



III Spotkanie  
Użytkowników



# Validation of spatial data in Python

Michał Pilarski

# Workshop Repository



[https://gitlab.com/michpil/geospatial\\_python/-/tree/qgis\\_workshop\\_geospatial\\_data\\_validation](https://gitlab.com/michpil/geospatial_python/-/tree/qgis_workshop_geospatial_data_validation)



\*All materials (env installers, presentation and code) will be available on repository

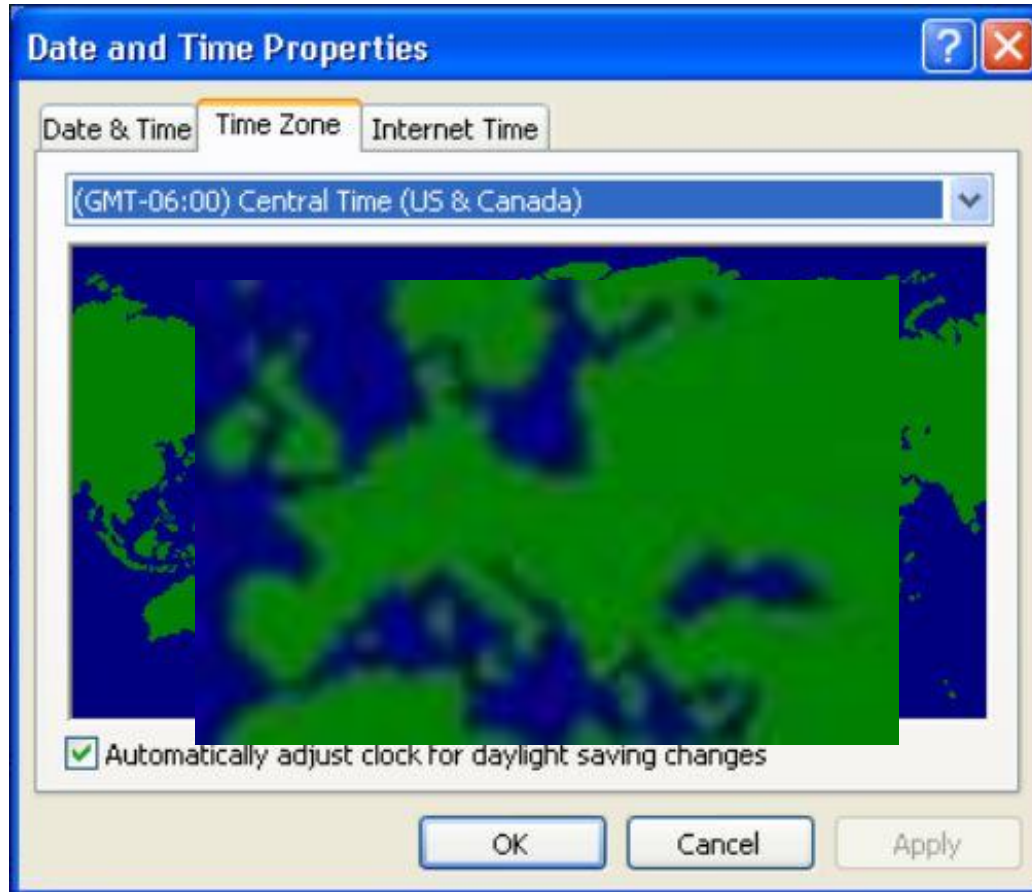
# Do we need testing spatial data?



# Motivation – 1<sup>st</sup> case



# Motivation – 2<sup>nd</sup> case



# Workshop takeaways



- Walk through full testing process of GIS data
- Integration of PyQGIS and PyTest
- Continuous Testing with Jenkins

# Data Validation - definitions

## Actual

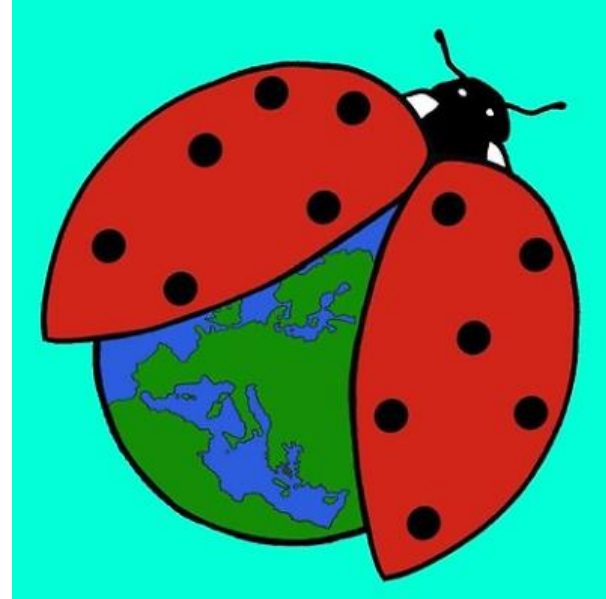
data collected from dev system based on reqs

## Expected

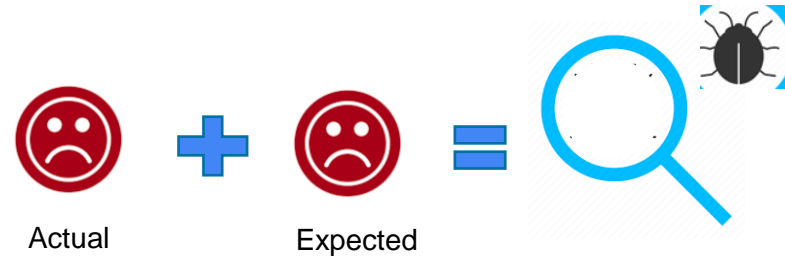
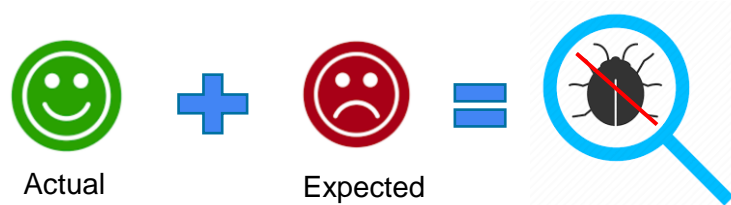
data prepared by tester based on reqs

## Actual vs. Expected

“... expected is usually compared with actual result, and if the actual result differs from the expected one, the difference is documented and called a **ladybug**”.



# Reqs and Actual vs. Expected





# Data Validation

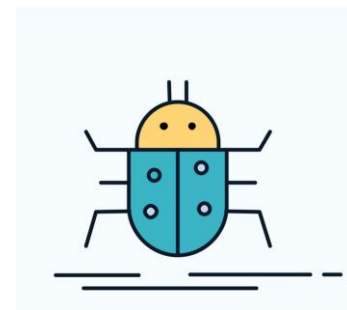
## What to validate?

### Data Structure

- data format
- count of columns and columns names
- rows amount, size of data
- checksum

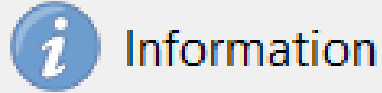
### Data Values

- mandatory, null and missing values
- duplicates
- data types
- range of values (numeric, dates), list (text)
- numeric precision
- data distribution (profiling) - outliers, histogram, variance
- attributes relationships and tables relationships
- pattern REGEX (numeric, text, date)
- primary keys and technical fields
- geometry accuracy, consistency, completeness, topology



# Tested Objects

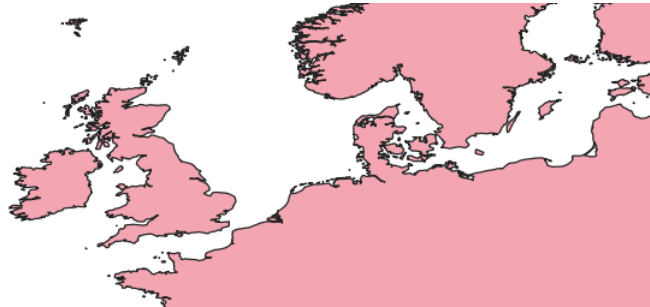
**TEST MetaData**



**TEST Data Attributes**



**TEST Data Geometry**



# Testing Env



PyCharm



pyQGIS



pytest



Jenkins



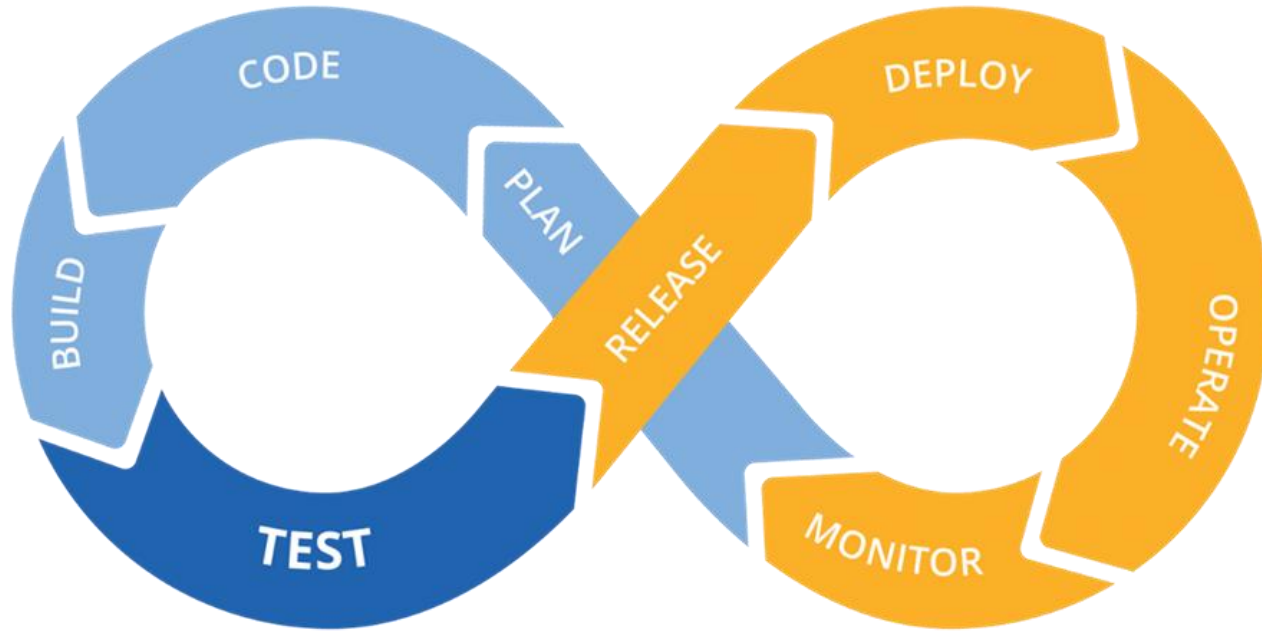
# Project – geospatial data tests

Find optimal location of local pubs in Poland based on below factors:

- Pub is located near regional brewery (distance  $\leq 10$  km)
- Pub is placed in city
  - Pub is located near major roads (distance  $\leq 1$  km on the left/right side of road)
- Pub is not too close to existing pubs (distance  $\geq 1$  km)



# CI/CD JENKINS











**Thank You!**

