

Exercise: Cleaning NYC Property Sales

This lesson will test the user on data cleaning.

WE'LL COVER THE FOLLOWING ^

- NYC Property Sales
 - 1. Change values
 - Input
 - Output
 - Exercise
 - 2. Missing values
 - Input
 - Output
 - Exercise
 - 3. Duplicate values
 - Input
 - Output
 - Exercise
 - 4. Outliers
 - Input
 - Output
 - Exercise

NYC Property Sales

In this lesson, you are going to clean the [NYC Property Sales Dataset](#). This dataset is a record of every building or building unit (apartment, etc.) sold in the New York City property market over a 12-month period.

In the below exercises, you will be writing functions for every task. The

functions will receive a dataframe `df` as an input argument. Your task will be to perform the required operations and return the dataframe.

1. Change values

In the `BOROUGH` column, the values range from 1 to 5. Change these according to the following

1 --> `Manhattan`

2 --> `Bronx`

3 --> `Brooklyn`

4 --> `Queens`

5 --> `Staten Island`

Input

A dataframe

Output

Dataframe with `BOROUGH` values changed.

Exercise

```
def change_values(df):  
    ## Write code here  
  
    pass
```



2. Missing values

Remove rows that have missing values in `SALE PRICE` column.

Input

A dataframe

Output

Dataframe with rows that have missing values in `SALE PRICE` column

Dataframe with rows that have missing values in `SALE PRICE` column removed.

Exercise

```
def remove_missing(df):  
    ## Write code here  
  
    pass
```



3. Duplicate values

Remove rows that have duplicates in each column.

Input

A dataframe

Output

Dataframe with duplicated rows removed.

Exercise

```
def remove_duplicates(df):  
    ## Write code here  
  
    pass
```



4. Outliers

Remove outliers from the dataframe by checking outliers using inter-quartile ranges in the following columns:

- `RESIDENTIAL UNITS`
- `COMMERCIAL UNITS`
- `TOTAL UNITS`
- `LAND SQUARE FEET`

- LAND SQUARE FEET
- GROSS SQUARE FEET
- YEAR BUILT

However, the quartile range should be between 0.10 and 0.90.

Input

A dataframe

Output

Dataframe with Outliers removed.

Exercise

```
def remove_outliers(df):  
    ## Write code here  
  
    pass
```



In the next lesson, we will review the solutions to these exercises.