

Spark – projekt – zestaw danych uk-traffic

Tematyka

Dane zawierają pomiary ruchu drogowego w Wielkiej Brytanii w latach 2000-2018. Zostały one uzupełnione o informacje dotyczące pogody.

Pochodzenie danych

Projekt realizowany jest tylko i wyłącznie w oparciu o dane dostarczone przez prowadzącego. Poniższe informacje wyjaśniają jedynie pochodzenie oryginalnych danych.

Główny trzon danych został uzyskany ze strony Departamentu Transportu Wielkiej Brytanii

<https://roadtraffic.dft.gov.uk/downloads>

Dane te dostępne są także na stronie <https://roadtraffic.dft.gov.uk/>, która przedstawia je na mapie Wielkiej Brytanii.

Dane powyższe zostały uzupełnione na temat pogody występującej w tamtym okresie uzyskanych ze strony

<https://www.kaggle.com/daveianhickey/2000-16-traffic-flow-england-scotland-wales>

Charakterystyka danych

Dane zostały dotyczące pomiarów ruchu drogowego zostały podzielone na 3 zestawy symulujące 3 niezależne bazy danych o podobnym schemacie. Podział dokonany został w oparciu o region pomiaru

1. "Scotland"
2. Północne regiony: "North West", "North East", "Yorkshire and The Humber", "Wales", "West Midlands", "East Midlands"
3. Południowe regiony: "South West", "South East", "London", "East of England"

Każdy zestaw składa się z trzech plików CSV (symulujących tabele) zawierających:

1. regions – dane dotyczące regionów
 - a. region_id,
 - b. region_name,
 - c. region_ons_code
2. authorities – dane dotyczące lokalnych obszarów administracyjnych
 - a. local_authority_ons_code,
 - b. local_authority_id,
 - c. local_authority_name,
 - d. region_ons_code
3. mainData – dane dotyczące pomiarów ruchu drogowego (poniżej najważniejsze atrybuty)
 - a. count_point_id,
 - b. direction_of_travel,
 - c. year,
 - d. count_date,
 - e. hour,
 - f. local_authoirty_ons_code,
 - g. road_name,
 - h. road_category,
 - i. road_type,

- j. start_junction_road_name,
- k. end_junction_road_name,
- . . .
- l. pedal_cycles,
- m. two_wheeled_motor_vehicles,
- n. cars_and_taxis,
- o. buses_and_coaches,
- p. lgvs,
- q. hgvs_2_rigid_axle,
- r. hgvs_3_rigid_axle,
- s. hgvs_4_or_more_rigid_axle,
- t. hgvs_3_or_4_articulated_axle,
- u. hgvs_5_articulated_axle,
- v. hgvs_6_articulated_axle,
- w. all_hgvs,
- x. all_motor_vehicles

Opis poszczególnych atrybutów został zawarty w poniższym dokumencie w paragrafie *v. Raw manual counts data – minor roads*

<http://data.dft.gov.uk.s3.amazonaws.com/road-traffic/all-traffic-data-metadata.pdf>

Dane dotyczące pogody są dostępne w pliku tekstowym `weather.txt`. Zawierają one informację na temat daty i czasu pomiaru, opisu pogody oraz miejsca (`local_authority_ons_code`), w którym pogodę odnotowano.