

CS 510 - Data Engineering

11 March 2021

Answers for Data Integration Activity

Part A - Aggregate Census Data to County Level

	County	State	Population	Poverty	PerCapitaIncome
0	Loudoun	Virginia	374558	3.884375	50391.015625
1	Washington	Oregon	572071	10.446154	34970.817308
2	Harlan	Kentucky	27548	33.318182	16010.363636
3	Malheur	Oregon	30421	24.414286	17966.428571

Part B - Simplify the COVID Data

	County	State	TotalCases	Dec2020Cases	TotalDeaths	Dec2020Deaths
0	Loudoun	Virginia	2496450	376223	35820.0	4729.0
1	Washington	Oregon	2157339	424620	22455.0	3860.0
2	Harlan	Kentucky	205984	38959	3994.0	506.0
3	Malheur	Oregon	453634	82916	7770.0	1465.0

Part C - Integrate COVID Data with ACS Data

	County	State	Population	...	Dec2020Cases	TotalDeaths	Dec2020Deaths
1	Washington	Oregon	572071	...	424620	3925.212080	3860.0
3	Malheur	Oregon	30421	...	82916	25541.566681	1465.0

Part D - Analysis

	State	County	TotalCases	...	Population	Poverty	PerCapitaIncome
0	Oregon	Washington	7.931876e+05	...	2564646	13.422243	30729.954380
1	Oregon	Jackson	7.416558e+05	...	1722755	18.493882	26107.727059
2	Oregon	Klamath	3.396892e+05	...	66018	18.930000	23712.400000
3	Oregon	Douglas	9.380390e+05	...	1440952	13.097015	33581.874627
4	Oregon	Marion	9.534415e+05	...	2018526	19.951429	24578.778022
5	Oregon	Multnomah	4.280168e+05	...	788459	15.730588	36739.558824

Analysis of all counties in the State of Oregon

COVID total cases vs. % population in poverty - 0.29747113439452827
COVID total deaths vs. % population in poverty - 0.20946190596896622
COVID total cases vs. Per Capita Income level - -0.18950461546988256
COVID total cases vs. Per Capita Income level - -0.07405908604882054
COVID cases during December 2020 vs. % population in poverty - 0.0570595387576331
COVID deaths during December 2020 vs. % population in poverty - 0.05571798839796443
COVID cases during December 2020 vs. Per Capita Income level - 0.2851092192782961
COVID cases during December 2020 vs. Per Capita Income level - 0.2664467664558328

	County	TotalCases	...	Poverty	PerCapitaIncome
0	Snohomish	4.587665e+05	...	8.934899	36148.134228
1	Cook	1.180935e+06	...	17.988948	33006.301287
2	Orange	8.418201e+05	...	13.075722	35621.433298
3	Maricopa	1.267895e+06	...	16.130955	30761.265642
4	Los Angeles	1.122602e+06	...	17.323803	31389.413867

Analysis of all counties in the United States

COVID total cases vs. % population in poverty - 0.16690037795100285
COVID total deaths vs. % population in poverty - 0.199728642737955
COVID total cases vs. Per Capita Income level - -0.1870637430384339
COVID total cases vs. Per Capita Income level - -0.1413631086967846
COVID cases during December 2020 vs. % population in poverty - -0.027528196894307654
COVID deaths during December 2020 vs. % population in poverty - -0.02187824035470475
COVID cases during December 2020 vs. Per Capita Income level - 0.2021547054177172
COVID cases during December 2020 vs. Per Capita Income level - 0.21840972654694812