

Opis charakteru danych

Dane prezentują losowe dane uzyskiwane przy użyciu "fotoradarów" (czujników) ustawionych w różnych miejscach w Polsce. Dla zasady przyjmuję, że dane dotyczące pojazdu, właściciela itp. są przechowywane nie tylko po przekroczeniu ustawowej prędkości 50 km/h, ale są one generowane dla wszystkich przejeżdżących w danym miejscu aut. Dlatego "fotoradar" pełni bardziej funkcję czujnika aniżeli narzędzia służącego do kontroli przepisów ruchu drogowego. Jednak przyjmuję, że czujniki są ustawione na obszarach zabudowanych, dlatego po przekroczeniu ustawowej prędkości 50 km/h dla właściciela pojazdu będą generowane mandaty na podstawie polityki obowiązującej w 2023 roku. Etykieta czasowa oznacza losowy moment wykonania pomiaru. Etykieta ta może się spóźniać w stosunku do czasu systemowego maksymalnie do minuty (60 sekund).

Opis atrybutów

Atrybuty w każdym zdarzeniu mają następujące znaczenie:

- car
 - typ: string
 - znaczenie: nazwa marki samochodu
 - kategoria: atrybut lub atrybuty dodatkowe, opisujące zdarzenie
- manufacturer
 - typ: string
 - znaczenie: nazwa producenta danego samochodu
 - kategoria: atrybut lub atrybuty dodatkowe, opisujące zdarzenie
- car_owner
 - typ: string
 - znaczenie: imię i nazwisko właściciela pojazdu
 - kategoria: atrybut lub atrybuty dodatkowe, opisujące zdarzenie
- car_type
 - typ: string
 - znaczenie: rodzaj nadwozia pojazdu
 - kategoria: atrybut, lub atrybuty, po których można dane grupować (zawierający ograniczoną liczbę możliwych wartości (5-10))
- fuel_type
 - typ: string
 - znaczenie: rodzaj paliwa wykorzystywanego przez dany pojazd
 - kategoria: atrybut, lub atrybuty, po których można dane grupować (zawierający ograniczoną liczbę możliwych wartości (5-10))
- velocity
 - typ: int
 - znaczenie: prędkość pojazdu w punkcie pomiarowym
 - kategoria: atrybut lub atrybuty, których wartości można agregować (wartości numeryczne)
- fine
 - typ: int
 - znaczenie: kara pieniężna za przekroczenie prędkości (w PLN)
 - kategoria: atrybut lub atrybuty, których wartości można agregować (wartości numeryczne)

- `penalty_points`
 - typ: `int`
 - znaczenie: punkty karne za złamanie przepisu ruchu drogowego (przekroczenie prędkości), przyjmuję że co najmniej 24 punkty oznaczają utratę prawa jazdy przez właściciela pojazdu
 - kategoria: atrybut lub atrybuty, których wartości można agregować (wartości numeryczne)
- `ts`
 - typ: `string`
 - znaczenie: czas rejestracji zdarzenia (pomiaru prędkości wraz z identyfikacją pojazdu)
 - kategoria: znacznik czasowy zdarzeń (inny niż czas systemowy)

Opis trzech przykładowych analiz

Propozycje możliwych analiz opartych na tym generatorze danych

Agregacja

Utrzymuj informacje o maksymalnych zarejestrowanych prędkościach przez poszczególne osoby lub marki pojazdów w ciągu ostatnich 5 minut.

Wykrywanie anomalii

Wykrywaj przypadki przekroczenia prędkości 100 km/h przez danego kierowcę w danym aucie.

Wykrywanie anomalii oparte na agregacji

Wykrywaj przypadki zdobycia przez właścicieli pojazdów liczby punktów karnych większej lub równej 24, która oznacza pozbawienie danego kierowcy prawa jazdy.