

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.ev.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 edycję 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.

Cron jobs

 REFRESH

A cron job is a scheduled task that runs at a specific time or at regular intervals.


Cron job ^	Description	Frequency	Last run	Status	Log
/tasks/update_dataset	Get updated datasets	every 8928 hours (GMT)	2020-06-11 (21:23:57) 371 days 23 hr early	Success	View Run now

Rysunek 3.1 Cron







testing_kaggle_transfer

[Objects](#) [Overview](#) [Permissions](#) [Bucket Lock](#)

[Upload files](#) [Upload folder](#) [Create folder](#) [Manage holds](#) [Delete](#)

 Filter by prefix...

[Buckets](#) / testing_kaggle_transfer

<input type="checkbox"/>	Name	Size	Type	Storage class	Last modified
<input type="checkbox"/>	 2015.csv	16.17 KB	text/csv	Standard	11.06.20, 19:33:05 UTC+2
<input type="checkbox"/>	 2016.csv	16.73 KB	text/csv	Standard	11.06.20, 19:33:05 UTC+2
<input type="checkbox"/>	 2017.csv	28.84 KB	text/csv	Standard	11.06.20, 19:33:05 UTC+2
<input type="checkbox"/>	 2018.csv	8.6 KB	text/csv	Standard	11.06.20, 19:33:05 UTC+2
<input type="checkbox"/>	 2019.csv	8.62 KB	text/csv	Standard	11.06.20, 19:33:04 UTC+2
<input type="checkbox"/>	 countries of the world.csv	37.41 KB	text/csv	Standard	11.06.20, 19:33:05 UTC+2

Rysunek 3.2 Zacięgnięte przez cron'a dane

4. Integracja z bq

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

[FEATURES & INFO](#)
[SHORTCUT](#)

Query history

Saved queries

Job history

Transfers

Scheduled queries

Reservations

BI Engine

Resources [+ ADD DATA](#)

Search for your tables and datasets

still-primer-271314

country_info

data_2015

data_2016

data_2017

data_2018

data_2019

surveys

survey_2020

bigquery-public-data

laboratorium-1-s16569

Query editor

```
1 SELECT * FROM `still-primer-271314.surveys.survey_2020` LIMIT 1000
```

Run

Save query

Save view

Schedule query

More

Query results

SAVE RESULTS

EXPLORE DATA

Query complete (0.3 sec elapsed, 602 B processed)

Job information

Results

JSON

Execution details

Row	country	economy	freedom	generosity	healthcare	overall_happiness	social	trustworthiness
1	Andorra	4	2	1	3	2	2	4
2	Mexico	2	1	1	1	3	6	1
3	Poland	7	8	3	7	6	3	4
4	Virgin Islands, U.s.	3	3	3	7	5	5	6
5	Bahamas	10	10	10	10	10	10	10
6	Sudan	1	1	1	1	1	1	1
7	Andorra	8	6	5	7	4	5	5
8	Afghanistan	1	1	1	1	1	1	1
9	Afghanistan	4	5	7	7	4	5	2
10	Afghanistan	1	1	1	1	2	1	1

Rysunek 4.1 Kilka przykładowych row'ów wrzuconych poprzez wypełnienie formularza