Dokumentace

Úkol jsem začala zpracovávat vytvoření selektů pro zadané proměnné. Po vytvoření jednotlivých selektů (viz dokument níže) jsem začali z daných selektů vytvářet view.

Nejprve jsem vybrala a spojila údaje z tabulky countries a economies do view (v\_countries\_economies). Následně jsem vytvořila spojení view covid19\_basic\_differences a covid19\_tests (v\_cdif\_ctests). Uvedená view jsem spojila do view v\_vce\_vcc.

Dále jsem vytvořila view s rozdílem mezi očekávanou délkou dožití (v\_life\_expectancy\_1965\_2015). Vypracované body z úkolu časové proměnné jsem spojila s view v\_vce\_vcc (v\_variables\_23).

Poté jsem vytvořila view pro úkoly vztahující se k počasí v\_weather\_variable.

Bohužel jsem podcenila následné spojování tabulek a jejich různorodých dat a nechala to na poslední chvíli. V ten moment jsem zjistila, že mi na localhost nefunguje vytvoření view, ale jen table, což se opravilo až provedením aktualizace DBeaver.Díky tomu výsledná tabulka nesplňuje moje očekávání, ale asi je lepší odevzdat něco než nic.

Části které bych upravila:

* Populaci bych dotahovala z tabulky countries (z economies jsou starší data)
* Pro náboženství bych vytvořila jednotlivé sloupce – takto to nedává smysl
* Gini a GDP bych použila z roku 2019, jelikož k roku 2020 nejsou ještě data

Napadají Vás ještě nějaké další proměnné, které bychom mohli použít? Pokud vím, měl(a) byste si vystačit s daty z následujících tabulek: countries, economies, life\_expectancy, religions, covid19\_basic\_differences, covid19\_testing, weather, lookup\_table.

Výsledná data budou panelová, klíče budou stát (country) a den (date). Budu vyhodnocovat model, který bude vysvětlovat denní nárůsty nakažených v jednotlivých zemích. Samotné počty nakažených mi nicméně nejsou nic platné - je potřeba vzít v úvahu také počty provedených testů a počet obyvatel daného státu. Z těchto tří proměnných je potom možné vytvořit vhodnou vysvětlovanou proměnnou. Denní počty nakažených chci vysvětlovat pomocí proměnných několika typů. Každý sloupec v tabulce bude představovat jednu proměnnou. Chceme získat následující sloupce:

1. Časové proměnné
   * binární proměnná pro víkend / pracovní den

**SELECT** c.country, c.**date**,

**case** **when** **WEEKDAY**(c.**date**) **in** (5, 6) **then** 1 **else** 0 **end** **as** weekend

**FROM** covid19\_basic\_differences c

**JOIN** lookup\_table lt

**on** c.country = lt.country

**and** **month**(c.**date**) = 11

**ORDER** **BY** c.**date** **desc**

* + roční období daného dne (zakódujte prosím jako 0 až 3)

**CASE** **WHEN** vcc.**date** <= '2020-03-19' **THEN** 3

**WHEN** vcc.**date** >= '2020-03-20' **AND** vcc.**date** <= '2020-06-19' **THEN** 0

**WHEN** vcc.**date** >= '2020-06-20' **AND** vcc.**date** <= '2020-09-21' **THEN** 1

**WHEN** vcc.**date** >= '2020-09-22' **AND** vcc.**date** <= '2020-12-20' **THEN** 2

**WHEN** vcc.**date** >= '2020-12-21' **AND** vcc.**date** <= '2021-03-19' **THEN** 3

**ELSE** 'error'

**END** **AS** season,

1. Proměnné specifické pro daný stát
   * hustota zalidnění - ve státech s vyšší hustotou zalidnění se nákaza může šířit rychleji

**SELECT**

country,

population\_density

**FROM** countries c

**WHERE** population\_density **IS** **NOT** **NULL**

**ORDER** **BY** population\_density

* + HDP na obyvatele - použijeme jako indikátor ekonomické vyspělosti státu
  + [GINI koeficient](https://en.wikipedia.org/wiki/Gini_coefficient) - má majetková nerovnost vliv na šíření koronaviru?
  + dětská úmrtnost - použijeme jako indikátor kvality zdravotnictví
  + medián věku obyvatel v roce 2018 - státy se starším obyvatelstvem mohou být postiženy více

**SELECT**

country,

gini,

gdp,

median\_age\_2018

mortaliy\_under5

**FROM** economies e

* + podíly jednotlivých náboženství - použijeme jako proxy proměnnou pro kulturní specifika. Pro každé náboženství v daném státě bych chtěl procentní podíl jeho příslušníků na celkovém obyvatelstvu

**SELECT** c.country, r.religion, c.population\_density , **ROUND** (r.population / c.population \*100, 2) **AS** share\_religion

**FROM** countries c

**JOIN** religions r

**ON** c.country = r.country

* + rozdíl mezi očekávanou dobou dožití v roce 1965 a v roce 2015 - státy, ve kterých proběhl rychlý rozvoj mohou reagovat jinak než země, které jsou vyspělé už delší dobu

**SELECT** a.country, a.life\_exp\_1965 , b.life\_exp\_2015,

**round**( b.life\_exp\_2015 - a.life\_exp\_1965, 2 ) **as** life\_exp\_ growth

**FROM** (

**SELECT** le.country , le.life\_expectancy **as** life\_exp\_1965

**FROM** life\_expectancy le

**WHERE** **year** = '1965'

) a **JOIN** (

**SELECT** le.country , le.life\_expectancy **as** life\_exp\_2015

**FROM** life\_expectancy le

**WHERE** **year** = '2015'

) b

**ON** a.country = b.country

;

1. Počasí (ovlivňuje chování lidí a také schopnost šíření viru)
   * průměrná denní (nikoli noční!) teplota
   * počet hodin v daném dni, kdy byly srážky nenulové
   * maximální síla větru v nárazech během dne

**SELECT**

city,

**replace** (**date**, '00:00:00', '') **as** **date**,

**COUNT**(**time**) \*3 **AS** rain\_hours,

(**SELECT** **AVG**(temp) **AS** avegera\_temp

**FROM** weather w

**where** **time** **IN** ('06:00', '09:00', '12:00', '15:00', '18:00', '21:00') ) **as** avg\_tep,

**MAX** (gust) **AS** max\_wind

**FROM** weather w

**WHERE** city **IS** **NOT** **NULL** **AND** rain != '0.0 mm'

**GROUP** **BY** city, **date**