

Analiza danych

Dataset zawiera 10 kolumn i 91523 wierszy, nie posiada brakujących wartości.

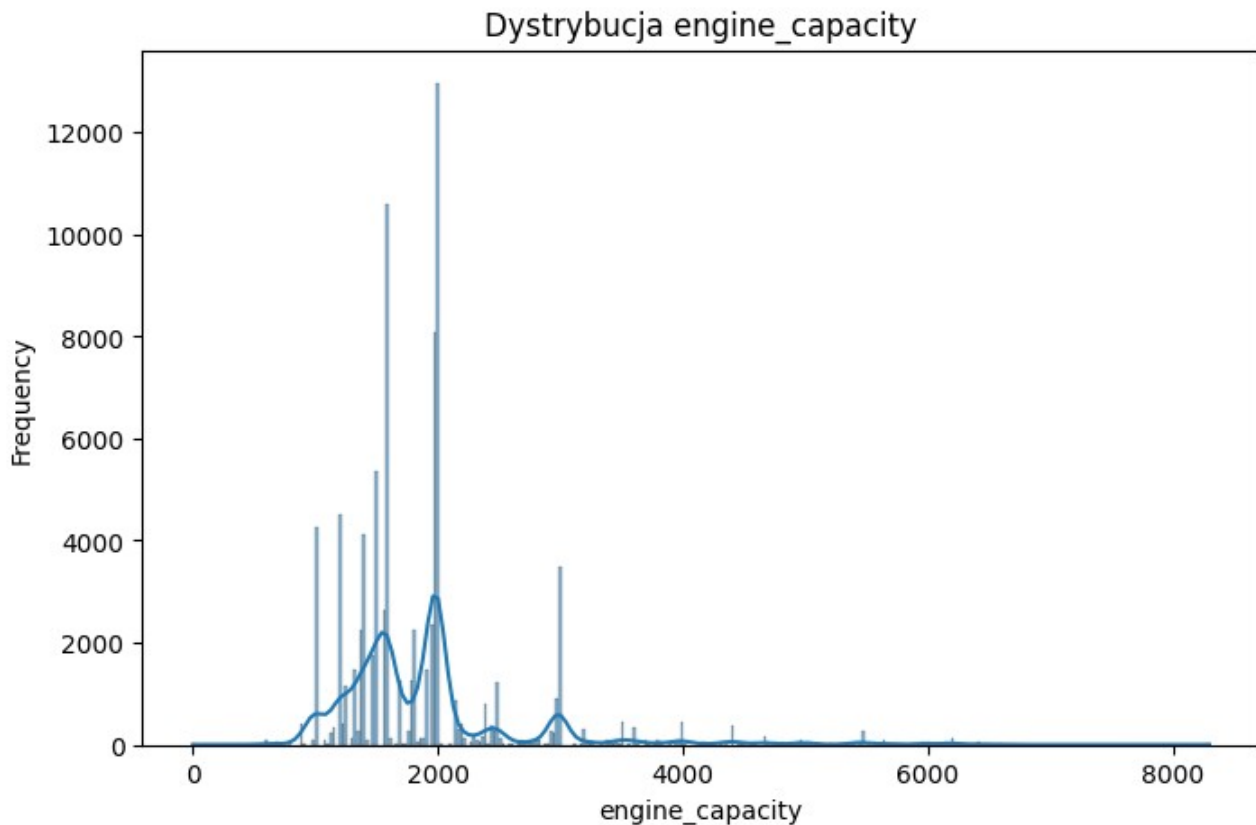
Kolumny:

1. brand – marka samochodu
2. model – nazwa konkretnego modelu
3. price_in_pln – cena w PLN
4. mileage – przebieg w kilometrach (np. „133 760 km”)
5. gearbox – rodzaj skrzyni biegów, manualna lub automatyczna
6. engine_capacity – pojemność silnika w centymetrach sześciennych (np. „1970 cm3”)
7. fuel_type – rodzaj paliwa (np. „Benzyna”, „Benzyna+LPG”, „Diesel”)
8. city – miejscowość wystawienia oferty
9. voivodship – województwo wystawienia oferty
10. year – rok produkcji samochodu

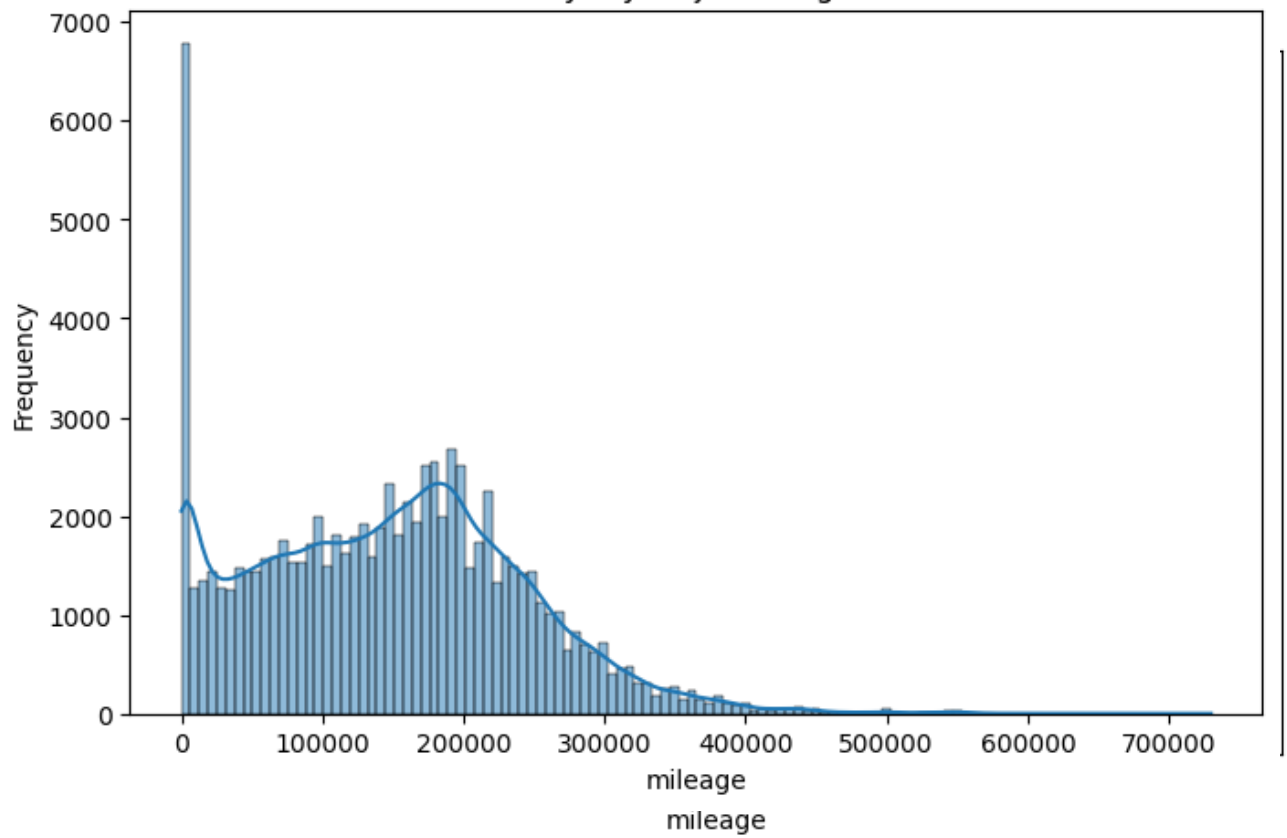
4 kolumny to kolumny numeryczne – price_in_pln, mileage, engine_capacity, year. Pozostałe kolumny to kolumny katagoryczne.

Po wstępnym formatowaniu kolumn okazało się że dataset zawiera wiersze w których dane w poszczególnych kolumnach są pomieszane. Po ich usunięciu zostało 85677 wierszy.

Rozkład zmiennych numerycznych



Dystrybucja mileage



Dystrybucja year

