**Průvodní listina pro SQL**

Jako první věc před tvorbou skriptu, jsem si prohlédl všechny tabulky a pokusil se pochopit, co reprezentují, jak s nimi pracovat a jaký může nastat problém při práci s nimi. Další krok byl, vytvořit **v\_covid19\_basic\_differences\_new** a **v\_countries\_new**, kde jsem upravil názvy Czechia na Czech Republic a Praha na Prague abych měl uniformní data..

Práci jsem si rozhodl rozdělit do ‚sub-views‘ podle využití tabulek.

* **v\_engeto\_projekt\_casove\_promene vepcp**
* **v\_engeto\_projekt\_tests**
* **v\_engeto\_projekt\_economie**
* **v\_engeto\_projekt\_life\_expectancy veple**
* **v\_engeto\_projekt\_nabozenstvi**
* **v\_engeto\_projekt\_weather**

Nejprve jsem vytvořil jakousi kostru, na kterou budu moct posléze ‚přilepovat‘ další tabulky.  
Začal jsem tedy s tabulkou **covid\_basic\_differences** ke které jsem připojil data o počasí s pomocí funkce *CASE WHEN* .

Jako další tabulka byla **v\_engeto\_projekt\_tests**, ze které jsem vybral pouze 3 relevantní sloupce   
a zajistil abych neměl NULLové hodnoty.

Následující tabulka **v\_engeto\_projekt\_economie** byla o něco složitější, musel jsem si pomocí funkce *WITH* vytvořit dvě tabulky. V první z nich jsem vytáhl data například: median\_age\_2018, population\_density, population.  
Ve druhé tabulce jsem si vytvořil sloupec last\_record. To proto že v tabulce spousta informací chyběla a nešlo vybrat jeden konkrétní rok, a proto jsem našel pro každou zemi, rok posledního zápisu.

Největší výzvou pro mě byl **v\_engeto\_projekt\_nabozenstvi** na kterém jsem strávil suverénně nejvíce času z celého projektu.

Problémy:

* nedařilo se mi dostat data z jednoho sloupce do vícero sloupců
* když jsem se pokoušel vzít údaje o populaci v dané zemi z tabulky countries, tak mi neseděl počet se součtem všech věřících (pravděpodobně vinnou zaokrouhlování a odhadování)
* dlouhý zápis funkce *WITH* znamenal velký prostor pro chybějící znaménko..

Nakonec se mi povedlo vytvořit sloupce pro každé náboženství z tabulky a odstranit všechny NULLové hodnoty.

Jsem si celkem jistý, že existuje mnohem jednodušší a elegantnější způsob, ale na nic lepšího jsem nepřišel.

Tabulka **v\_engeto\_projekt\_life\_expectancy** byla o poznání jednodušší, opět pomocí funkce *WITH* jsemvytvořil tabulku pro rok 1965 a pro tok 2015 a spojil je do jedné, ze které jsem si nechal vypočíst rozdíl mezi těmito roky pro každou zemi.

Druhým oříškem bylo vytvoření **v\_engeto\_projekt\_weather** které jsem se rozhodl rozdělit na 3 podskupiny.   
Samotné vytváření hledaných dat nebyl takový problém jako zjištění, že při SELECTu fungují, ale při vytvoření TABULKy už ne.

Kromě agregačních funkcí jsme musel zapojit funkce CAST a REPLACE.

Všechny tři pod-tabulky jsem spojit do jedné, pro lepší pozdější práci.

A jako velké finále přišlo spojit všechny zmíněné views do jedné finální tabulky.  
 Šel jsem na to již už známou funkcí WITH , kde jsem spojit všechny vše spojil do jedné tabulky a poté si SELECTem vybral jsem sloupce, které potřebuji.

Ve výsledné tabulce je spousta NULLových hodnot, což je způsobenou použitím LEFT JOINu a také tím, že nám některá data chybí.

Například:

1. sloupec tested\_performed, kde například nemáme žádná data pro zemi Afghanistán
2. ekonomické údaje například GDP, gini a pod také nejsou uvedeny pro spoustu zemí
3. pro data z tabulky weather bylo třeba spojit hlavní města se zeměmi, některá hlavní města ale chybí, a proto nebylo možné uvést data pro danou zemi.(Například nejsou data pro Kabul a proto nejsou data ani pro Afghanistan)