# Urban Food Production

## Week-3 – Integrated Analytical Exploration

CS896 Capstone Project | Lokesh Das

Team 9:
Boddupalli Navya
Mallikarjun Maguluri
Venkata Subba Rao Are

# Core Dataset: PLACES Census Tracts

The foundation of our health burden analysis is the PLACES Census Tract data, providing detailed local health metrics.

**Selected Health Indicators:**

- Obesity
- Diabetes
- High Blood Pressure
- Stroke
- Physical Inactivity
- Smoking
- Healthcare Access

*Visual representation of the seven core health metrics used to calculate community vulnerability.*

# Data Validation & Cleaning Workflow

To ensure the reliability of our health burden insights, we implemented a rigorous three-step data cleaning pipeline:

- **Uniqueness Verification:** Confirmed all TractFIPS are unique with zero duplicates detected in the final set.
- **Missing Value Handling:** Applied dropna() to remove incomplete records, ensuring analysis consistency.
- **Type Conversion:** Standardized all health indicators into numeric formats for accurate statistical computation.

*Methodological workflow for preparing raw PLACES data for analytical modeling.*

# Descriptive Statistics

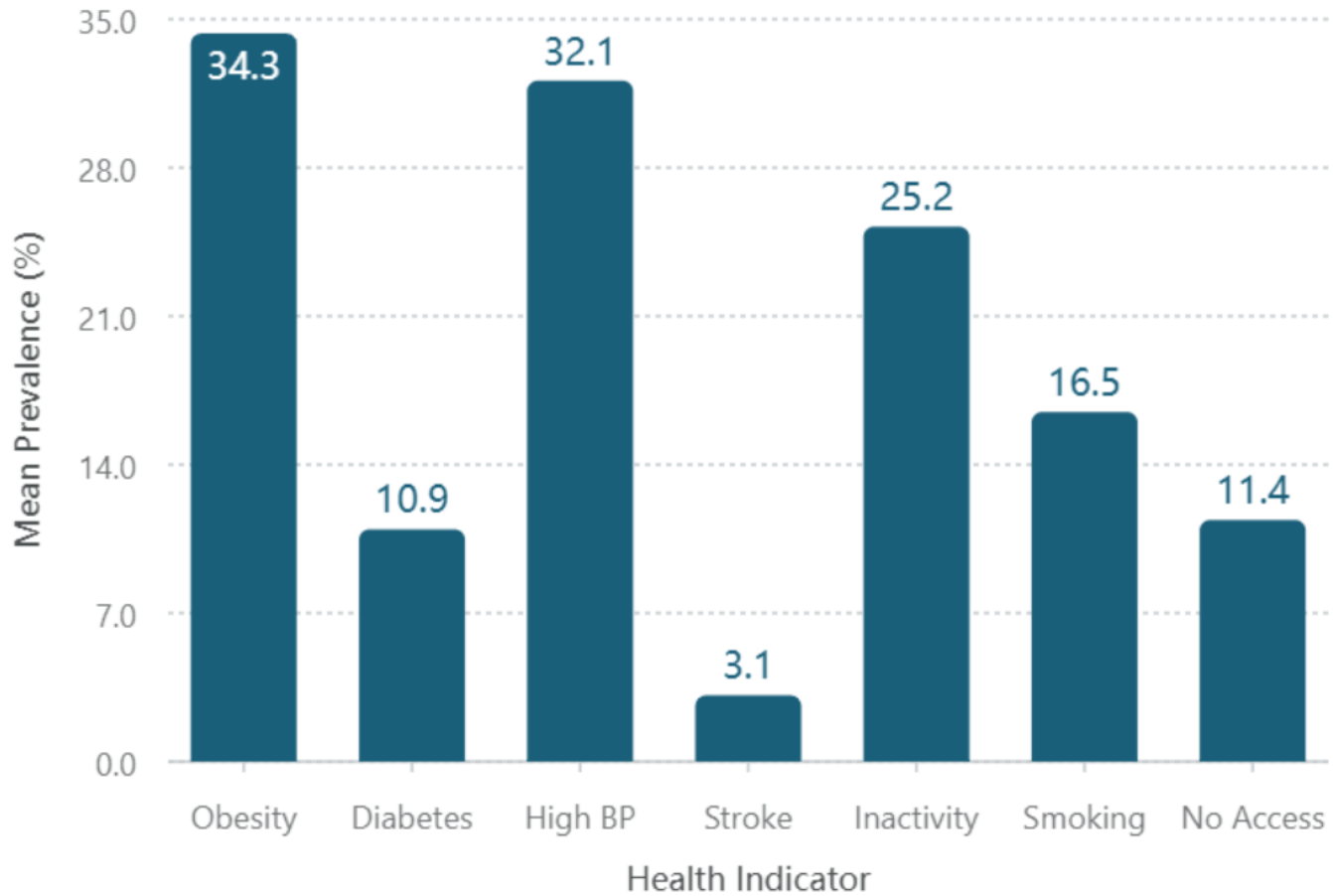
Total Census Tracts Analyzed

**68,172**

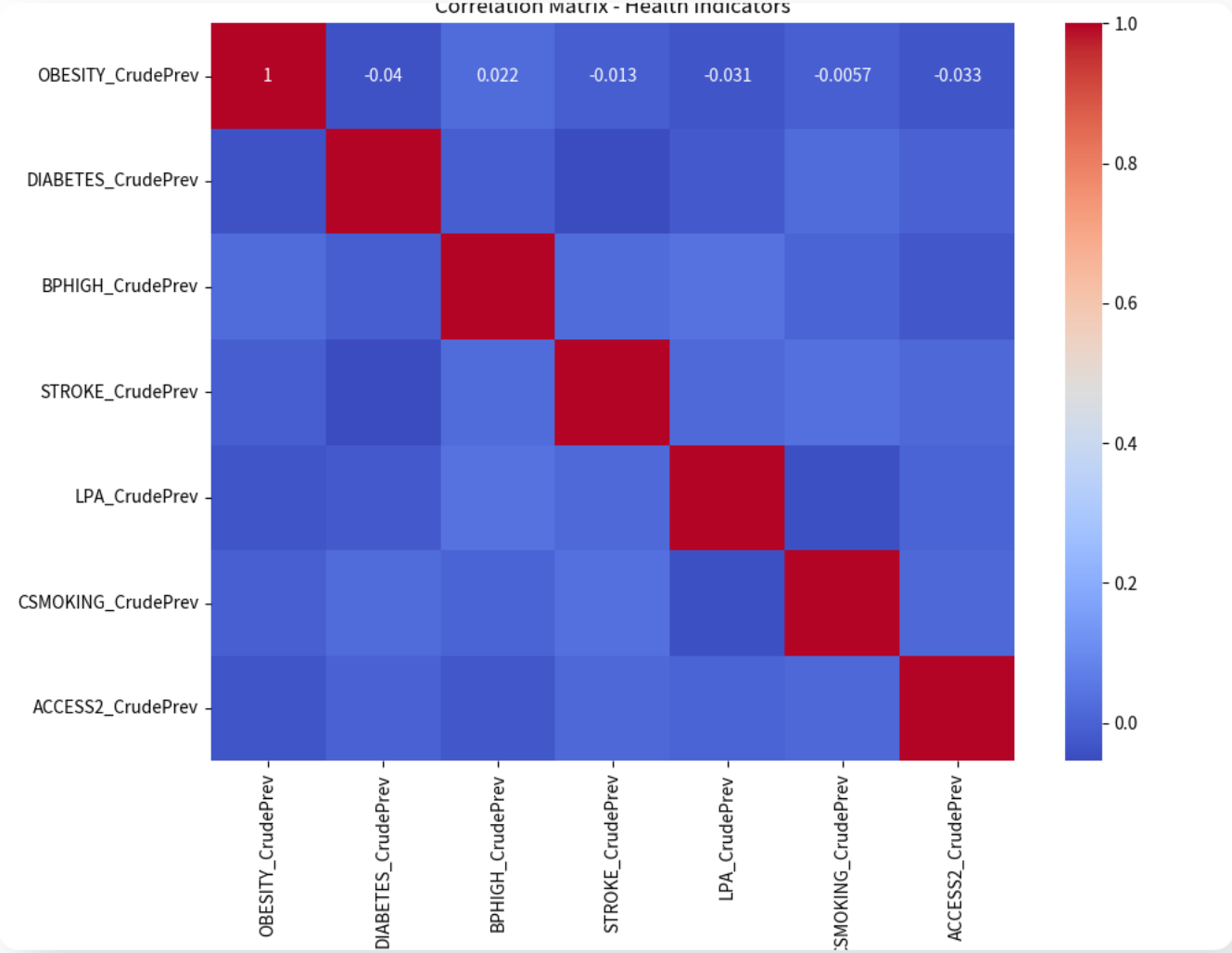
**High-Level Mean Metrics:**

- **Obesity:** 34.3%
- **High BP:** 32.1%
- **Physical Inactivity:** 25.2%

The data reflects a significant chronic disease burden across the United States, with obesity and hypertension being the most prevalent conditions.



# Correlation & Risk Factor Analysis

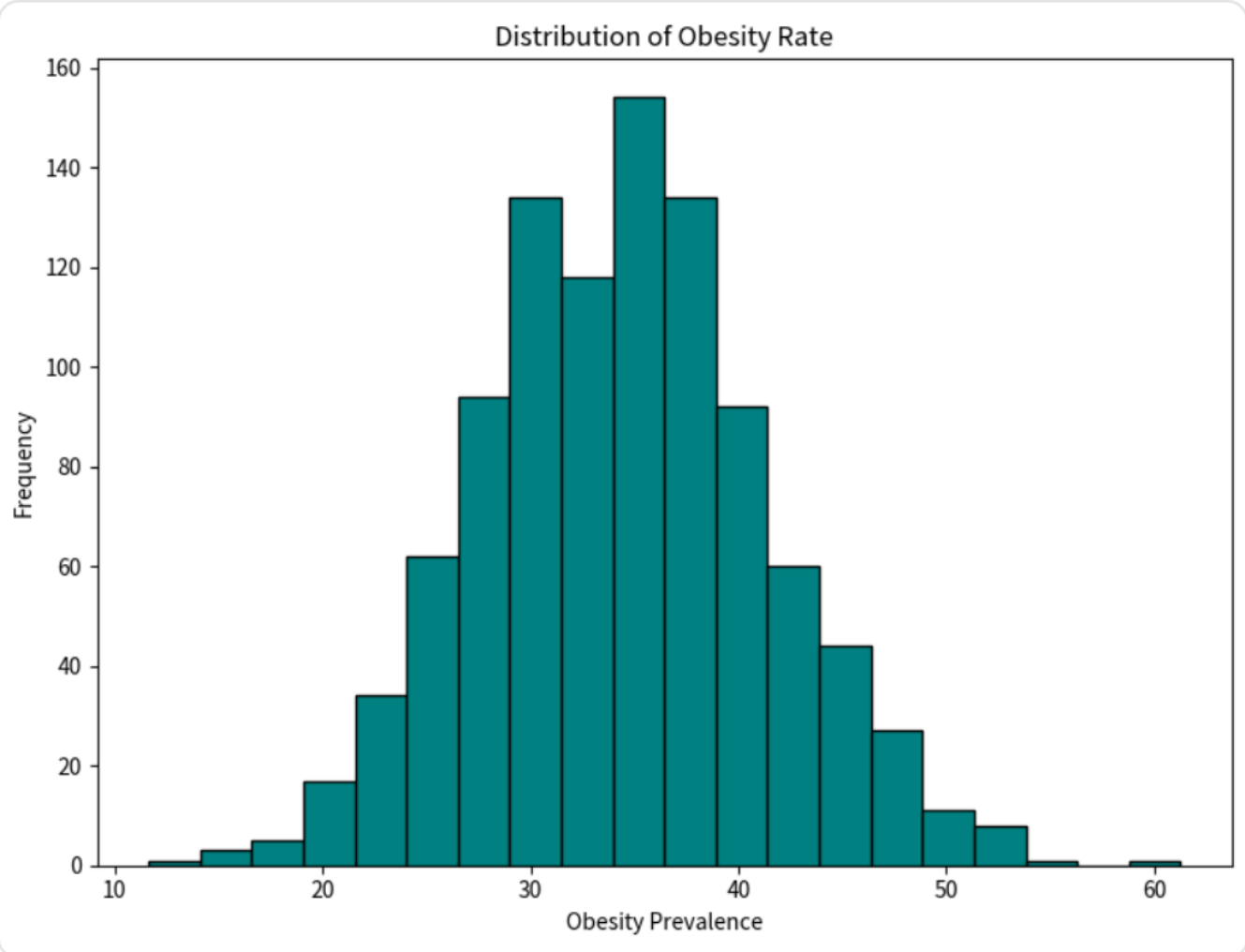



**Key Findings:**

Health risks do not occur in isolation but cluster strongly within specific tracts.

- **Obesity ↔ Diabetes:** High positive correlation indicating comorbid patterns.
- **Inactivity ↔ Obesity:** Strong link reinforcing physical activity as a key driver.
- **Smoking ↔ Chronic Conditions:** Consistent association across multiple disease states.

# Distribution Analysis & Normalization Need



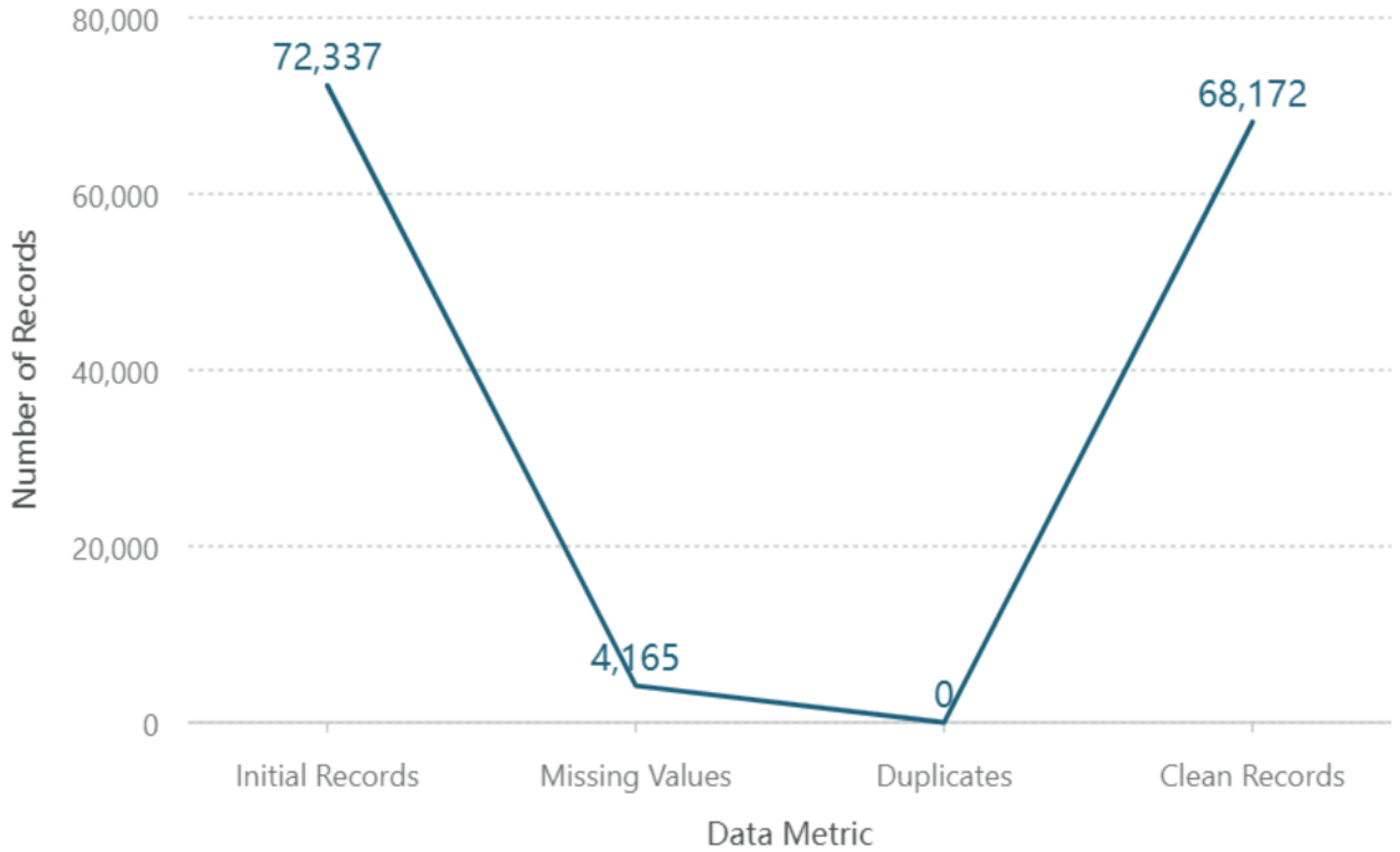

**Addressing Variable Skewness:**

Exploratory histograms revealed that several health indicators exhibit non-normal distributions and varying scales.

**Why Normalize?**

- Prevents variables with higher raw values from dominating the composite score.
- Ensures each indicator contributes proportionately to the final vulnerability assessment.
- Facilitates direct comparison across diverse health metrics.

# Data Integrity & Cleaning Results




**Quality Metrics:**

- **Initial Records:** 72,337
- **Missing Values:** 4,165
- **Duplicate IDs:** 0

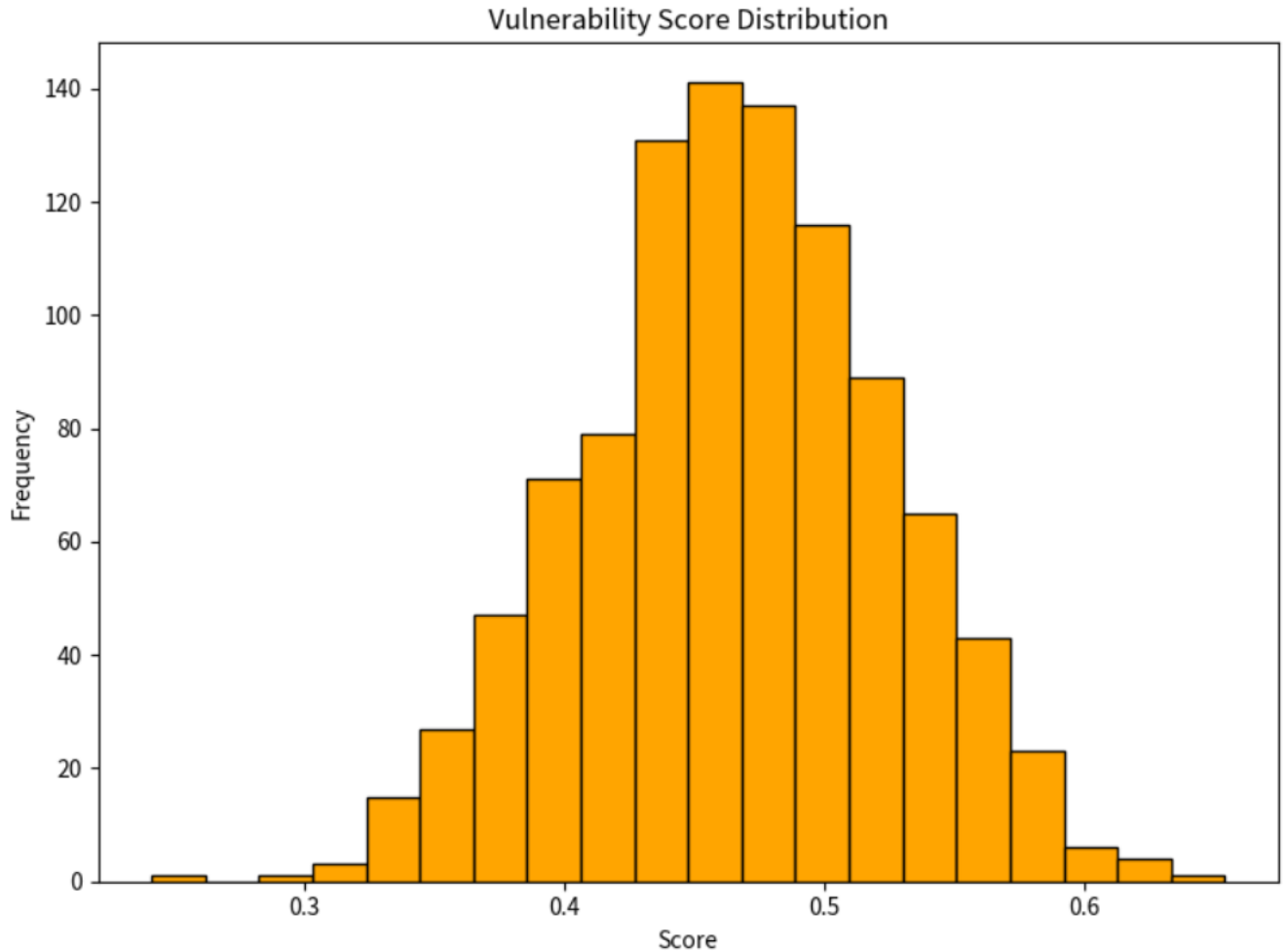
The rigorous cleaning process verified unique TractFIPS for all entries, resulting in a high-integrity analytical dataset.

# Vulnerability Score Development

**Quantifying Community Risk:**

We developed a Composite Vulnerability Score to synthesize multiple health burdens into a single actionable metric.

- **Method:** Averaged normalized health indicators per tract.
- **Ranking:** Tracts were ranked by score to identify priority areas.
- **Outcome:** Clear identification of high-burden zones for targeted urban farming interventions.



# Key Outcomes & Next Steps

**Week 3 Summary:**

- Cleaned and validated tract-level dataset.
- Identified significant health risk clusters.
- Developed preliminary Vulnerability Index.

**Looking Ahead:**

The foundation is set for **Week-4 Suitability Modeling**, where we will translate these scores into site-specific recommendations.

## Thank You!