**Zadání:** Od Vašeho kolegy statistika jste obdrželi následující email:

##########

Dobrý den,

snažím se určit faktory, které ovlivňují rychlost šíření koronaviru na úrovni jednotlivých států. Chtěl bych Vás, coby datového analytika, požádat o pomoc s přípravou dat, která potom budu statisticky zpracovávat. Prosím Vás o dodání dat podle požadavků sepsaných níže.

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
   * roční období daného dne (zakódujte prosím jako 0 až 3)
2. Proměnné specifické pro daný stát
   * hustota zalidnění - ve státech s vyšší hustotou zalidnění se nákaza může šířit rychleji
   * 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
   * 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
   * 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
3. 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

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.

V případě nejasností se mě určitě zeptejte.

S pozdravem, Student (a.k.a. William Gosset)

###############################

**Výstup:** Pomozte Vašemu kolegovi s daným úkolem. Výstupem by měla být tabulka na databázi, ze které se požadovaná data dají získat jedním selectem. Tabulku pojmenujte t\_{jméno}\_{příjmení}\_projekt\_SQL\_final. Na svém GitHub účtu vytvořte repozitář (může být soukromý), kam uložíte všechny informace k projektu - hlavně SQL skript generující výslednou tabulku, popis mezivýsledků, informace o výstupních datech (například kde chybí hodnoty apod.). Případné pomocné tabulky neukládejte na DB jako view! Vždy vytvořte novou tabulku (z důvodu anonymity).