

PRŮVODNÍ ZPRÁVA

Projektový plán: Analýza dostupnosti potravin a vývoje mezd v ČR a Evropě

Úvod

Na základě zadání budete zpracovávat datové podklady, které umožní analyzovat vývoj cen základních potravin a mezd v ČR a jejich vztah k HDP a dalším ukazatelům životní úrovně. Součástí projektu bude i porovnání dostupnosti základních potravin na základě mezd a vývoj ekonomických ukazatelů ve vybraných evropských zemích.

Cíle

1. **Primární data (Česká republika):** Spojení dat o mzdách a cenách potravin na stejné časové období pro Českou republiku.
2. **Sekundární data (Evropské státy):** Připravení dodatečných dat o HDP, GINI a populaci evropských států ve stejném období jako v primárních datech.
3. **Analytické dotazy:** Vypracování SQL dotazů pro zodpovězení předem definovaných výzkumných otázek.

Na základě zadání vytvoříme dvě hlavní tabulky:

1. `t_Mariia_Anaieva_project_SQL_primary_final` (pro mzdy a ceny potravin v ČR)
2. `t_Mariia_Ananieva_project_SQL_secondary_final` (pro data o evropských státech)

Tyto tabulky budou obsahovat data transformovaná z primárních zdrojů. K tomu musíme sjednotit a připravit data o mzdách a cenách potravin v ČR a dále data o HDP, GINI koeficientu a populaci dalších států. Soubory dat byly filtrovány a spojeny do dvou výsledných tabulek:

1. `t_Mariia_Ananieva_project_sql_primary_final`

Tabulka kombinuje údaje o mzdách v různých odvětvích s údaji o cenách některých potravinářských výrobků. Atributy:

- `id` – jedinečný identifikátor každého záznamu;
- `code` – kód kategorie (pro ceny - šestimístné číslo, pro platy - jednopísmenný kód);
- `name` – název kategorie (slovní popis kategorií odvozených z číselníků);
- `type` – typ kategorie (pro snadnější orientaci údaj o tom, zda se jedná o záznam popisující plat (payroll) nebo cenu (price));
- `date_year` – rozmezí let 2000 až 2020;
- `average_value` – průměrná cena/mzda v daném roce;
- `year1` – vypočtené ročního růstu pro každou cenovou a platovou kategorii.

2. `t_Mariia_Ananieva_project_SQL_secondary_final`

Druhá tabulka obsahuje dodatečná data o evropských státech: HDP, GINI koeficient a populace. Atributy:

- id – jedinečný identifikátor každého záznamu;
- country – název země;
- year – rozmezí let 2000 až 2020;
- tall_hdp – výše hdp;
- hdp_year1 – vypočítaný percentuální meziprocentní růst hdp pro každou zemi;
- gin – gini koeficient;
- population – počet obyvatel;
- taxes – výše daní.

Tato tabulka obsahuje údaje o evropských zemích, včetně HDP, GINI koeficientu a populace. Filtrujeme pouze evropské státy v letech odpovídajících dostupným datům z tabulky mezd pro ČR.