

# Průvodní listina – SQL projekt

## I. Tvorba tabulek:

- 1) Vytvoření tabulky „**t\_kristyna\_grouslova\_project\_SQL\_primary\_final**“ pro data mezd a cen potravin za Českou republiku sjednocených na totožné porovnatelné období – společné roky.
- 2) Vytvoření tabulky „**t\_kristyna\_grouslova\_project\_SQL\_secondary\_final**“ pro dodatečná data o dalších evropských státech.

### Popis řešení:

Ad 1) Klauzulí WITH byla vytvořena dočasná tabulka s názvem `table_payroll`, ve které byly získány vybrané informace o letech, kódu odvětví a názvu odvětví z databázových tabulek `czechia_payroll` a `czechia_payroll_industry_branch`, a to na základě jejich spojení (JOIN) prostřednictvím společných kódů (`cp.industry_branch_code` a `cpib.code`). V původní tabulce `czechia_payroll` byly průměrné výše mezd uváděny v jednotlivých letech dle kvartálů, proto byl vytvořen z těchto kvartálů průměr (`average_salary_czk`) za daný rok a zaokrouhlen na dvě desetinná místa. Záznamy byly seskupeny klauzulí GROUP BY na základě kódu odvětví a jednotlivých let. Vzhledem k tomu, že pro další práci s tabulkou byly potřeba pouze údaje o průměrných mzdách, byl výběr podmíněn hodnotou `cp.value_type_code`, která vybrala pouze informace s uvedením průměrné hrubé mzdy na zaměstnance a dále hodnotou `calculation_code` pro výběr hodnot s průměrnými fyzickými počty zaměstnanců. Data z tabulky `czechia_price` obsahují pouze období 2006–2018. Oproti tomu data z tabulky `czechia_payroll` obsahují zahrnují širší časové rozpětí (2000–2021). Z toho důvodu byla zvolena další podmínka výběru, a to jednotné srovnatelné období 2006–2018. Kontrolou zdrojových dat bylo zjištěno, že cena potravin „Jakostní víno bílé“ byla uvedena pouze v období let 2015–2018. Vzhledem k případnému zkreslení výsledků v dalších úkolech byla data týkající se této potravin odstraněna z výběru dat do výsledné tabulky vnořeným selectem.

Prostřednictvím klauzule WITH byla současně vytvořena dočasná tabulka s názvem `table_price`, do které byly následně získány informace o letech, kódech kategorií potravin, názvu potravin a ceně potravin za určitou jednotku z propojených tabulek `czechia_price` a `czechia_price_category`. Propojení tabulek bylo provedeno pomocí společných kódů kategorií potravin. Protože původní tabulka `czechia_price` obsahovala podrobná kalendářní data, byl použit příkaz YEAR k extrahování roků z dat (`date_from`). Informace o cenách potravin byly zprůměrovány za daný rok a zaokrouhleny na dvě desetinná místa. Tyto údaje byly seskupeny (GROUP BY) dle kódu kategorií potravin a jednotlivých let.

Výše uvedené dočasné tabulky `table_payroll` a `table_price` byly následně spojeny klauzulí JOIN na základě roků. Z uvedených příkazů byla vytvořena výsledná tabulka `t_kristyna_grouslova_project_SQL_primary_final`.

Ad 2) Tabulka `t_kristyna_grouslova_project_SQL_secondary_final` byla vytvořena s uvedením zemí, jejich HDP, populací, HDP na obyvatele (tj. podíl HDP a populace zaokrouhlený na dvě desetinná místa) a GINI koeficientu. Vnořeným dotazem byly pomocí tabulky `countries` vybrány pouze evropské země (tj. byly vybrány země s uvedením

kontinentu Evropa). Dále bylo jako podmínka zvoleno období 2006–2018, což je období srovnatelné s údaji v primární tabulce (přehled pro ČR).

## II. Výzkumné otázky:

- 1) Rostou v průběhu let mzdy ve všech odvětvích, nebo v některých klesají?
- 2) Kolik je možné si koupit litrů mléka a kilogramů chleba za první a poslední srovnatelné období v dostupných datech cen a mezd?
- 3) Která kategorie potravin zdražuje nejpomaleji (je u ní nejnižší procentuální meziroční nárůst)?
- 4) Existuje rok, ve kterém byl meziroční nárůst cen potravin výrazně vyšší než růst mezd (větší 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?

Popis řešení:

Ad 1) Tabulka `t_kristyna_grouslova_project_sql_primary_final` tkg byla napojena sama na sebe pomocí klauzule JOIN. Aby mohl být spočítán meziroční nárůst mezd, byla tabulka propojena pomocí klíče `industry_branch_code` a dále klíče s uvedením roku, tj. `tkg.payroll_year = tkg2.payroll_year + 1`, kdy k pravé tabulce byla připočtena 1 (porovnání mezd předchozího roku a roku, pro který je procentuální nárůst mezd počítán). Meziroční procentuální nárůst mzdy byl vypočítán jako procento rozdílu mzdy daného roku a mzdy roku předchozího. Hodnota byla dále zaokrouhlena na dvě desetinná místa. Data byla seskupena dle roku a kódu odvětví. K získání informací, ve kterých odvětvích a v jakých letech došlo k poklesu růstu mezd, byla následně použita klauzule HAVING s podmínkou výběru hodnot procentuálního meziročního nárůstu mezd méně než 0. Pro přehlednost byly výsledky seřazeny sestupně dle hodnoty procentuálního meziročního nárůstu mezd.

Výsledek: Z analýzy vyplývá, že ve všech odvětvích kromě kategorie C - zpracovatelský průmysl, H - doprava a skladování, N – administrativní a podpůrné činnosti, Q - zdravotní a sociální péče a S - ostatní činnosti byl zaznamenán alespoň jeden meziroční pokles mezd.

Ad 2) Za účelem stanovení odpovědi na zadanou otázku byly z tabulky `t_kristyna_grouslova_project_sql_primary_final` tkg vypočítány hodnoty celkové průměrné mzdy za konkrétní roky napříč všemi odvětvími. Množství konkrétních potravin, které lze pořídit za vypočtenou celkovou průměrnou mzdu, bylo vypočteno jako podíl celkové průměrné mzdy a průměrné ceny dané potraviny za jednotku v konkrétním roce, přičemž výsledek byl zaokrouhlen na celá čísla. Protože k výpočtu mělo být uvažováno s prvním a posledním srovnatelným obdobím v dostupných datech cen a mezd, byla stanovena podmínka výběru dat z let 2006 a 2018. Další podmínkou byl výběr kódu kategorie požadovaných potravin 114201 pro mléko a 111301 pro chléb. Vzhledem k tomu, že bylo počítáno s celkovými průměrnými mzdami, bylo potřeba seskupit data klauzulí GROUP BY dle roků a kódu kategorie potravin (`category_code`).

Výsledek: Ze získaných informací bylo zjištěno, že v roce 2006 bylo možné za průměrnou mzdu koupit 1 262 kg chleba a 1 409 l mléka, V roce 2018 to pak bylo 1 319 kg chleba a 1 614 l mléka.

Ad 3) K vyhodnocení otázky, která kategorie potravin zdražuje nejpomaleji, byla nejprve vytvořena dočasná tabulka `table_growth_rate` pomocí klauzule `WITH`. V této dočasné tabulce byl zjišťován procentuální meziroční nárůst cen jednotlivých kategorií potravin, a to pomocí napojení primární tabulky `t_kristyna_grouslova_project_sql_primary_final` tkg na sebe samu. Při spojování tabulky byly použity dva klíče – `tkg.price_year = tkg2.price_year + 1`, kdy k pravé tabulce byla připočtena 1 (porovnání cen předchozího roku a roku, pro který je procentuální nárůst cen počítán) a dále kód kategorie potravin. Meziroční procentuální nárůst cen potravin byl vypočítán jako procento rozdílu ceny potravin daného roku a ceny potravin roku předchozího. Hodnota byla dále zaokrouhlena na dvě desetinná místa. Údaje byly seskupeny dle roku a kódu kategorie potravin.

V další části bylo přistoupeno k výpočtu průměrného tempa růstu cen jednotlivých kategorií potravin za celé sledované období. Z dočasné tabulky byly vybrány informace o procentuálním meziročním nárůstu cen potravin, které byly zprůměrovány a zaokrouhleny na dvě desetinná místa (tj. průměrné tempo růstu cen). Výsledná data byla současně seskupena klauzulí `GROUP BY` dle kódu kategorie potravin a pro přehlednost seřazeny sestupně.

Výsledek: Z provedené analýzy dat vyplývá, že nejpomalejší tempo růstu cen za období 2006–2018 se týkalo cukru krystalového (průměrné tempo růstu ceny -1,92 %), následovaného rajskými jablky červenými kulatými (průměrné tempo růstu ceny -0,74 %) a banány žlutými (průměrné tempo růstu ceny 0,81 %).

Ad 4) K vyhodnocení daného úkolu byly vytvořeny tři dočasné tabulky `table_avg`, `table_avg2` a `table_comparison`, a to pomocí klauzule `WITH`. V první dočasné tabulce byly vypočítány pro každý rok průměrné hodnoty mezd (bez ohledu na odvětví) a průměrné hodnoty cen potravin (bez ohledu na kategorii potravin). Data v tabulce byla seskupena dle let.

V druhé dočasné tabulce bylo provedeno napojení první dočasné tabulky `table_avg` na sebe samu, přičemž ke spojení došlo na základě klíče `tg1.years = tg2.years + 1`, kdy k pravé tabulce byla připočtena 1 (porovnání hodnot předchozího roku a roku, pro který je procentuální nárůst počítán).

Třetí dočasná tabulka `table_comparison` byla použita k výpočtu meziročního růstu cen potravin a mezd. Následně byly vyselektovány informace o letech, ve kterých byl procentuální růst cen vyšší jak 10 a zároveň procentuální růst mezd byl nižší jak 10.

Výsledek: Dle popsané analýzy dat bylo zjištěno, že v rámci sledovaného období byl v roce 2017 zaznamenán výrazně vyšší průměrný meziroční nárůst cen oproti průměrnému meziročnímu nárůstu mezd.

Ad 5) V první dočasné tabulce `table_GDP` došlo ke spojení primární a sekundární tabulky na základě klíče – roků. V tabulce byly vypočítány průměrné hodnoty mezd a cen potravin dle jednotlivých let bez ohledu na odvětví či kategorii potravin. Vzhledem k tomu, že sekundární tabulka obsahovala údaje i o dalších evropských státech, byla přidána podmínka konkrétní země – Česká republika. Data byla seskupena dle roků.

V dočasné tabulce `table_growth_value` došlo k napojení tabulky `table_GDP` samu na sebe pomocí klíče `tg1.years = tg2.years + 1`, kdy k pravé tabulce byla připočtena 1 (porovnání hodnot předchozího roku a roku, pro který je procentuální nárůst počítán). Na základě tohoto spojení bylo možné vypočítat rozdíl mezi hodnotami v daném a předchozím roce u HDP na osobu, mzdy a ceny potravin. Data byla seskupena na základě roků. Pro srovnání byly do tabulky zároveň přidány procentuální hodnoty meziročního nárůstu HDP, mezd a cen potravin.

V další části byly z tabulky `table_growth_value` vyselektovány hodnoty meziročního růstu HDP na obyvatele, mezd a cen potravin v Kč v jednotlivých letech a také v procentech. Pomocí CASE byly hodnoty v Kč rozčleněny do kategorií dle výrazného růstu (u HDP a mezd se jednalo o rozdíl hodnot nad 1000 Kč, u cen nad 3 Kč), mírného růstu (u HDP a mezd šlo o rozdílovou hodnotu v rozmezí 0 až 999 Kč, u cen 0 až 2,99 Kč) či poklesu (rozdílová hodnota u HDP, mezd i cen potravin pod 0 Kč).

Výsledek: Z analýzy dat vyplývá, že výraznější nárůst HDP byl zaznamenán v roce 2015 a 2017. Výrazný nárůst mezd či cen potravin však nebyl v roce 2015 ani v roce následujícím zaznamenán. Výrazný nárůst mezd a cen potravin byl současně s nárůstem HDP shodně zaznamenán pouze v roce 2017, kdy výrazný nárůst mezd pokračoval i v roce následujícím. Vliv výše HDP na změny ve mzdách a cenách potravin tak nebyl jednoznačně prokázán.

Zpracovatel: Kristýna Groušlová

e-mail: [kristi.grousl@seznam.cz](mailto:kristi.grousl@seznam.cz)