In [39]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import scipy.stats as stats
from sklearn.preprocessing import StandardScaler, MinMaxScaler
from sklearn.impute import SimpleImputer
```

In [12]:
```
pip install pandas numpy matplotlib seaborn scikit-learn statsmodels plotly geopandas
```

```
Requirement already satisfied: pandas in c:\users\aadesh\anaconda3\lib\site-packages (2.1.4)Note: you may need to restart the kernel to use updated packages.


Requirement already satisfied: numpy in c:\users\aadesh\anaconda3\lib\site-packages (1.26.4)
Requirement already satisfied: matplotlib in c:\users\aadesh\anaconda3\lib\site-packages (3.8.0)
Requirement already satisfied: seaborn in c:\users\aadesh\anaconda3\lib\site-packages (0.12.2)
Requirement already satisfied: scikit-learn in c:\users\aadesh\anaconda3\lib\site-packages (1.2.2)
Requirement already satisfied: statsmodels in c:\users\aadesh\anaconda3\lib\site-packages (0.14.0)
Requirement already satisfied: plotly in c:\users\aadesh\anaconda3\lib\site-packages (5.9.0)
Collecting geopandas
  Downloading geopandas-1.0.1-py3-none-any.whl.metadata (2.2 kB)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\aadesh\anaconda3\lib\site-packages (from pandas) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\users\aadesh\anaconda3\lib\site-packages (from pandas) (2023.3.post1)
Requirement already satisfied: tzdata>=2022.1 in c:\users\aadesh\anaconda3\lib\site-packages (from pandas) (2023.3)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\aadesh\anaconda3\lib\site-packages (from matplotlib) (1.2.0)
Requirement already satisfied: cycler>=0.10 in c:\users\aadesh\anaconda3\lib\site-packages (from matplotlib) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\aadesh\anaconda3\lib\site-packages (from matplotlib) (4.25.0)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\aadesh\anaconda3\lib\site-packages (from matplotlib) (1.4.4)
Requirement already satisfied: packaging>=20.0 in c:\users\aadesh\anaconda3\lib\site-packages (from matplotlib) (23.1)
Requirement already satisfied: pillow>=6.2.0 in c:\users\aadesh\anaconda3\lib\site-packages (from matplotlib) (10.2.0)
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\aadesh\anaconda3\lib\site-packages (from matplotlib) (3.0.9)
Requirement already satisfied: scipy>=1.3.2 in c:\users\aadesh\anaconda3\lib\site-packages (from scikit-learn) (1.11.4)
Requirement already satisfied: joblib>=1.1.1 in c:\users\aadesh\anaconda3\lib\site-packages (from scikit-learn) (1.2.0)
Requirement already satisfied: threadpoolctl>=2.0.0 in c:\users\aadesh\anaconda3\lib\site-packages (from scikit-learn) (2.2.0)
Requirement already satisfied: patsy>=0.5.2 in c:\users\aadesh\anaconda3\lib\site-packages (from statsmodels) (0.5.3)
Requirement already satisfied: tenacity>=6.2.0 in c:\users\aadesh\anaconda3\lib\site-packages (from plotly) (8.2.2)
Collecting pyogrio>=0.7.2 (from geopandas)
  Downloading pyogrio-0.10.0-cp311-cp311-win_amd64.whl.metadata (5.6 kB)
Collecting pyproj>=3.3.0 (from geopandas)
  Downloading pyproj-3.7.0-cp311-cp311-win_amd64.whl.metadata (31 kB)
Collecting shapely>=2.0.0 (from geopandas)
  Downloading shapely-2.0.6-cp311-cp311-win_amd64.whl.metadata (7.2 kB)
Requirement already satisfied: six in c:\users\aadesh\anaconda3\lib\site-packages (from patsy>=0.5.2->statsmodels) (1.16.0)
Requirement already satisfied: certifi in c:\users\aadesh\anaconda3\lib\site-packages (from pyogrio>=0.7.2->geopandas) (2024.2.2)
Downloading geopandas-1.0.1-py3-none-any.whl (323 kB)
   ---------------------------------------- 0.0/323.6 kB ? eta -:--:--
   --------------- ------------------------ 122.9/323.6 kB 3.6 MB/s eta 0:00:01
   ---------------------------------------- 317.4/323.6 kB 4.9 MB/s eta 0:00:01
   ---------------------------------------- 323.6/323.6 kB 4.0 MB/s eta 0:00:00
Downloading pyogrio-0.10.0-cp311-cp311-win_amd64.whl (16.2 MB)
   ---------------------------------------- 0.0/16.2 MB ? eta -:--:--
    --------------------------------------- 0.3/16.2 MB 15.4 MB/s eta 0:00:02
   - -------------------------------------- 0.7/16.2 MB 7.8 MB/s eta 0:00:02
   -- ------------------------------------- 1.0/16.2 MB 7.0 MB/s eta 0:00:03
   --- ------------------------------------ 1.6/16.2 MB 9.2 MB/s eta 0:00:02
   ----- ---------------------------------- 2.1/16.2 MB 9.4 MB/s eta 0:00:02
   ------ --------------------------------- 2.6/16.2 MB 9.0 MB/s eta 0:00:02
   ------- -------------------------------- 3.2/16.2 MB 9.6 MB/s eta 0:00:02
   --------- ------------------------------ 3.7/16.2 MB 10.3 MB/s eta 0:00:02
   ---------- ----------------------------- 4.3/16.2 MB 10.6 MB/s eta 0:00:02
   ---------- ----------------------------- 4.9/16.2 MB 10.7 MB/s eta 0:00:02
   ----------- ---------------------------- 5.3/16.2 MB 10.6 MB/s eta 0:00:02
   ------------ --------------------------- 5.9/16.2 MB 10.9 MB/s eta 0:00:01
   ------------- -------------------------- 6.7/16.2 MB 11.3 MB/s eta 0:00:01
   -------------- ------------------------- 7.4/16.2 MB 11.5 MB/s eta 0:00:01
   ---------------- ----------------------- 8.0/16.2 MB 11.6 MB/s eta 0:00:01
   ---------------- ----------------------- 8.6/16.2 MB 11.7 MB/s eta 0:00:01
   ----------------- ---------------------- 8.7/16.2 MB 11.2 MB/s eta 0:00:01
   ------------------ --------------------- 9.4/16.2 MB 11.3 MB/s eta 0:00:01
   ------------------- -------------------- 10.0/16.2 MB 11.4 MB/s eta 0:00:01
   -------------------- ------------------- 10.6/16.2 MB 12.1 MB/s eta 0:00:01
   --------------------- ------------------ 11.0/16.2 MB 12.3 MB/s eta 0:00:01
   ---------------------- ----------------- 11.8/16.2 MB 12.1 MB/s eta 0:00:01
   ---------------------- ----------------- 12.2/16.2 MB 12.1 MB/s eta 0:00:01
   ----------------------- ---------------- 12.8/16.2 MB 12.6 MB/s eta 0:00:01
   ------------------------ --------------- 13.3/16.2 MB 12.1 MB/s eta 0:00:01
   ------------------------ --------------- 13.9/16.2 MB 12.4 MB/s eta 0:00:01
   ------------------------- -------------- 14.6/16.2 MB 12.4 MB/s eta 0:00:01
   -------------------------- ------------- 15.4/16.2 MB 12.6 MB/s eta 0:00:01
   --------------------------- ------------ 16.1/16.2 MB 12.8 MB/s eta 0:00:01
   ---------------------------- ----------- 16.2/16.2 MB 12.1 MB/s eta 0:00:00
Downloading pyproj-3.7.0-cp311-cp311-win_amd64.whl (6.2 MB)
   ---------------------------------------- 0.0/6.2 MB ? eta -:--:--
   --- ------------------------------------ 0.6/6.2 MB 17.5 MB/s eta 0:00:01
   ------ --------------------------------- 1.0/6.2 MB 12.1 MB/s eta 0:00:01
   -------- ------------------------------- 1.5/6.2 MB 11.8 MB/s eta 0:00:01
   ---------- ----------------------------- 1.9/6.2 MB 11.1 MB/s eta 0:00:01
   ----------- ---------------------------- 2.4/6.2 MB 10.8 MB/s eta 0:00:01
   ------------- -------------------------- 2.8/6.2 MB 10.4 MB/s eta 0:00:01
   --------------- ------------------------ 3.2/6.2 MB 10.3 MB/s eta 0:00:01
   ---------------- ----------------------- 3.6/6.2 MB 10.1 MB/s eta 0:00:01
   ------------------ --------------------- 4.1/6.2 MB 10.5 MB/s eta 0:00:01
   -------------------- ------------------- 4.5/6.2 MB 10.0 MB/s eta 0:00:01
   --------------------- ------------------ 4.9/6.2 MB 10.2 MB/s eta 0:00:01
   ----------------------- ---------------- 5.7/6.2 MB 10.7 MB/s eta 0:00:01
   ------------------------- -------------- 6.2/6.2 MB 11.0 MB/s eta 0:00:01
   ---------------------------------------- 6.2/6.2 MB 10.5 MB/s eta 0:00:00
Downloading shapely-2.0.6-cp311-cp311-win_amd64.whl (1.4 MB)
   ---------------------------------------- 0.0/1.4 MB ? eta -:--:--
   ---------- ----------------------------- 0.5/1.4 MB 16.2 MB/s eta 0:00:01
   ------------------- -------------------- 0.9/1.4 MB 11.6 MB/s eta 0:00:01
   ---------------------------------------  1.4/1.4 MB 11.4 MB/s eta 0:00:01
   ---------------------------------------- 1.4/1.4 MB 9.2 MB/s eta 0:00:00
Installing collected packages: shapely, pyproj, pyogrio, geopandas
Successfully installed geopandas-1.0.1 pyogrio-0.10.0 pyproj-3.7.0 shapely-2.0.6
```

In [13]:
```
df = pd.read_csv(r"C:\Users\Aadesh\Downloads\Public_Health_Statistics_-_Selected_public_health_indicators_by_Chicago_community_area_-_Historical.csv")
```

In [14]:
```
print(df.isnull().sum())
```

```
Community Area                          0
Community Area Name                     0
Birth Rate                              0
General Fertility Rate                  0
Low Birth Weight                        0
Prenatal Care Beginning in First Trimester      0
Preterm Births                          0
Teen Birth Rate                         0
Assault (Homicide)                      0
Breast cancer in females                0
Cancer (All Sites)                      0
Colorectal Cancer                       0
Diabetes-related                        0
Firearm-related                         0
Infant Mortality Rate                   0
Lung Cancer                             0
Prostate Cancer in Males                0
Stroke (Cerebrovascular Disease)        0
Childhood Blood Lead Level Screening    1
Childhood Lead Poisoning                1
Gonorrhea in Females                   12
Gonorrhea in Males                      0
Tuberculosis                            0
Below Poverty Level                     0
Crowded Housing                         0
Dependency                              0
No High School Diploma                  0
Per Capita Income                       0
Unemployment                            0
dtype: int64
```

In [23]:
```
#   - Impute missing values
df['Childhood Blood Lead Level Screening'].fillna(df['Childhood Blood Lead Level Screening'].mean(), inplace=True)
```

In [16]:
```
df['Childhood Lead Poisoning'].fillna(df['Childhood Lead Poisoning'].mean(), inplace=True)
```

In [17]:
```
df['Gonorrhea in Females'].fillna(df['Gonorrhea in Females'].mean(), inplace=True)
```

In [18]:
```
print(df.dtypes)
```

```
Community Area                          int64
Community Area Name                    object
Birth Rate                            float64
General Fertility Rate                float64
Low Birth Weight                      float64
Prenatal Care Beginning in First Trimester    float64
Preterm Births                        float64
Teen Birth Rate                       float64
Assault (Homicide)                    float64
Breast cancer in females              float64
Cancer (All Sites)                    float64
Colorectal Cancer                     float64
Diabetes-related                      float64
Firearm-related                       float64
Infant Mortality Rate                 float64
Lung Cancer                           float64
Prostate Cancer in Males              float64
Stroke (Cerebrovascular Disease)      float64
Childhood Blood Lead Level Screening    float64
Childhood Lead Poisoning              float64
Gonorrhea in Females                  float64
Gonorrhea in Males                     object
Tuberculosis                          float64
Below Poverty Level                   float64
Crowded Housing                       float64
Dependency                            float64
No High School Diploma                float64
Per Capita Income                       int64
Unemployment                          float64
dtype: object
```

In [22]:
```
#   - Convert data types if necessary
df['Gonorrhea in Males'] = pd.to_numeric(df['Gonorrhea in Males'], errors='coerce')
```

In [20]:
```
#   - Standardize (Z-score normalization)
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
df[['Birth Rate', 'Preterm Births']] = scaler.fit_transform(df[['Birth Rate', 'Preterm Births']])
```

In [21]:
```
#   - Min-Max Normalization
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
df[['Birth Rate', 'Preterm Births']] = scaler.fit_transform(df[['Birth Rate', 'Preterm Births']])
```

In [33]:
```
# Find out relation between Health indicators and Socio economic factors by correlation Analysis
```

In [26]:
```
# Calculate the correlation coefficient between 'Infant Mortality Rate' and 'Below Poverty Level'
correlation = df['Infant Mortality Rate'].corr(df['Below Poverty Level'])
```

In [27]:
```
print("Correlation coefficient:", correlation)
```

```
Correlation coefficient: 0.5907673433431874
```

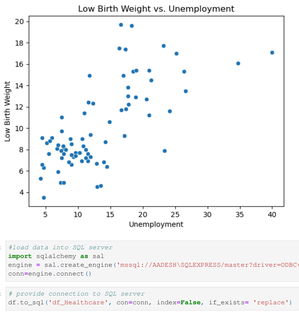
In [28]:
```
# Visualize the relationship using a scatter plot
sns.scatterplot(x='Below Poverty Level', y='Infant Mortality Rate', data=df)
plt.title('Infant Mortality Rate vs. Below Poverty Level')
plt.xlabel('Below Poverty Level')
plt.ylabel('Infant Mortality Rate')
plt.show()
```

In [29]:
```
# Calculate the correlation coefficient between 'Low Birth Weight' and 'Unemployment'
correlation = df['Low Birth Weight'].corr(df['Unemployment'])
```

In [30]:
```
print("Correlation coefficient:", correlation)
```

```
Correlation coefficient: 0.7022267993109089
```

In [31]:
```
# Visualize the relationship using a scatter plot
sns.scatterplot(x='Unemployment', y='Low Birth Weight', data=df)
plt.title('Low Birth Weight vs. Unemployment')
plt.xlabel('Unemployment')
plt.ylabel('Low Birth Weight')
plt.show()
```



In [44]:
```
#load data into SQL server
import sqlalchemy as sal
engine = sal.create_engine('mssql://AADESH\SQLEXPRESS/master?driver=ODBC+DRIVER+17+FOR+SQL+SERVER')
conn=engine.connect()
```

In [45]:
```
# provide connection to SQL server
df.to_sql('df_Healthcare', con=conn, index=False, if_exists= 'replace')
```

Out[45]: 5

In [ ]: