Instructions on how to run covid-poverty-final.py:
Please use the following commands to execute this script on Dumbo. Make sure to order inputs as shown below:

spark-submit --conf spark.pyspark.python=/share/apps/python/3.6.5/bin/python covid-poverty-final.py
/user/<netid>/COVID-19_Daily_Counts_of_Cases__Hospitalizations__and_Deaths.csv
/user/<netid>/DHS_Daily_Report.csv
/user/<netid>/Individual_Census_by_Borough__Community_District__and_Facility_Type.csv
/user/<netid>/New_York_City_Seasonally_Adjusted_Employment.csv /user/<netid>/dates.csv

Outputs:
The script will generate the following csv files that were used to generate visualization found in report.

excess_homelessness.csv
Schema: date, cases, cumulative cases, hospitalizations, cumulative hospitalizations, deaths, cumulative deaths,
4-yr avg. # of adults in homeless shelters, 4-yr avg. # of children in homeless shelters,
4-yr avg. # of individuals in homeless shelters, # of adults in homeless shelter in 2020,
# of children in homeless shelter in 2020, # of individuals in homeless shelter in 2020, adults p-score,
children p-score, individuals p-score

excess_homelessness_borough.csv*
Schema: date, cases, cumulative cases, hospitalizations, cumulative hospitalizations, deaths, cumulative deaths,
bx cases, bx cumulative cases, bx hospitalizations, bx cumulative hospitalizations, bx deaths, bx cumulative deaths,
bk cases, bk cumulative cases, bk hospitalizations, bk cumulative hospitalizations, bk deaths, bk cumulative deaths,
mn cases, mn cumulative cases, mn hospitalizations, mn cumulative hospitalizations, mn deaths, mn cumulative
deaths, qn cases, qn cumulative cases, qn hospitalizations, qn cumulative hospitalizations, qn deaths, qn
cumulative deaths, si cases, si cumulative cases, si hospitalizations, si cumulative hospitalizations, si deaths, si
cumulative deaths, 2-yr avg. # of individuals in bx homeless shelters, 2-yr avg. # of individuals in bk homeless
shelters, 2-yr avg. # of individuals in mn homeless shelters, 2-yr avg. # of individuals in qn homeless shelters,
2-yr avg. # of individuals in si homeless shelters, # of individuals in bx homeless shelter in 2020,
# of individuals in bk homeless shelter in 2020, # of individuals in mn homeless shelter in 2020, # of individuals in
qn homeless shelter in 2020, # of individuals in si homeless shelter in 2020, bx p-score, bk p-score, mn p-score,
qn p-score, si p-score

* bx: Bronx; bk = Brooklyn; mn: Manhattan; qn: Queens; si: Staten Island

excess_unemployment.csv*
Schema: date, cumulative cases, cumulative hospitalizations, cumulative deaths, 10-yr avg. non-farm, 10-yr avg.
private, 10-yr avg. education & health, 10-yr avg. professional & business services, 10-yr avg. trade, transportation,
& utilities, 10-yr avg. public admin, 10-yr avg. financial services, 10-yr avg. leisure & hospitality, 10-yr avg. retail,
10-yr avg. wholesale, 10-yr avg. manufacturing, 10-yr avg. information, 2020 non-farm, 2020 private, 2020
education & health, 2020 professional & business services, 2020 trade, transportation, & utilities, 2020 public
admin, 2020 financial services, 2020 leisure & hospitality, 2020 retail, 2020 wholesale, 2020 manufacturing, 2020
information, p-score non-farm, p-score private, p-score education & health, p-score professional & business
services, p-score trade, transportation, & utilities, p-score public admin, p-score financial services, p-score leisure
& hospitality, p-score retail, p-score wholesale, p-score manufacturing, p-score information

* each industry feature represents employment census for a given month