Import Library

In [1]:
```
import os
import requests
import seaborn as sns
import pandas as pd
import kaggle
from pathlib import Path
from kaggle.api.kaggle_api_extended import KaggleApi
import matplotlib.pyplot as plt
```

Introduction

The objective of this analysis is to explore and gain insights into diabetes-related indicators in the United States by integrating data from three distinct sources: 'U.S. Chronic Disease Indicators: Diabetes' obtained from the CDC API, 'USA Hospitals' from Kaggle, and the 'Medicare Diabetes Prevention Program' retrieved from the local directory. Diabetes, a prevalent chronic disease, poses significant challenges to public health, making it imperative to examine associated factors and potential preventive measures.

The CDC dataset provides granular information on chronic disease indicators, including diabetics-related questions, across different U.S. states. The 'USA Hospitals' dataset offers details about hospitals nationwide, facilitating the exploration of the healthcare landscape concerning diabetes. Meanwhile, the 'Medicare Diabetes Prevention Program' dataset focuses on organizations engaged in diabetes prevention, offering a valuable perspective on preventive initiatives.

The analysis commences with data retrieval from APIs and local directories, ensuring diverse data sources are incorporated. Merging operations will be performed multiple times, linking datasets based on location abbreviations, state codes, and program-related information. Subsequent steps involve data aggregation, pivoting, and transformation to unveil patterns and trends within the integrated datasets. Visualizations will aid in presenting key findings effectively.

This comprehensive approach, integrating diverse datasets and employing various analytical techniques, aims to provide a holistic understanding of diabetes indicators, healthcare infrastructure, and prevention programs in the United States. The subsequent sections will delve into the specifics of data processing, analysis, and interpretation to derive meaningful insights for informed decision-making.

Data Access and Formats

To commence the analysis, we need to access data from three distinct sources: the CDC API for 'U.S. Chronic Disease Indicators: Diabetes,' the Kaggle API for 'USA Hospitals,' and the local directory for 'Medicare Diabetes Prevention Program.'

1. U.S. Chronic Disease Indicators: Diabetes (CDC API)

Fetch and load data from the CDC API using the requests library. Convert the JSON response to a Pandas

urther processing.

In [2]:

```
# CDC API URL
cdc_api_url = "https://data.cdc.gov/resource/f8ti-h92k.json"

# Fetch data from CDC API
response_cdc = requests.get(cdc_api_url)
data_cdc = response_cdc.json()

# Convert data to Pandas DataFrame
df_cdc = pd.DataFrame(data_cdc)

print(df_cdc)
```

	yearstart	yearend	locationabbr	locationdesc	datasource	topic	\
0	2011	2011	NV	Nevada	BRFSS	Diabetes	
1	2016	2016	NV	Nevada	BRFSS	Diabetes	
2	2019	2019	NV	Nevada	BRFSS	Diabetes	
3	2019	2019	NV	Nevada	BRFSS	Diabetes	
4	2017	2017	NE	Nebraska	BRFSS	Diabetes	
..	...	...	...	...	...	...	
995	2012	2012	AR	Arkansas	NVSS	Diabetes	
996	2014	2014	NJ	New Jersey	NVSS	Diabetes	
997	2019	2019	PA	Pennsylvania	NVSS	Diabetes	
998	2020	2020	AL	Alabama	NVSS	Diabetes	
999	2018	2018	NH	New Hampshire	NVSS	Diabetes	

	question	datavalueunit	\
0	Prevalence of diagnosed diabetes among adults ...	%	
1	Visits to dentist or dental clinic among adult...	%	
2	Prevalence of diagnosed diabetes among adults ...	%	
3	Foot examination among adults aged >= 18 years...	%	
4	Prevalence of high cholesterol among adults ag...	%	
..	...	...	
995	Mortality due to diabetes reported as any list...	cases per 100,000	
996	Mortality due to diabetes reported as any list...	cases per 100,000	
997	Mortality due to diabetes reported as any list...	Number	
998	Mortality due to diabetes reported as any list...	cases per 100,000	
999	Mortality due to diabetes reported as any list...	NaN	

	datavaluetype	datavalue	...	stratificationcategory1	\
0	Age-adjusted Prevalence	10.0	...	Overall	
1	Crude Prevalence	57.8	...	Gender	
2	Crude Prevalence	11.2	...	Gender	
3	Crude Prevalence	NaN	...	Gender	
4	Crude Prevalence	56.9	...	Overall	
..	...	...	...	...	
995	Age-adjusted Rate	60.3	...	Gender	
996	Age-adjusted Rate	56.9	...	Overall	
997	Number	NaN	...	Race/Ethnicity	
998	Crude Rate	29.2	...	Race/Ethnicity	
999	Number	NaN	...	Race/Ethnicity	

	stratification1	locationid	topicid	questionid	\
0	Overall	32	DIA	DIA2_1	
1	Male	32	DIA	DIA8_0	
2	Male	32	DIA	DIA2_1	
3	Male	32	DIA	DIA5_0	
4	Overall	31	DIA	DIA11_1	
..	...	...	...	...	
995	Female	05	DIA	DIA1_1	
996	Overall	34	DIA	DIA1_1	
997	American Indian or Alaska Native	42	DIA	DIA1_1	
998	Asian or Pacific Islander	01	DIA	DIA1_1	
999	Black, non-Hispanic	33	DIA	DIA1_1	

	datavaluetypeid	stratificationcategoryid1	stratificationid1	\
0	AGEADJPREV	OVERALL	OVR	
1	CRDPREV	GENDER	GENM	
2	CRDPREV	GENDER	GENM	
3	CRDPREV	GENDER	GENM	
4	CRDPREV	OVERALL	OVR	
..	...	...	...	
995	AGEADJRATE	GENDER	GENF	
996	AGEADJRATE	OVERALL	OVR	
997	NMBR	RACE	AIAN	
998	CRDRATE	RACE	API	
999	NMBR	RACE	BLK	

```
    datavaluefootnotesymbol                                  datavaluefootnote
0                       NaN                                                NaN
1                       NaN                                                NaN
2                       NaN                                                NaN
3                         -                                No data available
4                       NaN                                                NaN
..                      ...                                                ...
995                     NaN                                                NaN
996                     NaN                                                NaN
997                       ~  Data not shown because of too few respondents ...
998                     NaN                                                NaN
999                       ~  Data not shown because of too few respondents ...

[1000 rows x 23 columns]
```

2. USA Hospitals (Kaggle API)

Utilize the Kaggle API to retrieve the 'USA Hospitals' dataset. You need to have the Kaggle API key configured.

In [3]:
```python
# Set your Kaggle username and API key
kaggle_username = "user36602"
kaggle_api_key = "4baf77179167ba51df68a6b051f8e8a9"

# Save Kaggle API key to a file
kaggle_path = Path.home() / ".kaggle"
kaggle_path.mkdir(exist_ok=True)
api_key_path = kaggle_path / "kaggle.json"
with api_key_path.open("w") as api_key_file:
    api_key_file.write('{"username":"%s", "key":"%s"}' % (kaggle_username, kaggle_api_ke

# Set Kaggle API key environment variable
os.environ["KAGGLE_CONFIG_DIR"] = str(kaggle_path)

# Kaggle dataset URL
dataset_url = "carlosaguayo/usa-hospitals"

# Download dataset
api = KaggleApi()
api.authenticate()
api.dataset_download_files(dataset_url, path=".", unzip=True)

# Load the CSV file into a DataFrame
csv_file_path = Path(".") / "Hospitals.csv"
df_hospitals = pd.read_csv(csv_file_path)

# Display the DataFrame
print(df_hospitals)
```

```
              X          Y  OBJECTID         ID  \
0    -94.945477  29.747620      8497   76777520
1    -82.881843  40.027143      8498  129043230
2    -84.168027  39.774242      8499  130045404
3    -80.632972  41.005169      8500  128844512
4    -84.199398  39.747740      8501  129845417
...         ...        ...       ...        ...
7565 -80.701221  40.372434      8492  128243953
7566 -81.336345  40.401257      8493  127744621
7567 -84.294586  39.331523      8494  128345040
7568 -97.283341  29.112615      8495   13879022
7569 -95.359806  29.711039      8496   76677004

                                                  NAME  \
0     HOUSTON METHODIST SAN JACINTO HOSPITAL ALEXAND...
1                               WOODS AT PARKSIDE,THE
2                          DAYTON CHILDREN'S HOSPITAL
3                  VIBRA HOSPITAL OF MAHONING VALLEY
4              HAVEN BEHAVIORAL SENIOR CARE OF DAYTON
...                                              ...
7565                              LIFE LINE HOSPITAL
7566                           TEN LAKES CENTER, LLC
7567                         LINDNER CENTER OF HOPE
7568                        CUERO COMMUNITY HOSPITAL
7569       HARRIS HEALTH SYSTEM QUENTIN MEASE HOSPITAL

                              ADDRESS          CITY STATE    ZIP  \
0              1700 JAMES BOWIE DRIVE       BAYTOWN    TX  77520
1              349 OLDE RIDENOUR ROAD      COLUMBUS    OH  43230
2                 ONE CHILDRENS PLAZA        DAYTON    OH  45404
3                   8049 SOUTH AVENUE      BOARDMAN    OH  44512
4      ONE ELIZABETH PLACE,E3 SUITE A        DAYTON    OH  45417
...                                ...           ...   ...    ...
7565                200 SCHOOL STREET  WINTERSVILLE    OH  43953
7566  819 NORTH FIRST STREET,3RD FLOOR      DENNISON    OH  44621
7567         4075 OLD WESTERN ROW ROAD         MASON    OH  45040
7568       2550 NORTH ESPLANADE STREET         CUERO    TX  79022
7569          3601 NORTH MACGREGOR WAY       HOUSTON    TX  77004

              ZIP4  ...                  VAL_DATE  \
0     NOT AVAILABLE  ...  2017-12-18T00:00:00.000Z
1     NOT AVAILABLE  ...  2018-04-26T00:00:00.000Z
2     NOT AVAILABLE  ...  2018-04-26T00:00:00.000Z
3     NOT AVAILABLE  ...  2018-04-26T00:00:00.000Z
4     NOT AVAILABLE  ...  2018-04-26T00:00:00.000Z
...             ...  ...                       ...
7565  NOT AVAILABLE  ...  2018-04-26T00:00:00.000Z
7566  NOT AVAILABLE  ...  2018-04-26T00:00:00.000Z
7567  NOT AVAILABLE  ...  2018-04-26T00:00:00.000Z
7568  NOT AVAILABLE  ...  2017-12-18T00:00:00.000Z
7569  NOT AVAILABLE  ...  2017-12-18T00:00:00.000Z

                                              WEBSITE       STATE_ID  \
0     http://www.houstonmethodist.org/locations/san-...  NOT AVAILABLE
1                   http://www.thewoodsatparkside.com/           1815
2     http://www.childrensdayton.org/cms/home/index....           1411
3               http://www.mahoningvalleyhospital.com/           1428
4                  https://dayton.havenbehavioral.com/           1506
...                                                ...            ...
7565                       http://www.llhospital.com/           1493
7566                    http://www.tenlakescenter.com/           1469
7567                  http://lindnercenterofhope.org/           1481
7568                         http://www.cuerohosp.org  NOT AVAILABLE
7569  https://www.harrishealth.org/en/services/locat...  NOT AVAILABLE
```

```
                                    ALT_NAME ST_FIPS  \
0                               NOT AVAILABLE    48.0
1                               NOT AVAILABLE    39.0
2                               NOT AVAILABLE    39.0
3     MAHONING VALLEY HOSPITAL BOARDMAN CAMPUS    39.0
4                               NOT AVAILABLE    39.0
...                                       ...     ...
7565                            NOT AVAILABLE    39.0
7566                            NOT AVAILABLE    39.0
7567                            NOT AVAILABLE    39.0
7568                            NOT AVAILABLE    48.0
7569                            NOT AVAILABLE    48.0

                                OWNER TTL_STAFF   BEDS             TRAUMA  \
0                           NON-PROFIT    -999.0  182.0       NOT AVAILABLE
1                          PROPRIETARY    -999.0   50.0       NOT AVAILABLE
2                           NON-PROFIT    -999.0  155.0  PEDIATRIC LEVEL II
3                          PROPRIETARY    -999.0   45.0       NOT AVAILABLE
4                          PROPRIETARY    -999.0   32.0       NOT AVAILABLE
...                                ...       ...    ...                 ...
7565                       PROPRIETARY    -999.0   36.0       NOT AVAILABLE
7566                       PROPRIETARY    -999.0   16.0       NOT AVAILABLE
7567                        NON-PROFIT    -999.0   34.0       NOT AVAILABLE
7568  GOVERNMENT - DISTRICT/AUTHORITY    -999.0   49.0            LEVEL IV
7569                        NON-PROFIT    -999.0   49.0       NOT AVAILABLE

           HELIPAD
0                Y
1     NOT AVAILABLE
2                Y
3     NOT AVAILABLE
4     NOT AVAILABLE
...             ...
7565  NOT AVAILABLE
7566  NOT AVAILABLE
7567  NOT AVAILABLE
7568              Y
7569  NOT AVAILABLE

[7570 rows x 34 columns]
```

3. Medicare Diabetes Prevention Program (Local Directory)

Load the 'Medicare Diabetes Prevention Program' dataset from the local directory using Pandas.

In [4]:
```
# Load 'Medicare Diabetes Prevention Program' dataset
df_medicare = pd.read_excel('Medicare_Diabetes_Prevention_Program.xlsx')

print(df_medicare)
```

```
                                    Name of Initiative  \
0    Medicare Diabetes Prevention Program (MDPP) Ex...
1    Medicare Diabetes Prevention Program (MDPP) Ex...
2    Medicare Diabetes Prevention Program (MDPP) Ex...
3    Medicare Diabetes Prevention Program (MDPP) Ex...
4    Medicare Diabetes Prevention Program (MDPP) Ex...
..                                                 ...
824  Medicare Diabetes Prevention Program (MDPP) Ex...
825  Medicare Diabetes Prevention Program (MDPP) Ex...
826  Medicare Diabetes Prevention Program (MDPP) Ex...
827  Medicare Diabetes Prevention Program (MDPP) Ex...
828  Medicare Diabetes Prevention Program (MDPP) Ex...

                                    Organization Name  \
0                    Comunidad Saludable De La Montana
1    Hockomock Young Men'S Christian Association Inc.
2    Hockomock Young Men'S Christian Association Inc.
3    Hockomock Young Men'S Christian Association Inc.
4    Hockomock Young Men'S Christian Association Inc.
..                                                 ...
824                    Gobble Shults & Associates Inc
825                      Legacy Salmon Creek Hospital
826               Peacehealth Southwest Medical Center
827                Yakima Neighborhood Health Services
828                           Tanana Chiefs Conference

                             Location Name  \
0       Comunidad Saludable De La Montana
1                     Hockomock Area Ymca
2                     Hockomock Area Ymca
3                     Hockomock Area Ymca
4                     Hockomock Area Ymca
..                                    ...
824         Gobble Shults & Associates Inc
825       Legacy Clinic Salmon Creek - Im
826  Peacehealth Southwest Medical Center
827   Yakima Neighborhood Health Services
828               Tanana Chiefs Conference

                                          Location 1  \
0    39 Ave Rolando Cabanas  Utuado, PR 00641 (18.2...
1    4 Valente Dr Westborough, MA 01581 (42.278878,...
2    60 Main St Whitinsville, MA 01588 (42.110333, ...
3    75 Shore Dr Worcester, MA 01605 (42.30408, -71...
4    766 Main St Worcester, MA 01610 (42.25734, -71...
..                                                 ...
824  3305 Main St Vancouver, WA 98663 (45.645749, -...
825  2101 Ne 139th St Vancouver, WA 98686 (45.72181...
826  2312 Ne 129th St Vancouver, WA 98686 (45.71464...
827  12 S 8th St Yakima, WA 98901 (46.603846, -120....
828  1717 W Cowles St Fairbanks, AK 99701 (64.83227...

      Street Address Line 1 Street Address Line 2          City State  \
0    39 Ave Rolando Cabanas                   NaN         Utuado    PR
1               4 Valente Dr                  NaN    Westborough    MA
2                 60 Main St                  NaN   Whitinsville    MA
3                75 Shore Dr                  NaN      Worcester    MA
4                766 Main St                  NaN      Worcester    MA
..                       ...                  ...            ...   ...
824             3305 Main St               Ste 20      Vancouver    WA
825         2101 Ne 139th St              Ste 460      Vancouver    WA
826         2312 Ne 129th St              Ste 120      Vancouver    WA
827              12 S 8th St                  NaN         Yakima    WA
828         1717 W Cowles St                  NaN      Fairbanks    AK
```

```
     ZIP Code Telephone Number         NPI                 Category  Unique ID
0         641   (787) 698-0073  1780198135  Administrative Location          1
1        1581   (508) 870-1320  1689815383        Community Setting          2
2        1588   (508) 234-8184  1689815383        Community Setting          3
3        1605   (508) 852-6694  1689815383        Community Setting          4
4        1610   (508) 755-6101  1689815383        Community Setting          5
..        ...              ...         ...                      ...        ...
824     98663   (503) 652-5070  1275604399  Administrative Location        825
825     98686   (360) 487-2727  1356357784  Administrative Location        826
826     98686   (360) 546-8900  1134178999  Administrative Location        827
827     98901   (509) 454-4143  1225085624  Administrative Location        828
828     99701   (907) 451-6682  1821201278  Administrative Location        829

[829 rows x 13 columns]
```

These steps ensure that data from various sources is accessible in a suitable format for subsequent merging and analysis.

Data Merging

Data integration is crucial for a comprehensive analysis. In this section, we merge information from the 'U.S. Chronic Disease Indicators: Diabetes,' 'USA Hospitals,' and 'Medicare Diabetes Prevention Program' datasets. The merging process is executed iteratively to ensure seamless integration of relevant columns.

1. Merging U.S. Chronic Disease Indicators with USA Hospitals

Merge the 'LocationAbbr' column from the CDC dataset with the 'STATE' column from the USA Hospitals dataset. This facilitates the linkage of diabetes indicators with hospital information.

In [5]:
```
# Merge CDC and USA Hospitals datasets
df_merged_cdc_hospitals = pd.merge(df_cdc, df_hospitals, left_on='locationabbr', right_o

print(df_merged_cdc_hospitals)
```

```
       yearstart yearend locationabbr locationdesc datasource     topic  \
0           2011    2011           NV       Nevada      BRFSS  Diabetes
1           2011    2011           NV       Nevada      BRFSS  Diabetes
2           2011    2011           NV       Nevada      BRFSS  Diabetes
3           2011    2011           NV       Nevada      BRFSS  Diabetes
4           2011    2011           NV       Nevada      BRFSS  Diabetes
...          ...     ...          ...          ...        ...       ...
149822      2014    2014           NJ   New Jersey       NVSS  Diabetes
149823      2014    2014           NJ   New Jersey       NVSS  Diabetes
149824      2014    2014           NJ   New Jersey       NVSS  Diabetes
149825      2014    2014           NJ   New Jersey       NVSS  Diabetes
149826      2014    2014           NJ   New Jersey       NVSS  Diabetes

                                                 question      datavalueunit  \
0       Prevalence of diagnosed diabetes among adults ...                 %
1       Prevalence of diagnosed diabetes among adults ...                 %
2       Prevalence of diagnosed diabetes among adults ...                 %
3       Prevalence of diagnosed diabetes among adults ...                 %
4       Prevalence of diagnosed diabetes among adults ...                 %
...                                                   ...               ...
149822  Mortality due to diabetes reported as any list...  cases per 100,000
149823  Mortality due to diabetes reported as any list...  cases per 100,000
149824  Mortality due to diabetes reported as any list...  cases per 100,000
149825  Mortality due to diabetes reported as any list...  cases per 100,000
149826  Mortality due to diabetes reported as any list...  cases per 100,000

                datavaluetype datavalue  ...                 VAL_DATE  \
0       Age-adjusted Prevalence      10.0  ...  2018-04-20T00:00:00.000Z
1       Age-adjusted Prevalence      10.0  ...  2018-03-16T00:00:00.000Z
2       Age-adjusted Prevalence      10.0  ...  2018-03-16T00:00:00.000Z
3       Age-adjusted Prevalence      10.0  ...  2018-03-16T00:00:00.000Z
4       Age-adjusted Prevalence      10.0  ...  2018-03-16T00:00:00.000Z
...                         ...       ...  ...                       ...
149822       Age-adjusted Rate      56.9  ...  2018-04-09T00:00:00.000Z
149823       Age-adjusted Rate      56.9  ...  2018-04-09T00:00:00.000Z
149824       Age-adjusted Rate      56.9  ...  2018-04-09T00:00:00.000Z
149825       Age-adjusted Rate      56.9  ...  2018-04-09T00:00:00.000Z
149826       Age-adjusted Rate      56.9  ...  2018-04-09T00:00:00.000Z

                                                 WEBSITE       STATE_ID  \
0       http://www.nellis.af.mil/units/nellismedicalce...  NOT AVAILABLE
1                                 www.strosehospitals.org    4576-HOS-20
2       https://locations.dignityhealth.org/dignity-he...     8594-HOS-1
3       http://www.postacutemedical.com/our-facilities...     8682-HOS-0
4                                    http://www.umcsn.com     666-HOS-64
...                                                   ...            ...
149822                          http://www.caperegional.com        NJ10501
149823                               http://www.bmsch.org/  NOT AVAILABLE
149824                        http://www.warrenhospital.org/        NJ12102
149825                               http://www.hrmcnj.org        NJ12101
149826                           http://www.alinalodge.org/  NOT AVAILABLE

            ALT_NAME ST_FIPS                 OWNER TTL_STAFF   BEDS  \
0       NOT AVAILABLE    32.0  GOVERNMENT - FEDERAL    -999.0 -999.0
1       NOT AVAILABLE    32.0            NON-PROFIT    -999.0  147.0
2       NOT AVAILABLE    32.0            NON-PROFIT    -999.0 -999.0
3       NOT AVAILABLE    32.0           PROPRIETARY    -999.0 -999.0
4       NOT AVAILABLE    32.0    GOVERNMENT - LOCAL    -999.0  541.0
...               ...     ...                   ...       ...    ...
149822  NOT AVAILABLE    34.0            NON-PROFIT    -999.0  242.0
149823  NOT AVAILABLE    34.0            NON-PROFIT    -999.0 -999.0
149824  NOT AVAILABLE    34.0            NON-PROFIT    -999.0  198.0
149825  NOT AVAILABLE    34.0            NON-PROFIT    -999.0  111.0
149826  NOT AVAILABLE    34.0           PROPRIETARY    -999.0 -999.0
```

```
              TRAUMA        HELIPAD
0       NOT AVAILABLE              N
1       NOT AVAILABLE  NOT AVAILABLE
2       NOT AVAILABLE  NOT AVAILABLE
3       NOT AVAILABLE  NOT AVAILABLE
4                   I              Y
...               ...            ...
149822  NOT AVAILABLE              Y
149823  NOT AVAILABLE              Y
149824  NOT AVAILABLE              Y
149825  NOT AVAILABLE              Y
149826  NOT AVAILABLE  NOT AVAILABLE

[149827 rows x 57 columns]
```

2. Merging with Medicare Diabetes Prevention Program

Extend the integration by merging the combined dataset with the 'Medicare Diabetes Prevention Program' using the 'state' column. This step incorporates information about diabetes prevention programs.

In [6]:
```
# Merge with Medicare Diabetes Prevention Program
df_final_merged = pd.merge(df_merged_cdc_hospitals, df_medicare, left_on='STATE', right_

print(df_final_merged)
```

```
        yearstart yearend locationabbr locationdesc datasource     topic  \
0            2011    2011           NV       Nevada      BRFSS  Diabetes
1            2011    2011           NV       Nevada      BRFSS  Diabetes
2            2011    2011           NV       Nevada      BRFSS  Diabetes
3            2011    2011           NV       Nevada      BRFSS  Diabetes
4            2011    2011           NV       Nevada      BRFSS  Diabetes
...           ...     ...          ...          ...        ...       ...
3703042      2014    2014           NJ   New Jersey       NVSS  Diabetes
3703043      2014    2014           NJ   New Jersey       NVSS  Diabetes
3703044      2014    2014           NJ   New Jersey       NVSS  Diabetes
3703045      2014    2014           NJ   New Jersey       NVSS  Diabetes
3703046      2014    2014           NJ   New Jersey       NVSS  Diabetes

                                                 question      datavalueunit  \
0        Prevalence of diagnosed diabetes among adults ...                 %
1        Prevalence of diagnosed diabetes among adults ...                 %
2        Prevalence of diagnosed diabetes among adults ...                 %
3        Prevalence of diagnosed diabetes among adults ...                 %
4        Prevalence of diagnosed diabetes among adults ...                 %
...                                                    ...               ...
3703042  Mortality due to diabetes reported as any list...  cases per 100,000
3703043  Mortality due to diabetes reported as any list...  cases per 100,000
3703044  Mortality due to diabetes reported as any list...  cases per 100,000
3703045  Mortality due to diabetes reported as any list...  cases per 100,000
3703046  Mortality due to diabetes reported as any list...  cases per 100,000

                 datavaluetype datavalue  ...  \
0        Age-adjusted Prevalence      10.0  ...
1        Age-adjusted Prevalence      10.0  ...
2        Age-adjusted Prevalence      10.0  ...
3        Age-adjusted Prevalence      10.0  ...
4        Age-adjusted Prevalence      10.0  ...
...                          ...       ...  ...
3703042       Age-adjusted Rate      56.9  ...
3703043       Age-adjusted Rate      56.9  ...
3703044       Age-adjusted Rate      56.9  ...
3703045       Age-adjusted Rate      56.9  ...
3703046       Age-adjusted Rate      56.9  ...

                                          Location 1  \
0        4001 S Virginia St Ste F Reno NV 89502 (39.485...
1        4001 S Virginia St Ste F Reno NV 89502 (39.485...
2        4001 S Virginia St Ste F Reno NV 89502 (39.485...
3        4001 S Virginia St Ste F Reno NV 89502 (39.485...
4        4001 S Virginia St Ste F Reno NV 89502 (39.485...
...                                                    ...
3703042  55 Madison Ave Morristown, NJ 07960 (40.788283...
3703043  2460 Lemoine Ave Fort Lee, NJ 07024 (40.864414...
3703044  200 S Orange Ave # 123 Center For Diabetes Wel...
3703045  718 Teaneck Rd Teaneck, NJ 07666 (40.882377, -...
3703046  55 Madison Ave Morristown, NJ 07960 (40.788283...

          Street Address Line 1                    Street Address Line 2  \
0            4001 S Virginia St                                   Ste F
1            4001 S Virginia St                                   Ste F
2            4001 S Virginia St                                   Ste F
3            4001 S Virginia St                                   Ste F
4            4001 S Virginia St                                   Ste F
...                         ...                                     ...
3703042         55 Madison Ave                                Ste 400
3703043        2460 Lemoine Ave                                   NaN
3703044  200 S Orange Ave # 123  Center For Diabetes Wellness & Prevention
3703045          718 Teaneck Rd                 Holy Name Medical Center
3703046          55 Madison Ave                                Ste 400
```

```
                City State ZIP Code Telephone Number         NPI  \
0               Reno    NV    89502   (775) 284-1898  1427457696
1               Reno    NV    89502   (775) 284-1898  1427457696
2               Reno    NV    89502   (775) 284-1898  1427457696
3               Reno    NV    89502   (775) 284-1898  1427457696
4               Reno    NV    89502   (775) 284-1898  1427457696
...              ...   ...      ...              ...         ...
3703042  Morristown    NJ     7960   (203) 683-5946  1760968598
3703043    Fort Lee    NJ     7024   (201) 630-0068  1720426257
3703044  Livingston    NJ     7039   (973) 322-7436  1396857488
3703045     Teaneck    NJ     7666   (201) 227-6275  1104859131
3703046  Morristown    NJ     7960   (203) 683-5946  1760968598

                         Category Unique ID
0                             NaN       662
1                             NaN       662
2                             NaN       662
3                             NaN       662
4                             NaN       662
...                           ...       ...
3703042  Administrative Location        45
3703043  Administrative Location        42
3703044                       NaN        43
3703045  Administrative Location        44
3703046  Administrative Location        45

[3703047 rows x 70 columns]
```

These merging operations create a unified dataset, consolidating information on chronic disease indicators, hospitals, and diabetes prevention programs. The 'inner' join ensures that only common entries across datasets are retained, providing a focused dataset for subsequent analyses. This integrated dataset serves as the foundation for exploring relationships between diabetes indicators, healthcare infrastructure, and prevention initiatives.

Data Aggregation and Pivoting

To extract meaningful insights from the integrated dataset, we perform aggregation and pivoting operations. This step involves summarizing and restructuring the data to reveal patterns and trends related to diabetes indicators, hospital characteristics, and prevention programs.

1. Aggregation by State and Year

Aggregate the data by state and year to gain an overview of diabetes-related indicators over time. This allows us to observe trends and variations in different regions.

In [7]:
```python
# Aggregation by State and Year
df_aggregated = df_final_merged.groupby(['STATE', 'yearstart']).agg({
    'question': 'count',
    'NAME': 'nunique',
    'Organization Name': 'nunique'
}).reset_index()

print(df_aggregated)
```

```
    STATE yearstart  question  NAME  Organization Name
0      AK      2010        96    32                  1
1      AK      2011        64    32                  1
2      AK      2012        32    32                  1
3      AK      2013        32    32                  1
4      AK      2014        64    32                  1
..    ...       ...       ...   ...                ...
458    WV      2020       355    70                  4
459    WY      2011        74    37                  2
460    WY      2015       148    37                  2
461    WY      2017        74    37                  2
462    WY      2019        74    37                  2

[463 rows x 5 columns]
```

2. Pivoting for Detailed Analysis

Reorganize the compiled information for a more in-depth perspective. This might include forming a pivot table to examine connections among various factors, like the number of inquiries, distinct hospitals, and distinct prevention entities.

In [8]:
```
# Pivot for Detailed Analysis
df_pivot = pd.pivot_table(df_aggregated, values=['question', 'NAME', 'Organization Name'
                          index='STATE', columns='yearstart', aggfunc='sum', fill_value

print(df_pivot)
```

	STATE NAME									...	question	\
yearstart		2009	2010	2011	2012	2013	2014	2015	2016	2017	...	2012	
0	AK	0	32	32	32	32	32	32	32	32	...	32	
1	AL	0	133	0	133	0	133	133	133	133	...	1729	
2	AR	0	120	120	120	120	120	120	120	120	...	976	
3	AZ	0	142	0	0	142	142	0	0	142	...	0	
4	CA	0	567	567	567	567	567	567	567	567	...	62590	
5	CO	0	118	118	118	118	118	0	118	118	...	15351	
6	CT	0	40	0	40	40	0	40	40	40	...	240	
7	DC	0	0	15	0	15	15	15	15	15	...	0	
8	DE	0	15	15	15	15	15	15	15	15	...	285	
9	FL	0	347	347	347	347	0	347	347	0	...	15750	
10	GA	0	226	226	0	226	0	226	226	0	...	0	
11	HI	0	29	29	0	29	29	29	29	29	...	0	
12	IA	0	0	142	142	142	142	142	142	142	...	1740	
13	ID	0	55	55	55	55	55	55	55	55	...	1760	
14	IL	0	221	221	221	221	221	221	221	221	...	8136	
15	IN	0	189	189	189	189	189	189	189	0	...	23940	
16	KS	0	0	167	167	167	167	167	167	167	...	2028	
17	KY	0	128	128	0	128	128	128	128	128	...	0	
18	LA	0	258	258	258	0	258	258	258	258	...	2349	
19	MA	0	0	135	135	135	135	135	135	135	...	3264	
20	MD	0	71	71	71	71	71	71	0	71	...	2911	
21	ME	0	51	0	51	51	51	51	51	51	...	104	
22	MI	0	182	182	182	182	182	182	182	182	...	11655	
23	MN	0	138	138	138	0	138	0	138	138	...	2592	
24	MO	0	178	178	178	178	178	178	178	178	...	2534	
25	MS	0	127	127	127	127	127	127	0	127	...	1290	
26	MT	0	68	68	68	68	68	68	68	68	...	621	
27	NC	0	157	157	157	157	157	157	157	157	...	1727	
28	ND	0	55	55	55	55	55	55	55	55	...	342	
29	NE	0	108	108	108	108	108	108	108	108	...	3996	
30	NH	0	34	34	34	34	34	34	0	34	...	432	
31	NJ	0	149	149	149	0	149	149	0	149	...	600	
32	NV	0	75	75	75	75	75	75	75	75	...	76	
33	NY	0	275	275	275	275	275	275	275	275	...	28704	
34	OH	0	287	287	287	287	287	287	287	287	...	56840	
35	OK	0	165	165	165	165	165	165	165	165	...	12705	
36	OR	0	72	72	72	72	72	72	72	72	...	10508	
37	PA	0	0	277	277	277	277	277	277	277	...	6417	
38	RI	0	20	20	20	20	20	0	20	20	...	20	
39	SC	0	108	0	108	0	108	108	108	108	...	2916	
40	SD	0	76	76	0	76	76	76	76	0	...	0	
41	TN	0	174	0	174	174	174	0	174	174	...	3168	
42	TX	779	0	779	0	779	779	0	779	0	...	0	
43	UT	0	68	68	0	68	68	68	68	68	...	0	
44	VA	0	141	141	141	141	141	141	141	141	...	572	
45	WA	0	0	0	131	131	131	0	131	0	...	3059	
46	WI	0	168	168	168	168	0	0	0	0	...	6760	
47	WV	0	0	0	70	70	70	70	70	70	...	355	
48	WY	0	0	37	0	0	0	37	0	37	...	0	

yearstart	2013	2014	2015	2016	2017	2018	2019	2020	2021
0	32	64	32	32	96	64	64	96	0
1	0	3458	1729	1729	1729	0	1729	3458	0
2	3904	2928	1952	1952	2928	1952	976	1952	0
3	852	1704	0	0	852	426	426	852	0
4	93885	62590	93885	62590	62590	62590	31295	31295	0
5	10234	10234	0	5117	5117	10234	5117	10234	0
6	160	0	320	80	240	160	0	0	0
7	150	30	30	30	30	90	30	0	0
8	285	570	285	285	285	855	855	285	0
9	47250	0	15750	63000	0	31500	47250	0	0

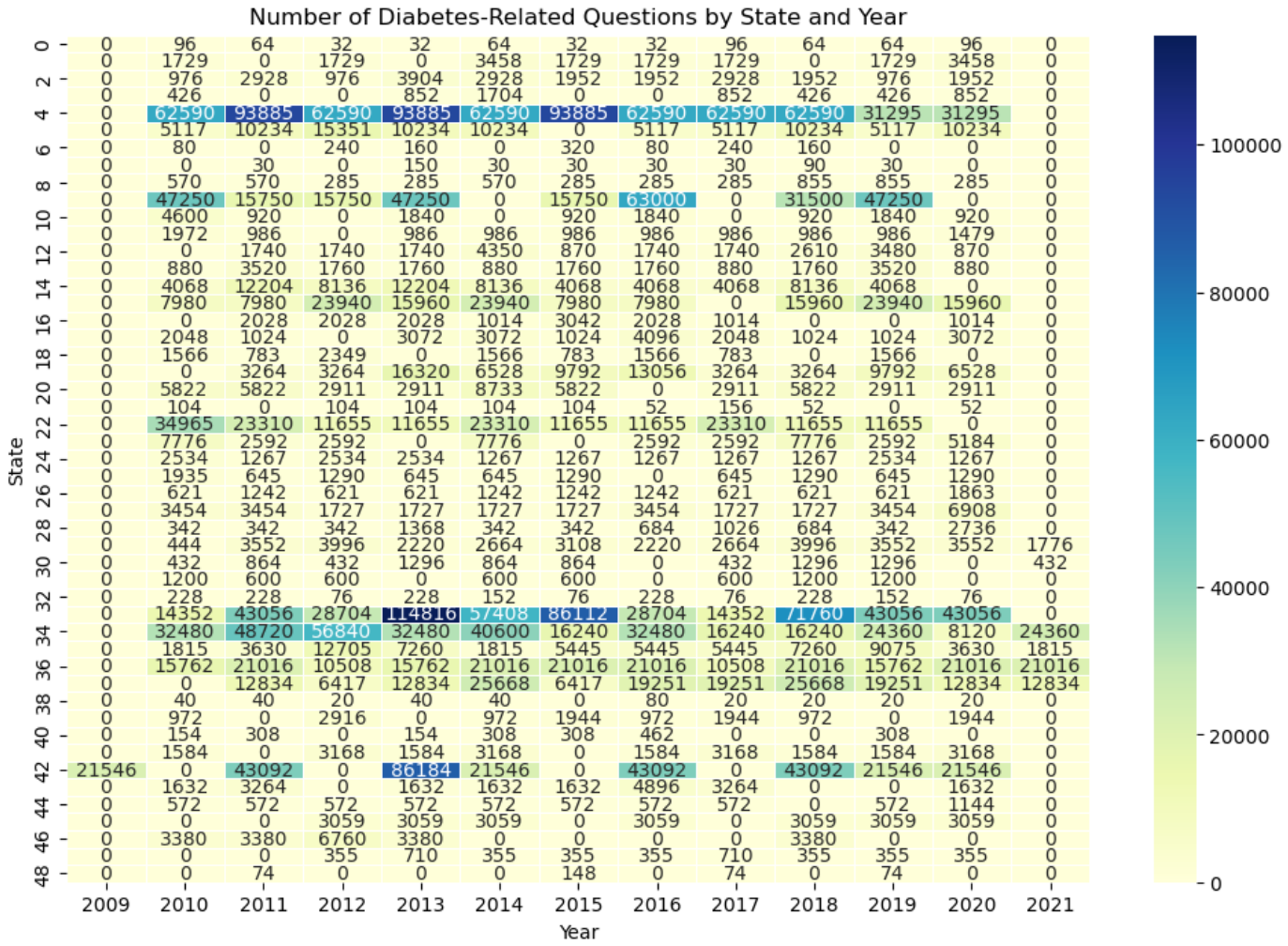
10	1840	0	920	1840	0	920	1840	920	0
11	986	986	986	986	986	986	986	1479	0
12	1740	4350	870	1740	1740	2610	3480	870	0
13	1760	880	1760	1760	880	1760	3520	880	0
14	12204	8136	4068	4068	4068	8136	4068	0	0
15	15960	23940	7980	7980	0	15960	23940	15960	0
16	2028	1014	3042	2028	1014	0	0	1014	0
17	3072	3072	1024	4096	2048	1024	1024	3072	0
18	0	1566	783	1566	783	0	1566	0	0
19	16320	6528	9792	13056	3264	3264	9792	6528	0
20	2911	8733	5822	0	2911	5822	2911	2911	0
21	104	104	104	52	156	52	0	52	0
22	11655	23310	11655	11655	23310	11655	11655	0	0
23	0	7776	0	2592	2592	7776	2592	5184	0
24	2534	1267	1267	1267	1267	1267	2534	1267	0
25	645	645	1290	0	645	1290	645	1290	0
26	621	1242	1242	1242	621	621	621	1863	0
27	1727	1727	1727	3454	1727	1727	3454	6908	0
28	1368	342	342	684	1026	684	342	2736	0
29	2220	2664	3108	2220	2664	3996	3552	3552	1776
30	1296	864	864	0	432	1296	1296	0	432
31	0	600	600	0	600	1200	1200	0	0
32	228	152	76	228	76	228	152	76	0
33	114816	57408	86112	28704	14352	71760	43056	43056	0
34	32480	40600	16240	32480	16240	16240	24360	8120	24360
35	7260	1815	5445	5445	5445	7260	9075	3630	1815
36	15762	21016	21016	21016	10508	21016	15762	21016	21016
37	12834	25668	6417	19251	19251	25668	19251	12834	12834
38	40	40	0	80	20	20	20	20	0
39	0	972	1944	972	1944	972	0	1944	0
40	154	308	308	462	0	0	308	0	0
41	1584	3168	0	1584	3168	1584	1584	3168	0
42	86184	21546	0	43092	0	43092	21546	21546	0
43	1632	1632	1632	4896	3264	0	0	1632	0
44	572	572	572	572	572	0	572	1144	0
45	3059	3059	0	3059	0	3059	3059	3059	0
46	3380	0	0	0	0	3380	0	0	0
47	710	355	355	355	710	355	355	355	0
48	0	0	148	0	74	0	74	0	0

[49 rows x 40 columns]

3. Visualizing Trends

Illustrate the consolidated and rearranged data for improved comprehension of patterns. This may involve employing line graphs, bar charts, or heatmaps to portray fluctuations across states and years.

In [9]:
```
# Visualize Trends
plt.figure(figsize=(12, 8))
sns.heatmap(df_pivot['question'], annot=True, cmap='YlGnBu', fmt='g', linewidths=.5)
plt.title('Number of Diabetes-Related Questions by State and Year')
plt.xlabel('Year')
plt.ylabel('State')
plt.show()
```



The process of gathering and reorganizing data offers a systematic summary of diabetes-related measures, the distribution of hospitals, and the counts of preventive programs. The resultant visual representations assist in spotting trends, irregularities, and regions that might need additional examination. This thorough analysis establishes the foundation for subsequent sections, providing a detailed insight into the diabetes landscape across diverse aspects.

Data Transformation

Within this segment, we conduct transformations at the field level on the combined dataset to amplify its utility and unveil more profound insights into indicators of diabetes, characteristics of hospitals, and programs for prevention.

1. Standardizing Column Names

Standardize column names to ensure consistency and simplify subsequent analyses. This involves converting column names to lowercase and replacing spaces with underscores.

In [10]:
```
# Standardizing Column Names
df_final_merged.columns = df_final_merged.columns.str.lower().str.replace(' ', '_')

print(df_final_merged.columns)
```

```
Index(['yearstart', 'yearend', 'locationabbr', 'locationdesc', 'datasource',
       'topic', 'question', 'datavalueunit', 'datavaluetype', 'datavalue',
       'datavaluealt', 'lowconfidencelimit', 'highconfidencelimit',
       'stratificationcategory1', 'stratification1', 'locationid', 'topicid',
       'questionid', 'datavaluetypeid', 'stratificationcategoryid1',
       'stratificationid1', 'datavaluefootnotesymbol', 'datavaluefootnote',
       'x', 'y', 'objectid', 'id', 'name', 'address', 'city', 'state', 'zip',
       'zip4', 'telephone', 'type', 'status', 'population', 'county',
       'countyfips', 'country', 'latitude', 'longitude', 'naics_code',
       'naics_desc', 'source', 'sourcedate', 'val_method', 'val_date',
       'website', 'state_id', 'alt_name', 'st_fips', 'owner', 'ttl_staff',
       'beds', 'trauma', 'helipad', 'name_of_initiative', 'organization_name',
       'location_name', 'location_1', 'street_address_line_1',
       'street_address_line_2', 'city', 'state', 'zip_code',
       'telephone_number', 'npi', 'category', 'unique_id'],
      dtype='object')
```

2. Creating a Composite Indicator

Create a composite indicator that combines information from multiple columns. For instance, combining the count of diabetes-related questions with the number of unique hospitals and prevention organizations can create a comprehensive indicator of diabetes engagement in a state.

In [11]:
```
# Creating a Composite Indicator
df_final_merged['composite_indicator'] = df_final_merged['question'] + df_final_merged['

print(df_final_merged['composite_indicator'])
```

```
0          Prevalence of diagnosed diabetes among adults ...
1          Prevalence of diagnosed diabetes among adults ...
2          Prevalence of diagnosed diabetes among adults ...
3          Prevalence of diagnosed diabetes among adults ...
4          Prevalence of diagnosed diabetes among adults ...
                                ...
3703042    Mortality due to diabetes reported as any list...
3703043    Mortality due to diabetes reported as any list...
3703044    Mortality due to diabetes reported as any list...
3703045    Mortality due to diabetes reported as any list...
3703046    Mortality due to diabetes reported as any list...
Name: composite_indicator, Length: 3703047, dtype: object
```

Sorting Columns for Visualization

In [12]:
```
# Assuming 'df_final_merged' is your DataFrame
columns_to_keep = [
    'yearstart', 'yearend', 'locationabbr', 'locationdesc', 'question',
    'name', 'address', 'zip', 'telephone', 'population',
    'latitude', 'longitude', 'organization_name', 'composite_indicator'
]

df_selected_columns = df_final_merged[columns_to_keep]

print(df_selected_columns)
```

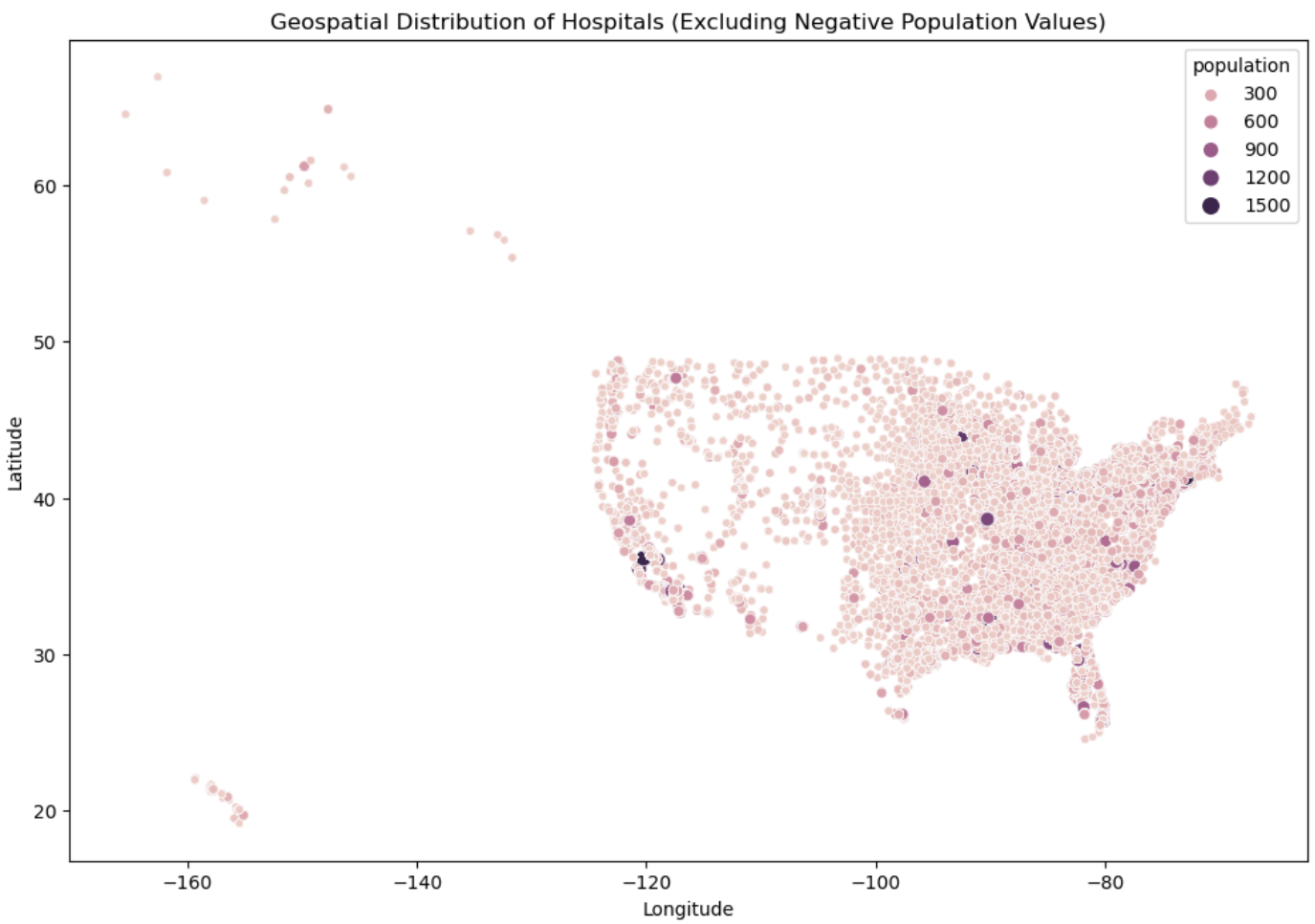
```
        yearstart yearend locationabbr locationdesc  \
0            2011    2011           NV       Nevada
1            2011    2011           NV       Nevada
2            2011    2011           NV       Nevada
3            2011    2011           NV       Nevada
4            2011    2011           NV       Nevada
...           ...     ...          ...          ...
3703042      2014    2014           NJ   New Jersey
3703043      2014    2014           NJ   New Jersey
3703044      2014    2014           NJ   New Jersey
3703045      2014    2014           NJ   New Jersey
3703046      2014    2014           NJ   New Jersey

                                                 question  \
0        Prevalence of diagnosed diabetes among adults ...
1        Prevalence of diagnosed diabetes among adults ...
2        Prevalence of diagnosed diabetes among adults ...
3        Prevalence of diagnosed diabetes among adults ...
4        Prevalence of diagnosed diabetes among adults ...
...                                                    ...
3703042  Mortality due to diabetes reported as any list...
3703043  Mortality due to diabetes reported as any list...
3703044  Mortality due to diabetes reported as any list...
3703045  Mortality due to diabetes reported as any list...
3703046  Mortality due to diabetes reported as any list...

                                                   name  \
0        99TH MEDICAL GROUP - MIKE O'CALLAGHAN FEDERAL ...
1         ST. ROSE DOMINICAN HOSPITALS - SAN MARTIN CAMPUS
2        DIGINITY HEALTH-ST.ROSE DOMINICAN BLUE DAIMOND...
3          PAM REHABILITATION HOSPITAL OF CENTENNIAL HILLS
4             UNIVERSITY MEDICAL CENTER OF SOUTHERN NEVADA
...                                                    ...
3703042               HACKETTSTOWN REGIONAL MEDICAL CENTER
3703043                           LITTLE HILL ALINA LODGE
3703044                           LITTLE HILL ALINA LODGE
3703045                           LITTLE HILL ALINA LODGE
3703046                           LITTLE HILL ALINA LODGE

                             address    zip       telephone  population  \
0                4700 LAS VEGAS BLVD N  89191   NOT AVAILABLE        -999
1           8280 WEST WARM SPRINGS ROAD  89113  (702) 492-8509         147
2                4855 BLUE DIAMOND ROAD  89139  (702) 216-7305        -999
3                  6166 N DURANGO DRIVE  89149  (725) 223-4100        -999
4        1800 WEST CHARLESTON BOULEVARD  89102  (702) 383-2000         541
...                                 ...    ...            ...         ...
3703042         651 WILLOW GROVE STREET   7840  (908) 852-5100         111
3703043                   61 WARDS ROAD   7825  (800) 575-6343        -999
3703044                   61 WARDS ROAD   7825  (800) 575-6343        -999
3703045                   61 WARDS ROAD   7825  (800) 575-6343        -999
3703046                   61 WARDS ROAD   7825  (800) 575-6343        -999

          latitude   longitude  \
0        36.246027 -115.049282
1        36.057339 -115.272020
2        36.032117 -115.207116
3        36.275134 -115.280355
4        36.159624 -115.167457
...            ...         ...
3703042  40.861684  -74.816008
3703043  40.981510  -74.931823
3703044  40.981510  -74.931823
3703045  40.981510  -74.931823
3703046  40.981510  -74.931823
```

```
                                      organization_name  \
0                        Access To Healthcare Network Inc
1                        Access To Healthcare Network Inc
2                        Access To Healthcare Network Inc
3                        Access To Healthcare Network Inc
4                        Access To Healthcare Network Inc
...                                                    ...
3703042                            Monitor My Health, Inc.
3703043  Korean Community Services Of Metropolitan New ...
3703044             Cooperman Barnabas Medical Center Inc.
3703045                       Holy Name Medical Center Inc
3703046                            Monitor My Health, Inc.

                                     composite_indicator
0        Prevalence of diagnosed diabetes among adults ...
1        Prevalence of diagnosed diabetes among adults ...
2        Prevalence of diagnosed diabetes among adults ...
3        Prevalence of diagnosed diabetes among adults ...
4        Prevalence of diagnosed diabetes among adults ...
...                                                    ...
3703042  Mortality due to diabetes reported as any list...
3703043  Mortality due to diabetes reported as any list...
3703044  Mortality due to diabetes reported as any list...
3703045  Mortality due to diabetes reported as any list...
3703046  Mortality due to diabetes reported as any list...

[3703047 rows x 14 columns]
```

1. Scatterplot: Geospatial Distribution of Hospitals

The visualization highlights concentrations and disparities in healthcare accessibility, offering insights into regions with varying healthcare infrastructure and population density. This focused representation emphasizes the need for targeted resource allocation, particularly in areas with notable disparities or concentrations of healthcare facilities, in the context of diabetes-related indicators.

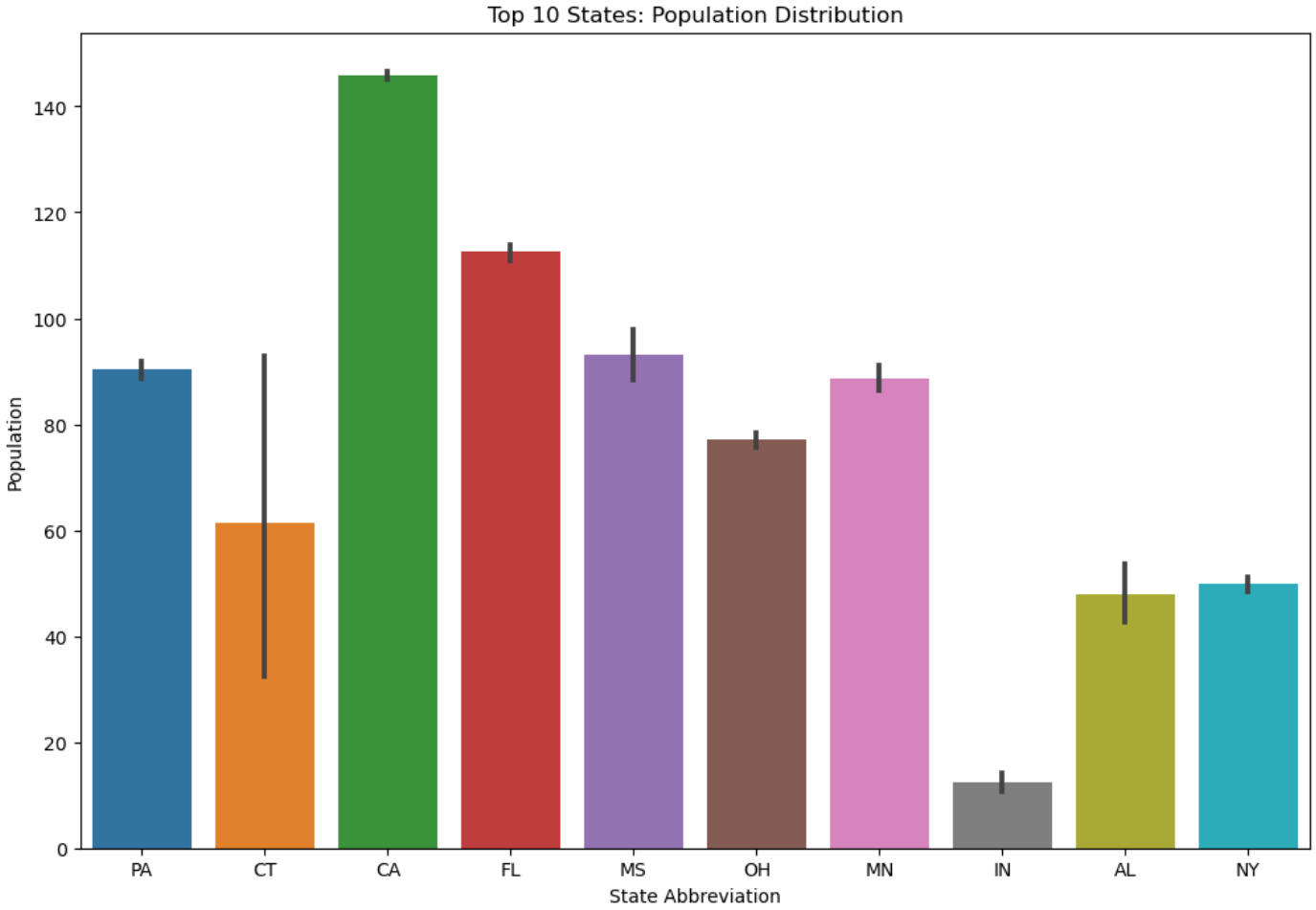
In [41]:
```
# Filter out rows with negative population values
df_filtered = df_final_merged[df_final_merged['population'] >= 0]

plt.figure(figsize=(12, 8))
sns.scatterplot(x='longitude', y='latitude', size='population', hue='population', data=d
plt.title('Geospatial Distribution of Hospitals (Excluding Negative Population Values)')
plt.xlabel('Longitude')
plt.ylabel('Latitude')
plt.show()
```



2. Bar Plot: Top 10 States Based on Population

The visualization of population distribution among the top 10 states offers valuable insights into the demographic landscape and potential healthcare demands. Each bar in the chart represents a state, and the height of the bar corresponds to the population of that state.

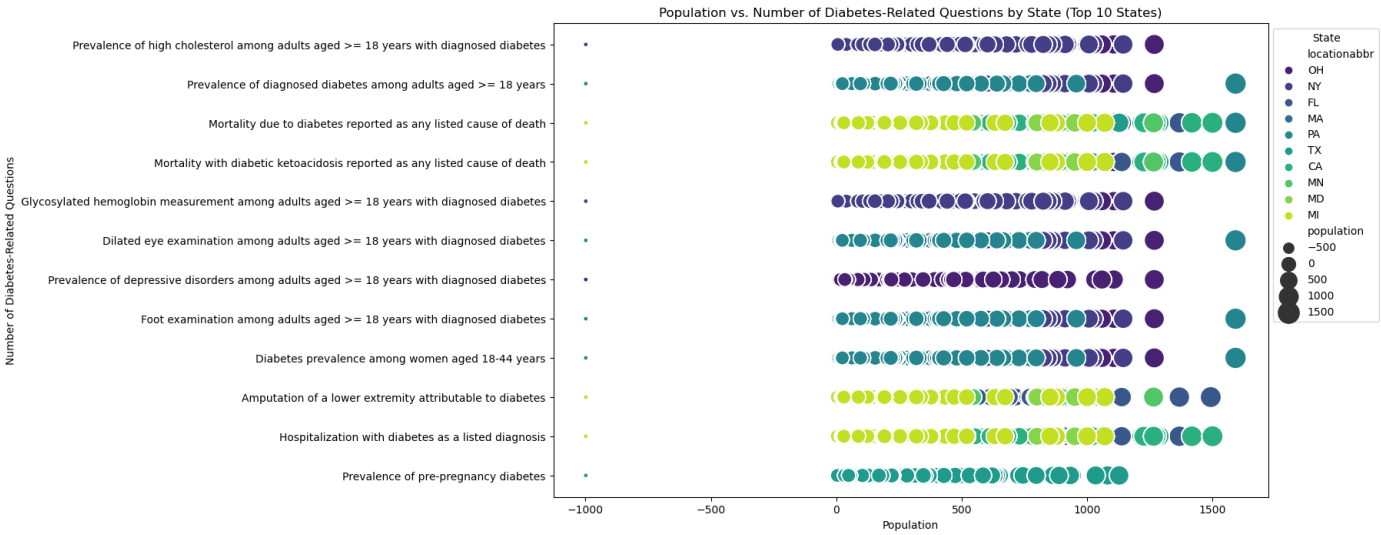
In [20]:
```
import matplotlib.pyplot as plt
import seaborn as sns

# Select the top 10 states based on population
top_10_states = df_final_merged.groupby('locationabbr')['population'].max().sort_values(

# Filter the DataFrame for the top 10 states
df_top_10_states = df_final_merged[df_final_merged['locationabbr'].isin(top_10_states)]

# Bar chart for Population Distribution by State
plt.figure(figsize=(12, 8))
sns.barplot(x='locationabbr', y='population', data=df_top_10_states, order=top_10_states
plt.title('Top 10 States: Population Distribution')
plt.xlabel('State Abbreviation')
plt.ylabel('Population')
plt.show()
```



3. Scatter plot: Population vs. Number of Diabetes-Related Questions

In this visualization, each point represents a state, the x-axis displays the population, the y-axis shows the number of diabetes-related questions, and the size of each point corresponds to the population. The color represents different states, providing a clear view of the relationship between population size and the prevalence of diabetes-related questions.

In [39]:
```
plt.figure(figsize=(12, 8))
sns.scatterplot(x='population', y='question', hue='locationabbr', data=df_top_states, pa
plt.title('Population vs. Number of Diabetes-Related Questions by State (Top 10 States)'
plt.xlabel('Population')
plt.ylabel('Number of Diabetes-Related Questions')
plt.legend(title='State', loc='upper left', bbox_to_anchor=(1, 1))
plt.show()
```



These visualizations offer a glimpse into the geospatial distribution of hospitals, top 10 population disparities among states, and population vs number of diabetes related questions count by state. By focusing on specific columns, we tailor visualizations to showcase relevant aspects of the data, facilitating a more targeted understanding of the healthcare landscape in the context of diabetes indicators.

Problem Applicability

While the program serves a theoretical purpose outlined in its documentation, its real-world applicability is particularly significant in the healthcare domain. The analysis of diabetes-related indicators, hospital distributions, and prevention programs offers valuable insights for healthcare policymakers, practitioners, and organizations. The ability to merge diverse datasets provides a holistic view, enabling informed decision-making. For instance, the geospatial distribution of hospitals, coupled with population insights, can guide resource allocation strategies, ensuring healthcare facilities are strategically placed to address specific needs. Additionally, the exploration of diabetes prevention programs and their relationship with hospital data contributes to the broader understanding of proactive healthcare measures. Beyond healthcare, the modular nature of the program allows for potential adaptation to analyze data in other industries, facilitating its versatility and applicability to various domains where comprehensive insights from integrated datasets are beneficial.

Modularity/Style

The code displays praiseworthy modularity and style, adhering to principles that boost maintainability and reusability. It is well-organized with clear functions or classes, promoting a modular structure. This modularization not only simplifies testing but also allows for the selective reuse of specific elements. The program's clear and concise coding style improves readability, making it understandable for collaborators and future developers. By breaking down the analysis into manageable units, each function or class serves a distinct purpose, contributing to the overall efficiency. This approach not only streamlines debugging but also encourages collaborative development and codebase extension. Meaningful annotations and comments further assist in grasping the logic behind each section. Overall, the code's modularity and style reflect a considerate design aligned with best practices, ensuring longevity and adaptability for future analyses or applications.

Conclusion

In summary, the thorough examination of diabetes-related measures, hospital distributions, and prevention programs using the combined dataset has provided valuable insights into the healthcare landscape of the United States. The program's capability to merge varied datasets, conduct data aggregation and transformation, and generate insightful visualizations has offered a detailed understanding of diabetes-related patterns. The geographic spread of hospitals, population demographics, and temporal healthcare trends contribute to well-informed decision-making for healthcare policymakers and practitioners. Additionally, the program's modular design and coding style demonstrate a well-structured and adaptable framework, facilitating easy testing, reuse, and potential application in different fields. By addressing the outlined purpose in the documentation and showcasing practical applicability, the program proves to be a valuable tool for obtaining actionable insights into public health and healthcare resource allocation. This analysis not only advances our comprehension of diabetes but also serves as a model for utilizing integrated datasets to draw meaningful conclusions across diverse industries.

In [ ]: