Data Assignment
Time assigned: May 1, 2023
Time due: May 3, 2023
Rita Qifan Yang

Websites that I attempted to scrape:
Wineonline.ca (used for test run) https://www.wineonline.ca/wine-cellar.html
Colliers Montreal Listings
https://www.collierscanada.com/en-ca/properties?gad=1&gclid=CjwKCAjwxr2iBhBJEiwAdXECw
w3RYsGWW1mv3XGmsVVmIcxg-F6oHsVQEUDBYF2tdGtjQucX3gCpQhoCxT8QAvD_BwE#so
rt=relevancy&f:listingtype=[All%20Listings]&f:recenttransactions=[0]&f:location=Quebec%20%3
E%20Greater%20Montreal
Apartments.com Montreal Listings https://www.apartments.com/montreal-qc/
EV Montreal Listings https://montreal.evrealestate.com/inventaire
LePage Montreal Listings 'https://www.royallepage.ca/en/qc/montreal/properties/1/
Louer.ca https://louer.ca/en/montreal?bbox=-73.66876,45.43711,-73.40319,45.55888&p=3

	Allows fetching	Verification token	API requests	Next embedded	Run code once	Results
wine	yes	no	N/A	html	All pages	**wine.csv**
Colliers	yes	yes	yes	html	1 page	**colliers.csv**
Apt	no	no	no	N/A	N/A	N/A
EV	yes	no	N/A	javascript	All pages	**ev.csv**
LePage	yes	no	N/A	html	All pages	**lepage.csv**
Louer	yes	no	no	N/A	N/A	N/A

Based on different obstacles, I used different methods and tools to scrape the websites, still
debugging some and working on the rest. For more details on methods and tools that I used,
you can find the current source code (python) in "**source_code.zip.**"

All of the result csv-files can be found in the folder "**results**."
Other files:
Properties.xlsx: rough notes I took in excel while gathering information in the first step
 Notes:
1. Included properties units = 1 so I don't repeat searches
2. Included properties < 40 units that belongs to a qualifying company (who has >40 units
   property somewhere)

Helper.sh: shell script to sort via column 1 and eliminate unnecessary content
sorted_final.csv: contains the final result (owner, address, contact)