



PRACA KOŃCOWA

zmiana klimatu (temperatury) vs emisja CO₂

Big Data - Inżynieria danych 2021/2022

grupa projektowa #3

Piotr Barczak

Przemysław Becella

Maciej Hercka

Tomasz Kieliszak

spis treści

- 1) założenia pracy końcowej
- 2) temat pracy
- 3) źródła danych
- 4) środowisko pracy
- 5) proces ETL – opis
- 6) dashboard – wizualizacja danych
- 7) trudności/problemy
- 8) wnioski końcowe

założenia pracy końcowej

☐ przedstawienie holistyczne procesu przetwarzania danych:

- uzasadnienie doboru narzędzi do postawionych celów przetwarzania danych, zarówno w warstwie prezentacji, jak i sposobu składowania i agregowania danych źródłowych
- uzasadnienia dobranych mechanizmów normalizacji i agregacji danych
- prezentacja zagadnień w formie dynamicznego panelu, który umożliwi zmianę kryteriów dla danych wejściowych

☐ zadania szczegółowe:

- pobranie danych źródłowych (min. 2 źródła) i załadowanie do własnej bazy (hurtowni)
- przygotowanie możliwości przeglądania danych źródłowych (MS Excel lub inne proste narzędzie)
- przygotowanie panelu prezentującego zaproponowane zagadnienia

temat pracy

- ☐ praca koncentrowała się na kwestii zmian klimatu reprezentowanych przez dane o wahaniach temperatury rok do roku na świecie
- ☐ dane dotyczące zmian temperatury stanowiły punkt wyjścia do dyskusji na temat potencjalnych czynników wpływających na m.in. ich występowanie, dynamikę, intensywność czy zróżnicowanie geograficzne i czasowe
- ☐ dane o zmianach klimatu zestawiono zatem z danymi o takich potencjalnych czynnikach wpływu, jak: emisja zanieczyszczeń, liczebność populacji, rozwój gospodarki światowej
- ☐ dobór źródeł danych, ich wzajemne powiązanie, przedstawienie graficzne i obliczenia statystyczne miały w założeniach doprowadzić do wyciągnięcia trafnych wniosków na temat tego, **co, w jaki sposób, gdzie i w jakiej skali wpływało na zmiany klimatu na świecie w ostatnich kilkudziesięciu latach**

źródła danych

zagadnienie	nazwa danych	źródło	typ pliku źródłowego
zmiany temperatury	wahania temperatury na świecie rok do roku (<i>Temperature change</i>)	fao.org/faostat/en/#data/ET	.CSV
emisja CO ₂	skumulowana emisja CO ₂ na świecie (<i>Annual CO₂ emissions</i>)	ourworldindata.org/grapher/annual-co2-emissions-per-country	.CSV
	emisja CO ₂ według sektorów na świecie (<i>Annual CO₂ emissions by sector</i>)	ourworldindata.org/emissions-by-sector	.CSV
PKB	skumulowane PKB na świecie (<i>GDP (constant 2015 US\$)</i>)	data.worldbank.org/indicator/NY.GDP.MKTP.KD	.CSV
	PKB <i>per capita</i> na świecie (<i>GDP per capita (constant 2015 US\$)</i>)	data.worldbank.org/indicator/NY.GDP.PCAP.KD	.CSV
populacja	skumulowana populacja na świecie (<i>Population, 10,000 BCE to 2021</i>)	ourworldindata.org/grapher/population	.CSV
emisja gazów cieplarnianych, ślad węglowy	environment transport, energy use, emmissions intensities	fao.org/faostat/en/#data/ET	.CSV

środowisko pracy

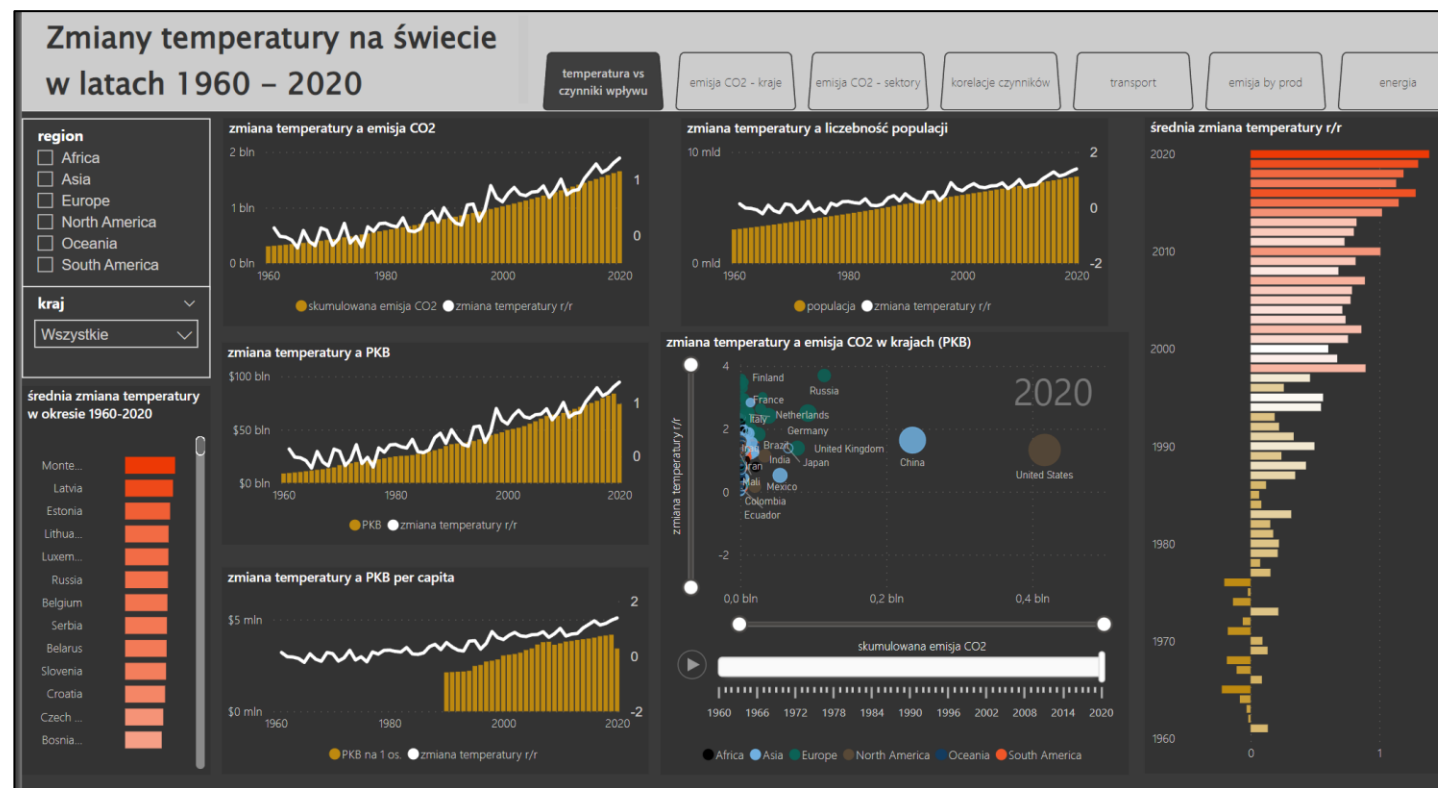
- ☐ Microsoft Excel (Power Query)
- ☐ Serwer NAS (Synology DS718+)
 - ☐ stworzenie współdzielonej przestrzeni dyskowej (Synology Drive) dla całego zespołu + utworzenie kont użytkowników z odpowiednimi uprawnieniami
 - ☐ instalacja serwera MariaDB w kontenerze Docker (https://hub.docker.com/_/mariadb)
- ☐ Power BI
- ☐ GitHub
- ☐ Backup
 - ☐ codzienny zrzut kontenerów za pomocą *Hyper Backup* i synchronizacja za pomocą *Cloud Sync*
 - ☐ eksport bazy danych do pliku *.sql*

proces ETL - opis

- ☐ Microsoft Excel, Power Query
- ☐ SQL model danych w bazie/hurtowni
- ☐ GitHub - https://github.com/Hercka/WSB_BigData
- ☐ model danych w Power BI, dodanie dodatkowych tabel i miar

dashboard – wizualizacja danych

❑ PowerBI



Trudności / Problemy

1. Połączenie z serwerem MariaDB (docker)

„(conn=617) Access denied for user 'root'@'172.17.0.1' (using password: NO)”

Rozwiązaniem było edycja pliku *custom.cnf* za pomocą komendy **vi** i „odkomentowanie” pozycji pozwalającej koncie root (SQL) na dostęp z każdego IP.

[vi command cheat sheet](#)

```
drwx--x--x 1 root root 10 May 25 11:36 ..
-rw-r--r-- 1 docker docker 3798 May 25 11:39 custom.cnf
drwxr-xr-x 1 root root 0 May 25 11:40 custom-cont-init.d
drwxr-xr-x 1 root root 0 May 25 11:40 custom-services.d
drwxr-xr-x 1 docker docker 290 May 25 22:16 databases
drwxr-xr-x 1 docker docker 10 May 25 11:39 log
#Server: /volume1/docker/volumes/MariaDB/_data$ chown root:root custom.cnf
chown: changing ownership of 'custom.cnf': Operation not permitted
#Server: /volume1/docker/volumes/MariaDB/_data$ sudo chown root:root custom.cnf
password:
total 4
#Server: /volume1/docker/volumes/MariaDB/_data$ ls -la
drwxr-xr-x 1 docker docker 114 May 25 11:40 .
drwx--x--x 1 root root 10 May 25 11:36 ..
-rw-r--r-- 1 root root 3798 May 25 11:39 custom.cnf
drwxr-xr-x 1 root root 0 May 25 11:40 custom-cont-init.d
drwxr-xr-x 1 root root 0 May 25 11:40 custom-services.d
drwxr-xr-x 1 docker docker 290 May 25 22:16 databases
drwxr-xr-x 1 docker docker 10 May 25 11:39 log
#Server: /volume1/docker/volumes/MariaDB/_data$ vi custom.cnf
#Server: /volume1/docker/volumes/MariaDB/_data$ ls -la
total 4
drwxr-xr-x 1 docker docker 114 May 25 11:40 .
drwx--x--x 1 root root 10 May 25 11:36 ..
-rw-r--r-- 1 root root 3798 May 25 11:39 custom.cnf
drwxr-xr-x 1 root root 0 May 25 11:40 custom-cont-init.d
drwxr-xr-x 1 root root 0 May 25 11:40 custom-services.d
drwxr-xr-x 1 docker docker 290 May 25 22:16 databases
drwxr-xr-x 1 docker docker 10 May 25 11:39 log
#Server: /volume1/docker/volumes/MariaDB/_data$ chmod 744 custom.cnf
chmod: changing permissions of 'custom.cnf': Operation not permitted
#Server: /volume1/docker/volumes/MariaDB/_data$ sudo chmod 744 custom.cnf
password:
total 4
drwxr-xr-x 1 docker docker 114 May 25 11:40 .
drwx--x--x 1 root root 10 May 25 11:36 ..
-rw-r--r-- 1 root root 3798 May 25 11:39 custom.cnf
drwxr-xr-x 1 root root 0 May 25 11:40 custom-cont-init.d
drwxr-xr-x 1 root root 0 May 25 11:40 custom-services.d
drwxr-xr-x 1 docker docker 290 May 25 22:16 databases
drwxr-xr-x 1 docker docker 10 May 25 11:39 log
#Server: /volume1/docker/volumes/MariaDB/_data$ sudo chown root:root custom.cnf
chown: changing ownership of 'custom.cnf': Operation not permitted
#Server: /volume1/docker/volumes/MariaDB/_data$ ls -la
total 4
drwxr-xr-x 1 docker docker 114 May 25 11:40 .
drwx--x--x 1 root root 10 May 25 11:36 ..
-rw-r--r-- 1 root root 3798 May 25 11:39 custom.cnf
drwxr-xr-x 1 root root 0 May 25 11:40 custom-cont-init.d
drwxr-xr-x 1 root root 0 May 25 11:40 custom-services.d
drwxr-xr-x 1 docker docker 290 May 25 22:16 databases
drwxr-xr-x 1 docker docker 10 May 25 11:39 log
#Server: /volume1/docker/volumes/MariaDB/_data$ sudo chown root:root custom.cnf
chown: changing ownership of 'custom.cnf': Operation not permitted
#Server: /volume1/docker/volumes/MariaDB/_data$ ls -la
total 4
drwxr-xr-x 1 docker docker 114 May 25 11:40 .
drwx--x--x 1 root root 10 May 25 11:36 ..
-rw-r--r-- 1 root root 3798 May 25 11:39 custom.cnf
drwxr-xr-x 1 root root 0 May 25 11:40 custom-cont-init.d
drwxr-xr-x 1 root root 0 May 25 11:40 custom-services.d
drwxr-xr-x 1 docker docker 290 May 25 22:16 databases
drwxr-xr-x 1 docker docker 10 May 25 11:39 log
#Server: /volume1/docker/volumes/MariaDB/_data$
```

```
192.168.1.10 - PuTTY
#relay_log_index = /config/log/mysql/relay-bin.index
#relay_log_info_file = /config/log/mysql/relay-bin.info
#log_slave_updates
#read_only
#
# If applications support it, this stricter sql_mode prevents some
# mistakes like inserting invalid dates etc.
#sql_mode = NO_ENGINE_SUBSTITUTION,TRADITIONAL
#
# * InnoDB
#
# InnoDB is enabled by default with a 10MB datafile in /var/lib/mysql/.
# Read the manual for more InnoDB related options. There are many!
default_storage_engine = InnoDB
# you can't just change log file size, requires special procedure
#innodb_log_file_size = 50M
innodb_buffer_pool_size = 256M
innodb_log_buffer_size = 8M
innodb_file_per_table = 1
innodb_open_files = 400
innodb_io_capacity = 400
innodb_flush_method = O_DIRECT
#
# * Security Features
#
# Read the manual, too, if you want chroot!
# chroot = /var/lib/mysql/
#
# For generating SSL certificates I recommend the OpenSSL GUI "tinyca".
#
# ssl-ca=/etc/mysql/cacert.pem
# ssl-cert=/etc/mysql/server-cert.pem
# ssl-key=/etc/mysql/server-key.pem
#
# * Galera-related settings
#
#[galera]
# Mandatory settings
#wsrep_on=ON
#wsrep_provider=
#wsrep_cluster_address=
#binlog_format=row
#default_storage_engine=InnoDB
#innodb_autoinc_lock_mode=2
#
# Allow server to accept connections on all interfaces.
#bind-address=0.0.0.0
#
# Optional setting
#wsrep_slave_threads=1
#innodb_flush_log_at_trx_commit=0

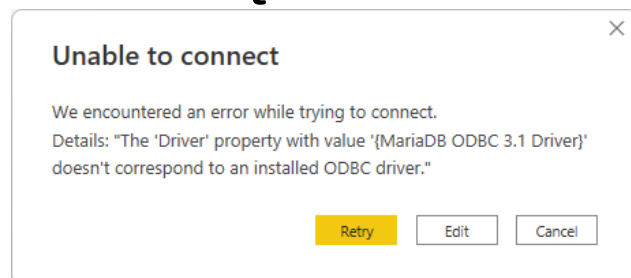
[mysqldump]
quick
quote-names
max_allowed_packet = 16M

[mysql]
#no-auto-rehash # faster start of mysql but no tab completion

[isamchk]
key_buffer_size = 16M
141.1 Bot
```

Trudności / Problemy

1. Połączenie PowerBi do MariaDB



Rozwiązanie:

Instalacja „connector’a”: <https://mariadb.com/downloads/connectors/connectors-data-access/odbc-connector/>