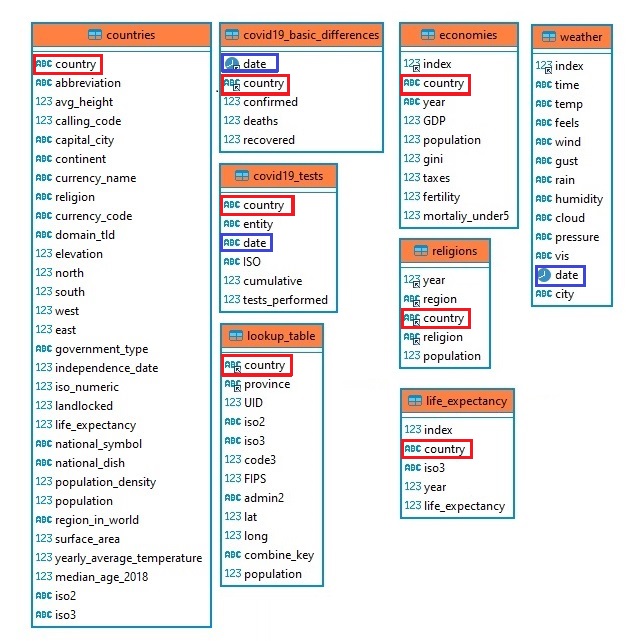
**DATOVÁ AKADEMIE 28.04.2021 – SQL PROJEKT**

**Alena Vítková**

*Zjednodušené zadání:* Zjistit faktory, které ovlivňují rychlost šíření koronaviru v jednotlivých státech.

Statistik bude vyhodnocovat model, který vysvětluje denní nárusty nakažených (**confirmed**). Je potřeba vzít v úvahu i počet provedených testů (**tests\_performed**) a počet obyvatel státu (**population**).

*Klíče****:* country**, **date**

1. Zobrazila jsem si všechny tabulky, abych věděla, co obsahují.

*Sloupce, které mají být ve výsledné tabulce:*

1. week (víkend = 0, prac. den = 1) (z covid19\_differences)
2. seasons\_of\_the\_year (jaro = 0, léto = 1, podzim = 2, zima = 3) (z covid19\_differences)
3. population\_density (v countries)
4. GDP (v economies)
5. GINI (v economies)
6. mortality\_under5 (v economies) …. v názvu je překlep mortaliy
7. median\_age\_2018 (v countries)
8. podíly náboženství v daných státech… procentní podíl příslušníků na celkovém počtu obyvatel (z religions)
9. rozdíl mezi očekávanou dobou dožití v roce 1965 a 2005 (z life\_expectancy)
10. průměrná denní (ne noční!) teplota (z weather)
11. počet hodin v daném dni, kdy byly srážky nenulové (z weather)
12. maximální síla větru v nárazech během dne (z weathers)

+ country, date, population, confirmed, test\_performed

*Poznatky:*

V tabulce weather jsou města, ne státy, v countries jsou státy i hlavní města.. potřeba nějak propojit a přiřadit výsledky ke státům. Měst je ale jen 34 a ještě některé napsané jinak než v tabulce countries, takže se nepropojí se státem. Výsledky týkající se počasí budou tedy dosti omezené.

Datumy se v tabulkách liší: covid19\_basic\_differences (2020-01-22 - 2021-05-03)

covid19\_tests (2020-01-01 – 2020-11-24)

weather (2016-05-01 – 2021-04-30)

V tabulce weather neodpovídá čas v date (datetime) s časem v time…pro úkol nevadí, ale date bych změnila na typ date, čas je tu nesmyslný.

K úkolu 10: Je potřeba se domluvit, co znamená denní teplota, lze ji počítat různě (7+14+2\*24)/4 ; 6hod-22hod, svítání až západ slunce). Pro úkol jsem počítala průměrnou denní teplotu (6+9+12+15+18+21)/6

*Postupy k jednotlivým úkolům (sloupcům):*

1. Vytvořila jsem flag week, kde jsou rozdělené datumy na weekend(0) a pracovní den(1). Z tabulky covid19\_basic\_differences .
2. Vytvořila jsem flag season, kde jsou rozdělené datumy na jaro (0) , léto (1), podzim (2), zima (3). Z tabulky covid19\_basic\_differences .

Spojila jsem do view **v\_base** klíčová data country, date, confirmed a získané flagy week a season, k tomu připojila sloupec capital\_city (z tabulky countries, pro pozdější spojení se city z tabulky weather).

1. Úkoly 3 – 7:

population\_density, median\_age\_2018, country a population vypíši z tabulky countries

GDP a mortality\_under5 …vypisuji nejaktuálnější hodnotu, tedy z roku 2019

Gini vypisuji z roku, který je nejaktuálnější, není stejný pro všechny země:

**SELECT** country, gini, **max**(**year**) **FROM** economies e **WHERE** gini **IS** **NOT** **NULL**

**GROUP** **by** country

Spojila jsem vše do view **v\_34567**.

1. Úkol č.8:

View **v\_8.**

1. Úkol č.9:

View **v\_9.**

1. Úkol č.10-12:

Pohled view **v\_10**, kde se vypíší hodnoty z úkolů 10-12.

Výsledná finální tabulka **t\_alena\_vitkova\_projekt\_SQL\_final** je výsledkem spojení všech předchozích (v\_base as v1, v\_34567 as v2, v\_8 as v8, v\_9 as v9, v\_10 as v) plus sloupec tests\_performed z tabulky covid19\_tests.