Ameya Shere

11/8/20

Data Ingest Details

Steps for Manufacturing Employment Data Ingest

1. Download data to local machine. To accomplish this, I navigated to the BLS.gov page for each state's manufacturing employment data, changed the year range to include data from 1990-2020 and clicked the blue ".CSV" button to download the data as a CSV file. All data was extracted on November 8th, 2020. Links to each state's manufacturing employment data BLS.gov page can be found in the following section.
2. Zip all files together on local machine into "Data.zip".
3. pscp (I am using a Windows computer) "Data.zip" file to DUMBO home directory: *pscp Data.zip as12366@dumbo.es.its.nyu.edu:/home/as12366*
4. Unzip "Data.zip" into folder called "Data": *unzip Data.zip*
5. Make new HDFS directory for data files called "Project_Data": *hdfs dfs -mkdir /user/as12366/Project_Data*
6. cd into "Data" folder on DUMBO and put all CSV files into "Project_Data": *hdfs dfs -put *.csv /user/as12366/Project_Data*

Links to Data Sources

State	Link to data source
Alabama	https://beta.bls.gov/dataViewer/view/timeseries/SMS01000003000000001
Alaska	https://beta.bls.gov/dataViewer/view/timeseries/SMS02000003000000001
Arizona	https://beta.bls.gov/dataViewer/view/timeseries/SMS04000003000000001
Arkansas	https://beta.bls.gov/dataViewer/view/timeseries/SMS05000003000000001
California	https://beta.bls.gov/dataViewer/view/timeseries/SMS06000003000000001
Colorado	https://beta.bls.gov/dataViewer/view/timeseries/SMS08000003000000001
Connecticut	https://beta.bls.gov/dataViewer/view/timeseries/SMS09000003000000001
Delaware	https://beta.bls.gov/dataViewer/view/timeseries/SMS10000003000000001
Florida	https://beta.bls.gov/dataViewer/view/timeseries/SMS12000003000000001
Georgia	https://beta.bls.gov/dataViewer/view/timeseries/SMS13000003000000001
Hawaii	https://beta.bls.gov/dataViewer/view/timeseries/SMS15000003000000001

Idaho	https://beta.bls.gov/dataViewer/view/timeseries/SMS16000003000000001
Illinois	https://beta.bls.gov/dataViewer/view/timeseries/SMS17000003000000001
Indiana	https://beta.bls.gov/dataViewer/view/timeseries/SMS18000003000000001
Iowa	https://beta.bls.gov/dataViewer/view/timeseries/SMS19000003000000001
Kansas	https://beta.bls.gov/dataViewer/view/timeseries/SMS20000003000000001
Kentucky	https://beta.bls.gov/dataViewer/view/timeseries/SMS21000003000000001
Louisiana	https://beta.bls.gov/dataViewer/view/timeseries/SMS22000003000000001
Maine	https://beta.bls.gov/dataViewer/view/timeseries/SMS23000003000000001
Maryland	https://beta.bls.gov/dataViewer/view/timeseries/SMS24000003000000001
Massachusetts	https://beta.bls.gov/dataViewer/view/timeseries/SMS25000003000000001
Michigan	https://beta.bls.gov/dataViewer/view/timeseries/SMS26000003000000001
Minnesota	https://beta.bls.gov/dataViewer/view/timeseries/SMS27000003000000001
Mississippi	https://beta.bls.gov/dataViewer/view/timeseries/SMS28000003000000001
Missouri	https://beta.bls.gov/dataViewer/view/timeseries/SMS29000003000000001
Montana	https://beta.bls.gov/dataViewer/view/timeseries/SMS30000003000000001
Nebraska	https://beta.bls.gov/dataViewer/view/timeseries/SMS31000003000000001
Nevada	https://beta.bls.gov/dataViewer/view/timeseries/SMS32000003000000001
New Hampshire	https://beta.bls.gov/dataViewer/view/timeseries/SMS33000003000000001
New Jersey	https://beta.bls.gov/dataViewer/view/timeseries/SMS34000003000000001
New Mexico	https://beta.bls.gov/dataViewer/view/timeseries/SMS35000003000000001
New York	https://beta.bls.gov/dataViewer/view/timeseries/SMS36000003000000001
North Carolina	https://beta.bls.gov/dataViewer/view/timeseries/SMS37000003000000001
North Dakota	https://beta.bls.gov/dataViewer/view/timeseries/SMS38000003000000001

Ohio	https://beta.bls.gov/dataViewer/view/timeseries/SMS39000003000000001
Oklahoma	https://beta.bls.gov/dataViewer/view/timeseries/SMS40000003000000001
Oregon	https://beta.bls.gov/dataViewer/view/timeseries/SMS41000003000000001
Pennsylvania	https://beta.bls.gov/dataViewer/view/timeseries/SMS42000003000000001
Rhode Island	https://beta.bls.gov/dataViewer/view/timeseries/SMS44000003000000001
South Carolina	https://beta.bls.gov/dataViewer/view/timeseries/SMS45000003000000001
South Dakota	https://beta.bls.gov/dataViewer/view/timeseries/SMS46000003000000001
Tennessee	https://beta.bls.gov/dataViewer/view/timeseries/SMS47000003000000001
Texas	https://beta.bls.gov/dataViewer/view/timeseries/SMS48000003000000001
Utah	https://beta.bls.gov/dataViewer/view/timeseries/SMS49000003000000001
Vermont	https://beta.bls.gov/dataViewer/view/timeseries/SMS50000003000000001
Virginia	https://beta.bls.gov/dataViewer/view/timeseries/SMS51000003000000001
Washington	https://beta.bls.gov/dataViewer/view/timeseries/SMS53000003000000001
West Virginia	https://beta.bls.gov/dataViewer/view/timeseries/SMS54000003000000001
Wisconsin	https://beta.bls.gov/dataViewer/view/timeseries/SMS55000003000000001
Wyoming	https://beta.bls.gov/dataViewer/view/timeseries/SMS56000003000000001