BU Spark!
Fuqing Wang, Tian Chen, Zixiang Wei, and Xiao Lu

Cambridge Eviction Study Project Proposal: Deliverable 1

Data collected:
1.Cambridge crime reports from 2009 to 2019.
2.Cambridge property dataset which provides much valuable information
3.Cambridge Housing code violation
4.Cambridge rent-occupied housing units census

We also collect some less relevant datasets such as the report of sanitary inspections, Cambridge rodent violation and so on. This data will be used later to help us have a better understanding of the case and find the correlation.

Questions answered:
1.Besides the arrears of rent, illegal action during the tenancy is another big factor that causes eviction.
We are attempting to resolve questions 4, 5 and 6 during our PR meeting.

Current Issues:
Something went wrong with the data collection process - the scraping tool is not usable, thus we are unable to scrap the new data. Based on our investigation, the webpage has been updating but the scraping tool has not. A new script - specifically a newer version of the pageparse.py file is needed to get new data.
Talked with PR, the scraper needs to be reworked if we do not find any further solutions.

```
C:\Windows\system32\cmd.exe                                                    —    □

Microsoft Windows [Version 10.0.18362.657]
(c) 2019 Microsoft Corporation. All rights reserved.

C:\Users\18656\LEARNING\CS506\project>py -3 -m scraper data/2020/Jan
Connected to database
Adding Files...
<_io.BufferedReader name='data/2020/Jan/20H84SP000019.html'>
Unable to load data from file 20H84SP000019.html: 'NoneType' object has no attribute 'find_all'
Traceback (most recent call last):
  File "C:\Users\18656\AppData\Local\Programs\Python\Python38-32\lib\runpy.py", line 193, in _run_module_as_main
    return _run_code(code, main_globals, None,
  File "C:\Users\18656\AppData\Local\Programs\Python\Python38-32\lib\runpy.py", line 86, in _run_code
    exec(code, run_globals)
  File "C:\Users\18656\LEARNING\CS506\project\scraper.py", line 167, in <module>
    main()
  File "C:\Users\18656\LEARNING\CS506\project\scraper.py", line 40, in main
    parse_files(directory, c, conn)
  File "C:\Users\18656\LEARNING\CS506\project\scraper.py", line 86, in parse_files
    p_list, d_list = parse_page(parser, file, c, conn, p_list, d_list)
TypeError: cannot unpack non-iterable NoneType object
```

Future Work:

If we are able to get a usable version of scraper, we plan to catch up with other groups as soon as possible, the following works are included:

1. Collect new data and combine them with existing data in .csv file

2. Clean all collected data, basically filter the noise and missing values

3. Submit our processed data and a complete report that outlines our questions answered and findings

In the worst case, we may have to rewrite the scraper on our own before these tasks