Task 3 – Spark SQL DataFrame Operations

Q1. Check for null values in each column in all the files and report the null value count for each.

```
Null/NaN values for csv/state-abbreviation-edited.csv
+-----+------------+
|state|abbreviation|
+-----+------------+
|    0|           0|
+-----+------------+

Null/NaN values for csv/state-abbreviation.csv
+-----+------------+
|state|abbreviation|
+-----+------------+
|    0|           0|
+-----+------------+

Null/NaN values for csv/state-area.csv
+-----+----+
|state|area|
+-----+----+
|    0|   0|
+-----+----+

Null/NaN values for csv/state-population.csv
+-----+----+----+----------+
|state|ages|year|population|
+-----+----+----+----------+
|    0|   0|   0|        20|
+-----+----+----+----------+
```

Q2. Report the total population of New Jersey from 2001-2010.

```
+-----+-----+----+----------+
|state| ages|year|population|
+-----+-----+----+----------+
|   NJ|total|2001|   8492671|
|   NJ|total|2002|   8552643|
|   NJ|total|2003|   8601402|
|   NJ|total|2004|   8634561|
|   NJ|total|2005|   8651974|
|   NJ|total|2006|   8661679|
|   NJ|total|2007|   8677885|
|   NJ|total|2008|   8711090|
|   NJ|total|2009|   8755602|
|   NJ|total|2010|   8802707|
+-----+-----+----+----------+
```

Q3. For the year of 199x (x being the last digit of your NetId), merge "states-area.csv" and "state-population.csv" to get the state wide area and population for the desired year, and compute the state wide area per person for each state. (last digit of net id = 1)

year	state	area	population	area_per_person
1991	AL	52423	4099156	0.01279
1991	AK	656425	570193	1.15123
1991	AZ	114006	3788576	0.03009
1991	AR	53182	2383144	0.02232
1991	CA	163707	30470736	0.00537
1991	CO	104100	3387119	0.03073
1991	CT	5544	3302895	0.00168
1991	DE	1954	683080	0.00286
1991	DC	68	600870	1.1E-4
1991	FL	65758	13369798	0.00492
1991	GA	59441	6653005	0.00893
1991	HI	10932	1136754	0.00962
1991	ID	83574	1041316	0.08026
1991	IL	57918	11568964	0.00501
1991	IN	36420	5616388	0.00648
1991	IA	56276	2797613	0.02012
1991	KS	82282	2498722	0.03293
1991	KY	40411	3722328	0.01086
1991	LA	51843	4253279	0.01219
1991	ME	35387	1237081	0.02861
1991	MD	12407	4867641	0.00255
1991	MA	10555	6018470	0.00175
1991	MI	96810	9400446	0.0103
1991	MN	86943	4440859	0.01958
1991	MS	48434	2598733	0.01864
1991	MO	69709	5170800	0.01348
1991	MT	147046	809680	0.18161
1991	NE	77358	1595919	0.04847
1991	NV	110567	1296172	0.0853
1991	NH	9351	1109929	0.00842
1991	NJ	8722	7814676	0.00112
1991	NM	121593	1555305	0.07818
1991	NY	54475	18122510	0.00301
1991	NC	53821	6784280	0.00793
1991	ND	70704	635753	0.11121
1991	OH	44828	10945762	0.0041
1991	OK	69903	3175440	0.02201
1991	OR	98386	2928507	0.0336
1991	PA	46058	11982164	0.00384
1991	RI	1545	1010649	0.00153
1991	SC	32007	3570404	0.00896

Q4. For the year of 199x (x being the last digit of your NetId), merge “states-area.csv” and “state-population.csv” with the help of “state-abbrievations-edited.csv” to get the state wide area and population for the desired year, and compute the state wide area per person for each state. (last digit of netid = 1)

year	state	area	population	area_per_person
1991	AL	52423	4099156	0.01279
1991	AK	656425	570193	1.15123
1991	AZ	114006	3788576	0.03009
1991	AR	53182	2383144	0.02232
1991	CA	163707	30470736	0.00537
1991	CO	104100	3387119	0.03073
1991	CT	5544	3302895	0.00168
1991	DE	1954	683080	0.00286
1991	DC	68	600870	1.1E-4
1991	FL	65758	13369798	0.00492
1991	GA	59441	6653005	0.00893
1991	HI	10932	1136754	0.00962
1991	ID	83574	1041316	0.08026
1991	IL	57918	11568964	0.00501
1991	IN	36420	5616388	0.00648
1991	IA	56276	2797613	0.02012
1991	KS	82282	2498722	0.03293
1991	KY	40411	3722328	0.01086
1991	LA	51843	4253279	0.01219
1991	ME	35387	1237081	0.02861
1991	MD	12407	4867641	0.00255
1991	MA	10555	6018470	0.00175
1991	MI	96810	9400446	0.0103
1991	MN	86943	4440859	0.01958
1991	MS	48434	2598733	0.01864
1991	MO	69709	5170800	0.01348
1991	MT	147046	809680	0.18161
1991	NE	77358	1595919	0.04847
1991	NV	110567	1296172	0.0853
1991	NH	9351	1109929	0.00842
1991	NJ	8722	7814676	0.00112
1991	NM	121593	1555305	0.07818
1991	NY	54475	18122510	0.00301
1991	NC	53821	6784280	0.00793
1991	ND	70704	635753	0.11121
1991	OH	44828	10945762	0.0041
1991	OK	69903	3175440	0.02201
1991	OR	98386	2928507	0.0336
1991	PA	46058	11982164	0.00384
1991	RI	1545	1010649	0.00153
1991	SC	32007	3570404	0.00896
1991	PR	3515	NaN	null