# Postup

## base\_1

Tabulku *covid19\_basic\_differences* jsem použil jako kostru pro vytvoření finální tabulky. Začal jsem s dotazem, který vytvořil tabulku base\_1. V tabulce base\_1 jsem napojil sloupce tests\_performed a population.

Pomocí funkce EXCEPT jsem si zjistil, které země se nachází v tabulce *covid19\_tests* pod jiným názvem než v tabulce *covid19\_basic\_differences.* Následně jsem pomocí CASE WHEN upravil názvy z tabulky *covid19\_basic\_differences*, tak aby byly stejné jako v tabulce *covid19\_tests*. Následně jsem si vybral sloupce, které jsem potřeboval a udělal LEFT JOIN.

Dále jsem pokračoval v dotazu s napojením sloupce population. Jelikož budu následně dělat úkol, kde budu počítat jednotlivé podíly náboženství v daných zemí, tak se mi hodí vzít data pomocí SUM a GROUP BY z roku 2020 z tabulky *religions.* Kdybych vzal hodnoty například z *countries* nebo *lookup\_table*, tak by se mi potom stalo, že někde budu mít například podíl vyznavačů Islamu vyšší, než je celkový počet obyvatel v zemi, což nedává smysl.

## base\_2

Napojení časových proměnných na tabulku base\_1.

Pomocí CASE WHEN a WEEKDAY jsem označil každý záznam podle datumu ‘Pracovni den’ nebo ‘Vikend’.

Pomocí CASE WHEN a MONTH jsem označil každý záznam podle datumu od 0 do 3.  
0 je jaro, 1 je léto, 2 je podzim a 3 je zima.

## base\_3

Napojení sloupce hustota\_zalidneni. Udělal jsem to pomocí vlastního výpočtu, protože jsem si celkovou populaci vytáhl z tabulky *religions*, tak aby to odpovídalo, populaci z religion, kterou jsem udělal v base\_1 jsem vydělil surface\_area z tabulky *countries*.

## base\_4

Napojení sloupce HDP\_na\_obyvatele. Vzal jsem HDP z let 2019 a 2020 z tabulky *economies*, které jsem následně vydělil sloupcem population z mojí tabulky base\_3. Vzal jsem HDP ze dvou ročníků úmyslně z důvodu, že pro rok 2020 v mojí tabulce jsem použil k výpočtu HDP z roku 2019 a pro rok 2021 HDP z roku 2020.

## base\_5

Napojení sloupce gini\_koeficient. V tomto sloupci chybí velké množství hodnot, některé země v tomto sloupci nemají žádné hodnoty nebo je mají několik let zpětně. Zvolil jsem tedy postup aby koeficient alespoň orientačně ukázat hodnotu gini, tím, že jsem pro každou zemi hodnoty gini zprůměroval. Je to opravdu pouze orientační, protože hodnoty v tomto sloupci v tabulce *economies* jsou uvedeny pouze několik let zpětně nebo vůbec.

## base\_6

Přidal jsem dva sloupce detska\_umrtnost a median\_age\_2018. Pro detskou úmrtnost jsem použil nejaktuálnější informace z roku 2019 a sloupec median\_age\_2018 jsem jednoduše napojil na moji tabulku base\_5.

## base\_7 – 14

Vytváření tabulek base\_7 až base\_14. Pro každé náboženství jsem vydělil počet vyznavačů s populací. Každý krok (base\_x) přidal jeden sloupec s podílem náboženství v dané zemi.

## base\_15

Přidání sloupce diff\_life\_expectancy. Pomocí vnořeného SELECTu ve vnořeném SELECTu a pomocí funkce LAG jsem vytvořil sloupec, který ukazuje rozdíl v očekávané délce života mezi rokem 2015 a 1965

## weather2

Jedná se o upravenou tabulku *weather.* Bylo potřeba upravit pomocí REPLACE aby ve sloupcích byly pouze čísla. Následně čísla převést na číselný datový typ. Potom jsem musel pomocí LEFT JOIN přidělit země k záznamům přes CASE WHEN a jednotlivá města. Teď bude tabulka sloužit jako pomocná tabulka pro poslední 3 úkoly.

## base\_16

Pomocí tabulky weather2 jsem připojil k moji tabulce base\_15 sloupec prum\_den\_tep. Úkol byl pouze denní teplota, nikoliv noční. V tabulce jsou data po 3h intervalech. Vzal jsem tedy dobu od 6:00 do 21:00 (`time` = 18:00). Tuhle dobu považuji za denní. Jinak jsem pomocí AVG, GROUP BY a následné podmínky vymezující čas vytvořil sloupec prum\_den\_tep.

## base\_17

Připojení sloupce počet\_h\_deste. Pomocí nejnovější verze mojí základní tabulky a pomocné tabulky weather2 jsem připojil další sloupec. V pomocné tabulce jsem měl připravený sloupec počet\_hodin\_deste, tento sloupec měl číslo 3 pokud byla srážka nenulová v časovém 3h intervalu. Pomocí SUM, GROUP BY jsem vytvořil nový sloupec, který ukazoval počet hodin deště v daném dni u dané země. Zjistil jsem, že někde pršelo více než 24h za den. Našel jsem záznamy, které byly asi duplikované, země měla stejný záznam ve stejný den 2x. Vyřešil jsem to pomocí toho, že jsem SUMu vydělil 2 a hodnoty v tu chvíli dávaly smysl. Nikde nepršelo více jak 24h za den.

## t\_Filip\_Sebesta\_projekt\_SQL\_final

Napojení posledního sloupce bylo jednodušší než dva předchozí, použil jsem akorát agregační funkci MAX společně s GROUP BY na sloupec gust. Následně jsem pomocí JOINU připojil data k jednotlivým záznamům.