

Průvodní listina k „Projektu z SQL“

Řešitel: Petr Podolinský (@Discord: „Petr P. / podolinsky.petr“)

1. Přehled dodaných souborů s jednotlivými SQL kódy

Název souboru	Příloha k průvodní listině – názvy sloupců / komentář
Kód_1__create_t_petr_podolinský_project_SQL_primary_final	Příloha č. 1
Kód_2__create_t_petr_podolinský_project_SQL_secondary_final	Příloha č. 2
Kód_3__úkol #1_petr_podolinský	kód pro 1. úkol
Kód_4__úkol #2_petr_podolinský	kód pro 2. úkol, Příloha č. 3
Kód_5__úkol #3_petr_podolinský	kód pro 3. úkol
Kód_6__úkol #4_petr_podolinský	kód pro 4. úkol
Kód_7__úkol #5_petr_podolinský	kód pro 5. úkol

2. Řešení jednotlivých úkolů včetně interpretace výsledků (zjištění)

Úkol 1	Rostou v průběhu let mzdy ve všech odvětvích, nebo v některých klesají?
---------------	---

Ve výsledné tabulce, která generuje 19 řádkových záznamů, jsou pro 21 meziročních „period“ v období let 2001 až 2021 (počáteční meziroční srovnání se týká let 2000/2001) zobrazeny tyto sloupce:

- industry_branch_code – kód průmyslového odvětví,
- branch – název odvětví,
- count_of_yearly_increase – počet „period“ meziročního růstu,
- pct_growth_periods_y2001_to_2021 – procentuální podíl „period“ meziročního růstu vztaženo k jejich celkovému počtu (celkem 21) → *formální vyjádření podílu růstových „period“ ve sledovaném období*,
- avg_yearly_payroll_change (*) – hodnota průměrné meziroční změny průměrné mzdy za každé jednotlivé odvětví ve sledovaném období.

Poznámka (*): data jsou řazena sestupně podle posledního sloupce.

Zjištění:

Z dostupných údajů je zřejmé, že **u všech sledovaných (průmyslových) odvětví dochází v období let 2000 až 2021 k růstu** (ve smyslu průměrného meziročního růstu) **průměrných mezd** v absolutní hodnotě na maximální úrovni +1980 Kč/rok (platí pro „informační a komunikační činnosti“), přičemž minimální růst průměrných mezd je typický pro odvětví „ubytování, stravování a pohostinství“ s hodnotou +594 Kč/rok – uvedené tvrzení však nezohledňuje meziroční vývoj inflace v relativním i absolutním vyjádření, pro kterou nejsou data z dostupných tabulek k dispozici. K největšímu průměrnému meziročnímu růstu dochází vedle již uvedených „informačních a komunikačních činností“ též:

- ve „výrobě a rozvodu elektřiny, plynu, tepla a klimatiz.“ (+1775 Kč/rok),
- peněžnictví a pojišťovnictví (+1738 Kč/rok) a
- ve „zdravotní a sociální péči“ (+1607 Kč/rok).

Z celkového počtu 21 sledovaných meziročních „period“ byl růst průměrných mezd v každém odvětví zaznamenán minimálně v 80,95 % případů¹ (údaj platí pro odvětví „těžba a dobývání“ s průměrným ročním růstem mezd na úrovni +1020 Kč/rok), přičemž mzdy vykázaly ve všech těchto „periodách“ systematický růst pouze ve 4 odvětvích z 19 sledovaných („zdravotní a sociální péče“, „zpracovatelský průmysl“, „doprava a skladování“ a „ostatní činnosti“, z nichž v posledním uvedeném oboru je parametr `avg_yearly_payroll_change` = +701 Kč/rok → pro ověření lze v klauzuli **ORDER BY** odkomentovat následující: `-- count_of_yearly_increase DESC`).

Úkol 2

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

Ve výsledné tabulce, která generuje 38 řádkových záznamů – 19× pro kategorii potravin „mléko polotučné pasterované“ a 19× pro „chléb konzumní kmínový“, jsou uvedeny tyto sloupce:

- `goods_category_name` – název kategorie zboží,
- `branch` – název (průmyslového) odvětví,
- `quantity_available_2006` – dostupné množství dle odvětví v prvním srovnatelném období (rok 2006),
- `quantity_available_2018` – dostupné množství dle odvětví v posledním srovnatelném období (rok 2018),
- `quantity_difference` – rozdíl dostupných množství mezi definovanými obdobími 2018/2006 → *tento sloupec byl použit pro sestupné řazení v rámci dílčích kategorií zboží.*

Poznámka: zjištěno jako podíl průměrné mzdy za dané odvětví a průměrné ceny dané kategorie zboží v konkrétním roce.

Zjištění (pro rychlé posouzení viz „přehled“ dole na této stránce):

V případě první kategorie zboží („mléko polotučné pasterované“) je možné si v roce 2006 průměrně koupit 2449 jednotek tohoto zboží pouze v případě, kdy zaměstnanec pracuje v odvětví „informační a komunikační činnosti“ (v „ubytování, stravování a pohostinství“ by se naopak jednalo pouze o 789 jednotek mléka) – ve srovnání s tímto výchozím obdobím je situace v IT sektoru roce 2018 z hlediska dostupnosti posuzovaného zboží lepší o +382 jednotek mléka a v tomto kontextu se jedná o nejvyšší úroveň zlepšení dostupnosti „mléka“ (prostřednictvím vývoje mezd v tomto sektoru) napříč jednotlivými průmyslovými odvětvími, a to na +2831 jednotek (+15,6 %).

U druhé kategorie zboží („chléb konzumní kmínový“) lze největší rozdíl (změnu) v dostupnosti za období 2018/2006 zaznamenat v odvětví „zdravotní a sociální péče“, kde je rozdíl +203 jednotek. Pokud srovnáme „mléko“ a „chléb“, lze z uvedených čísel (+382 vs. +203 jednotek) odhadovat, že dochází ke snížení dostupnosti chleba ve srovnání s mlékem a tuto skutečnost lze **souhrnně za všechna (průmyslová) odvětví** dokumentovat následujícím způsobem (viz doplňkový SQL dotaz uvedený v **příloze č. 3**):

<u>Název kategorie zboží</u>	<u>Průměr množství jednotek</u>		<u>Rozdíl</u>
	<u>ø(r. 2006)</u>	<u>ø(r. 2018)</u>	<u>dostupnosti (jedn. zboží)</u>
- Chléb konzumní kmínový	1287,368	1342,263	54,89
- Mléko polotučné pasterované	1437,316	1641,632	204,32

¹ Předpokládám, že pokud by v některých odvětvích byl růst mezd klesající, alespoň v polovině případů, tj. 11 „period“ z celkového počtu 21, by byl meziroční růst mezd nulový nebo záporný, což se nepotvrdilo. Opět však platí, že bez zohlednění **inlace** nelze stejným způsobem interpretovat vývoj reálných mezd.

Poznámka: možná by v této souvislosti stálo za to prozkoumat i podíly dovozu jednotlivých kategorií zboží – lze očekávat, že dovoz mléka významněji převyšuje dovoz chleba, který je spíše „domácí vyráběnou komoditou“, tím pádem se do ceny chleba tolik nepromítá konkurenční prostředí, a rozdíl dostupnosti mléka je tak markantnější než v případě chleba.

V některých odvětvích naopak dochází v porovnávaném období let 2018/2006 ke snížení dostupnosti obou položek zboží:

- v případě „mléka“ se jedná o odvětví „peněžnictví a pojišťovnictví“ s rozdílem dostupného množství -19 jednotek,
- v případě „chleba“ se tato situace týká „činností v oblasti nemovitostí“, „těžby a dobývání“, „administrativních a podpůrných činností“, „ostatních činností“ a paradoxně též „peněžnictví a pojišťovnictví“ (viz situace u „chleba“).

Poznámka: ve vazbě na změnu dostupnosti obou diskutovaných kategorií zboží je oblast „peněžnictví a pojišťovnictví“ jedinou, ve které je tato změna záporná.

Úkol 3

Která kategorie potravin zdražuje nejpomaleji (je u ní nejnižší percentuální meziroční nárůst)?

Názvy sloupců v generované tabulce:

- goods_category_name – název kategorie zboží (potravin),
- avg_annual_pct_change – průměrná percentuální roční změna ceny zboží.

Zjištění:

Na základě vyhodnocení 27 řádkových záznamů seřazených vzestupně dle sloupce „avg_annual_pct_change“ (vysvětlení viz názvy sloupců výše) lze jako kategorii zboží (potravin), která zdražuje nejpomaleji, uvést **„cukr krystalový“**. Výslednou úroveň avg_anual_pct_change = -1,92 % pro tuto kategorii zboží významně ovlivnily² dvě periody od počátku datové řady, tj. meziroční změna 2015/2014 (-19,16 %) a dále vývoj v letech 2018/2017 (-21,09 %).

Za touto kategorií „cukr krystalový“ pak vzestupně následují tyto kategorie zboží s hodnotami průměrné percentuální roční změny v závorkách: „rajská jablka červená kulatá“ (-0,74 %), „banány žluté“ (+0,81 %) a „vepřová pečeně s kostí“ (+0,99 %).

Úkol 4

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

Názvy sloupců v generované tabulce:

- year – rok (rozmezí let 2000 až 2021),
- food_price_change – meziroční změna cen potravin (%),
- payroll_change – meziroční změna úrovně mezd (%),
- abs_change_difference – absolutní hodnota rozdílu meziročních změn cen potravin a mezd (%).

² Zde uvedeno bez důkazu – ověřeno souběžně pomocí Excelu.

Zjištění

V případě kladných meziročních změn jak průměrných cen potravin, tak i průměrných mezd byla nejvyšší úroveň parametru `abs_change_difference` zaznamenána v roce 2018, a to na úrovni +5,46 % (zde však platí, že *payroll_change > food_price_change*, což *neodpovídá zadání*). Bez ohledu na kladnou či zápornou meziroční změnu cen potravin a/nebo mezd dosáhl parametr `abs_change_difference` hodnoty +9,57 % v roce 2009 (zde však byla meziroční změna cen potravin záporná -6,41 %, resp. u mezd kladná +3,16 %, tedy *opět* zde nastává *rozpor s podmínkami zadání*).

Lze tedy konstatovat, že v rámci dostupných dat **neexistuje rok, ve kterém (by) byl meziroční nárůst cen potravin výrazně vyšší³ nad úrovní 10 % než růst mezd.**

Úkol 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?
---------------	--

Názvy sloupců v generované tabulce:

- | | |
|---------------------------------------|--|
| ▪ <code>year</code> | – rok (rozmezí let 2000 až 2021), |
| ▪ <code>GDP_mld</code> | – HDP v mld. USD, |
| ▪ <code>avg_payroll</code> | – průměrné mzdy v jednotlivých letech, |
| ▪ <code>avg_price_goods</code> | – průměrné ceny zboží v jednotlivých letech, |
| ▪ <code>GDP_change_pct</code> | – meziroční změna HDP (%), |
| ▪ <code>payroll_change_pct</code> | – meziroční změna průměrných mezd (%), |
| ▪ <code>price_goods_change_pct</code> | – meziroční změna cen zboží (%). |

Zjištění:

Při nastavení podmínky meziroční změny **HDP > 5 %** v SQL dotazu, což by mohlo reprezentovat výraznější růst této ekonomické veličiny, se zobrazí tři roky (2007, 2015 a 2017). Rozhodl jsem se ponechat podmínku

WHERE (`GDP_mld - prev_GDP`) / `prev_GDP` * 100 > 5

zakomentovanou a podívat se na situaci s meziroční změnou ostatních parametrů, tj. „`avg_payroll`“ a „`avg_price_goods`“ v bezprostředně následujících letech, tedy s použitím podmínky

WHERE `year` **IN** (2007, 2008, 2015, 2016, 2017, 2018),

a to za účelem zjištění vývoje HDP, cen zboží a mezd ve stejném nebo následujícím roce.

Při porovnávání vývoje HDP vs. „(průměrné) ceny zboží“ a „(průměrné) mzdy“ lze tedy očekávat jistou adaptační setrvačnost – konkrétně:

- meziroční **změna HDP +5,57 %** v roce 2007 přinesla:
 - meziroční změnu +6,84 / +7,87 % ve „mzdách“ pro r. 2007 / 2008,
 - meziroční změnu +6,76 / +6,18 % v „cenách zboží“ pro r. 2007 / 2008;
- meziroční **změna HDP +5,39 %** v roce 2015 přinesla:
 - meziroční změnu +2,51 / +3,65 % ve „mzdách“ pro r. 2015 / 2016,
 - meziroční změnu -0,55 / -1,19 % v „cenách zboží“ pro r. 2015 / 2016;

³ Pozn.: v původním kódu v souboru „Kód_6__úkol #4_petr_podolinský“ byly v klauzuli `WHERE` záměrně ponechány zakomentované části kódu, se kterými byla tato zjištění ověřována.

- meziroční **změna HDP +5,17 %** v roce **2017** přinesla:
 - o meziroční změnu +6,28 / +7,62 % ve „mzdách“ pro r. **2017 / 2018**,
 - o meziroční změnu +9,63 / +2,17 % v „cenách zboží“ pro r. **2017 / 2018**.

Na základě předchozího srovnání lze závěrem uvést tyto skutečnosti:

- a) **změna HDP > 5 %** (dále jen „výraznější kladná změna“) v roce N_i , kde $N_i = 2007, 2015$ a **2017 výrazněji** (tj. $>6\%$) **ovlivnila meziroční změnu ve „mzdách“ i „cenách zboží“ ve dvou případech ze tří, a to spíše v rámci stejného roku N_i než v roce $N_i + 1$;**
- b) na výraznější kladnou změnu HDP se zdají být citlivější „mzdy“, a to v obou porovnávaných obdobích N_i a $N_i + 1$ jednotlivých let **2007, 2015 a 2017;**
- c) pro jednoznačné určení vlivu kladné změny HDP na vývoj (průměrných) cen potravin a (průměrných) mezd by bylo nutné provést srovnání pro větší počet časových období, případně určit, jakým způsobem je vhodné modifikovat podmínku pro stanovení „výraznější kladné změny“ HDP [viz **bod a)** výše].

PŘÍLOHY

Příloha č. 1 k průvodní listině – význam názvů sloupců pro tabulku získanou SQL dotazem ze souboru „Kód_1__create_t_petr_podolinský_project_SQL_primary_final“

Počet záznamů celkem: 6669

Název sloupce	Popis
year	rok (rozmezí let 2000 až 20 21)*
industry_branch_code	kód průmyslového odvětví
branch	název odvětví
avg_payroll_value_per_year	průměrná roční mzda v každém roce*
goods_category_code	kód kategorie zboží
goods_category_name	název kategorie zboží
avg_price_goods_per_year	průměrná cena zboží v každém roce*
GDP_rounded_mld	HDP (v miliardách USD)

Poznámky k tabulce:

Poznámka 1/3: Chybějící hodnoty v rozmezí let 2000 až 2005 & 2019 až 2021.

Poznámka 2/3: Chybějící hodnoty pro rok 2021.

Jednotka „mld. USD“ ověřena ze zdrojů <https://www.kurzy.cz/svet/nemecko/hdp/> & <https://www.google.com/search?client=firefox-b-d&q=hdp+%C4%8Dr> při doplnění SQL dotazu: **WHERE** country **IN** ('Czech Republic', 'Germany') **AND** `year` = 2020.

Poznámka 3/3: Viz další tabulka.

Příloha č. 2 k průvodní listině – význam názvů sloupců pro tabulku získanou SQL dotazem ze souboru „Kód_1__create_t_petr_podolinský_project_SQL_**secondary_final**“

Počet záznamů celkem: 945

Název sloupce	Popis
year	rok (rozmezí let 2000 až 2020)
country	země
GDP_mld	HDP (v miliardách USD) – viz tabulka výše
Gini_koef	Giniho koeficient (poměřuje rovnost bohatství mezi jednotlivci na stupnici od 0 do 1)
population_mid2018	velikost populace uprostřed roku 2018

Poznámky k tabulce:

Poznámka 1/3:	Dostupných (NOT NULL) je 882 záznamů, tj. cca 93 % z 945 záznamů.
Poznámka 2/3:	Dostupných (NOT NULL) je 578 záznamů, tj. cca 61 % z 945 záznamů.
Poznámka 3/3:	Uvedeno pouze pro r. 2018 ve vazbě na tabulku 'countries' – kontrola pro ČR: - počet obyvatel na počátku roku 2018 = 10 610 055, - počet obyvatel na konci roku 2018 = 10 649 800; - výběr přes SQL dotaz = 10 629 928, což odpovídá průměru uvedených čísel. Zdroj: ČSÚ

Příloha č. 3 – Doplněk kódu pro úkol 2

```
CREATE TABLE t_petr_podolinsky_project_sql_tertiary AS (
    Kód ze souboru „Kód_4__úkol #2_petr_podolinský“
);

SELECT
    goods_category_name,
    round(avg(quantity_available_2006), 3) AS prumer_mnozstvi_2006,
    round(avg(quantity_available_2018), 3) AS prumer_mnozstvi_2018,
    round(avg(quantity_available_2018) - avg(quantity_available_2006), 2)
    AS rozdil_dostupnosti
FROM t_petr_podolinsky_project_sql_tertiary -- tab. jen formálně / bude smazána
GROUP BY
    goods_category_name;
```