Original doc at:
https://docs.google.com/document/d/1njh2UBXySirf8ZyK8SHn9LnwG-CoyqJHG9sTcCZg03o/edit#

Original issues with the CSV-based parse/export

- Order is not always correct, even in the original xslx
    - In New York State individuals may vote if they are:
      Those who are currently on parole can still vote if they have a Certificate of Relief from Disabilities or a Certificate of Good Conduct that restores their voting rights. See pages 280–282 for more information.
    - Getting Out and Staying Out (GOSO) is a reentry program for justice- involved men 16–24 years old. Fewer than 10 percent of GOSO participants return to jail, as compared to a national average of 67 percent for the age group. GOSO uses early intervention within the
      555 Bergen Avenue, 3rd Floor, Bronx, NY 10455 Telephone: 347.584.8601 A www.networkssi.org
- Some things even seem to be missing from the original xslx
    - Cultural Resources
- Page numbers can get in the way
    - THE COLLEGE INITIATIVE 45
    - 62 COVENANT HOUSE / UNDER 21 NEW YORK
    - 26 131 West 25th Street, 12th Floor, New York, NY 10001 Telephone: 212.803.5700
- The 2 columns sometimes interfere
    - ARGUS COMMUNITY, INC. / ACCESS PROGRAM BOOM!HEALTH
    - QVCMH FOR JCAP, INC.
      ODYSSEY HOUSE
- Multi-line titles sometimes get split
    - BROOKLYN PUBLIC LIBRARY—
      ADULT LITERACY, PRE-HSE AND ESOL
    - 166 Organizations for Formerly
      167 None
      168 None
      169 Incarcerated People
      170 None
      171 None
- ~~One row can be added to multiple organizations (/categories etc) if orgs etc aren't reset properly~~
    - ~~FAMILY & CULTURAL PROGRAMMING including story hours for infants, toddlers, preschoolers, and school- age children, science and art workshops for children of all ages, homework help and tutoring, musical performances, and much more. Adults can find an array of cultural programs, including lectures, concerts, films, exhibitions, and author talks. Visit nypl.org to find~~
- ~~Descriptions sometimes start from the middle, and potentially of an unrelated organization…~~

  - ~~ming to adults and youth affected by the justice system. Exodus offers youth on probation ages 16–24 the ARCHES program, and also offers the NextSTEPS and CommonUnity programs for youth affected by the justice system. Participants receive mentoring, TASC preparation, job preparation, a stipend, hot meals, and MetroCards. For people returning from jail or prison, Exodus provides workshops and case management to help secure housing, substance treatment, health referrals and benefits, support groups, and employment. The Exodus Wellness Center is an 822 OASAS–licensed substance abuse outpatient program that provides mental health assessments, individual counseling, and groups such as Anger Management, Relapse~~

Script issues

- Multiple URLs appear a lot in the data, but I don't think are supported by our schema
- Multiple pieces of location information on the same line (different lines in the original PDF)
  - When the name and another piece of information (usually address) appear on the same line, the name is ignored
  - ~~When URL and address are together, they each appear in the other's field~~
  - Multiple phone numbers on the same line don't work
  - ~~It's possible if some of the description is on the same line as a URL, that part of the description will just be ignored [*Doesn't seem to happen*]~~
  - Phones should be parsed including the prefix and extensions etc, but that's a problem when the address, URL, etc are on the same line

Manual changes

Cleanup

Input

- ~~Add back Cultural Resources category~~
- Address all the script warnings
  - ~~Non-parsed addresses~~
    - ~~"One Pierrepont Plaza" -> "1 Pierrepont Plaza"~~
    - ~~"Boerum Place and Schermerhorn Street" -> "99 Schermerhorn Street"~~
  - ~~URL in unexpected format~~
    - ~~www.bronxdefenders.org/programs/new-york-immigrant-family-unity-project~~
    - ~~http://www1.nyc.gov/site/ocdv/programs/family-justice- centers.page~~
- ~~Delete the organization "JOHNNY PEREZ"...~~
- ~~Rejoin organization names (+maybe other kinds of rows) that were split due to being 2 lines~~

    - ~~Based on “Description: None”?~~
- ~~Rejoin split URLs~~
    - ~~www1.nyc.gov/site/dhs/shelter/families/families-with~~
    - ~~www.bronxdefenders.org/programs/new-york-immigrant-family-~~
    - ~~www.nypl.org/help/community-outreach/services-for-adults/~~
    - ~~www.grandcentralneighborhood.org/services/mainchance-~~

Output

- Regex find-replace fixes:
    - ~~Remove “({3,}[\d]+)|([\d]+ {3,})” - page numbers~~
    - ~~Replace “([a-z])\- ([a-z])” with $1$2 - multiline descriptions with split words~~

Post-parse additions

- Use geocoding API to add long/lats
- Maybe use same organization even when it appears under different categories