

# Zestaw 2 – nyc-taxi

Pochodzenie danych to <https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page>

Uwaga! Dane pobieramy z miejsca wskazanego w opisie projektu

## Dwa zbiory danych

1. datasource1 – informacje na przejazdów taksówek (1)

Dane mają format CSV. Pliki nie posiadają nagłówka.

Pola w pliku:

0. VendorID - Kod wskazujący dostawcę TPEP, który dostarczył rekord. 1= Creative Mobile Technologies, LLC; 2= VeriFone Inc.
1. tpep\_pickup\_datetime - Data i godzina włączenia taksometru.
2. tpep\_dropoff\_datetime - Data i godzina wyłączenia taksometru.
3. Passenger\_count - Liczba pasażerów w pojeździe. Jest to wartość wprowadzona przez kierowcę.
4. Trip\_distance - Odległość, w milach, podana przez taksometr.
5. RateCodeID - Ostateczny kod stawki obowiązujący na koniec podróży. 1= stawka standardowa 2=JFK 3=Newark 4=Nassau lub Westchester 5=Taryfa do negocjacji 6=Jazda grupowa
6. Store\_and\_fwd\_flag - Wskazane czy rekord podróży był przechowywany w pamięci pojazdu przed wysłaniem do dostawcy, czyli "store and forward", ponieważ pojazd nie miał połączenia z serwerem. Y = tak N = nie
7. PULocationID - TLC Taxi Zone, w której taksometr był włączony
8. DOLocationID - Strefa taksówek TLC, w której taksometr został wyłączony
9. Payment\_type - Kod numeryczny oznaczający sposób zapłaty za podróż przez pasażera. 1=Karta kredytowa, 2=Gotówka, 3=Bez opłat, 4=Spór, 5=Nieznany, 6=Przejazd anulowany
10. Fare\_amount - Opłata za przejazd czasowo-dystansowy obliczona przez licznik.
11. Extra - Dodatkowe dodatki i dopłaty. Obecnie obejmuje to tylko opłaty w godzinach szczytu w wysokości 0,50 USD i 1 USD oraz opłaty za nocleg.
12. MTA\_tax - Podatek MTA w wysokości 0,50 USD, który jest uruchamiany automatycznie na podstawie stawki licznika w użyciu.
13. Tip\_amount - Kwota napiwku — to pole jest automatycznie wypełniane w przypadku napiwków zapłaconych za pomocą kart kredytowych. Napiwki gotówkowe nie są uwzględniane.
14. Tolls\_amount - łączna kwota wszystkich opłat drogowych uiszczanych podczas podróży.
15. Improvement\_surcharge - \$0.30 improvement surcharge assessed trips at the flag drop. The improvement surcharge began being levied in 2015.
16. Total\_amount - Całkowita kwota pobierana od pasażerów. Nie obejmuje napiwków gotówkowych.

2. datasource4 – informacje na temat stref taksówek (4)

Dane mają format CSV, każdy z plików ma wiersz nagłówka.

Pola w pliku

- 0. LocationID – identyfikator lokalizacji TLC Taxi Zone
- 1. Borough – dzielnica
- 2. Zone – nazwa strefy taksówek
- 3. service\_zone – nazwa strefy usługowej