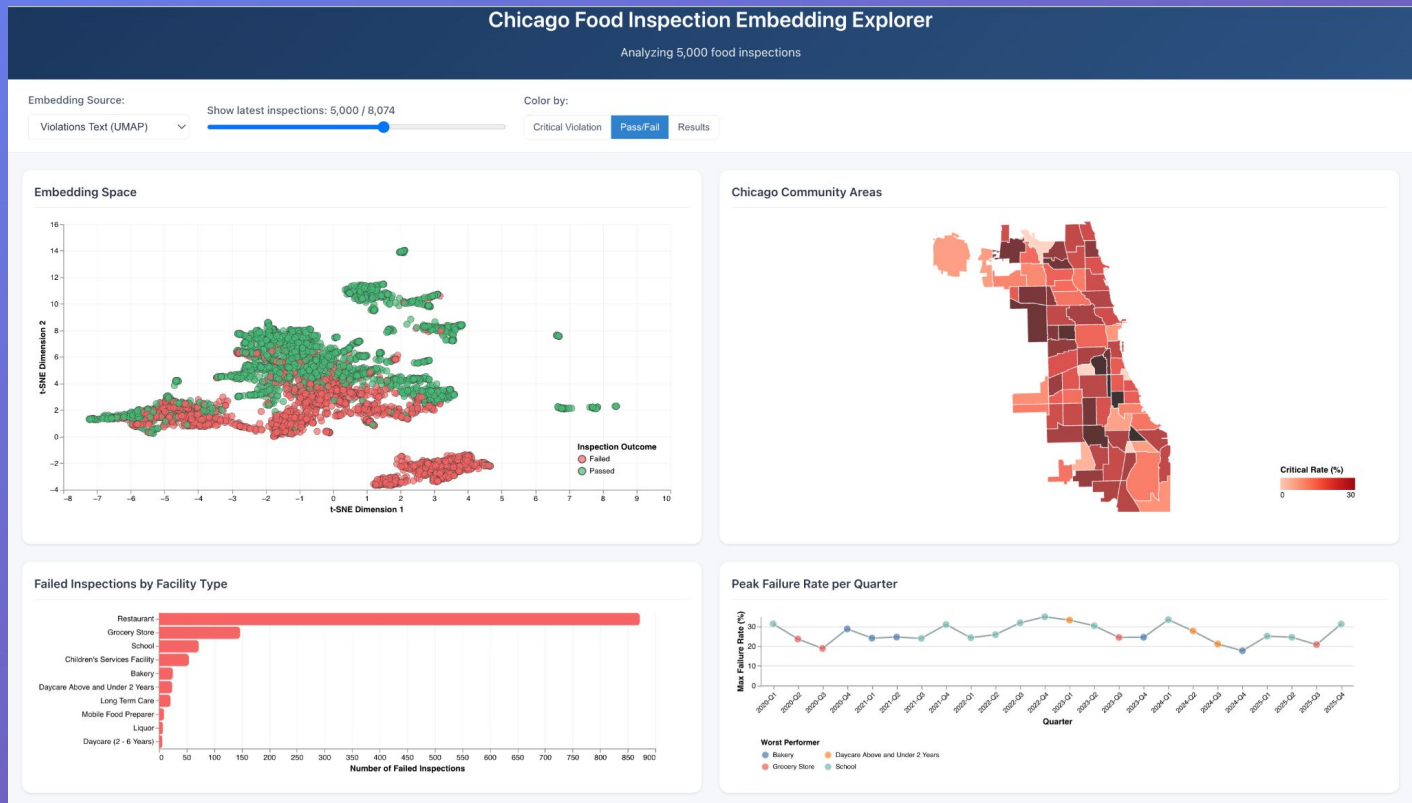


# Chicago Food Inspection Embeddings Explorer

Interactive Visualization System for Exploring Food Safety Pattern



# Dataset Overview

Food Safety & Affordable Housing Analysis in Chicago (2020-2025)

## Food Inspections

Restaurant and food establishment inspection records from 2010-present, reflecting public health performance and safety compliance.

### Key Attributes:

- Results: Pass/Fail outcomes
- Inspection Date: Temporal data
- Location: Lat/Long coordinates

## Affordable Housing

Affordable housing developments across Chicago, capturing distribution and diversity of housing types in different communities.

### Key Attributes:

- Community Area: Neighborhood
- Property Type: Senior/Family/Supportive
- Location: Lat/Long coordinates

## Community Areas

Boundaries of Chicago's 77 officially recognized community areas, serving as primary spatial units for neighborhood-level analysis.

### Key Attribute:

- the\_geom: Polygon geometries representing community boundaries

## Research Questions

What types of food establishments in different housing contexts show the most severe or frequent violations?

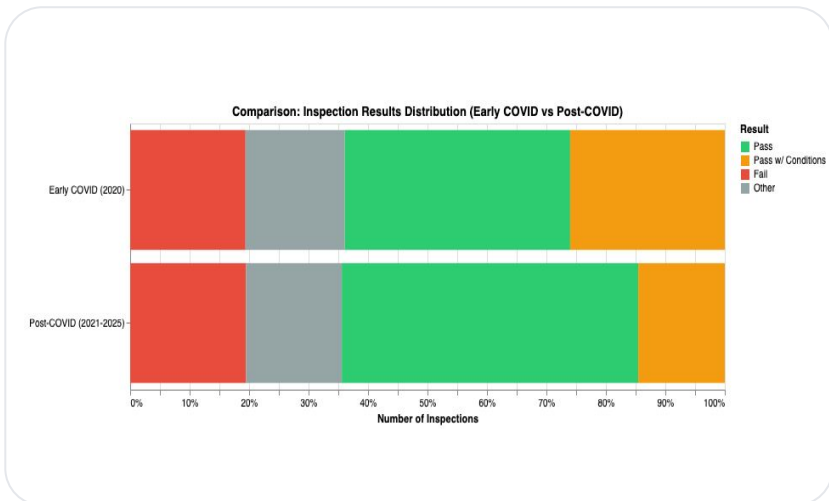
How have food safety patterns evolved over time in communities and which facility types consistently show the worst failure rates across different periods?

How are different inspection types spatially distributed across Chicago? Do certain inspection types concentrate in specific communities, and can we identify geographic patterns in inspection methodologies?

# Previous Interactive Visualizations Analyses

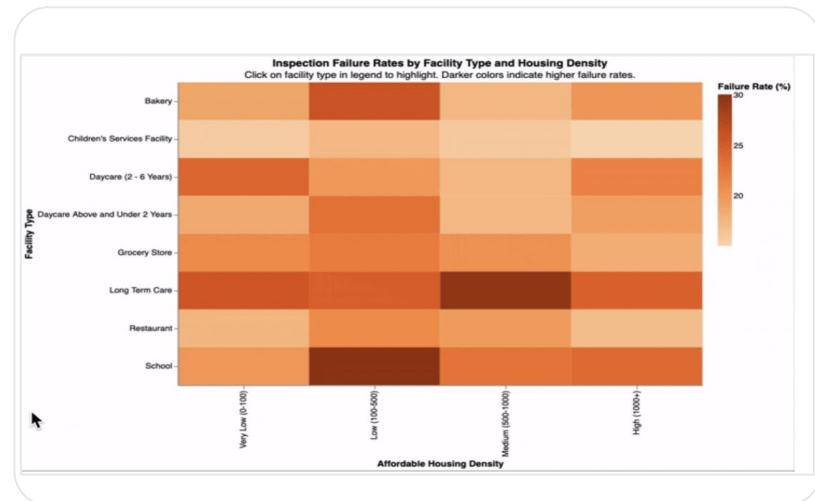
Interactive single-view visualizations developed across Assignment 2

## Temporal Comparison: Early COVID vs Post-COVID



**Normalized stacking** enables fair comparison of inspection outcomes despite different volumes between 2020 and 2021-2025, revealing how COVID-19 disrupted restaurant operations and inspection practices.

## Facility Types by Housing Context



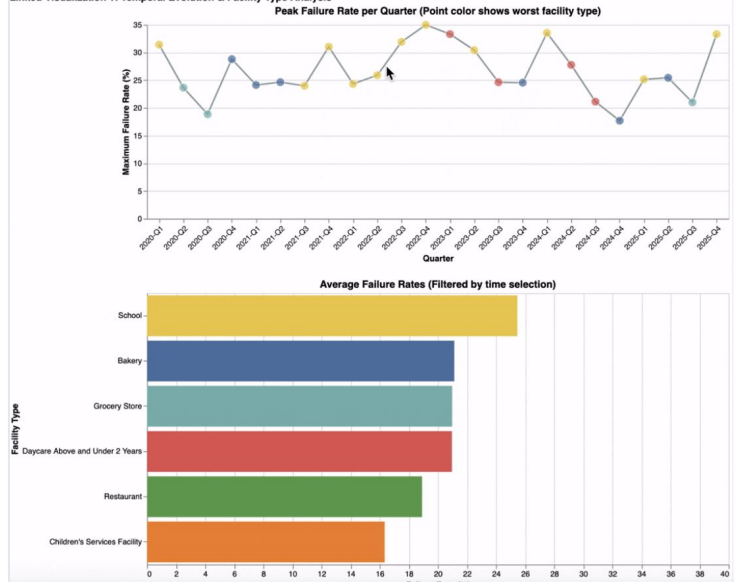
**Interactive heatmap** showing failure rates across top 8 facility types and four housing density levels (Very Low, Low, Medium, High). Darker colors indicate higher failure rates in specific housing contexts.

# Previous Interactive Visualizations Analyses

Interactive multiple-view visualizations developed across Assignment 3

## Temporal Evolution & Facility Type Analysis

Linked Visualization 1: Temporal Evolution & Facility Type Analysis

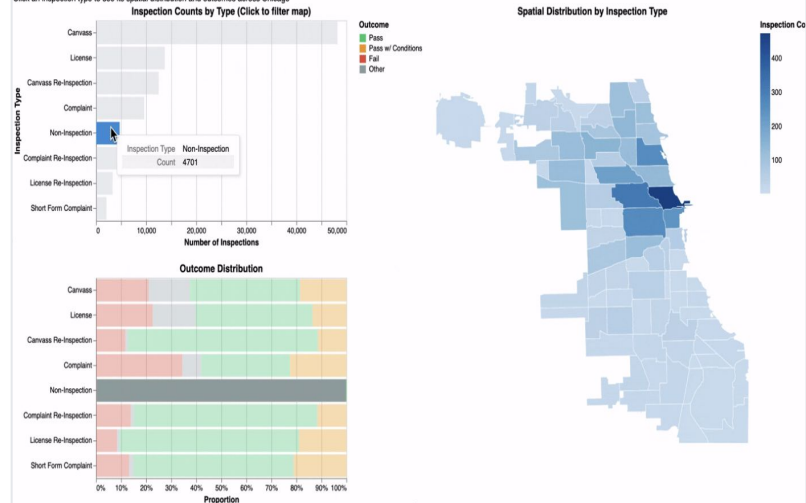


**Linked views with brushing:** Line chart shows quarterly peak failure rates with color-coded points indicating worst facility type. Bar chart updates to show aggregated rates for selected time range.

## Spatial Distribution of Inspection Types

Task 3: Linked Spatial Analysis of Inspection Types

Click an inspection type to see its spatial distribution and outcomes across Chicago



**Coordinated three-panel view:** Choropleth map with sequential blue encoding reveals geographic patterns in inspection deployment. Click inspection types to update all views simultaneously.

# Embeddings - Initial Approach

## Chicago's Food Inspection Evaluation

### Key Idea:

- Chicago's DOPH developed a predictive model
- Goal: Prioritize inspections to find critical violations faster

### Features Used:

- Past violation history (critical, serious, minor)
- Business characteristics (age, license types)
- Neighborhood context (crime density, sanitation complaints)
- Time since last inspection

## Feature Based Embedding

### Approach 1 - Direct t-SNE:

- 13 engineered features → StandardScaler → t-SNE → 2D

### Approach 2 - MLP Embeddings:

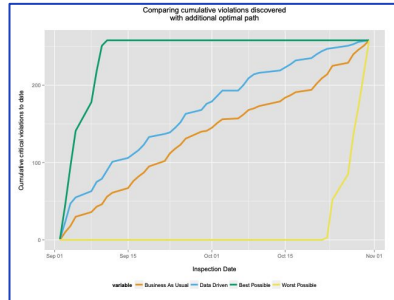
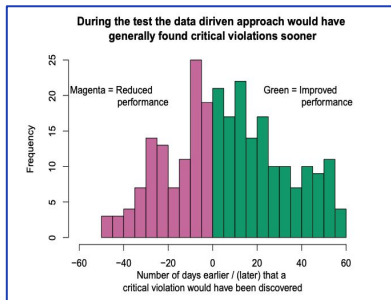
- Same features → Neural Network → Hidden Layer → t-SNE → 2D

## Forecasting restaurants with critical violations in Chicago

*Tom Schenk Jr. (City of Chicago), Gene Leynes (City of Chicago), and Aakash Solanki (City of Chicago), Stephen Collins (Allstate Insurance), Gavin Smart (Allstate Insurance), Ben Albright, and David Crippin (Allstate Insurance)*

*May 15, 2015*

The Chicago Department of Public Health (CDPH) inspects more than 15,000 restaurants with fewer than three dozen inspectors over the course of the year. This paper describes a predictive model designed to identify the presence of a critical violation in a particular food establishment. The goal of this model is to prioritize inspections by likelihood in order to identify the riskiest restaurants earlier, thereby reducing the length of exposure of risky restaurants to patrons. Critical violations were identified approximately 7.44 days earlier over a 60 day period compared to current operations in the out-of-sample test.



# Problems with Initial Approach

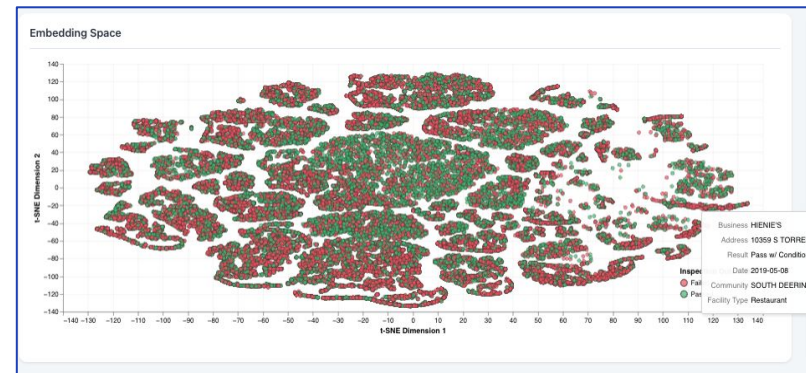
## ! The Embeddings Weren't Meaningful

### Key Issues:

- Points scattered randomly regardless of Pass/Fail status
- No clear clustering by violation type

## Why?

- Features capture **business context**, not **violation content**.
- Two restaurants with same age and neighborhood can have completely different violations.



## Our Realization

- *The 'Violations' field describes what actually went wrong.*
- Similar violation descriptions = similar inspection outcomes
- *Text embeddings can capture this semantic similarity*

### Example Violations:

"OBSERVED RODENT DROPPINGS IN STORAGE AREA"

"FOOD NOT HELD AT PROPER TEMPERATURE - COLD FOOD AT 52F"

[illegible]

# Final Approach: Violations Text Embeddings

## Text-Based Embeddings

### Outline:

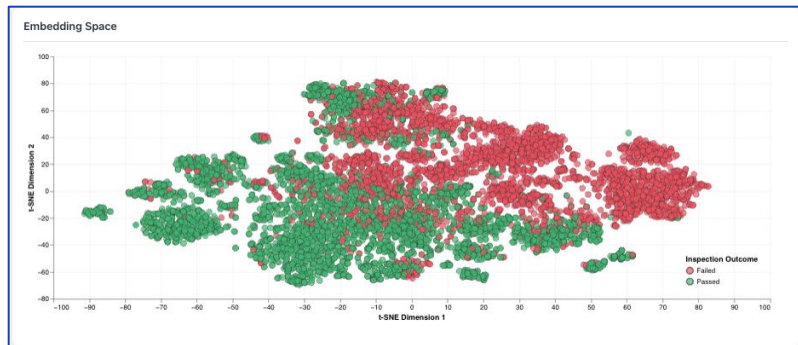
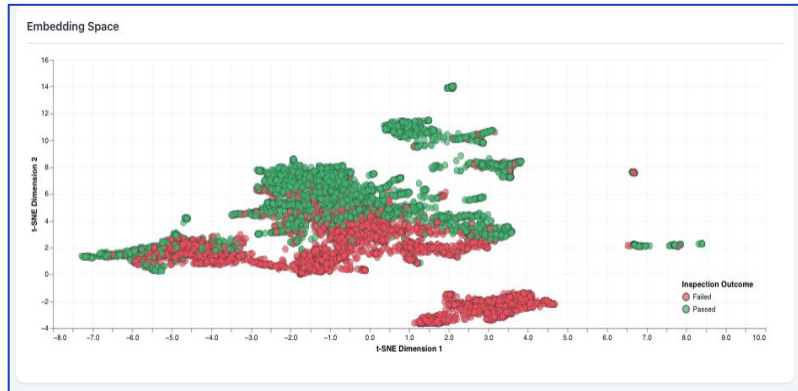
- Clean violations text (lowercase, remove special chars)
- Generate embeddings using Sentence Transformers (384D)
- Apply UMAP/t-SNE for 2D projection

### Pre-Processing:

- Convert to lowercase
- Remove special characters (except pipes separating violations)
- Collapse whitespace
- Calculate embeddings and aggregations offline, store to JSON and load in React/Vega-lite front-end.

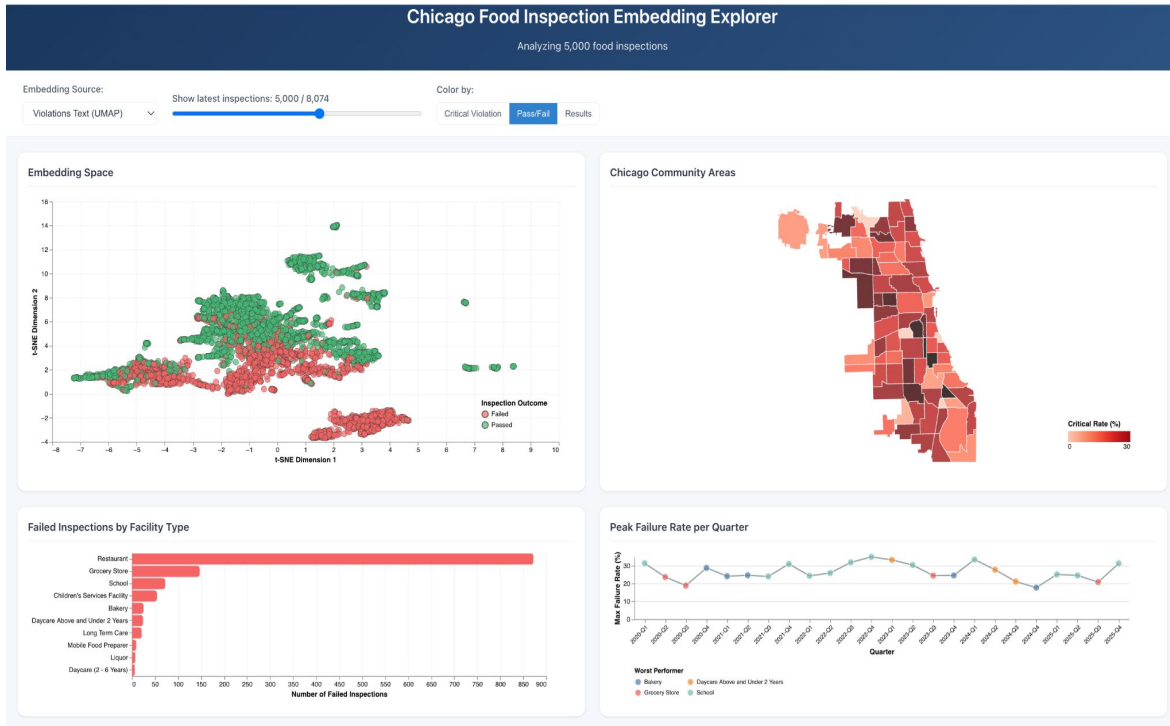
## Improved Results

- *Semantically meaningful clusters*
- Inspections actually cluster by the type of problem they had.
- *Consistent across both UMAP/T-SNE projections*



# Interactive Interface & Visualizations

## 🎯 Main Dashboard Interface



## 📊 What - Where - When

**Embedding Space:** t-SNE/UMAP 2D scatter plot colored by Pass/Fail or Critical Violations

**Chicago Community Map:** Choropleth showing critical violation rates by area

**Facility Type Analysis:** Bar chart of failed inspections by establishment type

**Temporal Evolution:** Line chart tracking peak failure rates per quarter

## 🔗 Linked Views & Interactions

**Embedding Methods:** UMAP, t-SNE for dimensionality reduction

**Feature Engineering:** Violations text embeddings + temporal + spatial features

**Dynamic Filtering:** Slider to control inspection volume (latest 5,000/8,074)

**Color Encoding:** Toggle between configurations and easily see how different facilities have performed over time