

Eksploracyjna analiza danych - laboratorium 4

December 7, 2023

```
[1]: import pandas as pd  
import numpy as np
```

1 Zadania do samodzielnego wykonania

1.1 Zadanie 1

Załaduj plik csv do obiektu DataFrame o nazwie df:

<https://storage.googleapis.com/esmartdata-courses-files/dash-course/data.csv>

Ustaw indeks na pierwszą kolumnę. Wydrukuj zawartość obiektu df do konsoli.

1.2 Zadanie 2

Znajdź listę zawierającą nazwy wszystkich kolumn obiektu df (zad 1). Wydrukuj listę do konsoli.

1.3 Zadanie 3

Z obiektu df usuń kolumnę o nazwie New_Price i wydrukuj pierwsze 5 wierszy.

1.4 Zadanie 4

Znajdź liczbę brakujących wartości dla każdej zmiennej w obiekcie df. Wydrukuj wynik do konsoli.

1.5 Zadanie 5

Usuń wszystkie wiersze z brakami danych w obiekcie df. Użyj parametru `inplace=True` i wypisz informacje o obiekcie df.

1.6 Zadanie 6

Zastąp wszystkie duże litery w nazwach kolumn małymi, np. - Name -> name - Fuel_Type -> fuel_type

Wyświetl pierwsze pięć wierszy za pomocą metody `pd.DataFrame.head()` i funkcji `print()`.

1.7 Zadanie 7

Znajdź rozkład zmiennej `owner_type` i wydrukuj do konsoli.

1.8 Zadanie 8

Spójrz na rozkład zmiennej `engine`. Zauważ, że zmienna `engine` jest typu `object`. Usuń ostatnie 3 znaki z każdego elementu tej zmiennej i przekonwertuj na typ `int`. Wydrukuj pierwsze pięć wierszy za pomocą metody `pd.DataFrame.head()` i funkcji `print()` z kolumn: `name` oraz `engine`.

1.9 Zadanie 9

Spójrz na rozkład zmiennej `power`. Zwróć uwagę, że w tej zmiennej występują brakujące wartości w postaci `null bhp`. Zamień te brakujące wartości na `np.nan`. W tym celu użyj funkcji `np.where()`. W odpowiedzi wydrukuj rozkład zmiennej `power` do konsoli (tylko pierwsze pięć elementów według częstotliwości).

1.10 Zadanie 10

Pogrupuj dane na poziomie roku (`year`) i policz liczbę wierszy dla każdego roku. Wydrukuj wynik do konsoli.