LA County Health Disparities Analysis Project

Sirtaj Singh

github.com/sirtajsingh8/LA-Health-Disparities-Project

Executive Summary

This project analyzes health disparities across Los Angeles County communities using a data-driven approach that integrates SQL Server, Python, and Power BI. By examining socioeconomic, environmental, and healthcare access factors, we identified significant disparities in health outcomes between different communities. Our analysis reveals that low-income and minority communities face substantially higher disease rates, reduced healthcare access, increased environmental challenges, and shorter life expectancies. The interactive dashboard created through this project enables stakeholders to explore these disparities visually and supports the development of targeted interventions to address health inequities.

Project Background & Objectives

Health disparities represent preventable differences in health outcomes experienced by socially disadvantaged populations. In Los Angeles County, with its diverse and geographically dispersed communities, these disparities manifest in significant variations in disease prevalence, life expectancy, healthcare access, and environmental conditions.

This project aims to:

1. Identify and quantify health disparities across LA County communities
2. Understand the relationships between socioeconomic factors and health outcomes
3. Map healthcare resource distribution and access barriers
4. Analyze environmental health factors contributing to disparities
5. Provide data-driven recommendations for targeted interventions

Public Health Data Sources

This project leverages multiple authoritative public health data sources to ensure comprehensive and accurate analysis:

LA County Department of Public Health

- **LA County Health Survey**: Provides community-level data on health behaviors, disease prevalence, and healthcare access
- **Mortality and Morbidity Reports**: Offers ZIP code-level data on causes of death, life expectancy, and disease burden
- **Community Health Assessment Data**: Includes information on preventable hospitalizations and health service utilization

California Health Interview Survey (CHIS)

- Provides detailed health status, health conditions, and health behaviors data
- Includes information on health insurance coverage and healthcare access
- Offers data on vulnerable populations and social determinants of health

American Community Survey (Census Data)

- Demographic information including population, race/ethnicity, and age distribution
- Socioeconomic data including income, poverty rates, education levels, and employment
- Housing information including housing costs, density, and quality indicators

CDC Social Vulnerability Index

- Measures community resilience to external stresses on human health
- Includes socioeconomic status, household composition, minority status, and housing/transportation data
- Helps identify communities that may need support in preparing for or recovering from disasters

CalEnviroScreen

- Identifies California communities most affected by pollution
- Provides data on exposure indicators (air quality, drinking water contaminants)
- Includes environmental effect indicators (toxic releases, traffic density)
- Offers sensitive population indicators (asthma rates, cardiovascular disease)

Healthcare Facility Data

- California Office of Statewide Health Planning and Development (OSHPD) facility database
- Centers for Medicare & Medicaid Services (CMS) provider information
- LA County Department of Health Services facility directory

Public Transit and Transportation Data

- LA Metro transportation accessibility data
- Southern California Association of Governments (SCAG) transportation analysis data
- Walk Score and transit accessibility metrics

Health Disparity Index: Methodology and Calculation

A central innovation of this project is the development of a comprehensive Health Disparity Index (HDI) that quantifies inequities across communities. This composite measure integrates multiple factors to provide a single metric for identifying areas with the greatest needs.

Components of the Health Disparity Index

The HDI incorporates four major categories of factors, each weighted based on their established impact on health outcomes:

1. **Health Outcomes (40% weight)**
   - Disease prevalence rates (diabetes, heart disease, asthma, hypertension, obesity)
   - Mental health disorder prevalence
   - Preventable hospitalization rates
   - Life expectancy
2. **Healthcare Access Barriers (35% weight)**
   - Percentage without regular checkups
   - Percentage reporting delayed care
   - Transportation barriers (percentage without transportation, distance to facilities)
   - Public transit accessibility
   - Digital divide indicators (affecting telehealth access)
3. **Environmental Factors (15% weight)**
   - Air pollution indices
   - Water quality measures
   - Food desert scores
   - Green space access
   - CalEnviroScreen composite scores
4. **Protective Factors (10% weight)**
   - Healthcare facilities per capita
   - Public transit access
   - Green space availability
   - Water quality measures

Calculation Methodology

The Health Disparity Index is calculated through a multi-step process:

1. **Standardization**: All component metrics are standardized using z-scores to enable comparison across different units of measurement.
2. **Directionality Adjustment**: Metrics are adjusted so that higher values consistently indicate worse disparities. For positive factors (like water quality or green space), the scale is inverted.
3. **Weighted Averaging**: The standardized scores are combined using the specified category weights (40%, 35%, 15%, 10%).
4. **Scaling**: The final index is scaled to a 0-100 range, where 100 represents the highest possible disparity and 0 represents no disparity.
5. **Categorization**: Communities are classified into five disparity levels: Very Low, Low, Moderate, High, and Very High, based on quintile distribution.

Mathematical Formula

The HDI is calculated using the following formula:

HDI = [(0.4 × average of standardized health outcome factors) + (0.35 × average of standardized access barrier factors) + (0.15 × average of standardized environmental factors) - (0.1 × average of standardized protective factors)] × Scaling Factor

Where the scaling factor normalizes the result to a 0-100 scale.

Validation and Benchmarking

The Health Disparity Index was validated through:

- Comparison with existing health equity measures
- Expert review by public health professionals
- Correlation analysis with established health indicators
- Geographic pattern analysis to confirm alignment with known disparity patterns

Key Findings

Our analysis revealed several significant insights about health disparities in LA County:

1. **Geographic Disparity Patterns**:
   - The most underserved communities are South Central LA, University Park, and Pico-Union
   - Health disparities follow clear geographic patterns, with higher disparities concentrated in south and east LA
2. **Income-Health Relationship**:
   - Low-income communities have 10.0% higher diabetes rates and 5.8% higher heart disease rates
   - Life expectancy is 10.6 years shorter in low-income areas compared to high-income areas
3. **Healthcare Resource Distribution**:
   - High-income areas have 1.6x more healthcare facilities per capita than low-income areas
   - The average distance to hospitals is significantly greater in low-income communities
4. **Environmental Justice Issues**:
   - Low-income communities face 28.0% higher air pollution levels
   - Food desert scores are 37.2% worse in low-income areas
   - Green space access is significantly limited in areas with higher disparity indices
5. **Healthcare Access Barriers**:
   - Residents in low-income areas are 21.3% more likely to skip regular checkups
   - Transportation barriers affect 17.0% more residents in low-income communities
   - Digital divide issues significantly impact telehealth access in underserved areas

6. **Community Profiles**:
    o Communities cluster into distinct profiles including:
        ▪ "High Resource / Good Health" communities with strong outcomes
        ▪ "Underserved / Poor Health" communities needing intervention
        ▪ "Environmental Justice Concerns" communities with specific environmental challenges
        ▪ "Food Access Challenges" communities with limited healthy food options

Recommended Interventions

Based on our analysis, we recommend the following targeted interventions to address health disparities:

For Highest Disparity Communities (South Central LA, University Park, Pico-Union):

1. **Healthcare Resource Expansion**:
    o Establish new community health centers in healthcare deserts
    o Deploy mobile health clinics to improve service coverage
    o Implement transportation assistance programs for medical appointments
2. **Environmental Health Initiatives**:
    o Increase green space through urban pocket parks and community gardens
    o Establish farmers markets and healthy food retail incentives
    o Implement targeted air quality improvement programs
3. **Preventive Care Programs**:
    o Fund community health worker programs focused on chronic disease prevention
    o Create incentive programs for regular health screenings
    o Implement school-based health education and screening programs

For Communities with Environmental Justice Concerns:

1. **Air Quality Monitoring and Mitigation**:
    o Install community air monitoring networks
    o Implement targeted emission reduction programs
    o Create buffer zones between industrial areas and residential communities
2. **Green Infrastructure Development**:
    o Plant urban tree canopies to reduce air pollution and heat islands
    o Develop green spaces in environmentally challenged communities
    o Implement rain gardens and other green infrastructure to address water quality

For Communities with Food Access Challenges:

1. **Food Access Expansion**:
   - Provide incentives for grocery stores in food deserts
   - Support farmers markets and community-supported agriculture
   - Implement healthy corner store initiatives
2. **Community Food Programs**:
   - Develop and support community gardens
   - Create food prescription programs through healthcare providers
   - Implement school and community nutrition education programs

System-Level Recommendations:

1. **Healthcare Integration**:
   - Develop integrated care models addressing social determinants of health
   - Implement telehealth programs with hardware support for digitally underserved areas
   - Create cross-sector partnerships between healthcare, housing, and transportation
2. **Policy and Funding Priorities**:
   - Direct healthcare resources based on the Health Disparity Index
   - Implement health impact assessments for new development projects
   - Create health equity zones with targeted funding and programs

Technical Implementation Details

SQL Server Implementation

The SQL Server database was designed with a star schema centering on ZIP codes as the primary geographical unit. The database implementation included:

- Creating normalized tables with appropriate relationships
- Implementing data validation rules and constraints
- Creating a unified view (CommunityHealthView) to simplify analysis
- Setting up appropriate indexing for performance optimization

The SQL Server component serves as the data foundation, enabling structured storage and efficient querying of health-related information.

Python Analysis Techniques

The Python analysis pipeline leveraged several key data science libraries:

- **pandas**: For data manipulation and analysis
- **sklearn**: For statistical modeling and clustering
- **matplotlib/seaborn**: For data visualization
- **pyodbc**: For database connectivity

Key analytical techniques included:

- Development of a weighted health disparity index incorporating multiple factors
- Standard scaling of diverse metrics to enable comparison
- K-means clustering to identify community typologies
- Statistical correlation analysis to identify key relationships
- Automated insight generation based on data patterns

Power BI Visualization Features

The Power BI dashboard was designed for maximum usability and insight generation:

- Interactive filtering using slicers for DisparityLevel, IncomeGroup, and CommunityProfile
- Cross-filtering between visualizations to show relationships
- Geographic visualization of health disparities
- Comparative analysis of health outcomes by community type
- Facility distribution and access analysis
- Environmental health factor visualization
- Mobile-responsive design for field use

Project Impact and Applications

This health disparities analysis project has several potential applications:

1. **Resource Allocation**: Helps healthcare administrators and policymakers direct resources to communities with the greatest needs
2. **Program Development**: Supports the development of targeted health programs based on community-specific challenges
3. **Policy Advocacy**: Provides evidence for advocacy efforts focused on health equity
4. **Progress Tracking**: Establishes baseline metrics that can be monitored over time to assess intervention effectiveness
5. **Community Engagement**: Offers transparent data visualization that can be shared with community stakeholders to drive local action

Future Development Opportunities

The current project establishes a foundation that can be expanded in several directions:

1. **Temporal Analysis**: Add time-series data to track how disparities change over time
2. **Predictive Modeling**: Develop predictive models to anticipate future health disparities based on current trends
3. **Intervention Simulation**: Create what-if analysis tools to simulate the impact of potential interventions
4. **Expanded Data Sources**: Integrate additional datasets such as education, housing, and transportation
5. **Real-time Updates**: Implement automated data pipelines for continuous updates as new health data becomes available

Conclusion

The LA County Health Disparities Analysis Project demonstrates the power of integrated data technology in addressing complex public health challenges. By combining SQL Server's data management capabilities, Python's analytical power, and Power BI's visualization strengths, we've created a comprehensive system for understanding and addressing health inequities.

The project's Health Disparity Index provides a powerful tool for identifying and prioritizing communities with the greatest needs. By quantifying disparities through this multidimensional metric, we can move beyond anecdotal approaches to health equity and toward data-driven decision-making.

The stark disparities revealed by our analysis call for immediate and targeted action, particularly in South Central LA, University Park, and Pico-Union. By addressing the specific factors contributing to disparities in these communities - whether related to healthcare access, environmental conditions, or socioeconomic challenges - we can work toward ensuring that all LA County residents have the opportunity to achieve optimal health outcomes regardless of their ZIP code, income level, or demographic background.