

Data Dictionary: aqi_project.db

1. Table: counties (Dimension / Master Table)

Source: U.S. Census Bureau

- fips (TEXT, Primary Key): 5-digit Federal Information Processing Standard code uniquely identifying the county.
- county_name (TEXT): Name of the county.
- state_name (TEXT): Name of the state.

2. Table: demographics (Fact Table)

Source: U.S. Census Bureau (ACS)

- fips (TEXT, Foreign Key): Connects to counties table.
- total_pop (INTEGER): Total estimated population.
- median_age (REAL): Median age of the county population.
- pop_under_5, pop_under_18, pop_15_to_44, pop_65_and_over (INTEGER): Population counts by specific age brackets.

3. Table: income (Fact Table)

Source: U.S. Census Bureau

- fips (TEXT, Foreign Key): Connects to counties table.
- median_income (REAL): Median household income in USD.

4. Table: air_quality (Fact Table)

Source: Environmental Protection Agency (EPA)

- fips (TEXT, Foreign Key): Connects to counties table.
- year (INTEGER): The year of the data snapshot (2025).
- median_aqi (INTEGER): The median Air Quality Index value for the year.
- aqi_p90 (INTEGER): The 90th percentile AQI value (represents peak pollution days).
- days_pm25 (INTEGER): Count of days where PM2.5 was the primary pollutant.

5. Table: health (Fact Table)

Source: Centers for Disease Control and Prevention (CDC)

- fips (TEXT, Foreign Key): Connects to counties table.
- incidence_rate (REAL): Age-adjusted cancer incidence rate per 100,000 population.
- avg_annual_count (INTEGER): Average total number of cancer cases annually.