Konsumpcja Danych

1. Konfiguracja VM

Konfiguracja

https://console.cloud.google.com/appengine/versions?serviceId=projekt&project=still -primer-271314&cloudshell=true&supportedpurview=project

Preview https://projekt-dot-still-primer-271314.ey.r.appspot.com/

2. Wprowadzanie danych / pobieranie danych z zewnątrz

Dane których używam są znormalizowanym zbiorem ocen. Normalizacja przebiega wg algorytmu organizatorów ankiety więc nie wiem jak dokładnie to liczą.

Jednakże znalazłem informacje o metodyce przeprowadzania badań i na jej podstawie (ocena 6 czynników i ogólnej satysfakcji w skali 1-10) stworzyłem formularz do zbierania danych ,które będzie można wysłać gdy organizatorzy zaczną pracę na edycją na rok 2020.

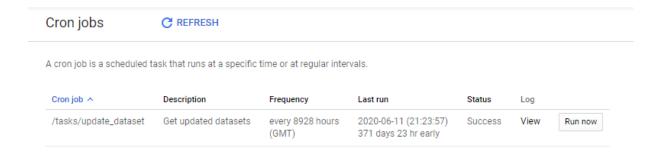


Rysunek 2.1 Formularz do wprowadzania danych

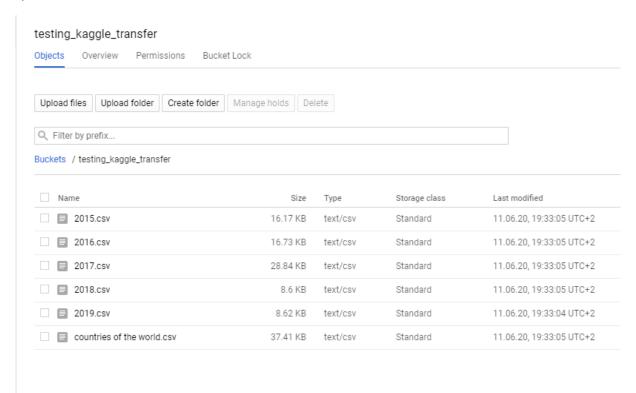
3. Integracja z gcs

Dataset o rankingu szczęśliwości jest update'owany w miare regularnie co rok około listopada / grudnia. Dataset o info o krajach został tylko raz wrzucony 2 lata temu i nikt od tego czasu go nie ruszał.

Ustawiłem więc cron'a ,który zaciąga dataset'y co 8928 godzin czyli rok gdyby każdy miesiąc miał po 31 dni.



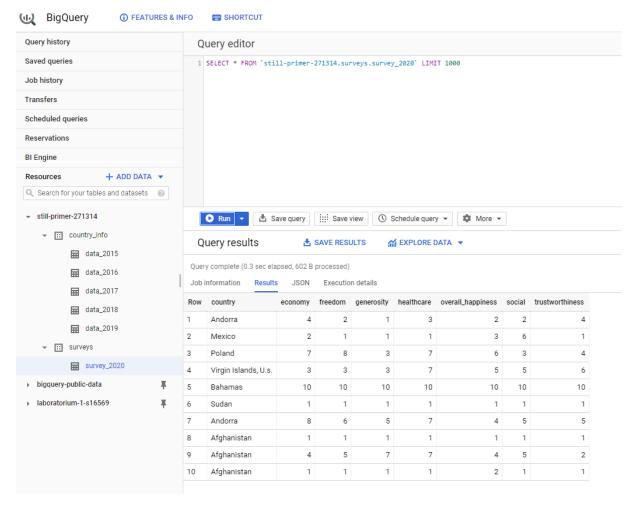
Rysunek 3.1 Cron



Rysunek 3.2 Zaciągniete przez cron'a dane

4. Integracja z bq

Formularz z punktu 2 od razu wrzuca zwalidowane dane do bq.



Rysunek 4.1 Kilka przykładowych row'ów wrzuconych poprzez wypełnienie formularza