

# MLS Data Project

This document provides an overview of the functions and objects currently built for the project. The structure of the project is likely to change as we add more functions and objects. The repository is located here: [https://github.com/Hadidata/MLS\\_Data\\_Project](https://github.com/Hadidata/MLS_Data_Project). You will see a total of 5 python files. This documentation will only focus on Functions.py, DataExtract.py and UnitTest.py, the other files are strictly for functionality testing.

## The individual Files

### *Functions.py*

The project was initially designed to have a series of functions do the data extractions for MLS listings, however as data requirements grew a need for classes became apparent. The functions module now holds general function used by the different classes in the project.

#### **Function: isNumber(Str)**

##### *Input*

Str: Input as a String or Integer

This function checks if an expression is a number or not. Returns true if it is a number and false if it is not.

Example:

```
import Functions as fn
print(fn.isNumber(12))
#True
print(fun.isNumber("aa"))
#False
```

### *DataExtract.py*

#### **Class: liveMls(Url,Pages)**

##### *input*

Url: The Url of the city quadrant based on the liveMls website

Pages: The number of pages to extract from the city quadrant if left blank goes to the max number of pages in the city quadrant

This is the main class to extract data from the website: <https://www.livrealestate.ca/>, the following properties and methods outline the data you can extract from the class

### Property: getPageUrl()

This property generates the URL for each page based on the number of pages provided. This property does not communicate with the real estate website

Example:

```
import DataExtract

Url = "https://www.livrealestate.ca/calgary-city-centre/"
Pages = 3
MlsData = DataExtract.liveMls(Url, Pages)
print(MlsData.getPageUrl)
# ['https://www.livrealestate.ca/calgary-city-centre/?pg=1',
'https://www.livrealestate.ca/calgary-city-centre/?pg=2',
'https://www.livrealestate.ca/calgary-city-centre/?pg=3']
```

### Method: getMlsUrl(progress=False)

*Progress*: Boolean input used for testing purposes to output the progress of the scrap default is false

This Method returns the list of MLS urls in a specific page.

Examples:

```
# When Progress is False
import DataExtract

Url = "https://www.livrealestate.ca/calgary-city-centre/"
Pages = 1
MlsData = DataExtract.liveMls(Url, Pages)
print(MlsData.getMlsUrl(progress=False))
# ['https://www.livrealestate.ca/idx/151-somme-manor-sw-calgary-ab-t2t-6j3/13065335_spid/', ..]

# When Progress is True
import DataExtract

Url = "https://www.livrealestate.ca/calgary-city-centre/"
Pages = 2
MlsData = DataExtract.liveMls(Url, Pages)
print(MlsData.getMlsUrl(progress=True))
Page 1 of 2 extracted
Page 2 of 2 extracted
# ['https://www.livrealestate.ca/idx/151-somme-manor-sw-calgary-ab-t2t-6j3/13065335_spid/', ..]
```

## Method: getPriceChangeHistory(progress=False)

*Progress*: Boolean input used for testing purposes to output the progress of the scrap default is false

This Method is used to extract the price change history of a particular property, if there is not price change all rows are empty except for the link row. Returns a dataframe of the link to the MLS listing the Date of the price change, the Old price, the New price and the percentage change

## Example

```
import DataExtract
# When Progress is False
Url = "https://www.livrealstate.ca/calgary-city-centre/"
Pages = 2
MlsData = DataExtract.liveMls(Url, Pages)
print(MlsData.getPriceChangeHistory(progress=False))
```

	Link	...	Percent	Change
0	https://www.livrealstate.ca/idx/151-somme-man...	...		None
1	https://www.livrealstate.ca/idx/650-10-st-sw-...	...		None
2	https://www.livrealstate.ca/idx/2923-14-ave-n...	...		None
3	https://www.livrealstate.ca/idx/1905-11-ave-s...	...		None
4	https://www.livrealstate.ca/idx/139-20-ave-ne...	...		None
5	https://www.livrealstate.ca/idx/823-19-ave-sw...	...		None
..				

```
import DataExtract
# When Progress is True
Url = "https://www.livrealstate.ca/calgary-city-centre/"
Pages = 2
MlsData = DataExtract.liveMls(Url, Pages)
print(MlsData.getPriceChangeHistory(progress=True))
```

Page 1 of 2 extracted  
Page 2 of 2 extracted  
Page 1 of 24 extracted -- Price History  
Page 2 of 24 extracted -- Price History  
Page 3 of 24 extracted -- Price History  
Page 4 of 24 extracted -- Price History  
..

	Link	...	Percent	Change
0	https://www.livrealstate.ca/idx/151-somme-man...	...		None
1	https://www.livrealstate.ca/idx/650-10-st-sw-...	...		None
2	https://www.livrealstate.ca/idx/2923-14-ave-n...	...		None
3	https://www.livrealstate.ca/idx/1905-11-ave-s...	...		None
4	https://www.livrealstate.ca/idx/139-20-ave-ne...	...		None
5	https://www.livrealstate.ca/idx/823-19-ave-sw...	...		None
..				

### Method: getRoomInfo(progress=False)

Progress: Boolean input used for testing purposes to output the progress of the scrap default is false

This Method returns information about an MLS listing rooms, if no information is provided about the room. A blank row is returned with only the link populated. The dataframe returned includes the Link, Rooms, Level, Size and Description.

```
import DataExtract
# When Progress is False
Url = "https://www.livrealstate.ca/calgary-city-centre/"
Pages = 2
MlsData = DataExtract.liveMls(Url, Pages)
print(MlsData.getRoomInfo(progress=False))
```

	Link	...	Description
0	https://www.livrealstate.ca/idx/151-somme-man...	...	
1	https://www.livrealstate.ca/idx/151-somme-man...	...	
2	https://www.livrealstate.ca/idx/151-somme-man...	...	
3	https://www.livrealstate.ca/idx/151-somme-man...	...	
4	https://www.livrealstate.ca/idx/151-somme-man...	...	
..	...	...	...
160	https://www.livrealstate.ca/idx/1817-11-ave-s...	...	
161	https://www.livrealstate.ca/idx/1817-11-ave-s...	...	
162	https://www.livrealstate.ca/idx/1817-11-ave-s...	...	
163	https://www.livrealstate.ca/idx/1817-11-ave-s...	...	
164	https://www.livrealstate.ca/idx/1817-11-ave-s...	...	

```
import DataExtract
# When Progress is True
Url = "https://www.livrealstate.ca/calgary-city-centre/"
Pages = 2
MlsData = DataExtract.liveMls(Url, Pages)
print(MlsData.getRoomInfo(progress=True))
```

Page 1 of 2 extracted  
Page 2 of 2 extracted  
Page 1 of 24 extracted -- Room Info  
Page 2 of 24 extracted -- Room Info  
Page 3 of 24 extracted -- Room Info  
..

	Link	...	Description
0	https://www.livrealstate.ca/idx/151-somme-man...	...	
1	https://www.livrealstate.ca/idx/151-somme-man...	...	
2	https://www.livrealstate.ca/idx/151-somme-man...	...	
3	https://www.livrealstate.ca/idx/151-somme-man...	...	
4	https://www.livrealstate.ca/idx/151-somme-man...	...	
..	...	...	...
160	https://www.livrealstate.ca/idx/1817-11-ave-s...	...	
161	https://www.livrealstate.ca/idx/1817-11-ave-s...	...	

```
162 https://www.livrealestate.ca/idx/1817-11-ave-s... ...
163 https://www.livrealestate.ca/idx/1817-11-ave-s... ...
164 https://www.livrealestate.ca/idx/1817-11-ave-s... ...
```

**Private Method: \_\_getLastPage(Link)**

Link: The link should be the first page of a city quadrant.

This Method is only used inside the class to get the page number of the last page in a city quadrant.