In [1]:
```
import pandas as pd
from sqlalchemy import create_engine
import numpy as np
```

In [2]:
```
legislators_file = "Resources/legislators-current-terms.csv"
legislators_df = pd.read_csv(legislators_file)
legislators_df.head()
```

Out[2]:

	address	bioguide	caucus	chamber	class	contact_form	district	end	fax	last	...	party_a
0	NaN	B000944	NaN	NaN	NaN	NaN	13.0	1995-01-03	NaN	NaN	...	
1	NaN	B000944	NaN	NaN	NaN	NaN	13.0	1997-01-03	NaN	NaN	...	
2	NaN	B000944	NaN	NaN	NaN	NaN	13.0	1999-01-03	NaN	NaN	...	
3	NaN	B000944	NaN	NaN	NaN	NaN	13.0	2001-01-03	NaN	NaN	...	
4	NaN	B000944	NaN	NaN	NaN	NaN	13.0	2003-01-03	NaN	NaN	...	

5 rows × 23 columns

In [3]:
```
no_health_ins_file = "Resources/Population_Without_Health_Insurance_Coverag
no_health_ins_df = pd.read_csv(no_health_ins_file)
no_health_ins_df.head()
```

Out[3]:

	State	Medicaid Expansion State? Yes (Y) or No (N)1	2013 Uninsured - Number in thousands	2013 Uninsured - Number - Margin of Error2 (±)	2013 Uninsured - Percent	2013 Uninsured - Percent - Margin of Error2 (±)	2014 Uninsured - Number	2014 Uninsured - Number - Margin of Error2 (±)	2 Uninsu - Per
0	United States	NaN	45,181	200	14.5	0.1	36,670	190	
1	Alabama	N	645	17	13.6	0.4	579	17	
2	Alaska	N	132	7	18.5	1.0	122	6	
3	Arizona	Y	1,118	24	17.1	0.4	903	18	
4	Arkansas	Y	465	14	16.0	0.5	343	13	

5 rows × 22 columns

In [4]:
```
states_file = "Resources/states.csv"
states_df = pd.read_csv(states_file)
states_df.head()
```

Out[4]:

	state_name	abbreviation
0	Alabama	AL
1	Alaska	AK
2	Arizona	AZ
3	Arkansas	AR
4	California	CA

In [5]:
```
no_health_df = pd.merge(no_health_ins_df, states_df, how = "inner", left_on
no_health_df.head()
```

Out[5]:

	State	Medicaid Expansion State? Yes (Y) or No (N)1	2013 Uninsured - Number in thousands	2013 Uninsured - Number - Margin of Error2 (±)	2013 Uninsured - Percent	2013 Uninsured - Percent - Margin of Error2 (±)	2014 Uninsured - Number	2014 Uninsured - Number - Margin of Error2 (±)	2 Uninsu - Per
0	Alabama	N	645	17	13.6	0.4	579	17	
1	Alaska	N	132	7	18.5	1.0	122	6	
2	Arizona	Y	1,118	24	17.1	0.4	903	18	
3	Arkansas	Y	465	14	16.0	0.5	343	13	
4	California	Y	6,500	57	17.2	0.2	4,767	47	

5 rows × 24 columns

In [6]:
```
new_no_health_df = no_health_df[['abbreviation', '2014 Uninsured - Number']
new_no_health_df.head()
```

Out[6]:

	abbreviation	2014 Uninsured - Number
0	AL	579
1	AK	122
2	AZ	903
3	AR	343
4	CA	4,767

In [7]:
```
new_no_health_df = new_no_health_df.rename(columns={'abbreviation': 'State'
new_no_health_df.head()
```

Out[7]:

	State	Uninsured Population
0	AL	579
1	AK	122
2	AZ	903
3	AR	343
4	CA	4,767

In [8]:
```
legislators_df = legislators_df[['party', 'state']].copy()
legislators_df.head()
```

Out[8]:

	party	state
0	Democrat	OH
1	Democrat	OH
2	Democrat	OH
3	Democrat	OH
4	Democrat	OH

In [9]:
```
dem_df = legislators_df.query('party == "Democrat"')
print(len(legislators_df))
print(len(dem_df))
```

```
3092
1566
```

In [10]:
```
dem_df = dem_df.groupby('state').count().reset_index()
dem_df.head()
```

Out[10]:

	state	party
0	AL	9
1	AZ	14
2	CA	238
3	CO	25
4	CT	44

In [11]:
```
dem_df = dem_df.rename(columns={'state': 'State', 'party': 'Democrats'})
dem_df.head()
```

Out[11]:

	State	Democrats
0	AL	9
1	AZ	14
2	CA	238
3	CO	25
4	CT	44

In [12]:
```
rep_df = legislators_df.query('party == "Republican"')
print(len(legislators_df))
print(len(rep_df))
```

```
3092
1450
```

In [13]:
```
rep_df = rep_df.groupby('state').count().reset_index()
rep_df.head()
```

Out[13]:

	state	party
0	AK	28
1	AL	42
2	AR	21
3	AS	2
4	AZ	34

In [14]:
```
rep_df = rep_df.rename(columns={'state': 'State', 'party': 'Republicans'})
rep_df.head()
```

Out[14]:

	State	Republicans
0	AK	28
1	AL	42
2	AR	21
3	AS	2
4	AZ	34

In [15]:
```
ind_df = legislators_df.query('party == "Independent"')
print(len(legislators_df))
print(len(ind_df))
```

```
3092
12
```

In [16]:
```
ind_df = ind_df.groupby('state').count().reset_index()
ind_df.head()
```

Out[16]:

	state	party
0	ME	1
1	MP	1
2	VT	10

In [17]:
```
ind_df = ind_df.rename(columns={'state': 'State', 'party': 'Independent'})
ind_df.head()
```

Out[17]:

	State	Independent
0	ME	1
1	MP	1
2	VT	10

In [18]:
```
dem_rep_merge = pd.merge(dem_df, rep_df,  on = "State", how='outer')
dem_rep_merge.head()
```

Out[18]:

	State	Democrats	Republicans
0	AL	9.0	42.0
1	AZ	14.0	34.0
2	CA	238.0	91.0
3	CO	25.0	20.0
4	CT	44.0	NaN

In [19]:
```
dem_rep_ind_merge = pd.merge(dem_rep_merge, ind_df,  on = "State", how = "l
dem_rep_ind_merge.head().sort_values('State')
```

Out[19]:

	State	Democrats	Republicans	Independent
0	AL	9.0	42.0	NaN
1	AZ	14.0	34.0	NaN
2	CA	238.0	91.0	NaN
3	CO	25.0	20.0	NaN
4	CT	44.0	NaN	NaN

In [20]:
```
dem_rep_ind_merge = dem_rep_ind_merge.fillna(0)
```

In [21]:
```
state_df = pd.merge(dem_rep_ind_merge, new_no_health_df,  on = "State", how

state_df.head().sort_values('State')
```

Out[21]:

	State	Democrats	Republicans	Independent	Uninsured Population
0	AL	9.0	42.0	0.0	579
1	AZ	14.0	34.0	0.0	903
2	CA	238.0	91.0	0.0	4,767
3	CO	25.0	20.0	0.0	543
4	CT	44.0	0.0	0.0	245

In [22]:
```
state_df = state_df.fillna(0)
state_df.head()
```

Out[22]:

	State	Democrats	Republicans	Independent	Uninsured Population
0	AL	9.0	42.0	0.0	579
1	AZ	14.0	34.0	0.0	903
2	CA	238.0	91.0	0.0	4,767
3	CO	25.0	20.0	0.0	543
4	CT	44.0	0.0	0.0	245

In [23]:
```
state_df = state_df.set_index("State")
```

In [24]:
```
state_df = state_df.drop(['PR', 'GU', 'MP', 'AS', 'VI', 'DC'])
state_df.head()
```

Out[24]:

State	Democrats	Republicans	Independent	Uninsured Population
AL	9.0	42.0	0.0	579
AZ	14.0	34.0	0.0	903
CA	238.0	91.0	0.0	4,767
CO	25.0	20.0	0.0	543
CT	44.0	0.0	0.0	245

In [25]:
```
sate_df = state_df.reset_index(inplace=True)
state_df.head()
```

Out[25]:

	State	Democrats	Republicans	Independent	Uninsured Population
0	AL	9.0	42.0	0.0	579
1	AZ	14.0	34.0	0.0	903
2	CA	238.0	91.0	0.0	4,767
3	CO	25.0	20.0	0.0	543
4	CT	44.0	0.0	0.0	245

In [26]:
```
state_df.dtypes
```

Out[26]:
```
State                   object
Democrats              float64
Republicans            float64
Independent            float64
Uninsured Population    object
dtype: object
```

In [27]:
```
state_df['Uninsured Population'] = pd.to_numeric(state_df['Uninsured Popula
state_df.head()
```

Out[27]:

	State	Democrats	Republicans	Independent	Uninsured Population
0	AL	9.0	42.0	0.0	579
1	AZ	14.0	34.0	0.0	903
2	CA	238.0	91.0	0.0	4767
3	CO	25.0	20.0	0.0	543
4	CT	44.0	0.0	0.0	245

In [28]:
```
state_df.dtypes
```

Out[28]:
```
State                   object
Democrats              float64
Republicans            float64
Independent            float64
Uninsured Population     int64
dtype: object
```

In [29]:
```
state_df['Uninsured Population'] = state_df['Uninsured Population'] * 1000
state_df.head()
```

Out[29]:

	State	Democrats	Republicans	Independent	Uninsured Population
0	AL	9.0	42.0	0.0	579000
1	AZ	14.0	34.0	0.0	903000
2	CA	238.0	91.0	0.0	4767000
3	CO	25.0	20.0	0.0	543000
4	CT	44.0	0.0	0.0	245000

In [30]:
```
state_df['Uninsured Population'] = state_df['Uninsured Population'].astype(
state_df['Independent'] = state_df['Independent'].astype(int)
state_df['Republicans'] = state_df['Republicans'].astype(int)
state_df['Democrats'] = state_df['Democrats'].astype(int)
state_df.head()
```

Out[30]:

	State	Democrats	Republicans	Independent	Uninsured Population
0	AL	9	42	0	579000
1	AZ	14	34	0	903000
2	CA	238	91	0	4767000
3	CO	25	20	0	543000
4	CT	44	0	0	245000

In [31]:
```
state_df.dtypes
```

Out[31]:
```
State                   object
Democrats                int32
Republicans              int32
Independent              int32
Uninsured Population     int32
dtype: object
```

Connect to local database

In [39]:
```
state_df = state_df.rename(columns={'State': 'States', 'Uninsured Populatic
state_df.head()
```

Out[39]:

	States	Democrats	Republicans	Independent	Uninsured_Population
0	AL	9	42	0	579000
1	AZ	14	34	0	903000
2	CA	238	91	0	4767000
3	CO	25	20	0	543000
4	CT	44	0	0	245000

In [51]:
```
state_df.columns = ['states','democrats','republicans','independent','unins
```

In [52]:
```
rds_connection_string = "postgres:postgres@localhost:5432/etl_project_db"
engine = create_engine(f'postgresql://{rds_connection_string}')
```

Check for tables

In [53]:
```
# Confirm tables
engine.table_names()
```

Out[53]:
```
['etl_project']
```

Use pandas to load csv converted DataFrame into database

In [54]:
```
state_df.to_sql(name='etl_project', con=engine, if_exists='append', index=F
```

In [55]:
```
pd.read_sql_query('select * from etl_project', con=engine).head()
```

Out[55]:

	id	states	democrats	republicans	independent	uninsured_population
0	1	AL	9	42	0	579000
1	2	AZ	14	34	0	903000
2	3	CA	238	91	0	4767000
3	4	CO	25	20	0	543000
4	5	CT	44	0	0	245000

In [ ]:
```

```