Texas A&M University

Python Toolbox for Analyzing Food Deserts and Identifying Locations for Potential New Stores

Joel Arndt, Brooke Bennett, Clay Davlin, Braden Hannagan, Jennyfer Hoang, Fio Llerena, Niú, Yuán

2025 Fall GIS Programming
Dr. Zhe Zhang
November 30, 2025

Food Desert Analysis and Solutions

Introduction

A growing problem throughout the United States, especially in low-income communities, is proper access to nutritious food. These areas are formally defined as a food desert, "a geographic area that lacks sufficient access to grocery stores (Karpyn et al., 2019). Looking closer at Texas, there is a significant number of food deserts, about 12.5% of the state is considered a food desert (Abdullah et al., 2025). This statistic includes both rural areas and major cities in Texas such as Dallas and Houston. This project creates a toolbox in ArcGIS Pro that can identify food deserts and detect premium areas for new grocery stores to solve this problem based on the user's specified county. The tool works with any county in Texas but for discussion purposes our report will focus on San Patricio county in Texas.

Data Description

Both the county and population density data came from a feature layer in the ArcGIS online portal titled "USA 2020 Census Population Characteristics" created by esri_demographics. This feature layer splits up the entire United States into block groups in a polygon format. These block groups contain a population density attribute which is used in calculating priority areas for new grocery stores. Along with this, each block group contains a county attribute which is used to filter the results down to the user's specified county when the tool runs.

The grocery store data came from a feature layer in the ArcGIS online portal, created using OpenStreetMaps and was processed with the help of a notebook. All shops in North America were exported from OpenStreetMaps and added as a layer in the project. Each store has a "shop" attribute that specifies the category of the store. To select only the grocery stores needed for the analysis, a notebook was run that used a SQL statement to select the supermarkets and convenience stores which was then made into a new feature layer to be used in later calculations.

Methodology

To generate the layers that show the possible food deserts within the specified county, our program goes through multiple steps. The first step is to create a buffer around the grocery stores, and clip that buffer with the block group layer from the specified county. This buffer is set at 1.6 kilometers, as that is the standard defined radius associated with a food desert.

After doing the clip operation, centroids are generated for each block group in the county. These centroids will be how a block group is determined to be inside or outside the 1.6km buffer.

In order to calculate the priority, two base statistics must be calculated for each block group. The first calculation is the population density, which is very basically how many people live in the block group divided by the area of the block group. After that, the distance to the nearest grocery store is calculated.

The final calculation done to the block groups is the urgency score. In order to prioritize both higher distance *and* the density of the block groups, the distance and density are multiplied, meaning a higher number means both high density and long distance from the nearest grocery store. This means that both the close and the low density block groups are much less likely to be selected for export.

After the final calculation, the final steps are to sort the block groups based on score, export the block groups as a whole, and select and export the most urgent block groups from the data set.

Results

The tool produces five main feature layers as output, all clipped to the boundaries of the user's selected county. The first layer is the "BlockGroups_FoodDesert", this contains all of the block groups in the areas determined to be a food desert. The second layer produced is the "GroceryStores_Buffer" which creates the one-mile radius around all grocery stores. These two layers help provide context for the candidate layers. The last three layers produced are all variations of new grocery store candidates. The first of these layers is the full candidate results. This layer provides a recommended location for a grocery store in each block group within the food desert without using the scoring system. Both the top three and top one candidate layers use the scoring system to determine an optimal location for a grocery store based on both population density and distance from a grocery store, outputting either one or three options.

When adding all of these feature layers to the map, the areas where residents face the greatest burden of distance from a food source while taking population density into account are clearly shown. With the "BlockGroups_FoodDesert" and "GroceryStores_Buffer" visible on the map with any of the candidate layer options, locations for new store construction can be easily identified to aid in effectively reducing food desert conditions within the selected county.

Discussion

After running the tool, several areas become identified as underserved areas within San Patricio County. Particularly in Mathis where the top three locations lay within. Other areas of note are almost all in or around city centers within the county. This includes the aforementioned Mathis, Sinton, Odem, Gregory, Taft, Portland, Aransas Pass, and Ingleside.

While Mathis contains the three most underserved areas, it is also worth noting that the highest concentration of clusters of underserved areas are in the coastal area of the county around Ingleside and Aransas Pass. Furthermore, there are also points roughly equidistant from the city centers or edges of the counties identified as underserved areas. These are mostly rural areas of the county and rank lower for urgency, but are nonetheless points to consider for future planning of food centers.

Future iterations of the tool could utilize a more refined “Urgency Score”, network analysis, and intercounty data to more accurately determine real food desert areas. Refining these vectors would help reveal optimal locations for food centers. Additionally, increasing scope may reveal broader regional trends, deficiencies that only some regions possess, or ultimately reveal trends present across all locations. Later iterations of the tool can also be structured into different versions that are designed to target specific variables like socioeconomic or geographic boundaries.

Conclusion

Food deserts represent a significant barrier to public health, and efficient planning is required to mitigate their impact. This project culminated in the creation of a specialized analytical tool designed to bridge the gap between spatial data and community action. By integrating population density with proximity analysis, our tool provides a quantitative "Urgency Score" that objectively prioritizes areas for development.

The output of our tool—specifically the identification of the Top 1 optimal site—serves as a powerful recommendation engine for stakeholders. As shown in the San Patricio County analysis, the tool filters out noise and highlights the exact location where a new grocery store would maximize community benefit.

Moving forward, this Python Toolbox offers a flexible foundation for further expansion. While currently optimized for county-level analysis, the logic can be adapted for larger regional studies. By transitioning to a desktop-based toolbox, we ensured that the analysis remains rigorous, accurate, and capable of supporting real-world infrastructure planning decisions.

Limitations

While the Food Desert Analysis Tool successfully identifies high-need areas, there are several limitations to the current implementation:

1. Simplified Scoring Model: The current "Urgency Score" is calculated based solely on population density and linear distance to the nearest grocery store. While effective as a baseline, it does not account for other critical socioeconomic factors that define food insecurity, such as median household income, vehicle ownership rates, or public transit accessibility. A high-density area might be distant from a store but wealthy enough to own cars, whereas a lower-density area might be poorer and more dependent on walking.
2. Geographic Scope & Scalability: The tool is currently optimized to process one county at a time. While this allows for detailed local analysis, conducting a state-wide or national assessment would require iterative manual execution, limiting its immediate scalability for large-regional studies.
3. Distance Calculation: The analysis relies on geodesic buffers and distances. It does not currently utilize Network Analysis (drive-time or walk-time), which would provide a more realistic measure of accessibility by accounting for road networks and traffic barriers.

Future Work

Moving forward, there are several key areas for development to enhance the tool's impact and usability:

1. Integration of Socioeconomic Variables: Future versions of the scoring algorithm will incorporate poverty rates and vehicle access data from the Census Bureau. Weighting these factors into the "Urgency Score" will help prioritize disadvantaged communities more accurately than population density alone.
2. Transition to Web Deployment: We aim to revisit our original goal of a Web Application. By deploying the Python logic as a Geoprocessing Service (via ArcGIS Enterprise or a custom backend), we can make the tool accessible via a web browser, allowing policymakers and the public to use it without needing GIS software.
3. Network-Based Analysis: Replacing simple buffers with Service Areas (Drive-time/Walk-time polygons) would significantly improve accuracy. This would account for physical barriers (like highways or rivers) that might make a "nearby" store actually difficult to reach.
4. Batch Processing Capabilities: To address the scalability limitation, we plan to implement an iterator within the toolbox that allows users to input a list of counties or a state, automatically generating reports for multiple regions in a single run.

Works Cited

Abdullah, Youssef et al. “SUN-706 Geospatial Analysis of Food Deserts and Obesity in Texas: A CDC Data-Driven Heatmap Study.” *Journal of the Endocrine Society* vol. 9,Suppl 1 bvaf149.213. 22 Oct. 2025, doi:10.1210/jendso/bvaf149.213

Karpyn, Allison E et al. “The changing landscape of food deserts.” *UNSCN nutrition* vol. 44 (2019): 46-53.

“OSM_NA_Shops” (2022) [feature layer] Steven Moore. URL: https://www.arcgis.com/home/item.html?id=2191257ce798459e9f649dba3c3f86ea

“USA 2020 Census Population Characteristics” (2025) [feature layer]. Esri. URL: https://www.arcgis.com/home/item.html?id=ebeb65deb5c14f4d8849fd68944b7ee6