

Průvodní listina k projektu SQL – Petra Klimešová

1. Popis projektu

Jsme společnost zabývající se životní úrovní občanů. Dohodli jsme se, že zodpovíme pár definovaných výzkumných otázek, které adresují dostupnost základních potravin široké veřejnosti.

Vyšli jsme z dokumentace datové sady obsahující časovou řadu statistických údajů o průměrných spotřebitelských cenách 27 výrobků potravinářského zboží za Českou republiku a jednotlivé kraje. Průměrná cena položky se počítá, jako aritmetický průměr z cen za kraje. Nejvíce šetření probíhalo v hl. m. Praze. V jednotlivých krajích se spotřebitelské ceny zjišťují u malého souboru prodejen. Uváděné ceny mají proto orientační character. Tyto data máme za období 2006 až 2018.

Dále jsme využili datovou sadu obsahující časovou řadu počtu zaměstnanců a průměrných měsíčních mezd podle odvětví. Průměrné měsíční mzdy jsou hrubé měsíční mzdy, tj. před snížením o pojistné, zálohové splátky daně a další zákonné nebo dohodnuté srážky. Do odvětví jsou data tříděna podle Klasifikace ekonomických činností CZ-NACE, která byla vypracována podle Satistické klasifikace ekonomických činností v Evropském společenství. Tyto údaje máme za roky 2000 až 2019, přičemž roky 2017-2019 jsou jen předběžné, proto s nimi dále raději nepočítáme.

Pro porovnání cen zboží s průměrnými výdělky jsme zvolili pouze roky 2006 – 2016, které nám definují totožné porovnatelné období.

Navíc jsme vyšli z datové sady obsahující časovou řadu statistických údajů o vývoji jednotlivých států světa, základních popisných informací, jejich obyvatel a ekonomických ukazatelích. Pro porovnání a zjištění odpovědí na výzkumné otázky jsme pracovali pouze s daty týkající se České republiky a dalších evropských států v letech 2006 – 2016 (tzv. společné roky).

2. Vvorba primární a sekundární tabulky a omezení dat

Tvorba primární tabulky

Při tvorbě primární tabulky jsem vyšla z těchto dostupných tabulek dat:

- czechia_price
- czechia_price_category
- czechia_region
- czechia_payroll
- czechia_payroll_industry_branch
- czechia_payroll_value_type

Tyto tabulky obsahují data, která potřebujeme ke zjištění odpovědí na výzkumné otázky. Do primární tabulky jsem vložila data za období 2006 až 2016, ve kterých jsou data obsažená.

Z `czechia_price` jsem pro srovnatelné období získala rok extraktem ze sloupce `date_from`. Zde pro mne byla klíčová informace, kdy `region_code` vyplněný, označuje jednotlivé kraje, je-li nevyplněný (NULL) jedná se o průměr za celou Českou republiku.

Z `czechia_payroll` jsem období, definované 4 čtvrtletí každého roku, zjednodušila a použila pouze roky a průměrné mzdy za rok jsem vypočetla jako průměr z jednotlivých čtvrtletí. Důležitá informace je ve sloupci `value_type_code`, který definuje, že pod kódem 316 jde o statistiku počtu zaměstnanců, zatímco 5958 jde o průměrnou mzdu.

U tabulky `czechia_payroll_industry_branch` je obdobná logika jako výše, kde je definované odvětví, týká se to konkrétního odvětví, kde není vyplněno, jedná se o úhrn za všechna odvětví.

Tvorba sekundární tabulky

Při tvorbě sekundární tabulky jsem vyšla z tabulek:

- `countries`
- `economies`

Nejdříve jsem se zaměřila na sloupce, které budu potřebovat pro výpočet výzkumných otázek, jako jsou:

- roky
- HDP
- průměrné mzdy v ČR
- průměrné ceny potravin

Následně jsem zahrнула i požadované proměnné, které mohou být zajímavé pro srovnání výše uvedeného:

- další evropské země
- Giniho koeficient
- počet obyvatel

Když jsem výše uvedený výběr zadávala pouze pro ČR, nebyl to problém a tabulka se vytvořila za pár vteřin. Jakmile jsem zadala všechny státy Evropy, zadáním výběru ve sloupci `continent` "Evropa" z tabulky `countries`, tvorba trvala velmi dlouho. Zkusila jsem tedy přes zadání `region` ve světě (jižní Evropa, východní Evropa, Západní Evropa, severské státy, baltské státy, britské ostrovy, střední a jihovýchodní Evropa), dotaz byl na databázi příliš náročný a vůbec se mi tabulku nepodařilo vytvořit. Zkusila jsem tedy zvolit pro optimalizaci pouze `region` východní Evropa, kde je zahrnuta i Česká republika a sousední státy. Nicméně ani tady nešlo tabulku vytvořit. Ponechala jsem tedy podmínku přes sloupec `continent` a počkala na vytvoření tabulky několik minut. Vyšla jsem ze stejného srovnatelného období, jako v primární tabulce, 2006-2016, kde mám všechny dostupné reálné hodnoty. V jednotlivých sloupcích jsem aplikovala zaokrouhlení, aby data byla přehlednější. Hrubý domácí produkt jsem zadala v miliardách, průměrné mzdy v celých číslech a průměrné ceny potravin jsem zaokrouhlila na 2 desetinná místa.

3. Otázky a odpovědi, popis dat – zdůvodnit, na co jsme přišli

1. *Rostou v průběhu let mzdy ve všech odvětvích, nebo v některých klesají?*

Pro zjištění odpovědi jsem si připravila dočasnou tabulku a z ní následně dopočítala, vyhodnocení zda-li mzdy klesají, rostou, popřípadně další informace a zobrazila tento výsledek v samostatné sloupci wage_growth pomocí CASE expression. Data jsem následně seřadila od nejnižších po nejvyšší, tedy od těch co klesají nejvíce až po ty co rostou nejvíce.

Ve většině případů mzdy meziročně rostou. Nicméně jsou roky a odvětví, kdy to neplatí.

V roce 2009 klesly mzdy ve 3 odvětvích:

- Těžba a dobývání
- Ubytování, stravování a pohostinství
- Zemědělství, lesnictví, rybářství

V roce 2010 mzdy klesly také ve 3 odvětvích, ale v jiných:

- Profesní, vědecké a technické činnosti
- Veřejná správa a obrana; povinné sociální zabezpečení
- Vzdělávání

V roce 2011 také 3 odvětví snížili meziročně své průměrné mzdy, nicméně 2 z nich se již opakují z předchozích let:

- Kulturní, zábavní a rekreační činnosti
- Ubytování, stravování a pohostinství
- Veřejná správa a obrana; povinné sociální zabezpečení

V roce 2013 meziroční pokles vykazovalo nejvíce odvětví, celkem 11 odvětví v České republice:

- Administrativní a podpůrné činnosti
- Informační a komunikační činnosti
- Kulturní, zábavní a rekreační činnosti
- Peněžnictví a pojišťovnictví
- Profesní, vědecké a technické činnosti
- Stavebnictví
- Těžba a dobývání
- Velkoobchod a maloobchod; opravy a údržba motorových vozidel
- Výroba a rozvod elektřiny, plynu, tepla a klimatiz. Vzduchu
- Zásobování vodou; činnosti související s odpady a sanacemi
- Činnosti v oblasti nemovitostí

Následující roky již docházelo k poklesu mezd pouze výjimečně:

- v roce 2014 u Těžba a dobývání,
- v roce 2015 v sektoru Výroba a rozvod elektřiny, plynu, tepla a klimatiz. vzduchu
- a v roce 2016 Těžba a dobývání.

Nejčastěji mzdy klesaly, a to 4x v segmentu Těžba a dobývání.

Naopak v odvětví Doprava a skladování a Zpracovatelský průmysl meziroční pohyb průměrných mezd zaznamenal pouze nárůst mezd.

Ve většině případů se jedná o pokles do 2%, někdy do 3 %. Nicméně v roce 2013 nastal i propad v sektoru Peněžnictví a pojišťovnictví, kdy průměrné mzdy meziročně propadly dokonce o cca 9 %.

Největší meziroční poklesy v průměrných mzdách byli pravděpodobně důsledkem recese české ekonomiky a poklesu HDP ve stejném roce. Hlavní událostí roku 2013 byli devizové intervence České národní banky, která se snažila oslabit korunu a podpořit export. Vliv na Českou republiku měla určitě i globální ekonomická krize a klesající domácí poptávka.

2. *Kolik je možné si koupit litrů mléka a kilogramů chleba za první a poslední srovnatelné období v dostupných datech cen a mezd?*

Udělala jsem si výběr, který obsahuje sloupce food_category, year, food_price, avg_wages a dopočítala jsem quantity_affordable (jako poměr průměrné mzdy a ceny potravin), a vyjela to za první a poslední srovnatelné období, tedy za roky 2006 a 2016. Výpočet quantity_affordable jsem zaokrouhlila striktně dolů, protože nelze v obchodě nakupovat třeba jen 2 dcl mléka nebo dva krajíce chleba.

Výsledkem je, že ač ceny zdražují, platy rostou v porovnání s cenami vybraných potravin rychleji. Za mzdu v roce 2006 jsem si mohla koupit méně zboží, konkrétně 1 191 ks chleba a 1 331 l mléka, než v roce 2016, kde jsem si mohla koupit 1 253 chlebů a 1524 l mléka.

3. *Která kategorie potravin zdražuje nejpomaleji (je u ní nejnižší procentuální meziroční nárůst)?*

Vypočítala jsem si nejdříve, vnitřním selectem, průměrnou cenu za rok a potravinovou kategorii a meziroční změnu. Dalším výpočtem jsem si spočítala průměrnou procentuální změnu napříč všemi roky pro každou kategorii.

Výsledek ukázal, že dokonce 2 kategorie potravin, z 27 sledovaných, průměrně za sledovaná léta zlevňují, a to rajská jablka červená kulatá a cukr krystalový.

Kategorie s nejnižším průměrným procentuálním meziročním nárůstem mi vyšla jablka konzumní s průměrným meziročním zdražením o 0,43 %.

Mezi další patřili vepřová pečeně s kostí o 0,51 % a eidamská cihla s 0,66 %. Nejvíce potravin se pohybovalo v rozmezí od 1 do 4 %. Naopak k nejrychleji zdražujícím potravinám patří pšeničná mouka hladká s nárůstem v průměru o 5,91 % a rýže loupaná dlouhozrná s 6,1 %. Výrazně nejrychleji zdražují papriky a to průměrným meziročním růstem o 9,39 %.

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

Vyšla jsem z primární tabulky, kde jsem měla veškeré potřebné hodnoty pro zodpovězení této otázky. Čekalo mne několik kroků, bohužel jsem však neměla práva tvořit tabulky, view, materializované view, kvůli zablokovanému přístupu. Vyšla jsem tedy z tvorby dočasných tabulek TEMP TABLE. Tu potřebuji vždy znova vytvořit rozkliknutím ctrl+enter, ale jinak funguje jako jiná tabulka. V prvním kroku jsem vytvořila první dočasnou tabulku, kde jsem si vypočítala a zobrazila průměrné ceny potravin a průměrné mzdy v letech 2006 až 2016. Vyjela jsem to vcelku jednoduchým selectem, kdy jsem využila znalosti, že je-li industry_name nevyplněné, jsou hodnoty průměrných mezd vyjádřeny za celou Českou republiku. A z obdobného předpokladu jsem vyšla u region_code, je-li nevyplněn (IS NULL), jsou průměrné ceny uváděny opět jako průměr za Českou republiku. Stačilo zadat spojení na letech a byla připravená krátká jasná tabulka o 11 řádcích.

V druhém kroku jsem vytvořila druhou dočasnou tabulku, kde jsem dopočetla meziroční růst cen potravin a meziroční růst průměrných mezd.

V třetím kroku jsem selektem z druhé dočasné tabulky vyjela požadované sloupce, vypočetla rozdíl v meziročním růstu cen potravin versus mezd a přes CASE expression nechala vypsát a vysvětlit, jak velká změna v jednotlivých letech v růstu těchto dvou sledovaných veličinách je.

Kdybych měla přesně odpovědět na otázku, da-li existuje rok, ve kterém byl meziroční nárůst cen potravin výrazně vyšší než růst mezd, více jak o 10 %? Pak nebyl. Nejvyšší růst potravin versus mzdy byl v roce 2013 o 6,05 %. Ve více případech se jednalo dokonce o nižší růst cen potravin než mezd, někde dokonce o pokles průměrných cen potravin. Nejvýraznější pokles cen potravin oproti mzdám byl v roce 2009, kdy potraviny zaznamenaly meziroční změnu o více jak 9 % v porovnání s růstem mezd. Ve skutečnosti totiž v tomto roce ceny potravin klesly o více jak 6 % oproti předchozímu roku, zatímco mzdy rostly o více jak 3 % v porovnání s předchozím rokem.

Vývoj meziročních průměrných cen potravin byl ve společných letech v 6 případech růst cen, zatímco v 5 případech nastal dokonce pokles průměrných cen. U meziročního růstu mezd byl trend stabilnější, v letech 2007 a 2008 byl meziroční nárůst značný, o cca 7 %. Následující roky se již pohyboval stabilně mezi 2 až 3 procenty meziročního růstu. Pouze v roce 2013 došlo k mírnému meziročnímu poklesu mezd, o cca 1 %.

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?*

Ze sekundární tabulky jsem si vytvořila dočasnou tabulku, kde jsem uvedla požadované hodnoty k výpočtu, rok, HDP, průměrné mzdy a průměrné ceny. K jednotlivým ekonomickým ukazatelům jsem dopočetla procentuální nárůst oproti předchozímu roku a zaznamenala do sloupců: gdp_comparison, wage_comparison a f_prices_comparison. Všechny tyto nárůsty jsem seřadila podle roků 2007 až 2016.

Porovnáme-li vývoj HDP a mezd, pak to vypadá, že změna v HDP se projeví změnou stejným směrem v průměrných mzdách v následujícím roce. S tím, že nárůst ve mzdách je výraznější (o více procent) než je samotná změna HDP. Naopak, pokud jde o pokles hrubého domácího produktu, pak pokles mezd je mírnější, nebo dokonce ani k poklesu nedochází. Z pohledu zaměstnanců je vývoj mezd oproti vývoji HDP příznivější.

Srovnáme-li vývoj HDP a průměrných cen vybraných potravin, tak do roku 2011 včetně byla změna v podobném směru i výši tentýž rok. Tedy nárůst HDP byl doprovázen růstem průměrných cen potravin. Nicméně od roku 2012 byla procentuální změna přibližně stejná, avšak opačným směrem. Jinými slovy, zatímco HDP v letech 2012 a 2013 klesalo, průměrné ceny stoupaly a roky 2015 a 2016 HDP rostlo a průměrné ceny klesaly.