

Projekt SQL – Data o mzdách a cenách potravin a jejich zpracování pomocí SQL. Průvodní listina

I. Požadavky

V rámci výzkumného úkolu byl definován požadavek na vznik datového podkladu, který by poskytl informace týkající se dostupnosti základních potravin spotřebitelům z řad zaměstnanců společně s vývojem HDP v ČR.

Datový podklad, který bude plnit vstupní požadavek, musí zahrnovat data několika ekonomických oblastí, a to:

- Data monitorující vývoj mezd
- Data HDP ČR
- Data monitorující ceny vybraných základních potravin

Spojením výše uvedených sad dat na základě stejných časových období vznikne sada dat, která bude vhodným podkladem pro monitorování definovaných ekonomických oblastí, stejně jako pro získání odpovědí na otázky položené v části III.

II. Datové podklady

Primární tabulka

Na základě požadavku definovaného v I. kapitole necháme vzniknout **primární tabulku**. Ta bude těžit data z níže uvedených základních tabulek:

- **czechia_payroll**

Tabulka, která zachycuje mzdy a počty zaměstnanců v odvětvích v letech 2000 – 2021. Data selektujeme a agregujeme následujícím způsobem:

- je žádoucí selektovat pouze data mezd, nikoli data s počty zaměstnanců. Tedy **value_type_code = 5958**
- je žádoucí pracovat s přepočtenými hodnotami mezd (hodnotami, které nejsou zkresleny různými poměry plných a zkrácených úvazků napříč obory). Tedy **calculation_code = 200**
- budeme pracovat pouze s časovým úsekem let 2006 - 2018, protože pouze pro tyto roky máme relevantní data cen potravin, tedy **payroll_year mezi roky 2006 a 2018**
- budeme pracovat s ročními intervaly, proto hodnoty mezd zachycené v čtvrtletních průměrech převedeme na roční průměry
- tabulka **czechia_payroll** obsahuje pouze kódová označení pracovních oborů, proto ji spojíme s **czechia_payroll_industry_branch** (tabulka obsahuje výčet kódů pracovních

oborů i jejich název). Pro větší přehled v primární tabulce nahradíme kódová označení pracovních oborů jejich názvem

- **czechia_price**

Tabulka, která zachycuje ceny potravin v letech 2006 - 2018. Data selektujeme a agregujeme následujícím způsobem:

- budeme pracovat s celorepublikovými daty, nikoli s daty jednotlivých krajů, proto selektujeme data, kde **region_code IS NULL**
- budeme pracovat s ročními intervaly, proto hodnoty mezd zachycené v rámci týdenních průzkumů cen potravin přepočítáme na roční průměry
- tabulka **czechia_price** obsahuje pouze kódová označení druhu potravin, proto spojíme s **czechia_price_category** (tabulka obsahuje výčet kódů potravin i jejich název). Pro větší přehled v primární tabulce nahradíme kódová označení druhů potravin jejich názvem

- **economies**

Tabulka, která zachycuje data o výši HDP, populaci a další ekonomické informace o zemích světa v letech 1960 – 2020. Data selektujeme následujícím způsobem:

- budeme pracovat pouze s daty pro Českou republiku, tedy **country = Czechia**
- budeme pracovat pouze s časovým úsekem 2006 - 2018, protože pouze pro tyto roky máme relevantní data cen potravin, tedy **_year mezi roky 2006 a 2018**
- pro potřeby primární tabulky selektujeme pouze data o HDP

Všechny tři základní tabulky spojíme na základě shodných časových úseků v jednu primární tabulku, která bude v ročních intervalech sledovat vývoj mezd, HDP i cen potravin.

Provedeme kontrolu úplnosti dat, ověříme, že tabulka neobsahuje NULL hodnoty.

Sekundární tabulka

Doplňková **sekundární tabulka** bude sledovat vývoj HDP, Giniho koeficient a populační data v dalších zemích Evropy v letech 2006 - 2018. Bude těžit data ze základní tabulky **economies**, která poskytuje ekonomické informace o všech zemích světa. Dalším podkladem bude základní tabulka **countries**, která mezi řadou zeměpisných dat poskytuje i kontinentální zařazení zemí:

- **economies**

Tabulka, která zachycuje data o výši HDP, populaci a další ekonomické informace o zemích světa v letech 1960 – 2020. Data selektujeme následujícím způsobem:

- budeme pracovat pouze s časovým úsekem 2006 - 2018, protože pouze toto období zachycuje primární tabulka, tedy **_year mezi roky 2006 a 2018**
- budeme selektovat hodnoty **HDP, populace** a **Giniho koeficient** (rozložení bohatství v populaci)

- **countries**

- budeme selektovat pouze data, kde **continent = Europe**, tedy evropské země

Sekundární tabulka vznikne sloučením výše uvedených tabulek na základě názvů zemí. Na závěr jsou z dat odstraněny záznamy týkající se ČR tak, aby se sekundární tabulka týkala jen dat dalších evropských států.

III. Výzkumné otázky

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

Ve sledovaném období je 25 případů, kdy v různých odvětvích nominální hodnota mzdy meziročně klesla.

Z tohoto pohledu byl nejhorší rok 2013, kdy se hodnota nominální mzdy snížila oproti předchozímu roku hned v 11 odvětvích.

V pětině všech případů byl meziroční pokles velmi malý, menší než 100 Kč. Naopak velmi výrazný meziroční pokles větší než 1000 Kč byl zaznamenán pouze v roce 2013, a to ve 3 odvětvích.

Za sledované období bylo nejčastěji zasaženo poklesem nominální mzdy odvětví "těžba a dobývání" (4x) a "výroba a rozvod elektřiny, plynu a tepla" (3x).

Kolik je možné si koupit litrů mléka a kilogramů chleba za první a poslední srovnatelné období?

Cena obou produktů ve sledovaném období rostla, cena chleba rostla meziročním průměrným tempem téměř 4 %, cena mléka rostla meziročním průměrným tempem 3 %.

Přesto došlo díky růstu průměrné mzdy k zvýšení koupěschopnosti těchto dvou sledovaných produktů v roce 2018 oproti roku 2006. Zatímco v roce 2006 bylo možné za průměrnou měsíční mzdu koupit 1211 ks chleba nebo 1352 l mléka, v roce 2018 to bylo již 1321 ks chleba nebo 1616 l mléka. V roce 2018 tak bylo možné si za průměrnou měsíční mzdu koupit o 110 ks chleba více nebo o 264 l mléka více než v roce 2006.

Která kategorie potravin zdražuje nejpomaleji?

Z 27 sledovaných potravin došlo ve sledovaném období u 25 z nich k růstu ceny, přičemž nejvyšší tempo přes 7 % meziročního růstu jsme zaznamenali u papriky (7,29 %). Přes 6 % meziročního růstu ceny pozorujeme u másla (6,67 %), 5 % hranici překonala vejce (5,55 %), těstoviny (5,26 %), mouka hladká (5,24 %) a mrkev (5,23 %).

Naopak tempo meziročního růstu menší než 1 % se hlásí u vepřové pečeně (0,99 %) a banánů (0,81 %), dokonce záporné tempo meziročního růstu pozorujeme u rajských jablek (-0,74 %) a cukru (-1,92 %).

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

Ne, takový rok ve sledovaném období nenajdeme. Nejvyšší rozdíl byl v roce 2013, kdy tempo růstu cen sledovaných potravin bylo o 5 % vyšší než tempu růstu mezd. Nicméně hned v 8 sledovaných letech z 12 bylo meziroční tempo růstu potravin nižší než meziroční tempo růstu mezd.

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?

Ano, vztah mezi růstem HDP a růstem mezd v následujícím roce je pozorovatelný. Ve sledovaném období došlo ve 3 případech k mimořádně vysokému růstu HDP o více než 5 %. Mzdy následně vzrostly o 8 %, 7 % a 4,5 %, což jsou v rámci sledovaného období 3 ze 4 nejvyšších hodnot.

Naopak v prvních letech po hypoteční krizi v roce 2008, kdy docházelo k propadům HDP nebo jen k jeho velmi pomalému růstu rostly mzdy jen nízkým tempem blízko 2 %.

Nelze pozorovat vztah mezi HDP a cenami vybraných potravin. Vzorek vybraných potravin je patrně příliš malou částí celkového koše statků a služeb. Změny v tempu růstu HDP se do něj výrazně nepodepsaly. Navíc se jedná o pouze úzce zaměřený segment, na který působí další vlivy mimo vývoj HDP. Jedná se například o inflační vývoj, aktuální zemědělská úroda, vývoj komodit na světových trzích, konkurenční prostředí, státní a evropské dotace a další.