|  |  |
| --- | --- |
| Date | Activity |
| 2022-04-09 | * Accessing [ReferenceUSA (now Data Axle) - Carnegie Mellon University (exlibrisgroup.com)](https://cmu.primo.exlibrisgroup.com/discovery/fulldisplay?docid=alma991019889173004436&context=L&vid=01CMU_INST:01CMU&lang=en&search_scope=MyInst_and_CI&adaptor=Local%20Search%20Engine&tab=Everything&query=any,contains,data%20axle&mode=Basic) * Click “US Historical Business” 2019-2021, Major Industry Group = 58-EATING & DRINKING PLACES, Pittsburgh PA. 🡪 6423 results. * There are 100 pages. I must download at max 3 pages at a time (MAX 250 records allow at a time) P1-3,4-6,6-9,10-12,13-15,16-18,19-21,22-24,25-27 … 40-42. I stop at 42 because the website starts to think I am a bot and asking to complete Captcha every navigation. * I sign up for an account and save my search. * Experiencing lagging and long loading time 🡪 1 page download at a time |
| 2022-04-10 | Finally obtained all 100 pages. 40 csv file |
| 2022-04-13 | * Begin writing code “RestaurantSeperator.py” * Attributes to Keep = Version Year, IUSA Number,Company Name, Address, City, State, ZIP Code, |
| 2022-04-14 | * FIX: Add field into the code   df\_tempo = df\_tempo[["Version Year","IUSA Number", "Company Name", "Address", "City", "State", "ZIP Code","Primary SIC Description"]]   * Obtain output: * Begin writing code “PivotYear.ipynb” * Create ArcGIS project file * Set projection to be “WGS 1984 Web Mercator (Auxiliary Sphere)” for future story map * Back to the cleaning process. I notice there are some strange “Primary SIC Description” like Automobile Repairing & Service. |
|  |  |