|  |  |
| --- | --- |
| 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 1 field into the code, therefore … * df\_tempo = df\_tempo[["Version Year","IUSA Number", "Company Name", "Address", "City", "State", "ZIP Code","Primary SIC Description"]] * Obtain output:The “PghResturantAllYear.csv” is to be manipulated later. * Finish <https://github.com/korawichkavee/GIS-Restaurant_COVID_Walkable_PGH/blob/main/RestaurantSeperator.py> * Begin writing code “PivotYear.ipynb” to brings year rows to columns * Create ArcGIS project file * Set projection to be “WGS 1984 Web Mercator (Auxiliary Sphere)” for future story map * Back to the cleaning process. Result df.to\_csv("IUSAYear" + ".csv",index=False) * I notice there are some strange “Primary SIC Description” like “Automobile Repairing & Service”. TO BE REMOVE LATER * Push to Github my code <https://github.com/korawichkavee/GIS-Restaurant_COVID_Walkable_PGH/blob/main/PivotYear.ipynb> * Data Join: df\_join = pd.merge(df, df\_rest, on='IUSA Number', how='left') * Get more rows than expected. This is because some restaurant update their name in the same year🡪 Drop a question on Stack overflow <https://stackoverflow.com/questions/71879993/drop-duplicates-even-more-for-a-specific-column-with-latest-value/71880191#71880191> |
| 2022-04-15 | * Update my code. Use answer from stack overflow 🡪 Perfectly 2497 places (rows) as it should be = len(IUSAYear.csv). * Clean up Round 1 remove fields ls = ['Museums','Churches','Golf Courses','Convention & Meeting Facilities & Svc','Doors-Repairing','Book Dealers-Retail', 'Bags-Paper Plastic & Fabric Suppliers','Physicians & Surgeons Equip & Supls-Whls','Home Demonstration-Merchandise', 'Gift Shops', 'Fire Departments', 'Golf Courses-Miniature', 'Theatres-Live', 'Animal Hospitals','Real Estate','Business Management Consultants','Furniture-Dealers-Retail', 'Website Design Service','Video Games', 'Kitchen Cabinets & Equipment-Household', 'Crafts', 'Mailing & Shipping Services'] * Round 2 ls = ['Dairy Products-Wholesale','Hotels & Motels','Grocers-Retail','Bowling Centers','Convenience Stores','Candy & Confectionery-Manufacturers','Caterers','Non-Profit Organizations','Ice Cream & Frozen Desserts-Distr (Whls)','Restaurant Management','Brewers (Mfrs)','Wholesale Clubs','Food Products & Manufacturers', 'Farm Markets','Ham Specialty Stores','Food Service-Distributors (Whls)','Bartending Service','Florists-Wholesale','Bakers-Wholesale', 'Delivery Service','Poultry Processing Plants (Mfrs)', 'Skating Rinks','Music & Live Entertainment','Restaurant Equipment & Supplies (Whls)','Laundry & Garment Services NEC', 'Wedding Supplies & Services','Social Service & Welfare Organizations', 'Inns','Arts Organizations & Information', 'Liquors-Retail','Department Stores', 'Halls & Auditoriums', 'Casinos','Canned & Cured Fish & Seafoods (Mfrs)', 'Attorneys','Chemical Research', 'Resorts', 'Smoke Shops & Supplies','Home Builders', 'Miscellaneous Retail Stores NEC','Wineries (Mfrs)', 'Pet Washing & Grooming','General Merchandise-Retail', 'Garbage Collection', 'Retail Shops','Fruit Baskets-Gift','Motels & Hotels Reservations', 'Entertainers'] * Round 3 ls = ['Arcades' ,'Party Centers'] * 2497 🡪 2467 🡪 2247🡪 2245 |