

SQL PROJEKT ENGETO

AUTOR: Jiří Valášek

1) Základní data

Základem pro výslednou tabulku budou záznamy o denních nárůstech nakažených, provedených testech a počtem obyvatel daného státu. Údaje o denních nárůstech nakažených získám z tabulky covid19_basic_differences, která sleduje nárůsty nakažených za každý den od začátku pandemie. Na tyto data navážu záznamy o provedených testech. Záznamy o testech nejsou ve všech zemích dostupné. Zároveň tyto záznamy nejsou dostupné pro každý den záznamu o počtu nakažených. Názvy zemí z obou tabulek bude potřeba upravit tak, aby se výsledná data navázala podle klíčů země a data.

V záznamech tabulky covid19_basic_differences chybí tyto velké země: Austrálie, Čína, Kanada.

2) Dětská úmrtnost

Rok záznamu o dětské úmrtnosti se liší dle země. Zároveň ne všude jsou tyto údaje dostupné. Rozhodl jsem se použít poslední dostupný údaj.

3) HDP na obyvatele

Záznamy o HDP jsou vidět v tabulce economies. Poslední dostupné z roku 2020. Rozhodl jsem se použít údaje i z roku 2019 u zemí, kde chybí za rok 2020.

4) GINI

Koeficient GINI není u všech zemí dostupný a u některých zemí je starý. Použil jsem poslední dostupné údaje a zároveň ne starší, než 10 let pro jednotlivé země.

5) Population density

Tento údaj vypočítám pouze u zemí, kde je vidět záznam o celkové populaci i rozloze. V tabulce countries u některých zemí je sice vypočtená hodnota populační hustoty, ale chybí některý ze záznamů o celkové populaci, nebo rozloze.

6) Podíly náboženství

Vypočítávám z tabulky religions.

7) Rozdíl očekávané délce dožití

Vypočítávám z tabulky life_expectancy.

8) Údaje o počasí

Všechny požadované údaje vychází z tabulky weather, kde chybí název země. Ten doplním podle hlavního města, kde jsou záznamy o počasí. Záznamy v tabulce weather jsou textové. Potřebuji je převést na integer. Textové záznamy před převodem na integer očistit o textové znaky. Údaje o počasí jsou jen pro evropské země.

9) Časové proměnné

Záznamy o víkendových dnech a roční období doplním rovnou do základní tabulky se záznamy o denních přírůstcích a testech.

U všech tabulek, které navazují na základní tabulku `t_covid_confirmed_tests_cas` zkontroluji názvy zemí a upravím tak, aby se podle tohoto klíče data spojila.

Po vytvoření finální tabulky jsem filtroval země, kde chybí údaje a zjistil, že v primárních datech jsou ještě nějaké rozdíly v názvech zemí. To jsem v primárních tabulkách upravil a finální tabulku znovu vytvořil.