# Cíl projektu SQL

Cílem projektu bylo dosáhnout zobrazení tabulky s požadovanými daty jedním selektem.

## Použité tabulky

Countries, economies, life\_expectancy, religions, covid19\_basic\_differences, covid19\_testing, weather, lookup\_table, religions\_by\_percentage\_share.

# Postup

## příprava

Datasety (tabulky) z databáze “data.engeto.com” jsem si postahoval k sobě ve formě csv souborů, protože mi za žádnou cenu nešel udělat db dump.

Tabulku **countries** jsem očistil od duplicit a nesmyslů typu „moje\_zeme“…

Poté jsem je naimportoval k na svou lokální mariadb databázi. A vytvořil tak lokální kopie těchto tabulek.

## První část zadání – Časové proměnné

Pro první část zadání jsem vytvořil jednoduchý select z tabulky **covid19\_basic\_differences**.

Na githubu je tento script v souboru: **CasovePromenne.sql**

## Druhá část zadání – Proměnné specifické pro daný stát.

Od některého ze spolužáků jsem okoukl vnořený select a tak jsem doslova „spatlal“ vnořený select s tvrdým joinem tabulek **countries** a **covid19\_basic\_differences**.

Totálně jsem se zasekl na náboženstvích neb mi ta tabulka přišla nějak poškozená. Hlavně výpočty poměrů jednotlivých náboženství byly podivné. Např. více než 100 procent.

Problém byl v různých hodnotách polí „populace“ u tabulek **countries** a **religions**.

Kontrolou za pomoci strejdy googla jsem zjistil, že špatně to asi bude u **religions**. Tak jsem si na:

<https://datahub.io/sagargg/world-religion-projections#resource-by_number_of_population>

stáhl podobnou tabulku a kupodivu to tam bylo úplně stejně. Nicméně součástí tabulky už byly i procentuální poměry jednotlivých náboženství, tak jsem tuto tabulku přidal do své lokální db jako

**religions\_by\_percentage\_share.** Apoužil ji přímo v selektu**.** Na githubu je tento script v souboru : **PromenneSpecifickeProDanyStat.sql**.

## Třetí část a konečný script.

Třetí část už jsem z časových důvodů rovnou vložil do konečného scriptu. Průměrnou teplotu jsem udělal roční doby, protože mi to přišlo lepší. Jinak tady už jsem se v tom tedy „plácal“ opravdu značně. Query finálového scriptu běželo klidně i 17 minut. Vytvořit tabulku z finálního skriptu se mi nedařilo. Chybové hlášení naznačovalo problém s typem dat (truncated incorrect DOUBLE value) teploty ale ani po přecastování a přeformátování proměnných kdy script už zmíněný error nehlásil mi create table stejně běžel celou noc, ale tabulka se nevytvořila.

Podařilo se mi vytvořit alespoň view **vi\_wolf\_final\_sql** kde lze jedním selectem získat požadovaná data i když se u toho jeden načeká.

# Závěr

Zadání s odřenýma ušima snad splněno ale chápu, že v reálném světě by mne s tím pan šéf asi vyhodil. To čekání až query proběhne je fakt dlouhé.

Příště bych už tolik vnořených selektů nedělal a snažil se použít jiný způsob.