Engeto – Popis SQL projektu

Študent: Mariana Cagáňová

Pre zjednodušenie som si vytvorila 3 pohľady:

* v\_MC\_weather\_year – ktorý obsahuje údaje o počasí v roku 2020 a 2021 a upravuje hodnoty o počasí a nárazoch vetra aby sa s tým dalo pracovať ako s číselnou hodnotou.
* v\_MC\_weather – ktorý obsahuje údaje a medzivýpočty potrebné pre počasie. Počet hodín kedy pršalo je zrátaný počet hodín kedy boli zrážky nenulové keďže inak sa nedalo s presnosťou určiť koľko hodín medzi jednotlivými údajmi pršalo. Aby sa dáta správne napárovali bolo potrebné zmeniť typ Date z Datetime na Date a upraviť názvy krajín a miest z tabuľky countries a weather, ktoré sa používali na spojenie s tabuľkou countries a neskôr CBD.
* v\_MC\_rel\_per\_country – obsahuje podiely náboženstiev jednotlivých krajín za rok 2020
* v\_MC\_life\_expectancy\_diff – obsahuje údaje o rozdieloch medzi life expectancy v rokoch 2015 a 1965

Ako hlavnú tabuľku som použila covid19\_basic\_differencies (ďalej ako CBD), takže všetky ostatné dáta sú napojené na tieto údaje. Keďže názvy krajín medzi jednotlivými tabuľkami sa mierne líšia brala som ako hlavné názvy, názvy krajín z CBD tabuľky a s tými pracovala ďalej. Nie všetky krajiny sa ale nachádzali v potrebných tabuľkách takže údaje k 4-5 záznamom chýbajú. Taktiež tabuľka CBD obsahovala údaje o výletných lodiach, takže tieto záznamy neobsahujú údaje o počasí, krajinách atď.

Údaje v tabuľke CBD nie sú časovo obmedzené keďže obsahujú len údaje z roku 2020 a 2021 a pri rozhodnutí použiť iba rok 2020 je možné tieto údaje neskôr jednoducho vyfiltrovať. Mierne obmedzenie by to mohlo spôsobiť pri údajoch o štátoch alebo náboženstvách.

Ďalej je vo veľa krajinách chýbajúci údaj o GINI indexe. Pre určenie úmrtností detí som použila záznamy z roku 2019, keďže údaje z roku 2020 neboli dostupné.