

ENGETO SQL projekt
Průvodní listina

Autor: Matouš Kocháň

Datum vypracování: 11. Června 2023

1) Obsah	
2) Úvod do projektu	3
3) Úkol 1	5
4) Úkol 2	7
5) Úkol 3	8
6) Úkol 4	9
8) Úkol 6	11
9) Závěr.....	12
10) Seznam použitých tabulek	13
11) Seznam použitých grafů	13

2) 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

Primární tabulky:

1. `czechia_payroll` – Informace o mzdách v různých odvětvích za několikaleté období. Datová sada pochází z Portálu otevřených dat ČR.
2. `czechia_payroll_calculation` – Číselník kalkulací v tabulce mezd.
3. `czechia_payroll_industry_branch` – Číselník odvětví v tabulce mezd.
4. `czechia_payroll_unit` – Číselník jednotek hodnot v tabulce mezd.
5. `czechia_payroll_value_type` – Číselník typů hodnot v tabulce mezd.
6. `czechia_price` – Informace o cenách vybraných potravin za několikaleté období. Datová sada pochází z Portálu otevřených dat ČR.
7. `czechia_price_category` – Číselník kategorií potravin, které se vyskytují v našem přehledu.

Číselníky sdílených informací o ČR:

1. `czechia_region` – Číselník krajů České republiky dle normy CZ-NUTS 2.
2. `czechia_district` – Číselník okresů České republiky dle normy LAU.

Dodatečné tabulky:

1. `countries` - Všechné informace o zemích na světě, například hlavní město, měna, národní jídlo nebo průměrná výška populace.
2. `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.

Na svém GitHub účtu vytvořte repozitář (může být soukromý), kam uložíte všechny informace k projektu – hlavně SQL skript generující výslednou tabulku, popis mezivýsledků (průvodní listinu) a informace o výstupních datech (například kde chybí hodnoty apod.).

3) Úkol 1

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

Výchozí tabulky k řešení prvního úkolu jsou zaznačeny v Tabulce 1.

Tabulka 1: Výchozí tabulky pro úkol 1

Tabulka	Alias
czechia_payroll	pay
czechia_payroll_value_type	val
czechia_payroll_unit	unit
czechia_payroll_industry_branch	bra
czechia_payroll_calculation	cal

Nově vzniklé tabulky jsou zaznačeny v Tabulce 2.

Tabulka 2: Nové tabulky pro úkol 1

Tabulka
income_1
income_2

Tabulky **pay**, **val**, **unit**, **bra**, **cal** byly sloučeny a přetvořeny do tabulky **income_1**.

Následně byla tabulka **income_1 i1** sloučena s tabulkou **income_1 i2**. Tímto spojením vznikla tabulka **income_2**.

Závěr úkolu 1

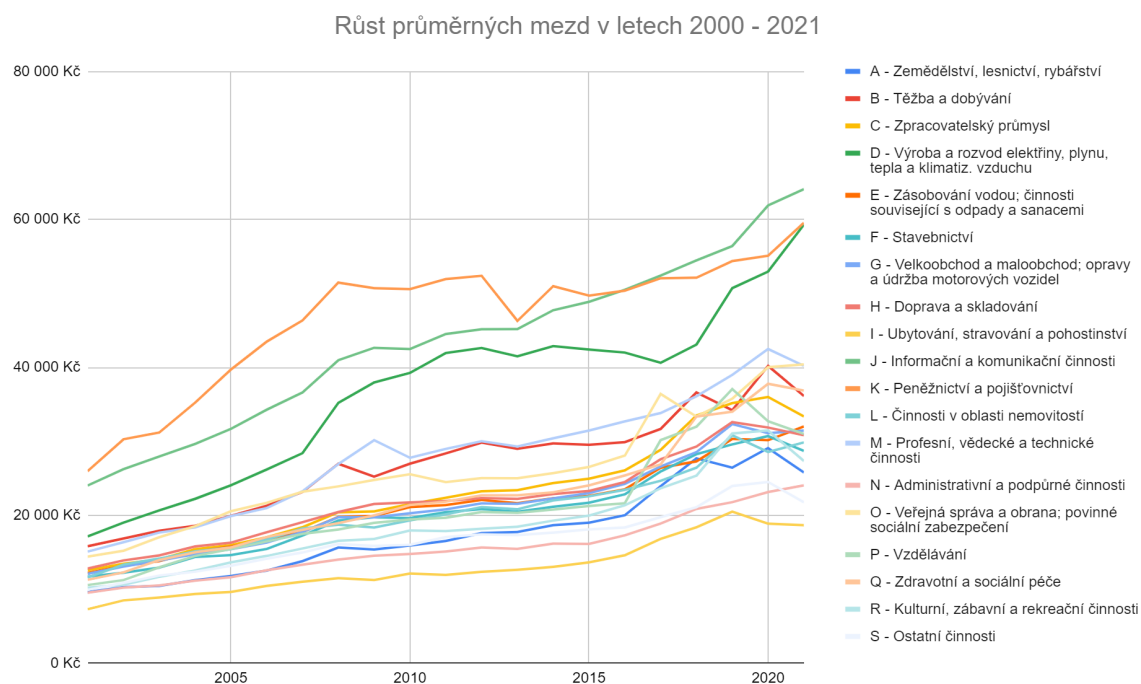
Sloupec **x** tabulky **income_2** nabývá hodnot 0, nebo 1.

Pokud je řádku přisouzena hodnota 1, pak je rozdíl průměrné mzdy v letech 2001–2000, 2002–2001 až 2021–2020 nezáporný, průměrná mzda roste.

Pokud je přisouzená hodnota rovna 0, pak je rozdíl záporný, průměrná mzda klesá.

Většina řádků ve sloupci **x** nabývá hodnot 1. To naznačuje, že mzdy v průběhu let rostou napříč všemi odvětvími. Mírný pokles byl pozorován u některých odvětví mezi roky 2020 a 2021, jak je znázorněno na Grafu 1.

Graf 1: Růst průměrných mezd v letech 2000 až 2021



4) Úkol 2

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

Vstupní tabulky, které byly použity k řešení druhého úkolu jsou zaznačené v Tabulce 3.

Tabulka 3: Výchozí tabulky pro úkol 2

Tabulka	Alias
czechia_price	pri
czechia_price_category	cat

Nově vzniklé tabulky jsou zaznačeny v tabulce 4.

Tabulka 4: Nové tabulky pro úkol 2

Tabulka
commodity
commodity_2
commodity_3

Tabulka **pri** se sloučila s tabulkou **cat** a vznikla tabulka **commodity**.

Data z **income_2** byla propojena s **commodity**. Vznikla tabulka **commodity_2**.

Závěr úkolu 2

V tabulce **commodity_3** je znázorněno, kolik je možné zakoupit litrů mléka a kilogramů chleba v letech 2006 a 2018. V průměru si zaměstnanci všech odvětví mohli v roce 2006 koupit 1194 kilogramů chleba a 1333 litrů mléka. V roce 2018 pak 1320 kilogramů chleba a 1614 litrů mléka.

5) Úkol 3

Zadání: Která kategorie potravin zdražuje nejpomaleji (je u ní nejnižší procentuální meziroční nárůst)?

V tabulce 5 jsou uvedeny nově vzniklé tabulky pro úkol 3.

Tabulka 5: Nové tabulky pro úkol 3

Tabulka
comm_grow
comm_grow_2

Závěr úkolu 3

Sloučením tabulky **commodity_2 com2a** a **commodity_2 com2b** vznikla tabulka **comm_grow**. V tabulce **comm_grow** byl vytvořen nový atribut **commodity_grow_per**, který udává, o kolik procent se zdražila komodita mezi konkrétní roky.

Nejpomaleji zdražily banány, meziroční procentuální nárůst je 0,81 %. U cukru lze sledovat meziroční procentuální snížení o 1,92 %.

6) Úkol 4

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

V tabulce 6 jsou uvedeny nově vzniklé tabulky pro úkol 4

Tabulka 6: Nové tabulky pro úkol 4

Tabulka
comm_grow_3

Závěr čtvrtého úkolu

Neexistuje rok, kde by byl rozdíl průměrného meziročního nárůstu cen potravin a průměrného meziročního růstu mezd větší než 10 %.

7) Úkol 5

Zadání: 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

V tabulce 7 jsou uvedeny výchozí tabulky k úkolu 5.

Tabulka 7: Výchozí tabulky pro úkol 5.

Tabulka
economies

V tabulce 8 jsou uvedeny nově vzniklé tabulky a pohledy pro úkol 5.

Tabulka 8: Nové tabulky pro úkol 5

Tabulka
GDP
t_matous_kochan_project_SQL_primary_final
GDP_grow

V tomto úkolu nebylo explicitně zadáno, co se myslí výrazným růstem HDP. Autor práce se rozhodl zvolit růst nad 5 % HDP jako výrazný.

Závěr pátého úkolu

Výrazný růst HDP byl zaznamenán v letech 2006, 2007, 2015 a 2017.

Pro rok 2006 nejsou dostupná data pro meziroční nárůst cen potravin, avšak mzdy vzrostly o 6,22 %.

V roce 2007 potraviny zdražily o 9,26 %, mzdy vzrostly o 10,15 %.

V roce 2008 potraviny zdražily o 8,92 %, mzdy vzrostly o 9,65 %.

V roce 2015 potraviny zlevnily o 0,69 %, mzdy vzrostly o 1,82 %.

V roce 2016 potraviny zlevnily o 1,41 %, mzdy vzrostly o 4,1 %.

V roce 2017 potraviny zdražily o 7,06 %, mzdy vzrostly o 19,66 %.

V roce 2018 potraviny zdražily o 2,41 %, mzdy vzrostly o 8,01 %.

Z dat vyplývá, že výška HDP má vliv na změny ve mzdách a cenách potravin. Když roste HDP, tak většinou roste cena potravin i mzdy v daném a následujícím roce.

8) Úkol 6

Tento úkol nebyl explicitně zadán. Lektorem bylo doporučeno vybrat sloupce z tabulek pro dodatečná data o dalších evropských státech tak, aby souvisely s ekonomickými ukazateli.

Vstupní tabulky, které byly použity k řešení šestého úkolu jsou zaznačené v Tabulce 9.

Tabulka 9: Vstupní tabulky pro úkol 6

Tabulka
economies
countries

Nově vzniklé tabulky jsou zaznačeny v tabulce 10.

Tabulka 10: Nové tabulky pro úkol 6

Tabulka
country_info
t_matous_kochan_project_SQL_secondary_final

Závěr šestého úkolu

Tabulka **economies e** se sloučila s tabulkou **countries c** a vznikla tabulka **country_info**. Data z **country_info x1** byla propojena s **country_info x2**. Vznikla tabulka **t_matous_kochan_project_SQL_secondary_final**, která byla obohacena o sloupce vyjadřující meziroční procentuální nárůst. Konkrétní atributy jsou uvedeny v tabulce 11.

Tabulka 11: Nově vzniklé atributy popisující meziroční procentuální nárůst

Název atributu	Popis atributu
GDP_grow	Meziroční růst HDP v %
population_grow	Meziroční růst populace v %
gini_grow	Meziroční růst ukazatele gini v %
taxes_grow	Meziroční nárůst daní v %
fertility_grow	Meziroční růst plodnosti v %
mortality_under5	Meziroční růst úmrtí dětí do 5 let v %

9) Závěr

V rámci vypracování tiskové zprávy zabývající se dostupností základních potravin široké veřejnosti bylo odpovězeno na pět výzkumných otázek.

Zprvce bylo zjištěno, že mzdy v průběhu let rostou napříč všemi odvětvími. Mírný pokles byl pozorován u některých odvětví mezi roky 2020 a 2021, jak je znázorněno na Grafu 1.

V druhém bodě bylo zjišťováno, kolik je možné zakoupit litrů mléka a kilogramů chleba v letech 2006 a 2018. V průměru si zaměstnanci všech odvětví mohli v roce 2006 koupit 1194 kilogramů chleba a 1333 litrů mléka. V roce 2018 pak 1320 kilogramů chleba a 1614 litrů mléka.

V třetím úkolu bylo zjišťováno, která kategorie potravin zdražuje nejpomaleji. Nejpomaleji zdražily banány, meziroční procentuální nárůst je 0,81 %. U cukru lze sledovat meziroční procentuální snížení o 1,92 %.

Výstupem čtvrtého úkolu je zjištění, že neexistuje rok, kde by byl rozdíl průměrného meziročního nárůstu cen potravin a průměrného meziročního růstu mezd alespoň 10 %.

V pátém úkolu autor práce zvolil růst nad 5 % HDP jako výrazný. Z dat vyplývá, že výrazný růst HDP má vliv na změny ve mzdách a cenách potravin. Když roste HDP, tak většinou roste cena potravin i mzdy v daném a následujícím roce.

Výstupy úkolů 1–5 lze získat z dat v tabulce `t_matous_kochan_project_SQL_primary_final`.

Poslední úkol nebyl explicitně zadán. Byly vybrány sloupce z tabulek pro dodatečná data o dalších evropských státech tak, aby souvisely s ekonomickými ukazateli. Výstupem je tabulka `t_matous_kochan_project_SQL_secondary_final`.

10) Seznam použitých tabulek

Název	Strana
Tabulka 1: Výchozí tabulky pro úkol 1	5
Tabulka 2: Nové tabulky pro úkol 1	5
Tabulka 3: Výchozí tabulky pro úkol 2	7
Tabulka 4: Nové tabulky pro úkol 2	7
Tabulka 5: Nové tabulky pro úkol 3	8
Tabulka 6: Nové tabulky pro úkol 4	9
Tabulka 7: Výchozí tabulky pro úkol 5	10
Tabulka 8: Nové tabulky pro úkol 5	10
Tabulka 9: Vstupní tabulky pro úkol 6	11
Tabulka 10: Nové tabulky pro úkol	11
Tabulka 11: Nově vzniklé atributy popisující meziroční procentuální nárůst	11

11) Seznam použitých grafů

Název	Strana
Graf 1: Růst průměrných mezd v letech 2000 až 2021	6