# Průvodní listina k SQL projektu – Marian Koutný

1. Výpis datových setů potřebných k projektu

Jako první jsem si vypsal všechny potřebné datové sety, které jsou pro projekt relevantní, abych si je postupně všechny prošel, osahal si je a viděl, jak jsou spolu vzájemně propojeny.

2. Vytvoření pomocných tabulek, pomoci kterých se dostaneme k první finální tabulce - t\_mk\_wage, t\_mk\_price, t\_mk\_price\_general

Vytvořil jsem si pomocnou tabulku t\_mk\_wage, která obsahuje průměrné roční mzdy v jednotlivých odvětvích od roku 2000 do roku 2021. Využil jsem k tomu tabulku czechia\_payroll spolu s tabulkou czechia\_payroll\_industry\_branch. Zprůměroval jsem mzdy z jednotlivých čtvrtletí skrze kód 5958 (tedy hodnotou mzdy) a napříč neprázdnými hodnotami industry\_branch\_code (u kterých bych nevěděl, k čemu mzda patří). Výsledek je 418 hodnot. (22 let naskrz 19 odvětvími)

Dále jsem si vytvořil tabulku t\_mk\_price, která obsahuje průměrné ceny jednotlivých potravin v daných letech a krajích. Použity byly sety czechia\_price\_category, czechia\_price a czechia\_region. Pro výpočet jsem používal datovou hodnotu date\_from a následně průměroval týdenní měření na celý rok přes neprázdné hodnoty regionu. Výsledek 4788 hodnot. (roky 2006-18 pro 14 krajů a 26 potravin, položka jakostní víno se objevuje pouze v letech 2015-18 -> celkem tedy 14\*26\*13 + 14\*4 = 4788 hodnot)

Poté jsem si vytvořil poslední tabulku t\_mk\_price\_general, kde jsem využil předchozí tabulku a jednotlivé ceny jsem zprůměroval pro všechny kraje. Ve výsledku tedy vznikla tabulka, která obsahuje ceny 26 potravin v letech 2006-18 a jakostního vína v letech 2015-18, celkem tedy 342 hodnot.

3. Vytvoření první finální tabulky t\_Marian\_Koutny\_project\_SQL\_primary\_final:

Nakonec jsem vytvořil konečnou tabulku t\_marian\_koutny\_project\_sql\_primary\_final , ke které jsem využil pomocné tabulky t\_mk\_wage a t\_mk\_price\_general. Zde jsem obě tabulky propojil na proměnné roku. Výsledkem je tabulka s 5 sloupci (odvětví, rok, příslušný průměrný plat v daném roce, název potraviny a její průměrná cena v daném roce – celkem 6669 hodnot. (26 potravin naskrz 19 odvětvími a 13 lety a jakostní víno naskrz 4 roky a 19 odvětvími a poté roky 2000-05 a 2019-21, tedy 9 let naskrz 19 odvětvími, které mají u potravin hodnotu NULL, jelikož cena není k dispozici -> suma sumárum 26\*19\*13 + 4\*19 + 9\*19 = 6669). Důležité pro zachování údajů o platech v letech 2000-05 a 2019-21 bylo použití LEFT JOIN. Nakonec proběhlo pár úprav tabulky.

4. Vytvoření pomocné tabulky pro sekundární tabulku projektu - tabulka t\_ec

Vytvořil jsem si pomocnou tabulku t\_ec, Využil jsem k tomu datovou sadu economies spolu se sadou countries. Ty jsem propojil na základě parametru country. Získal jsem tak tabulku s potřebnými údaji o jednotlivých zemích (název, GDP, gini, population, kontinent).

5. Vytvoření druhé finální tabulky t\_marian\_koutny\_project\_sql\_secondary\_final

Nakonec jsem vytvořil finální tabulku t\_marian\_koutny\_project\_sql\_secondary\_final za pomocí předchozí tabulky. Propojení bylo provedeno na základě country a roku. Vzniklá tabulka obsahuje data o 41 evropských zemích z let 2000 až 2021 (jsou-li k dispozici). Byla vyloučena závislá, okupovaná a další území, kde byl např. nedostatek dat. (Vatikán, Svalbard, S. Irsko, Faerské o., Gibraltar, Lichtenštejnsko a Montenegro). Výsledná tabulka obsahuje údaje o zemi, GDP, jeho vývoji, populaci a gini koeficientu. Tabulka obsahuje 818 rádků.

Jako poslední krok jsem odstranil všechny pomocné tabulky a zobrazil si obě finální tabulky k projektu.[[1]](#footnote-1)

# Část druhá – výzkumné otázky pro analytické oddělení

1. ROSTOU V PRŮBĚHU LET MZDY VE VŠECH ODVĚTVÍCH, NEBO V NĚKTERÝCH KLESAJÍ?

1.1

1. Pozn. autora: Je mi jasné, že existují určitě elegantnější řešení než tvorba pomocných tabulek, nicméně tento postup nebyl nikde zakázán, tudíž mi připadá korektní. Na konci postupu jsou výsledkem požadované 2 tabulky a všechny pomocné tabulky jsou odstraněny. [↑](#footnote-ref-1)