**Szymon Jabłoński**

**Kacper Obrębski**

**Dane: NY-House-Dataset**

Celem jest predykcja ceny nieruchomości.

Zbiór danych zawiera 4802 instancje i kolumny określające:

BROKERTITLE: Title of the broker

TYPE: Type of the house

PRICE: Price of the house

BEDS: Number of bedrooms

BATH: Number of bathrooms

PROPERTYSQFT: Square footage of the property

ADDRESS: Full address of the house

STATE: State of the house

MAIN\_ADDRESS: Main address information

ADMINISTRATIVE\_AREA\_LEVEL\_2: Administrative area level 2 information

LOCALITY: Locality information

SUBLOCALITY: Sublocality information

STREET\_NAME: Street name

LONG\_NAME: Long name

FORMATTED\_ADDRESS: Formatted address

LATITUDE: Latitude coordinate of the house

LONGITUDE: Longitude coordinate of the house

Skrypt importuje niezbędne biblioteki, wczytuje dane z pliku CSV do DataFrame, zmienia nazwy kolumn, definiuje typy kolumn i konwertuje je na odpowiednie typy. Następnie sprawdza kształt DataFrame, sprawdza obecność brakujących wartości, usuwa duplikaty i sprawdza, czy istnieją wartości mniejsze od zera w określonych kolumnach. Skrypt również usuwa znak '#' z kolumny 'FORMATTED\_ADDRESS' i na końcu koduje pewne kolumny za pomocą LabelEncoder.