

## **Průvodní listina k projektu**

**Autor:** Stanislav Weissmann  
**email:** stana.ws@gmail.com  
**discord:** Standa W.

### **Zadání projektu**

#### **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 použít pro získání vhodného datového podkladu**

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.

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

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

countries - Všechné 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 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?

## Výstup 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.

## **Realizace projektu:**

### **Primární tabulka**

V SQL skriptu k první výzkumné otázce vytvářím primární tabulku `t_stanislav_weissmann_project_SQL_primary_final`.

Tabulka vzniká spojením dotazů na tabulky `czechia_price` (ceny) a `czechia_payroll` (mzdy a zaměstnanci) prostřednictvím množinové operace sjednocení (UNION). K číselnému kódu zboží jsem přiřadil název zboží, k číselnému kódu odvětví jsem přiřadil název odvětví. Výsledná tabulka již číselné kódy neobsahuje a je tak přehlednější.

Z tabulky cen jsem vynechal řádky, ve kterých měl kód regionu hodnotu NULL (jedná se o průměrné hodnoty měření ze všech regionů). Dále jsem výsledky měření cen zprůměroval u každého zboží na roční hodnoty.

Z tabulky mezd (a zaměstnanců) jsem vybral pouze hodnoty mezd pro přepočtený stav zaměstnanců. Dále jsem data také zprůměroval u každého odvětví na roční hodnoty. Vynechal jsem řádky, ve kterých kód odvětví měl hodnotu NULL.

Původní tabulky byly rozsáhlé (mzdy – 6.880 řádků, ceny – 108.249 řádků). Výsledná tabulka má 589 řádků. Obsahuje sloupce – rok (`at_year`), název zboží (`name_of_product`), průměrná cena (`avg_price`), název odvětví (`name_of_industry_branch`), průměrné mzdy (`avg_wages`).

Nejprve jsou data o zboží (název odvětví a průměrné mzdy mají hodnotu NULL) a pak o mzdách (název zboží a průměrná cena mají hodnotu NULL). Z tabulky se dobře provádějí výběry (ceny nebo mzdy – viz dále v jednotlivých SQL skriptech) a vzhledem k počtu pěti sloupců je uživatelsky přívětivá – není zapotřebí scrollovat zleva doprava.

Původní tabulka cen má data za období 2006 až 2018, tabulka mezd za období 2000 až 2021. Dle zadání nová tabulka má časové období 2006 až 2018 (máme data o mzdách i o cenách).

Primární tabulku jsem použil k zodpovězení výzkumných otázek 1 až 5 (v 5. otázce zprostředkovaně z dalších tabulek, které z primární tabulky vychází).

### **Sekundární tabulka**

V SQL skriptu dodatečného materiálu vytvářím sekundární tabulku `t_stanislav_weissmann_project_SQL_secondary_final`.

Tabulku vytvářím propojením dvou tabulek – `economies` a `countries` pomocí příkazu JOIN. Tabulka dle zadání obsahuje údaje pro jednotlivé evropské země (44 různých zemí) v letech 2004 až 2018 – název, rok, HDP, počet obyvatel, GINI koeficient. Ještě jsem navíc přidal dle mého názoru zajímavý údaj HDP na jednoho obyvatele.

Roky 2004 až 2018 jsem zvolil z důvodu nepřetržité řady dat pro Českou republiku – máme údaje o HDP, počtu obyvatel a GINI koeficientu.

Výsledná tabulka obsahuje i údaje v letech, kdy GINI koeficient má hodnotu NULL, avšak vždy máme údaj o HDP a populaci. Všechny údaje (HDP, počet obyvatel, GINI koef.) má v cca 82 % řádků tabulky.

## Výzkumné otázky:

### 1) Rostou v průběhu let mzdy ve všech odvětvích, nebo v některých klesají?

V původní tabulce czechia\_payroll - jsou 2 údaje o výši mezd pro dané odvětví – pro fyzický a přepočtený počet zaměstnanců. Průměrná mzda pro přepočtený stav zaměstnanců bývá vyšší než pro fyzický (viz vzorek ve skriptu – odvětví A a rok 2000).

Rozhodl jsem se využívat přepočtený stav zaměstnanců, dle mého názoru je tento údaj objektivnější. Do přepočteného počtu zaměstnanců se zahrnují: všichni stálí i dočasní zaměstnanci, kteří jsou v pracovním poměru, zaměstnanci v době nemoci, brigádníci. Do přepočteného počtu zaměstnanců nepatří: osoby pracující na základě dohody o pracovní činnost nebo na základě dohody o provedení práce, ženy na mateřské dovolené a rodičovské dovolené (do 3 let věku dítěte), muži na rodičovské dovolené (do 3 let věku dítěte), zaměstnanci uvolnění k výkonu funkce (starosta obce), učni a studenti na provozní praxi.

#### Závěr:

Jako „případ“ pro následující závěry označuji výpočet pro jedno odvětví v jednom roce. Celkem ve 24 případech poklesla průměrná přepočtená mzda oproti předcházejícímu roku. Meziroční pokles příjmů dle mého názoru není ojedinělý jev – vyskytl se v cca 11 % sledovaných případů (24 z celkového počtu 228).

Z jednotlivých případů nejvíce poklesly průměrné mzdy v oboru peněžnictví a pojišťovnictví – o 8,83 % v r. 2013 a oboru výroby a rozvodu elektřiny, plynu, tepla a klimatiz. vzduchu – o 4,44 % v r. 2013. V r. 2013 poklesly mzdy v největším počtu odvětví – celkem v 11 případech.

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

Z řešení vyplývá, že v roce 2006 bylo možné zakoupit 1.313 chlebů a 1.466 l mléka. V roce 2018 v obou případech bylo možné zakoupit více – 1.365 chlebů a 1.670 l mléka. Obou zboží si mohli tedy lidé v pozdějším roce 2018 koupit více.

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

Nejpomaleji zdražuje krystalový cukr. Postup výpočtu: nejdříve jsem vytvořil tabulku vývoje cen jednotlivých kategorií zboží v jednotlivých letech (porovnání průměrné ceny v daném a předchozím roce). Pak jsem tyto pohyby cen v následujícím dotazu pro každou kategorii zboží zprůměroval. V některých letech průměrná cena zboží vzrostla, v některých letech poklesla. U krystalového cukru tento průměr vývoje cen za celé sledované období 2006 až 2018 dokonce dosahuje záporné hodnoty: -1,92 %. Kontrolní dotaz na ceny cukru potvrzuje, že v r. 2006 jeho průměrná cena byla 21,73 Kč, zatímco v roce 2018 stál průměrně 15,75 Kč.

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

Neexistuje. Největší rozdíl v meziročních nárůstech je v roce 2013, kdy rozdíl v cenovém vývoji činil cca 6,65 %. Mzdy v tomto roce totiž poklesly o 1,55 % a ceny potravin se zvýšily o 5,1 %.

**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?**

Nedostatek tohoto zadání je, co se myslí „výraznějším růstem“. Jako hranici pro „výrazněji“ jsem zvolil meziroční růst větší než 5 %.

V letech 2006 až 2018 rostlo HDP o více než 5 % ve čtyřech letech - 2006, 2007, 2015 a 2017. Ve třech letech (2006, 2007 a 2017) vzrostly o více než 5 % také mzdy nebo ceny (nebo obojí) v daném roce nebo v roce následujícím.

Takže z dat byla prokázána souvislost ve třech ze čtyř let.

Pokud je 5 %-ní hranice „výraznějším růstem“, pak lze tvrdit, že většinou se významnější růst HDP projeví výraznějším růstem cen či mezd v daném roce či následujícím.

**Dodatečný materiál - tabulka s HDP, GINI koeficientem a populací dalších evropských států ve stejném období jako primární přehled pro ČR.**

Více v popisu vytvoření sekundární tabulky. Jako významný ukazatel vývoje země jsem přidal sloupec GDP/1 obyvatele. Prvních 15 „příček“ tabulky obsadilo Monako, kdy i HDP/1 obyvatele v roce 2004 (15. příčka v tabulce) je vyšší hodnota tohoto ukazatele pro Lichštejnsko v roce 2010 (16. příčka v tabulce).

Zajímavý je i rozdíl v hodnotě tohoto ukazatele v r. 2018 mezi jednotlivými zeměmi – první Monako má 196 tis. USD HDP/1 obyvatele, druhé Lucembursko – 111 tis. USD. Česká republika v tomto roce měla 24 tis. USD.