

PROPERTY TAX PROJECT

REPORT

BI TEAM:

Manager/Supervisor:

Mr. Shane Hubbard

Senior Analyst:

Mr. Robert Borger

KEERTHI JAJULA

Initial Approach to project:

1. To get the property tax titles and tables for each url.
2. Automate the process of checking for each state county in a way that the domain name remains same, just the state and county names change.
3. Use County_list.csv for getting the links in following form:

<https://www.propertyshark.com/mason/info/Property-Taxes/State/County-county/>

Write those links to csv and check if state-counties have property tax tables if no, return error.

4. As not all state counties have property tax tables separate the existing links from others, use it for automating the process of finding the tax tables

Issues faced during the process:

Reading and writing files:

Issue: Reading the writing files is major part of this process and data was not in a usable format most times as data is pulled in and out of the code

Solution: required data cleaning to get the right format, data cleaning is very important here.

| | | | | | | | | | | | | | | | | | | | | | |
|----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | h | t | t | p | s | : | / | / | w | w | w | . | p | r | o | p | e | r | t | y | s |
| 2 | | | | | | | | | | | | | | | | | | | | | |
| 3 | h | t | t | p | s | : | / | / | w | w | w | . | p | r | o | p | e | r | t | y | s |
| 4 | | | | | | | | | | | | | | | | | | | | | |
| 5 | h | t | t | p | s | : | / | / | w | w | w | . | p | r | o | p | e | r | t | y | s |
| 6 | | | | | | | | | | | | | | | | | | | | | |
| 7 | h | t | t | p | s | : | / | / | w | w | w | . | p | r | o | p | e | r | t | y | s |
| 8 | | | | | | | | | | | | | | | | | | | | | |
| 9 | h | t | t | p | s | : | / | / | w | w | w | . | p | r | o | p | e | r | t | y | s |
| 10 | | | | | | | | | | | | | | | | | | | | | |
| 11 | h | t | t | p | s | : | / | / | w | w | w | . | p | r | o | p | e | r | t | y | s |
| 12 | | | | | | | | | | | | | | | | | | | | | |
| 13 | h | t | t | p | s | : | / | / | w | w | w | . | p | r | o | p | e | r | t | y | s |
| 14 | | | | | | | | | | | | | | | | | | | | | |
| 15 | h | t | t | p | s | : | / | / | w | w | w | . | p | r | o | p | e | r | t | y | s |
| 16 | | | | | | | | | | | | | | | | | | | | | |
| 17 | h | t | t | p | s | : | / | / | w | w | w | . | p | r | o | p | e | r | t | y | s |
| 18 | | | | | | | | | | | | | | | | | | | | | |
| 19 | h | t | t | p | s | : | / | / | w | w | w | . | p | r | o | p | e | r | t | y | s |
| 20 | | | | | | | | | | | | | | | | | | | | | |
| 21 | h | t | t | p | s | : | / | / | w | w | w | . | p | r | o | p | e | r | t | y | s |
| 22 | | | | | | | | | | | | | | | | | | | | | |
| 23 | h | t | t | p | s | : | / | / | w | w | w | . | p | r | o | p | e | r | t | y | s |
| 24 | | | | | | | | | | | | | | | | | | | | | |

1.a Figure: Links not in a proper format



| |
|---|
| https://www.propertyshark.com/mason/info/Property-Taxes/state/county/ |
| https://www.propertyshark.com/mason/info/Property-Taxes/AL/Bibb-County/ |
| https://www.propertyshark.com/mason/info/Property-Taxes/AL/Blount-County/ |
| https://www.propertyshark.com/mason/info/Property-Taxes/AL/Calhoun-County/ |
| https://www.propertyshark.com/mason/info/Property-Taxes/AL/Cherokee-County/ |
| https://www.propertyshark.com/mason/info/Property-Taxes/AL/Chilton-County/ |
| https://www.propertyshark.com/mason/info/Property-Taxes/AL/Clay-County/ |
| https://www.propertyshark.com/mason/info/Property-Taxes/AL/Cleburne-County/ |
| https://www.propertyshark.com/mason/info/Property-Taxes/AL/Coosa-County/ |
| https://www.propertyshark.com/mason/info/Property-Taxes/AL/Cullman-County/ |
| https://www.propertyshark.com/mason/info/Property-Taxes/AL/DeKalb-County/ |
| https://www.propertyshark.com/mason/info/Property-Taxes/AL/Etowah-County/ |

1.b Figure: Usable format of links

Before (1.a) and after (1.b) data cleaning

Working with huge dataset:

Issue: Reading the complete data into the lists, gave some errors and hard to find where I went wrong.

Solution: created a small dataset and it was easier to find the errors and rectify them, use the code for huge dataset, it was a good idea.

Exceptions/Errors:

Issue: Index Error list out of range is caused during the code execution.

Solution: Used Try-Except code block to catch exceptions and proceed with the code after error.

Connection requests:

Issue: As lot a links from the same domain are requesting for data, faced connection timed out or connection error issues

Solution: Using Beautiful Soup and requests libraries worked very well in this case, other libraries such as urllib can be used as well.

Future Use:

As there was no previous data for property tax, can be a good use for Data Analysis, Visualization & Business Insights.

Final Thoughts:

Used the initial approach, got the expected and desired results. This data is reliable for future use.