

Big Data

# Organizacyjnie

Prowadzący:

dr Mariusz Rafało

[mrafalo@sggw.edu.pl](mailto:mrafalo@sggw.edu.pl)

<http://mariuszrafalo.pl> (hasło: BIG)

# IDENTYFIKACJA NADUŻYĆ

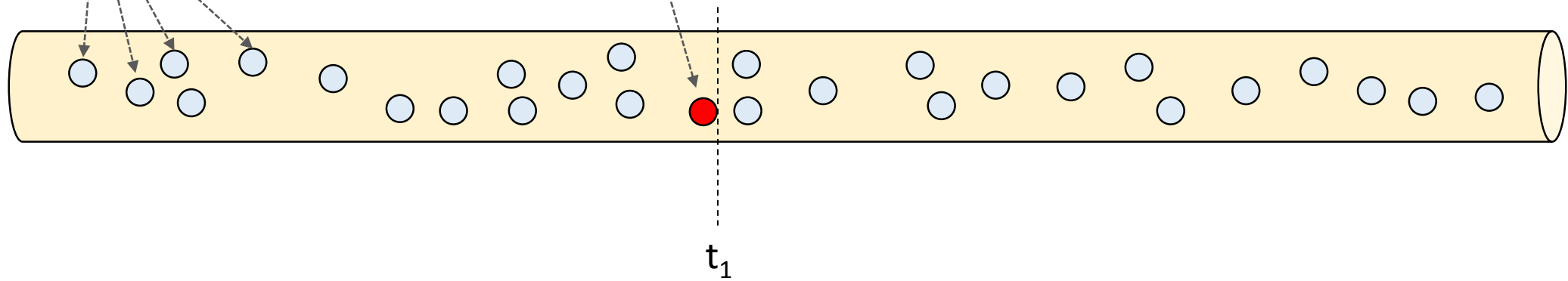
# Problem

- Identyfikacja nadużyć w płatnościach kartą kredytową

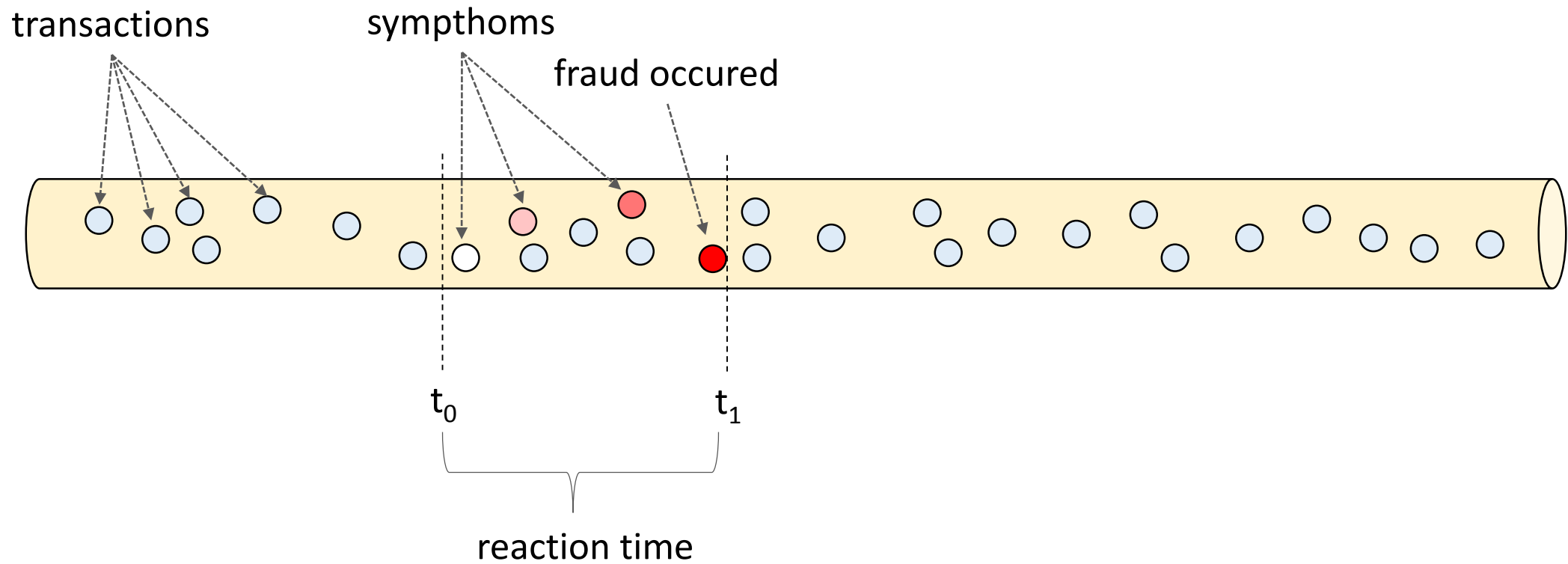
# Koncepcja

transactions

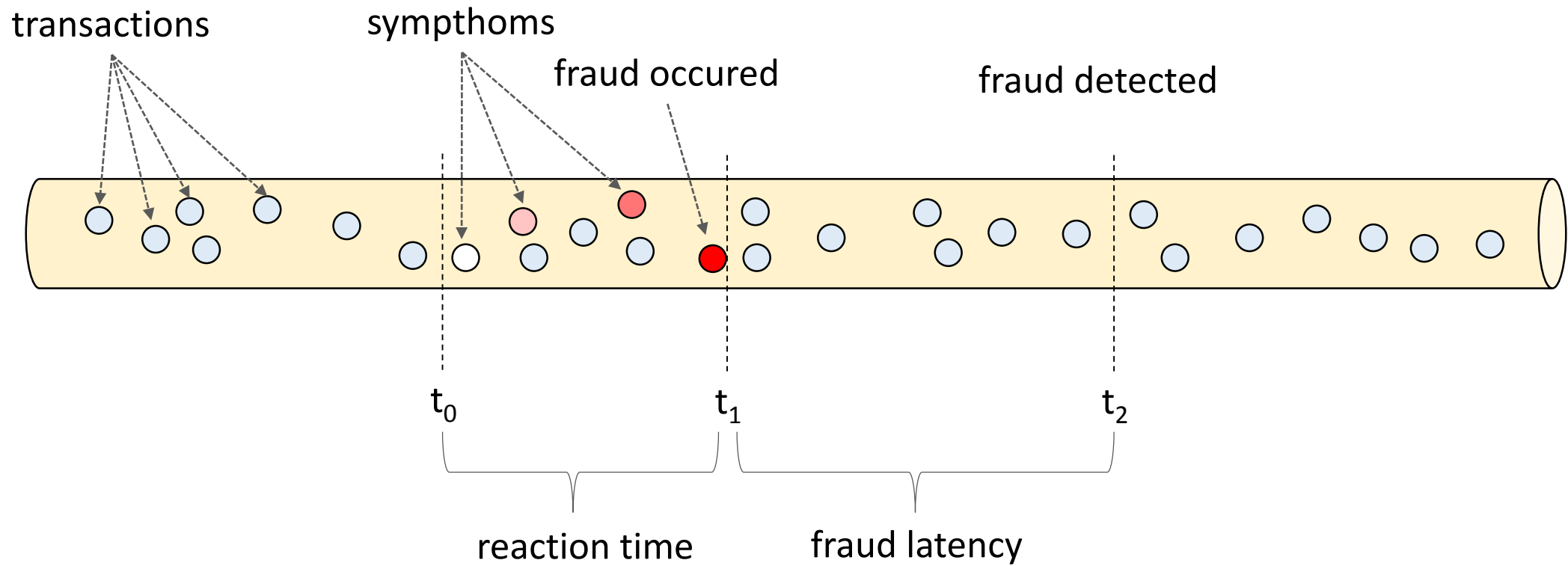
fraud occurred



# Koncepcja



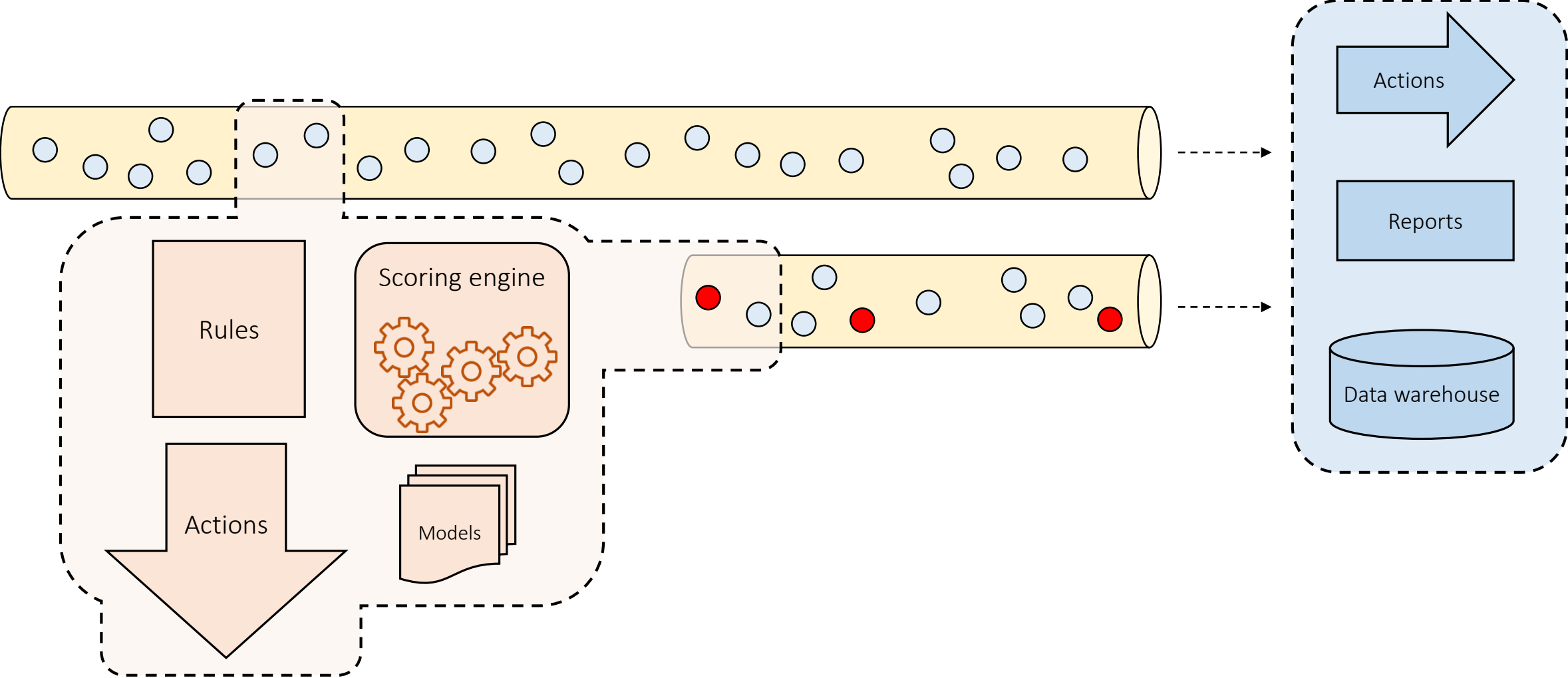
# Koncepcja



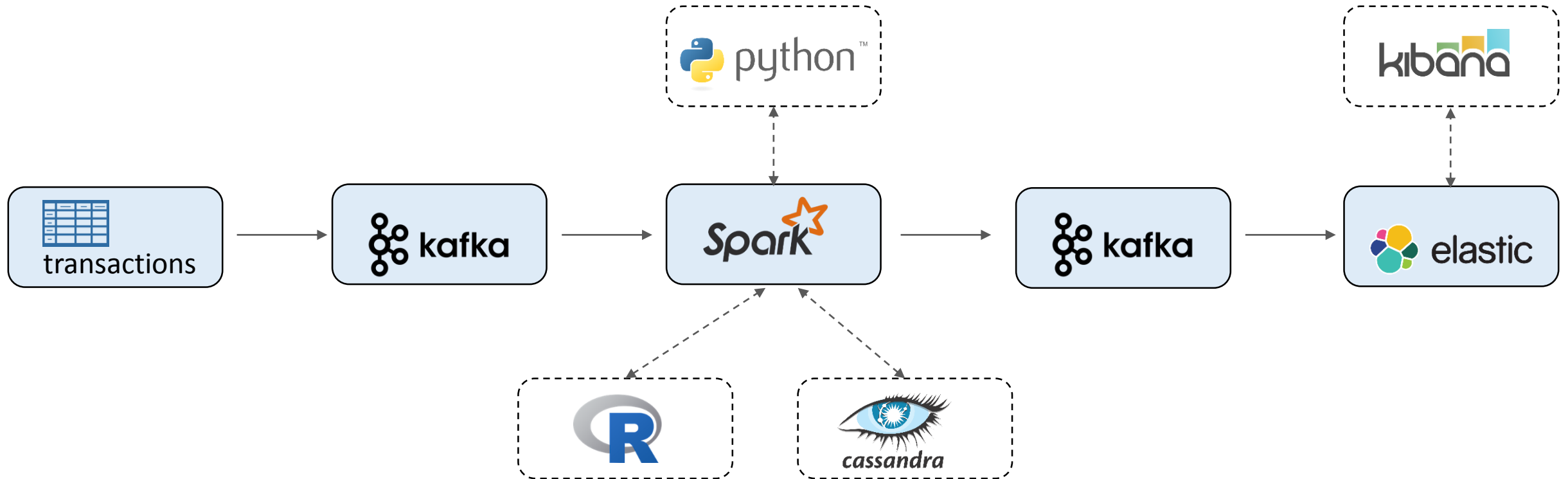
# ARCHITEKTURA



# High level concept



# Wykorzystane technologie

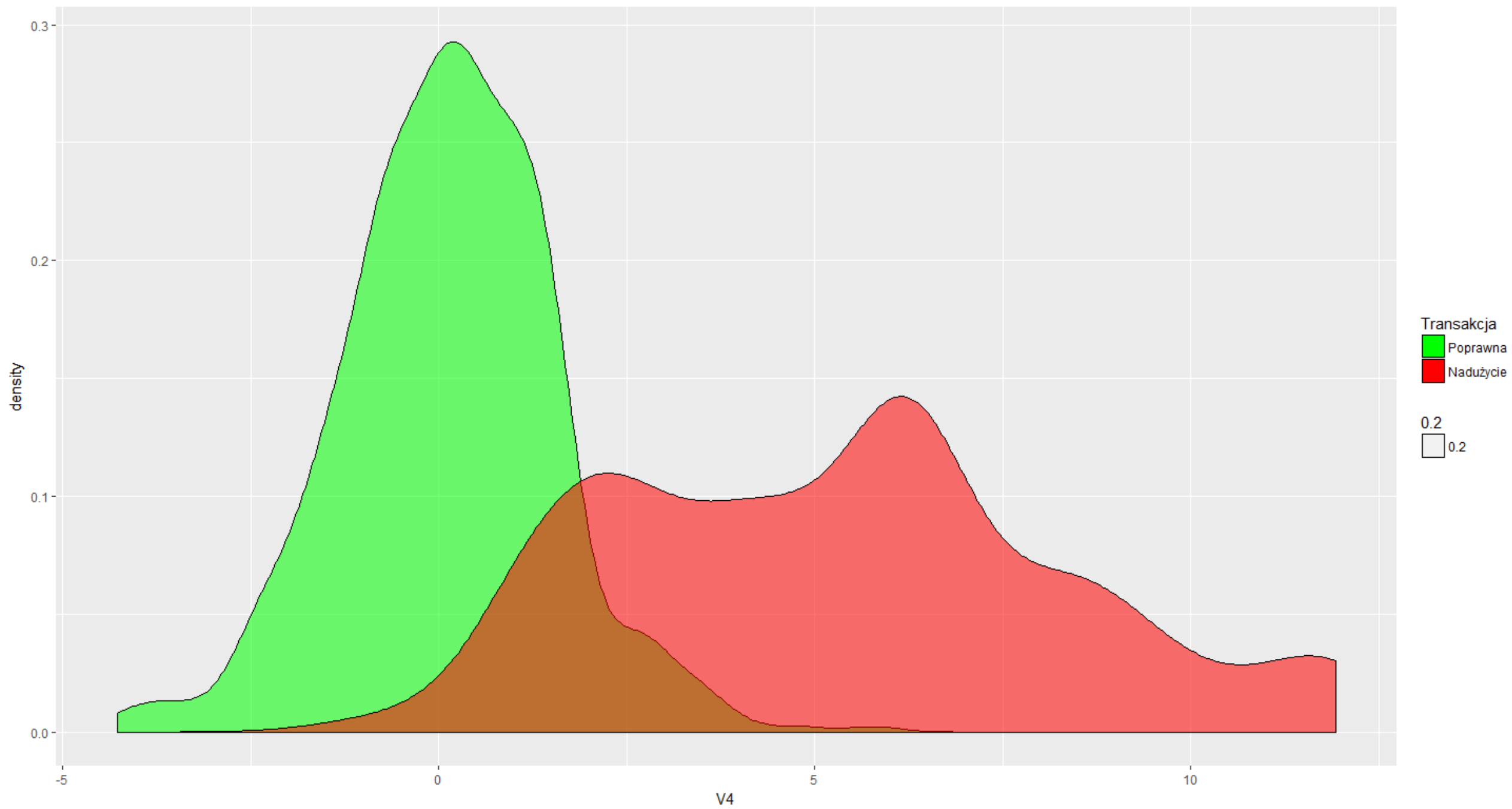


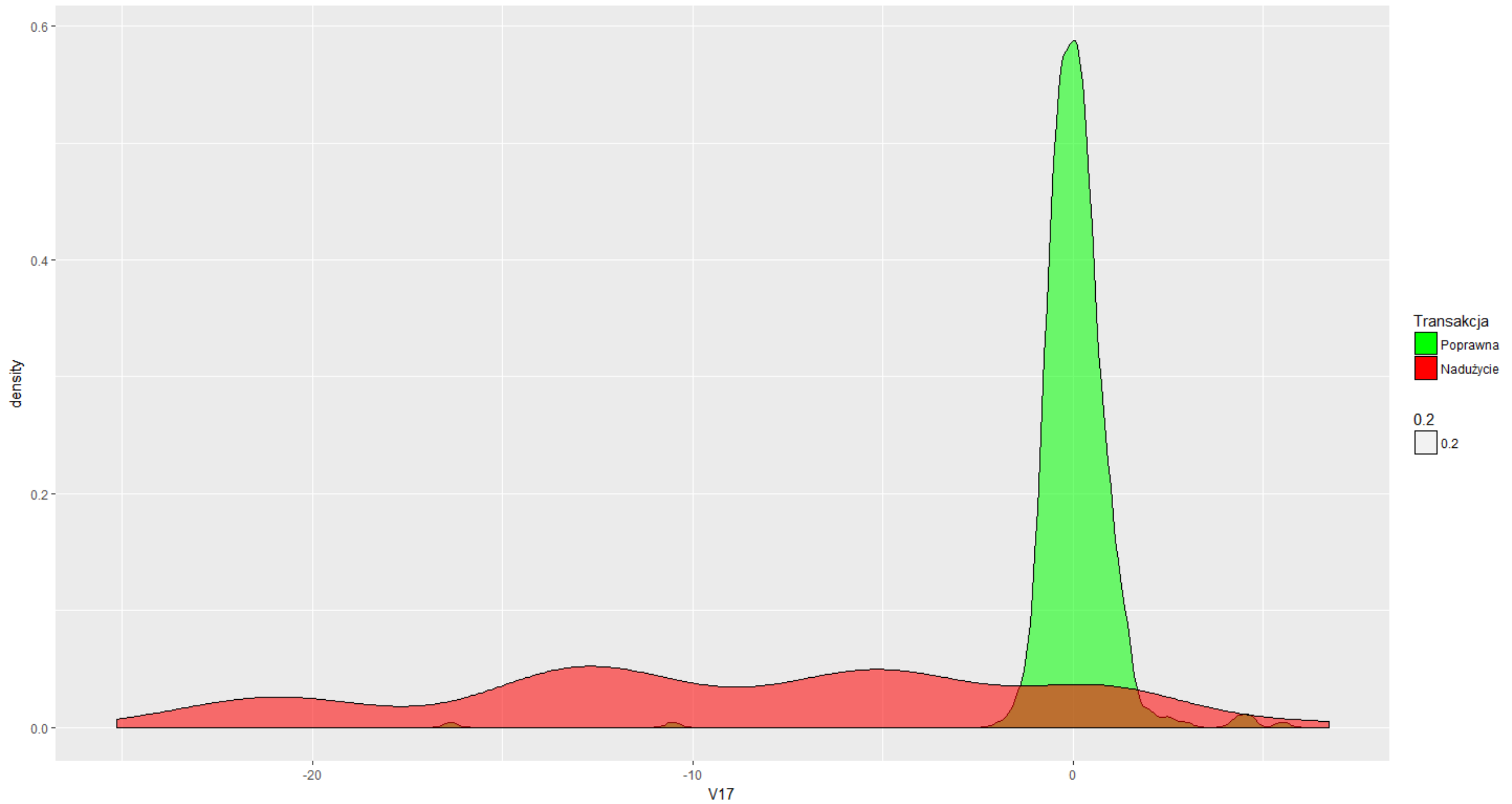
DANE

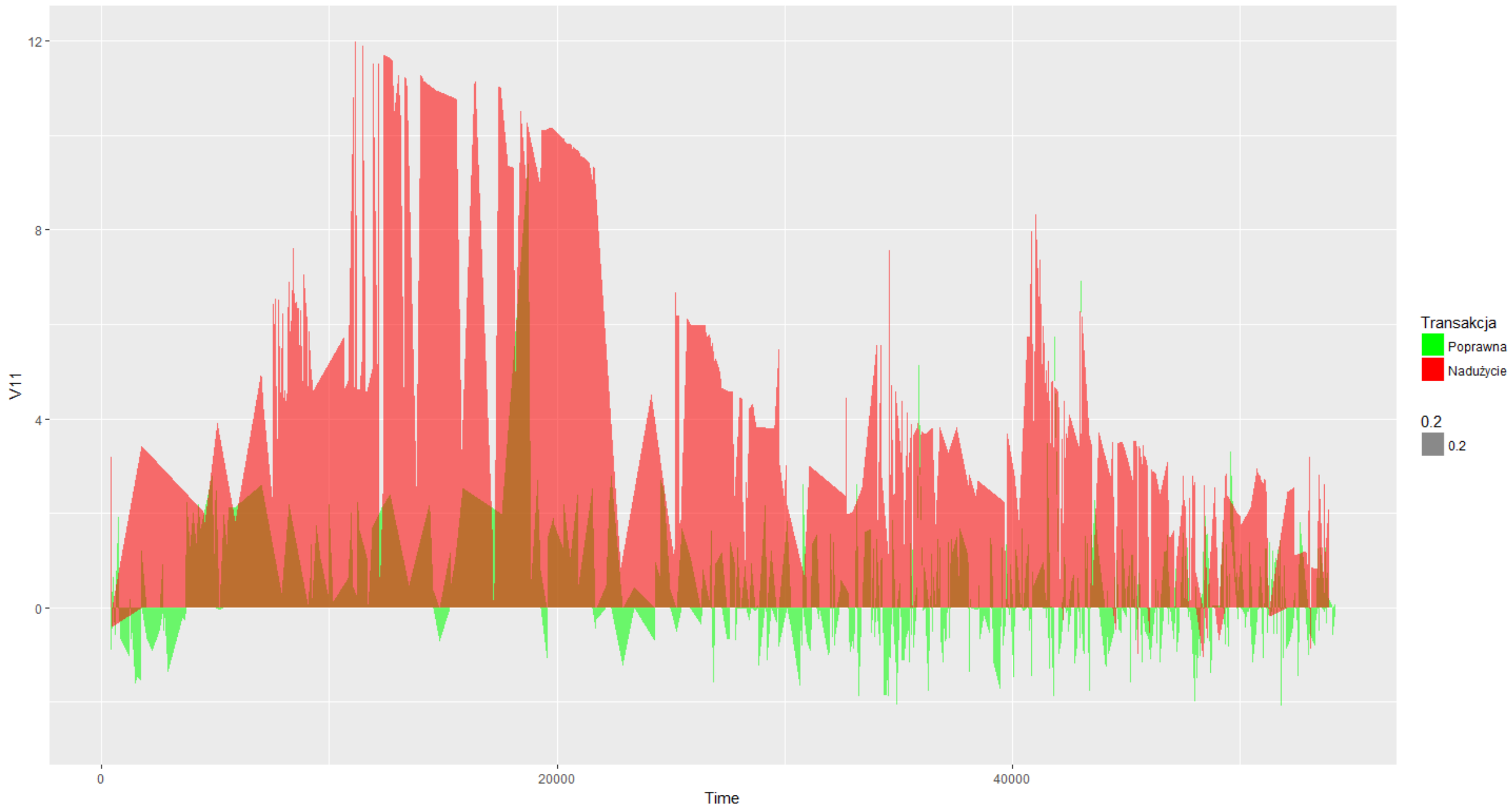
# Zawartość zbioru danych

- 284 807 obserwacji
- 32 zmienne
- Zbiór nie zbilansowany: 0.17% nadużyć (492 vs 284 315)

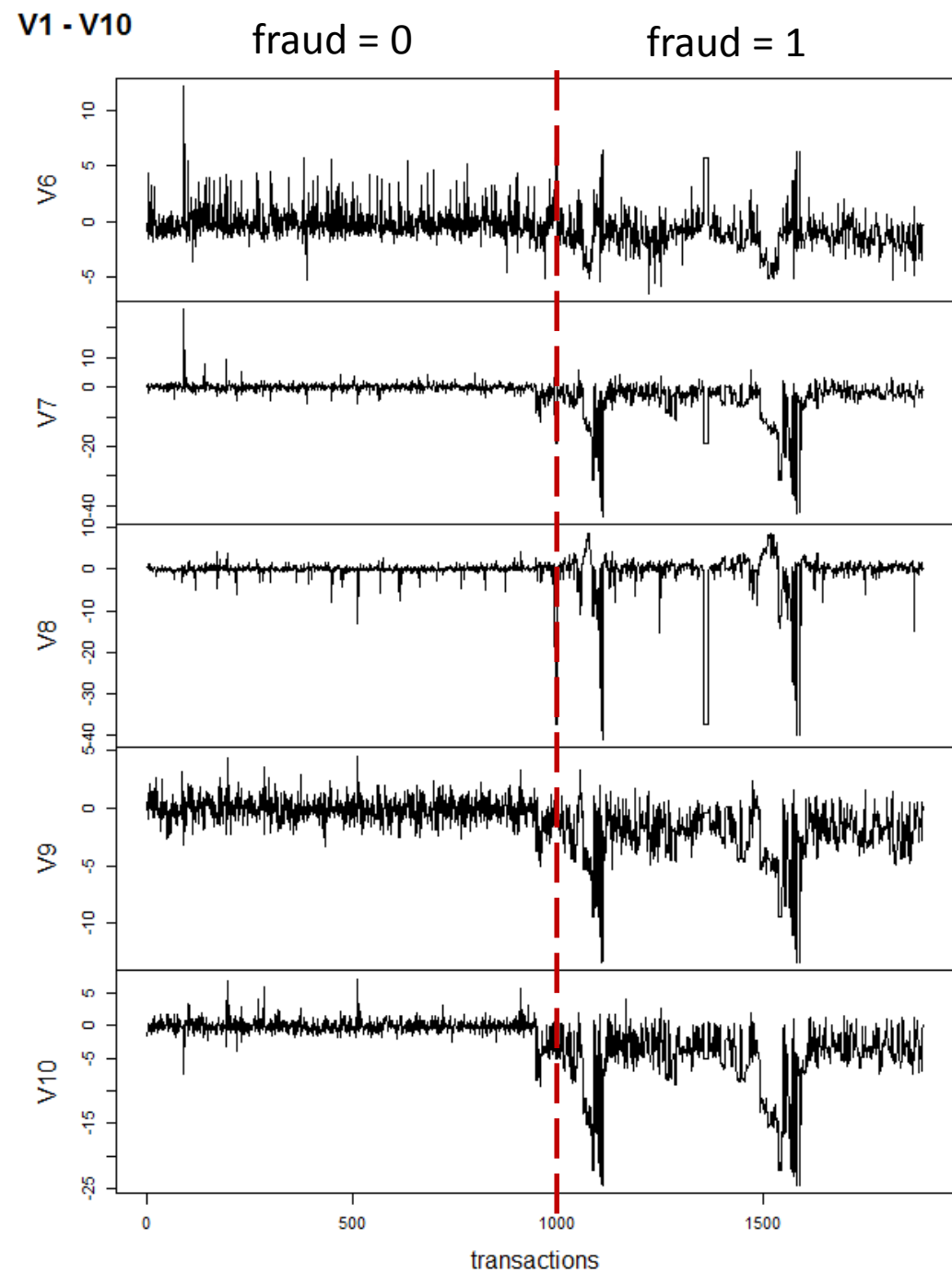
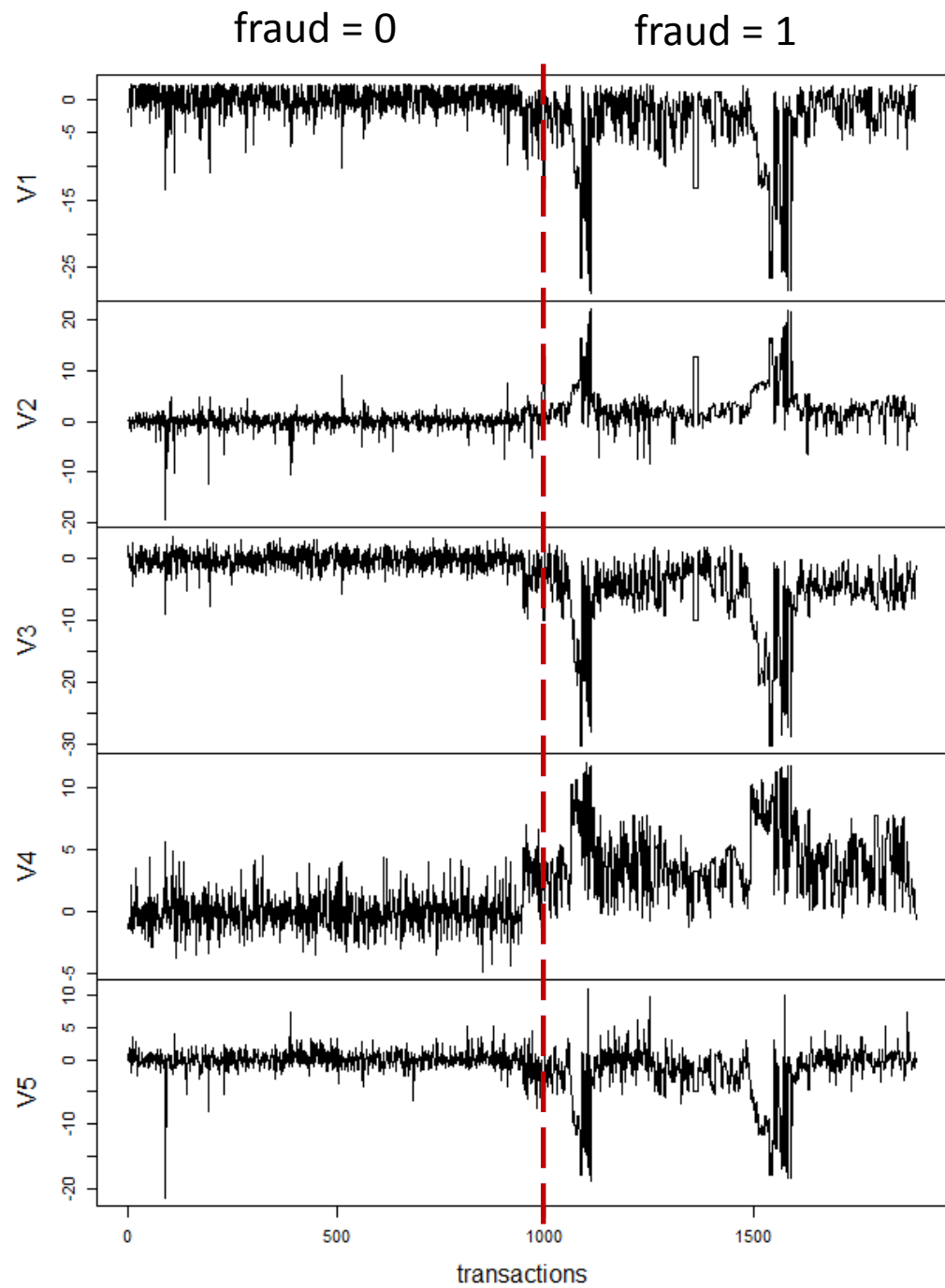
# WIZUALIZACJA





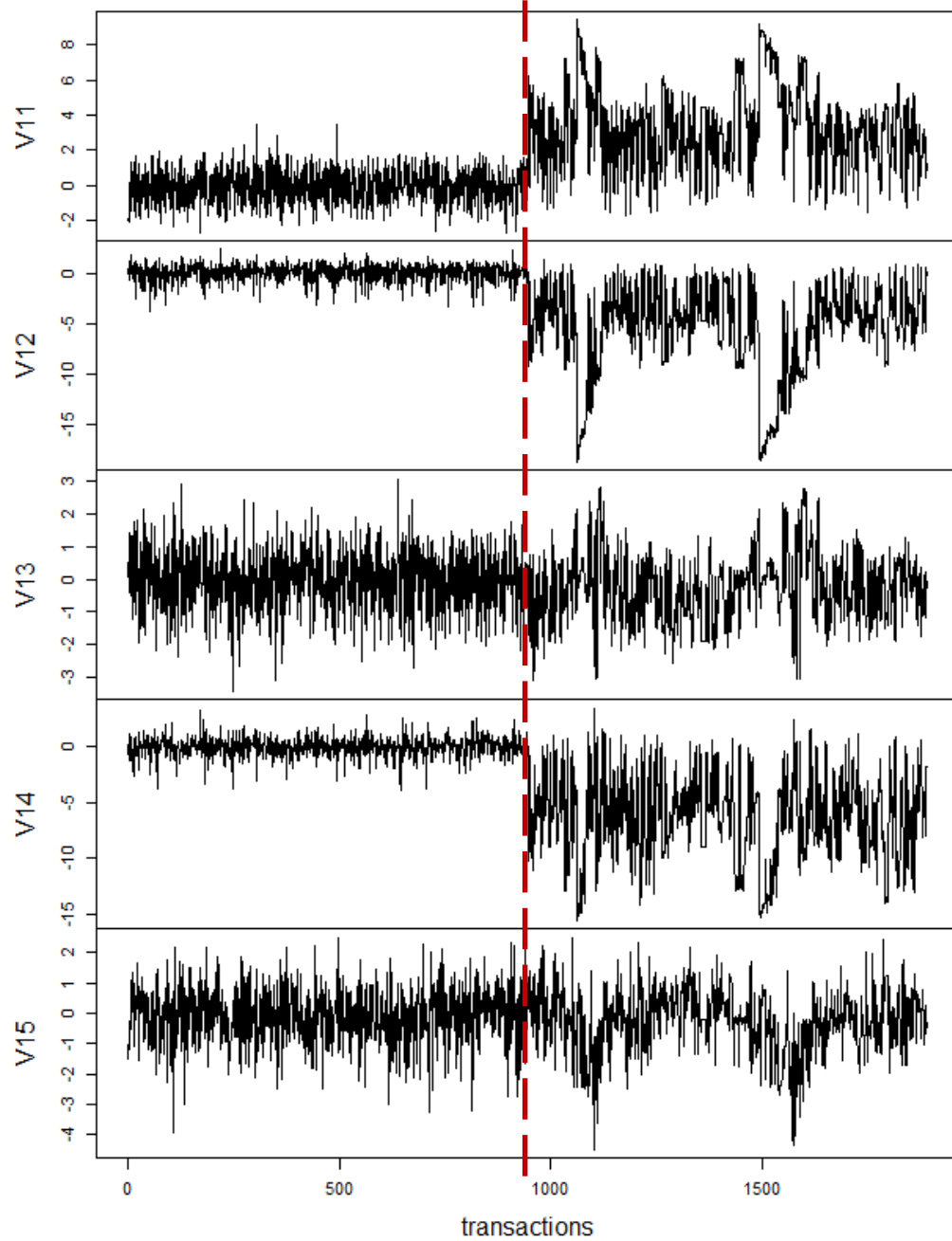






fraud = 0

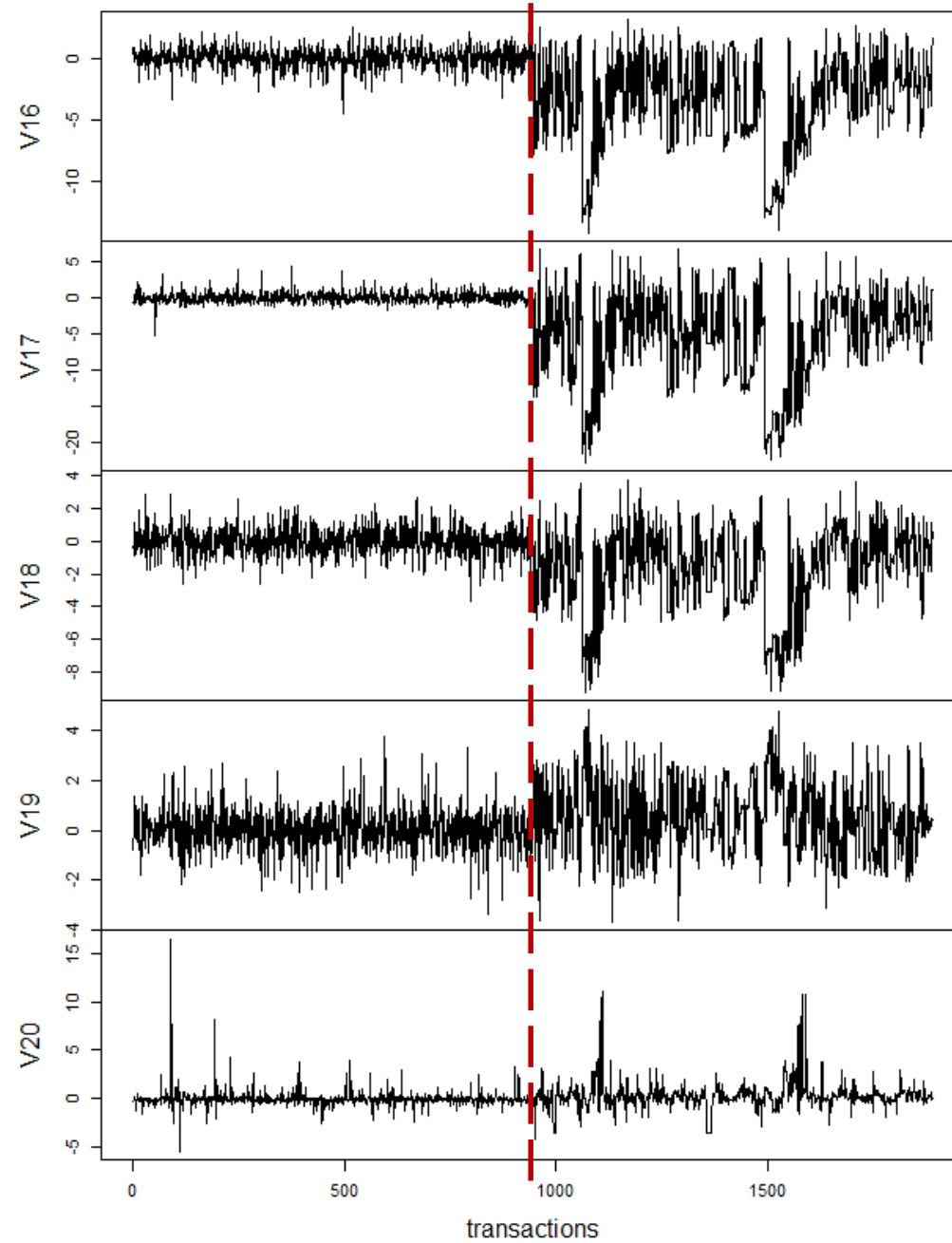
fraud = 1



V11 - V20

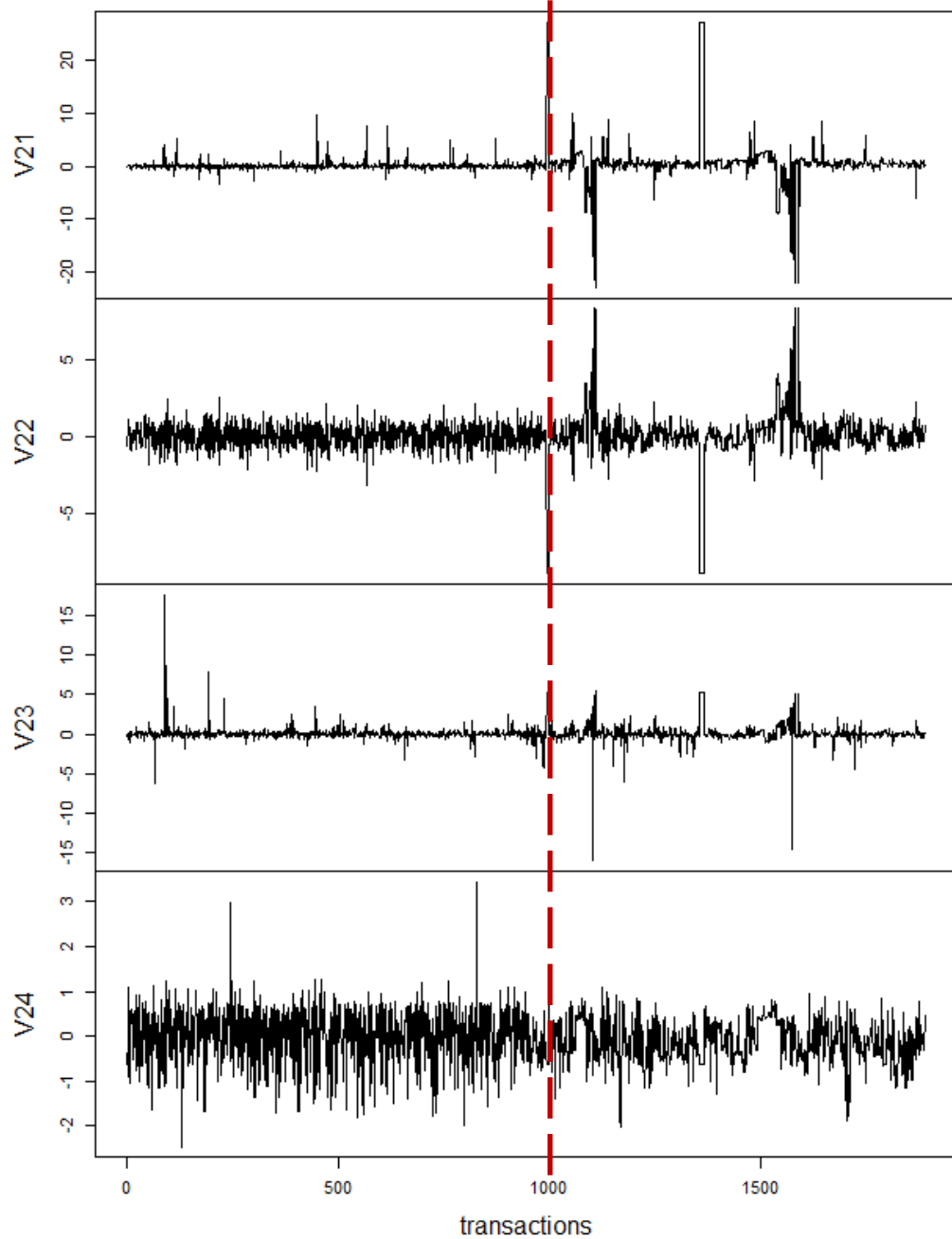
fraud = 0

fraud = 1



fraud = 0

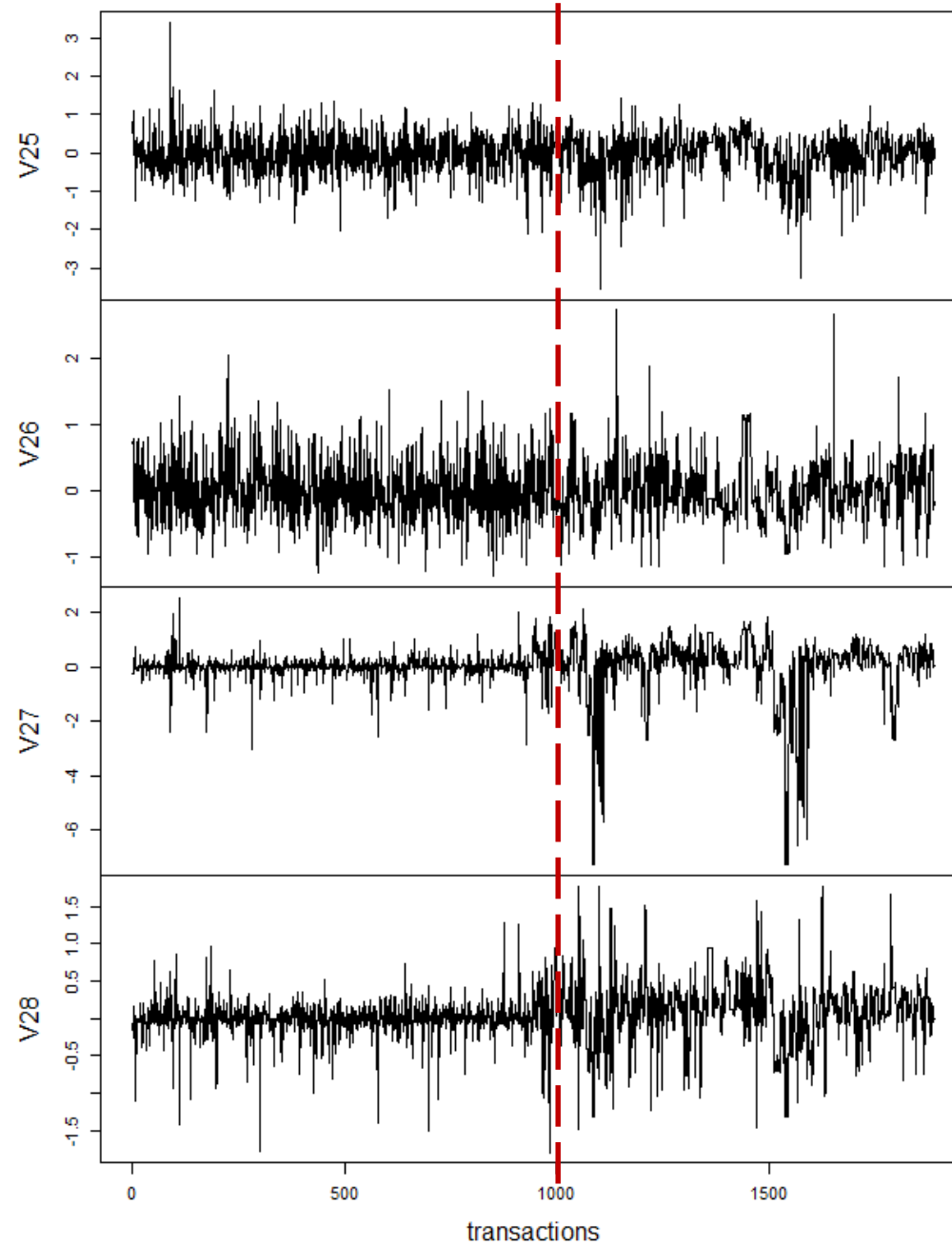
fraud = 1



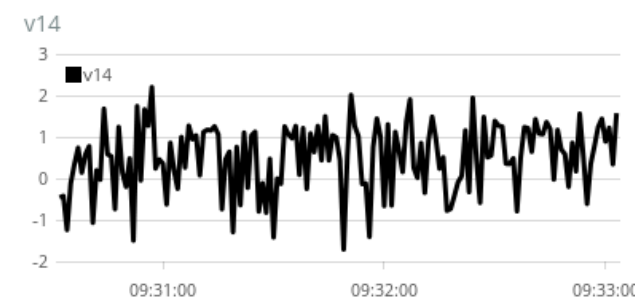
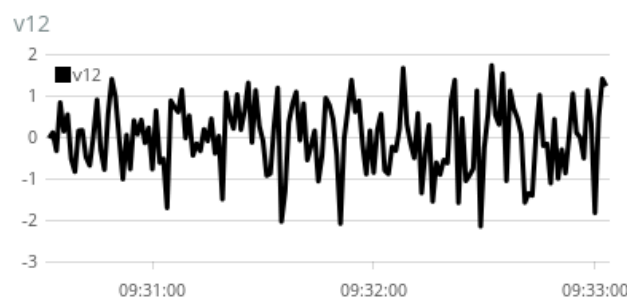
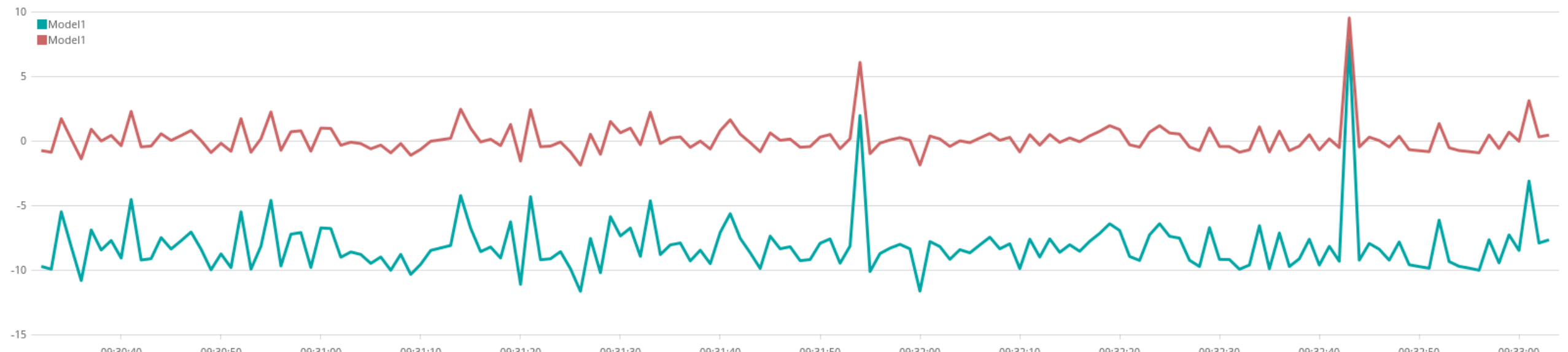
V21 - V18

fraud = 0

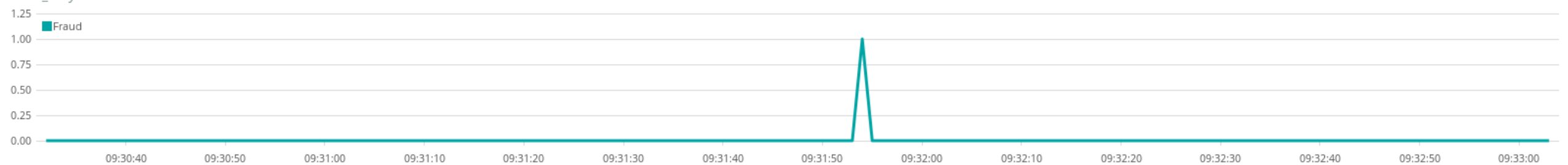
fraud = 1



Fraud on line2

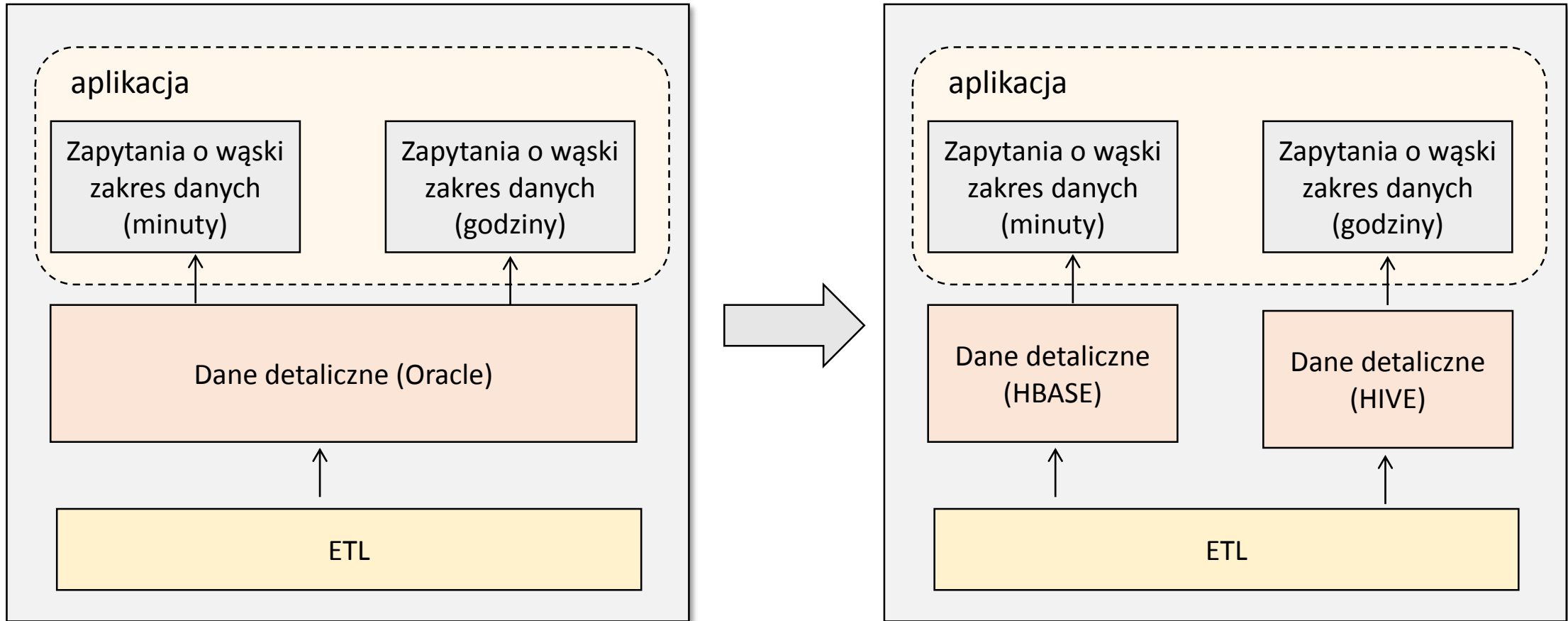


fraud\_only

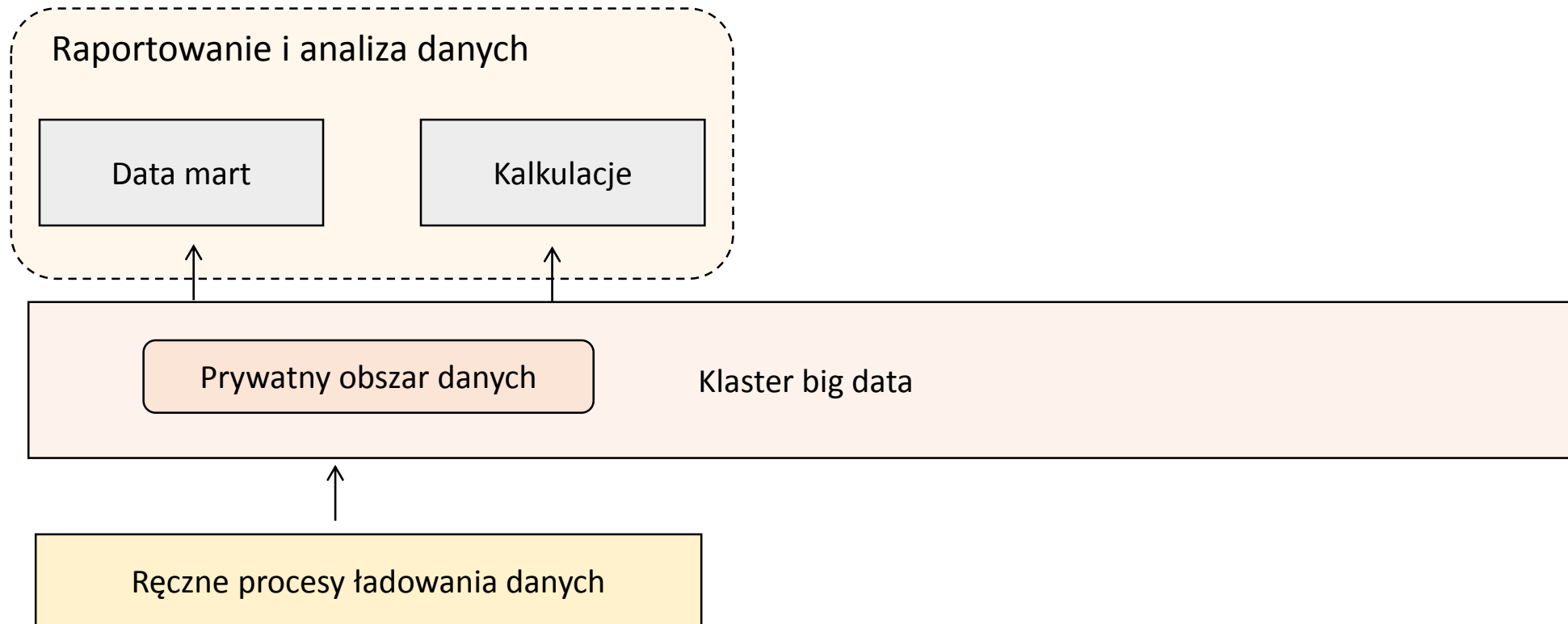


# INNE PRZYKŁADY

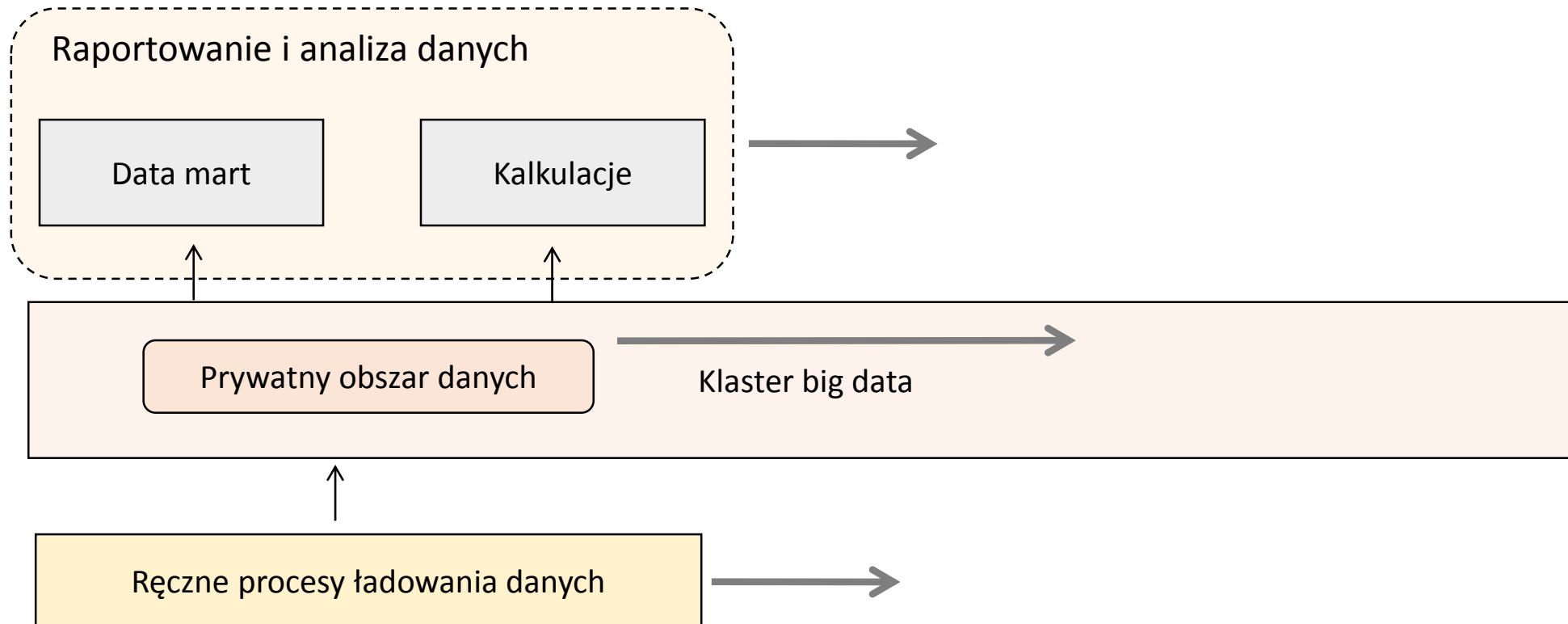
# Przykład: big data jako repozytorium danych



# Przykład: środowisko self-service BI

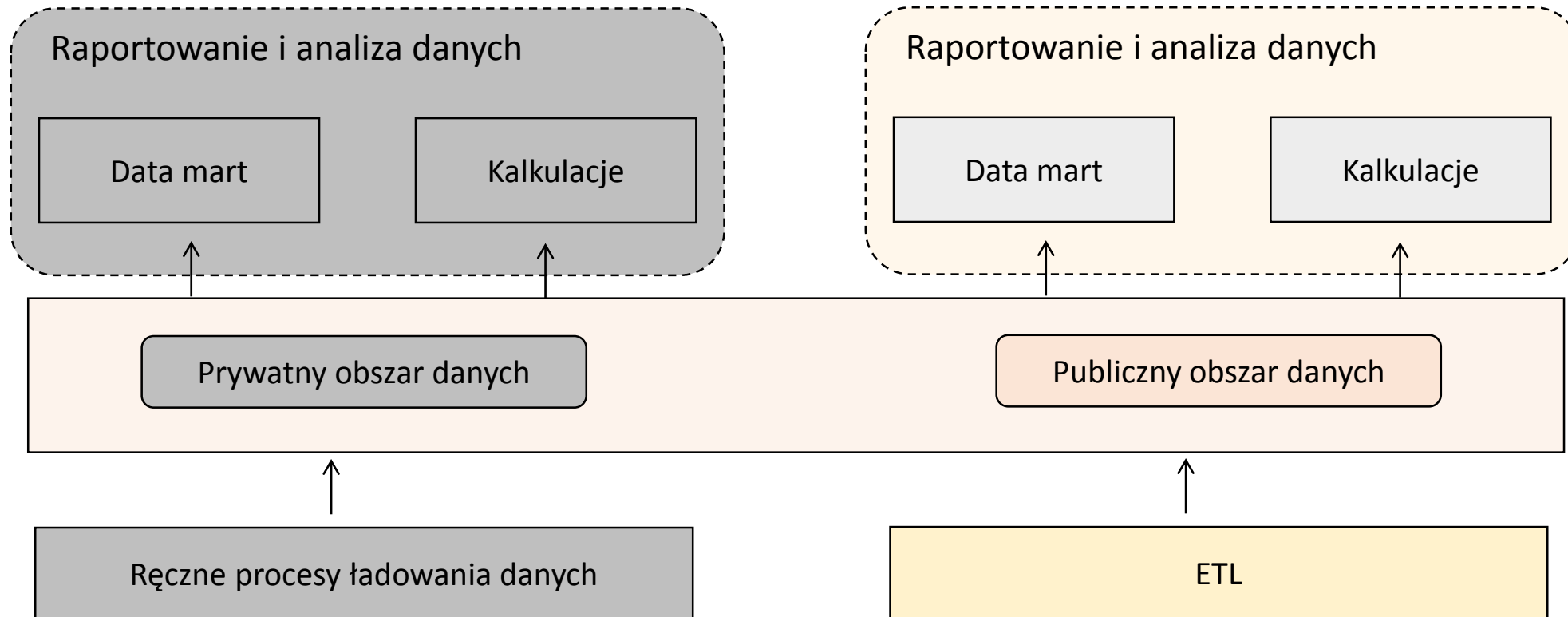


# Przykład: środowisko self-service BI





# Przykład: środowisko self-service BI



Dziękuję za uwagę