt SQL - průvodní listina

Pro tento projekt jsem využila tyto nabízené datové sady, které jsem si nastudovala a vybrala jsem z nich informace, které mi sloužily k zodpovězení daných výzkumných otázek:

- Czechia_payroll – výchozí tabulka, kde byly informace o mzdách v různých odvětvích, o průměrném počtu zaměstnaných osob.
- Czechia_payroll_calculation – číselník mezd, z něhož jsme se dozvěděla, zda se jedná o hrubou mzdu fyzickou, která znamená to, že jsou v ní zahrnuty i příjmy z vedlejšího pracovního poměru. Nebo, že se jedná přepočtenou hrubou mzdu, kde se výpočet průměrné mzdy koriguje podle vlivu různých pracovních úvazků a poměrů.
- Czechia_payroll_industry_branch – v této tabulce bylo uvedeno pod jakých kódem je konkrétních odvětví.
- Czechia_payroll_unit – v této tabulce byla důležitá informace o tom, že jsou průměrné mzdy uvedeny v korunách českých (Kč)
- Czechia_payroll_value_type – zde jsem se dozvěděla to, co znamená číselná hodnota v tabulce Czechia_payroll. Jestli se jedná o hodnotu průměrné mzdy nebo průměrném počtu zaměstnaných osob. Jelikož mě zajímaly hodnoty průměrných mezd, tak kód 5958 hrál důležitou roli.
- Czechia_price – informace o cenách potravin v určitém časovém rozmezí.
- Czechia_price_category – v této tabulce jsou uvedeny konkrétní kategorie potravin, které najdeme v tabulce Czechia_price pod číselným kódem.
- Countries – zde jsem se zaměřila na data o zemích, jejich populaci, jejich zkratky.
- Economies – tady byla pro mě důležitá data o HDP(GDP), o gini koeficient a roky v jakých jsou data uvedena.

Zpracování první finální tabulky

Pro zpracování první finální tabulky jsem použila tabulky z nabízeného datového setu a jednalo se to tyto tabulky: Czechia_payroll, Czechia_payroll_unit, Czechia_payroll_calculation, Czechia_payroll_industry_branch, Czechia_payroll_value_type, Czechia_price, Czechia_price_category.

Tabulky jsem sjednotila na roky, aby se z nich dala data mezd a potravin porovnávat. Důležité pak bylo stanovit jednotnou průměrnou mzdu (avg_salary), ale tabulka neobsahuje dostatek informací o počtu zaměstnanců za jednotlivá období. Výsledkem je průměr již zprůměrovaných hodnot, který může reálnou hodnotu zkreslovat. V tabulce Czechia_price taky bylo nutné udělat průměr cen u každé kategorie potravin (avg_price). Ve finální tabulce jsou i vzorce pro výpočet mzdy z předchozích roku (avg_salary_ly), rozdíly v mzdách (avg_salary_diff) i meziroční procentuální nárůst mezd (avg_salary_growth_percent), které jsem posléze využila v SQL dotazech pro zodpovězení výzkumných otázek. Podobný postup byl využit i v případě cen potravin (avg_price_ly), avg_price_growth_percent). Při vytváření tabulky byla použita funkce WITH. Pro reálnější výsledky zde byl i ponechán sloupec, který pojednával o tom, zda se jedná o fyzickou či přepočtenou hrubou mzdu.

Zpracování druhé finální tabulky

Pro zpracování této tabulky jsem spojila dvě tabulky Countries a Economies, kde jsem definovala, že mě zajímají informace pouze o Evropských státech. Prověřila jsem si, jestli je v obou tabulkách uveden název České republiky jako Czech republic, jelikož už se můžeme v jiných datových setech setkat s tím, že je Česká republika uvedena pod názvem Czechia. A dále jsem určila konkrétní roky k tomu, abych mohla tuto tabulku porovnávat daty z první finální tabulky.

Zpracování otázek

Otázka č.1

Aby se dala zodpovědět tato otázka, tak zde byl vytvořen další sloupec s výpočtem rozdílu (`avg_salary_diff`) ve mzdách oproti předchozímu roku, který můžeme najít v primární tabulce. Ve výsledné tabulce pak byly zobrazeny roky, v jaké roce a v jakém oboru došlo ke snížení mzdy, což bylo pro mě přehlednější.

Otázka č. 2

Nejdříve jsem ověřila, pod jakým celým názvem v tabulce je uvedeno mléko (Mléko polotučné pasterované) a chleba (Chléb konzumní kmínový). Dále pak bylo nutné zjistit, z jakého roku je první sledovaný záznam a zároveň poslední rok záznamu jedné nebo druhé výše uvedené kategorii potravin. Poté se musely zprůměrovat mzdy ze všech oborů za jeden konkrétní rok a následně udělat výpočet z průměrné mzdy a vydělit ji aktuální cenou (`buyable_salary`), které v prvním a pak posledním roce byla aktuální výše jmenovanou kategorii potravin. A tím pádem jsme zjistili, kolik jsme si mohli litrů mléka či kilogramů chleba v té době za celkovou průměrnou mzdu.

Otázka č. 3

Nejdříve se musel opět stanovit první a poslední záznam o průměrných cenách pro kategorii potravin (`first_year`, `last_year`). Pak dále určit o kolik se procentuálně zvedla cena mezi prvním a posledním rokem mezi nimi (`avg_price_growth_percent`), pak pro zajímavost zde byl vytvořen sloupec, kde vidíme, o jaký rozdíl let se jedná (`tracked_years`), nakonec se spočítalo průměrný procentuální roční nárůst cen (`avg_price_growth_percent_yearly`) a z toho pak bylo poznat, která potravina zdražuje nejpomaleji.

Otázka č.4

Tady se nejdříve musela stanovit průměrná z mzda již z uvedených zprůměrovaných cen ze všech ze všech odvětví (`avg_salary`), zde se i využil vzorec pro výpočet vytvořený v primární tabulce pro zobrazení zprůměrované průměrné mzdy z předchozího roku (`avg_salary_ly`) taky pro všechny odvětví pro jeden konkrétní rok. Tento samý postup byl i udělán pro záznam kategorii potravin, kdy se taky zprůměrovala průměrná cena všech kategorií potravin (`avg_price`) a zprůměrovala cena všech kategorií potravin z předchozího roku (`avg_price_ly`). Pro zjednodušený výstup byl udělán průměr z přepočtených hrubých mezd. Poté se z výše uvedených průměrovaných hodnot vypočítal meziroční procentuální nárůst mezd (`salary_growth_percent`) a cen potravin (`price_growth_percent`).

Otázka č.5

Pro zodpovězení této otázky byla využit stejný dotaz jako v otázce číslo 4, protože jsme potřebovali porovnat meziroční procentuální nárůst průměrných mezd a meziroční procentuální nárůst průměrných cen všech kategorií potravin za jeden určitý rok. Pouze chybělo meziroční

procentuální nárůst HDP. Informace o HDP byla uvedena v druhé finální tabulce, proto bylo nutné tabulky spojit pomocí funkce LEFT JOIN a vyselektovat údaje týkající se České republiky a zjistit výpočtem meziroční procentuální nárůst HDP (GDP_growth_percent). Poté se dala zjištěná data mohla mezi sebou porovnat a vyvodit výsledek.