Projekt SQL – Průvodní list

ZADÁNÍ :

**Úvod do projektu**

Na vašem analytickém oddělení nezávislé společnosti, která se zabývá životní úrovní občanů, jste se dohodli, že se pokusíte odpovědět na pár definovaných výzkumných otázek, které adresují **dostupnost základních potravin široké veřejnosti**. Kolegové již vydefinovali základní otázky, na které se pokusí odpovědět a poskytnout tuto informaci tiskovému oddělení. Toto oddělení bude výsledky prezentovat na následující konferenci zaměřené na tuto oblast.

Potřebují k tomu **od vás připravit robustní datové podklady**, ve kterých bude možné vidět **porovnání dostupnosti potravin na základě průměrných příjmů za určité časové období**.

Jako dodatečný materiál připravte i tabulku s HDP, GINI koeficientem a populací **dalších evropských států** ve stejném období, jako primární přehled pro ČR.

﻿

**Datové sady, které je možné požít pro získání vhodného datového podkladu**

**Primární tabulky:**

czechia\_payroll – Informace o mzdách v různých odvětvích za několikaleté období. Datová sada pochází z Portálu otevřených dat ČR.

czechia\_payroll\_calculation – Číselník kalkulací v tabulce mezd.

czechia\_payroll\_industry\_branch – Číselník odvětví v tabulce mezd.

czechia\_payroll\_unit – Číselník jednotek hodnot v tabulce mezd.

czechia\_payroll\_value\_type – Číselník typů hodnot v tabulce mezd.

czechia\_price – Informace o cenách vybraných potravin za několikaleté období. Datová sada pochází z Portálu otevřených dat ČR.

czechia\_price\_category – Číselník kategorií potravin, které se vyskytují v našem přehledu.

**Číselníky sdílených informací o ČR:**

czechia\_region – Číselník krajů České republiky dle normy CZ-NUTS 2.

czechia\_district – Číselník okresů České republiky dle normy LAU.

**Dodatečné tabulky:**

countries - Všemožné informace o zemích na světě, například hlavní město, měna, národní jídlo nebo průměrná výška populace.

economies - HDP, GINI, daňová zátěž, atd. pro daný stát a rok.

﻿

**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 percentuální 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?

﻿

**Výstupy z projektu**

Pomozte kolegům s daným úkolem. Výstupem by měly být dvě tabulky v databázi, ze kterých se požadovaná data dají získat. Tabulky pojmenujte t\_{jmeno}\_{prijmeni}\_project\_SQL\_primary\_final (pro data mezd a cen potravin za Českou republiku sjednocených na totožné porovnatelné období – společné roky) a t\_{jmeno}\_{prijmeni}\_project\_SQL\_secondary\_final (pro dodatečná data o dalších evropských státech).

Dále připravte sadu SQL, které z vámi připravených tabulek získají datový podklad k odpovězení na vytyčené výzkumné otázky. Pozor, otázky/hypotézy mohou vaše výstupy podporovat i vyvracet! Záleží na tom, co říkají data.

Postup zpracování :

Pro zodpovězení dotazů bylo třeba vysledovat trendy ve třech oblastech

1. Vývoj cen jednotlivých druhů zboží v ČR
2. Vývoj mezd v jednotlivých odvětvích v ČR
3. Vývoj HDP v ČR

Pro vyřešení tohoto úkolu již bylo potřeba propojit tabulku cen a tabulku mezd do požadované tabulky : t\_vladimir\_smid\_project\_SQL\_primary\_final

Ad I ) Vývoj cen jednotlivých druhů zboží v ČR

|  |
| --- |
| **create** **temporary** **table** tmp\_vsm\_projekt\_ceny **as** |
| **select** |
| cp.category\_code , |
| cpc.**name** **as** nazev, |
| cpc.price\_unit, |
| **date\_part**('year', date\_from) **as** rok , |
| **avg** (cp.value) **as** prumerna\_cena, |
| **lag** ( **avg** (cp.value)) **over** (**order** **by** cpc.**name** )**as** prumcena\_minulyrok , |
| **avg** (cp.value) - **lag** ( **avg** (cp.value)) **over** (**order** **by** cpc.**name** ) **as** narust\_cen\_v\_Kc, |
| (**avg** (cp.value) - **lag** ( **avg** (cp.value)) **over** (**order** **by** cpc.**name** ))/**lag** ( **avg** (cp.value)) **over** (**order** **by** cpc.**name** )\*100 **as** narust\_cen\_v\_procentech |
| **from** czechia\_price cp |
| **join** czechia\_price\_category cpc **on** cp.category\_code = cpc.code |
| **group** **by** rok, cp.category\_code, nazev, cpc.price\_unit |
| **order** **by** nazev, rok ; |

Ad II ) Vývoj mezd v jednotlivých odvětvích v ČR

|  |
| --- |
| **create** **temporary** **table** tmp\_vsm\_projekt\_mzdy **as** |
| **select** |
| cpib.**name**, |
| cp.payroll\_year **as** rok\_vyplaty, |
| cp.industry\_branch\_code, |
| **avg**(cp.value) **as** prumerny\_plat, |
| **lag** (**avg**(cp.value)) **over** (**order** **by** cp.industry\_branch\_code,cp.payroll\_year) **as** prumerny\_plat\_predchozi\_rok, |
| **avg**(cp.value) - **lag** (**avg**(cp.value)) **over** (**order** **by** cp.industry\_branch\_code,cp.payroll\_year) **as** narůst\_platu\_v\_Kc, |
| (**avg**(cp.value) - **lag** (**avg**(cp.value)) **over** (**order** **by** cp.industry\_branch\_code,cp.payroll\_year))/**avg**(cp.value)\*100 **as** narůst\_platu\_v\_procentech, |
| **case** |
| **when** **avg**(cp.value) - **lag** (**avg**(cp.value)) **over** (**order** **by** cp.industry\_branch\_code,cp.payroll\_year) > 0 **then** 'rostoucí' **else** 'klesající' |
| **end** **as** trend |
| **from** czechia\_payroll cp |
| **join** czechia\_payroll\_industry\_branch cpib **on** cp.industry\_branch\_code = cpib.code |
| **where** (cp.value\_type\_code=5958 **and** cp.unit\_code = 200 **and** cp.calculation\_code = 200 ) |
| **group** **by** cpib.**name**,cp.payroll\_year,cp.industry\_branch\_code |
| **order** **by** cp.industry\_branch\_code,cp.payroll\_year ; |

|  |
| --- |
| Spojením výše uvedených tabulek jsem pak vytvořil : |
| **create** **table** t\_vladimir\_smid\_project\_SQL\_primary\_final **as** |
| **select**\* |
| **from** tmp\_vsm\_projekt\_ceny |
| **join** tmp\_vsm\_projekt\_mzdy **on** tmp\_vsm\_projekt\_ceny.rok = tmp\_vsm\_projekt\_mzdy.rok\_vyplaty ; |

Ad III ) Vývoj HDP v ČR

**create** **or** **replace** view w\_vsm\_projekt\_vyvoj\_HDP **as**

**select**

\*,

(gdp-hdp\_minuly\_rok)/hdp\_minuly\_rok \*100 **as** narust\_HDP\_v\_procentech

**from**

( **select**

\*,

**lag** (gdp) **over** ( **order** **by** **year**) HDP\_minuly\_rok

**from** economies

**where** country = 'Czech Republic'

**and** **year** **between** 2006 **and** 2018

**order** **by** economies.**"year"** )

**where** **year** **between** 2007 **and** 2018;

V zadání je požadavek na vytvoření „a t\_{jmeno}\_{prijmeni}\_project\_SQL\_secondary\_final (pro dodatečná data o dalších evropských státech).“

Vzhledem k tomu, že v otázkám není žádný výzkumný úkol zkoumající dat v jiných státech,

tak jsem tabulku s názvem „t\_vladimir\_smid\_project\_SQL\_secondary\_final“ vytvořil,

ale pouze za účel propojení vybraných dat z „t\_vladimir\_smid\_project\_SQL\_secondary\_final “ a dat o vývoji HDP .

Odpověď na každou otázku jsem zpracoval jako samostatný soubor s vysvětlením, jakým způsobem jsem se dopracoval k výstupu .

Předem děkuji za posouzení projektu

S pozdravem Vladimír Šmíd