I. TITLE: Food Deserts in Chicago, IL

II. INTRODUCTION/BUSINESS PROBLEM

This is an investigation into the availability of fresh produce/groceries in Chicago's most impoverished areas and will analyze for improvement in these areas.

Many studies have shown that the lack of readily available access to fresh produce/groceries is directly detrimental to diet and health but can also indirectly impact education and career progression. The results of this study will assist civic leaders and urban planners in the strategic planning for the success of Chicago residents in impoverished areas, as well as provide a strategic analysis to potential national grocery chain stores.

A food desert is defined as "regions where people have limited access to healthful and affordable food. This may be due to having a low income or having to travel farther to find healthful food options." *Source: Medical News Today*

According to the same article in Medical News Today: *The United States Department of Agriculture (USDA) defines a food desert as an area that has either a poverty rate greater than or equal to 20% or a median family income not exceeding 80% of the median family income in urban areas, or 80% of the statewide median family income in nonurban areas.*

In order to qualify as a food desert, an area must also meet certain other criteria. In urban areas, at least 500 people or 33% of the population must live more than 1 mile from the nearest large grocery store. In rural areas, at least 500 people or 33% of the population must live more than 10 miles from the nearest large grocery store.

In this analysis we will focus in on the urban 1-mile radius but will let Data Science algorithms group Zip Codes by the K-Means squared algorithm.

III. DATA

The analysis will rely on data extracted from various sources as reflected in Table 1.

	SOURCE NAME/LINK	DATA RETRIEVED	EXAMPLE
1	Foursquare City Guide	CATEGORY = Grocery Stores per Zip Code in Chicago, IL	See Table 2
2	Zip Atlas	Median household income in Chicago, IL per Zip Code	See Table 3
3	Medical News Today	Definitions and summary of a Food Desert	See Link in SOURCE NAME/LINK column
4	Personal knowledge of Chicago's neighborhoods and geography	Verification of Data Science results	N/A

Table 1: Summary of data sources

a. SOURCE 1: Foursquare City Guide

Application that allows users to search venues/categories by geographic location. In this study we input geographic location as latitude and longitude and search for grocery stores in a 1-mile radius.

	name	categories	lat	lng
0	sweetgreen	Salad Place	41.884964	-87.624728
1	Roti Modern Mediterranean	Mediterranean Restaurant	41.886048	-87.624948
2	Wildberry Pancakes & Cafe	Breakfast Spot	41.884412	-87.623047
3	Chicago Architecture Center	Tour Provider	41.887720	-87.623650
4	Virgin Hotels Chicago	Hotel	41.886065	-87.625853
5	St. Jane Chicago	Hotel	41.886573	-87.624902
6	Giordano's	Pizza Place	41.885130	-87.623761
7	Chicago Cultural Center	Museum	41.883640	-87.624671
8	LondonHouse Chicago, Curio Collection by Hilton	Hotel	41.887832	-87.625426
9	Intelligentsia Coffee	Coffee Shop	41.884517	-87.625783
10	The Chicago Theatre	Theater	41.885539	-87.627151
11	Lakefront Trail	Trail	41.883470	-87.625750
12	Garrett Popcorn Shops - Millennium Park Plaza	Snack Place	41.884854	-87.624499
13	Cloud Gate by Anish Kapoor	Outdoor Sculpture	41.882668	-87.623319
14	Millennium Park	Park	41.882598	-87.624126
15	Jay Pritzker Pavilion	Amphitheater	41.882614	-87.621782
16	Chicago Riverwalk	Waterfront	41.887280	-87.627217
17	Radisson Blu Aqua Hotel, Chicago	Hotel	41.886308	-87.619921
18	Mariano's	Grocery Store	41.885326	-87.618923
19	Maggie Daley Ice Skating Ribbon	Skating Rink	41.884093	-87.619374

Table 2: Foursquare sample data as retrieved and parsed in Python

b. Source 2: Zip Atlas

Zip Atlas is a structured collection of zip code, area code, city and state demographic, social and economic profiles.
Source: Zipatlas.com

In this study we correlate median household income with selected Chicago, IL Zip Codes. Source data is from calendar year 2020.

#	Zip Code	Location	City	Population	Avg. Income/H/hold	National Rank
1.	60606	41.881988, -87.637329	Chicago, Illinois	1,682	$100,377.00	#259
2.	60601	41.885805, -87.622911	Chicago, Illinois	5,591	$77,374.00	#961
3.	60611	41.893024, -87.618998	Chicago, Illinois	26,522	$69,889.00	#1,558
4.	60614	41.922667, -87.652537	Chicago, Illinois	65,474	$68,324.00	#1,712
5.	60603	41.880614, -87.629311	Chicago, Illinois	378	$61,815.00	#2,591

Table 3: Zip Atlas sample data as shown on website

IV. METHODOLOGY/EXPLORATORY DATA ANALYSIS

a. Zip Atlas: Median Income Per Household

i. Retrieved from Zip Atlas (See Sect III) using the Beautiful Soup Web Scraper in Python
RATIONALE: *This methodology allows you to easily retrieve data from published sites*

ii. Identify the correct table from the multiple links on the site and from the HTML page source content
RATIONALE: *Webscraping methodology requires that you correctly identify the table*

iii. Latitude and Longitude were retrieved as a single value (x,y) and these were separate to x= latitude and y = longitude
RATIONALE: *Will need separate lat/long values to pass on to Foursquare*

iv. Table field were renamed with simpler naming and to eliminate spaces and special characters
RATIONALE: *Allows searching and indexing of table fields*

v. An ordered sequence number was dropped from the table and we thus relied on the default table numbering in Python
RATIONALE: *Eliminate duplicate numbering*

vi. Dropped the following columns: CITY, NATIONAL RANK
RATIONALE: *Field not used in the analysis*

vii. Removed the dollar sign ($)
RATIONALE: *Data cleaning to simplify data manipulation*

viii. Dropped zero values and Naan values
RATIONALE: *Data cleaning*

ix. Converted datatypes as follows: Income = *float*, latitude = *float*, longitude = float, and ZipCode = *int*
RATIONALE: *Data cleaning to simplify data manipulation*

x. Used Python Folium maps to create several geographic plots
RATIONALE: *Useful technique within Python to plot data on geographic maps*

xi. Utilized the K-means Clustering algorithm to create k=4 clusters of Chicago Zip Codes, Median Income, and Population
RATIONALE: *Although this is a small dataset, it was useful to use statistical tools in Python as a grouping methodology*

b. Foursquare: Venues per Zip Code

i. Retrieved list of venues (Grocery Stores) per Zip Code using Foursquare developer credentials
RATIONALE: *Foursquare provides updates venue information per zip code*

ii. A 1-mile radius was selected for the search parameter
RATIONALE: *As per the definition of a Food Desert (See Table 1-3) – note that a small side trial was performed with a 2-mile radius to observe the difference*

iii. Foursquare center points (Lat/Long) were selected in three different ways:
    1. Based on personal knowledge of geography
    2. Based on the x5 lowest median incomes
    3. Based on the x5 highest median incomes

RATIONALE: *Investigation included looking into the most economically challenged areas and it was also useful to compare this to perceived Food Deserts based on observation*

iv. One-hot encoding was used to group results and depict the x10 venues per zip code
RATIONALE: *One-hot encoding assists in analyzing categorical data*

V. RESULTS

a. With K-Means analysis we segmented the data into K=4 groups and observed that most of the areas with the lowest median income are in the south, southwest, and west sides of the city as depicted by the red dots in Figure 1.

DOT COLOR	CLUSTER #	MED INCOME
Purple	2	$68,300-$100,380
Blue	1	$47,510 - $61,820
Yellow	3	$33,630-$45,180
Red	0	$14,200 - $32,620

Table 4: Keycode explanation of Figure 1

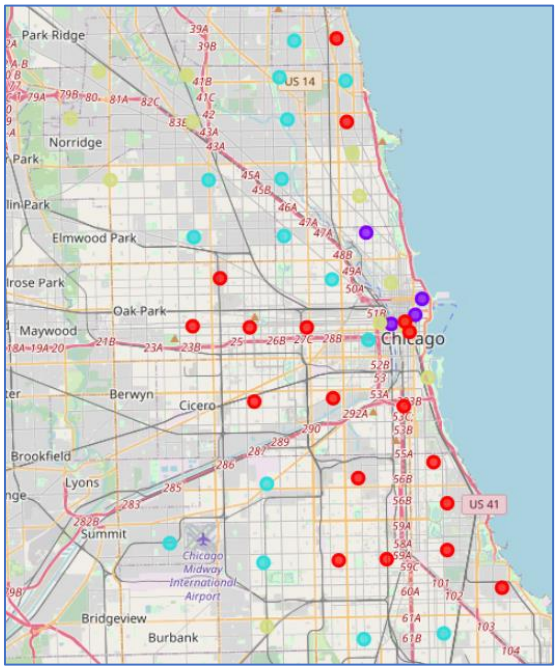

Figure 1: K-Means grouping of Chicago Zip Codes by Median Income per Household

b. The x5 zip codes with the lowest median income were selected as the central points for analysis – See Table 5. These will be known as "Analysis Zones"

Cluster Labels	ZipCode	Income	latitude	longitude	Pop	%of Tot Pop
0	60602	23750.0	41.883067	-87.629187	70	0.002432
0	60637	23228.0	41.780673	-87.603500	57090	1.983289
0	60624	22426.0	41.880048	-87.722352	45647	1.585763
0	60621	19718.0	41.776404	-87.640058	47514	1.650622
0	60653	14205.0	41.819722	-87.611809	34502	1.198589

Table 5: x5 most impoverished zip codes (Analysis Zones)

c. The x10 Most Common Venues of the Analysis Zones are depicted in Table 6
   i. It is interesting to note the prevalence of Fast-Food Restaurants in the top x3 spots for two of the zip codes – an indication of a continued problem with lack of access to healthy food
   ii. It is also interesting to observe the prevalence of Art Galleries and Museums – an indication that efforts are being made to improve the communities

Zip	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
60602	Hotel	Snack Place	Gastropub	Steakhouse	Concert Hall	Salad Place	Bar	Coffee Shop	Theater	Amphitheater
60621	Fast Food Restaurant	Sandwich Place	Train Station	Café	Gas Station	Food	Pharmacy	Park	Mexican Restaurant	Light Rail Station
60624	Train Station	Fast Food Restaurant	Fried Chicken Joint	Sandwich Place	Discount Store	Food	Mobile Phone Shop	Park	Garden	Grocery Store
60637	Coffee Shop	Sandwich Place	History Museum	Bookstore	Park	Grocery Store	Food Truck	Café	Noodle House	Cosmetics Shop
60653	Beach	Art Gallery	Discount Store	BBQ Joint	Historic Site	Seafood Restaurant	Southern / Soul Food Restaurant	Performing Arts Venue	Coffee Shop	African Restaurant

Table 6: Top x10 venues in Analysis Zones

d. Figure 2 depicts the grocery stores per Analysis Zones
    i. The blue circles are a 1-mile radius from the Analysis Zone
    ii. There are x5 Analysis Zones (blue circles) as described in Sect V-b
    iii. The GPS Markers depict a grocery store

Most of the Analysis Zones have at least 1 Grocery Store per 1-mile radius – eliminating the tag of "Food Desert".

Analysis Zone 60602 is an interesting case study as:

- This does not have any Grocery Stores within a 1-mile radius: This is mainly a tourist/office building location
- It encompasses purple and blue dots, which correspond to the #1 and #2 most wealthy cluster points: Further analysis would be required to determine the impact of those in the red dot within this Analysis Zone

A comparison with the x5 wealthiest zones clearly shows a stark difference in the density of grocery stores per zone – with the wealthiest zones having multiple grocery stores within a 1-mile radius.

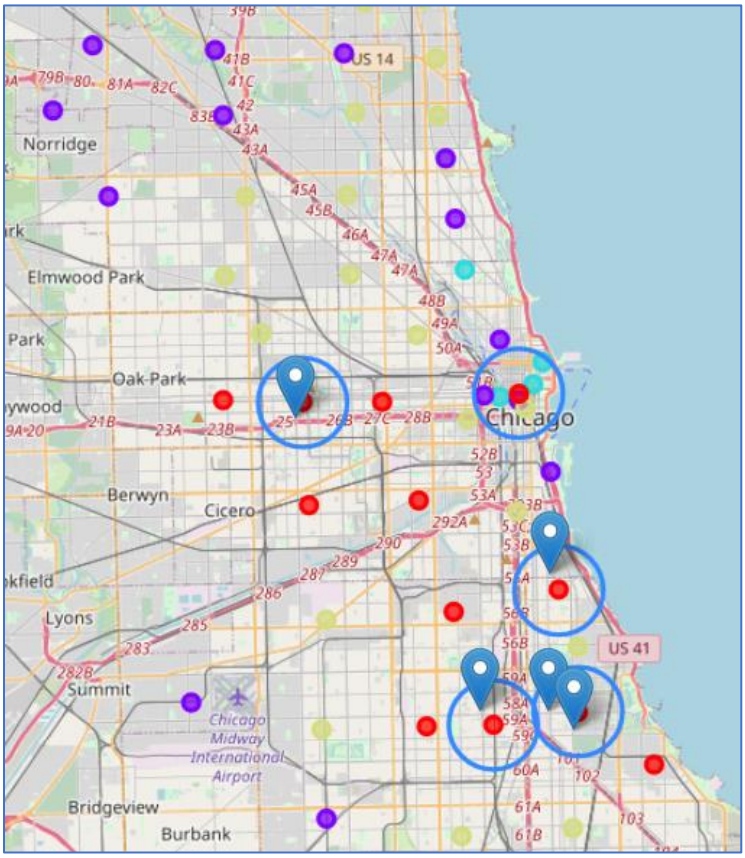

Figure 2: Map of Grocery Stores in the Analysis Zones

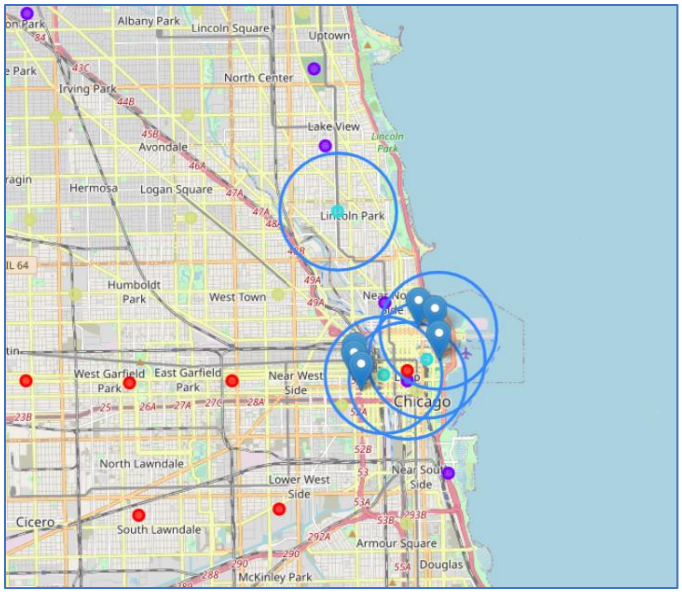

Figure 3: Map of Grocery Stores in x5 Highest Medium Income Zones

e. Fig 4 depicts the number of grocery stores per zip code in the Analysis Zones. This reinforces the points made in Section V-d

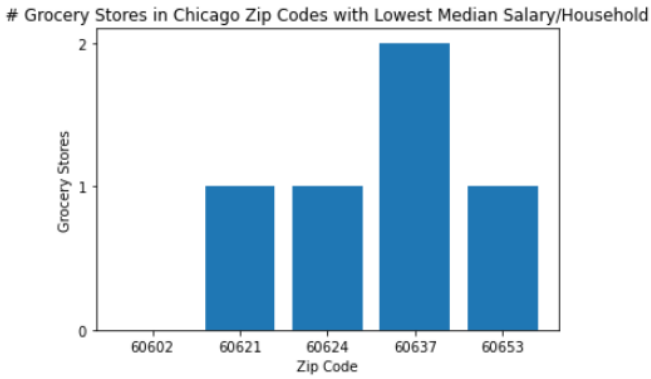

Figure 4: # Grocery Stores per Zip Code in Analysis Zones

f. Fig 5 depicts the number of grocery stores per zip code in the x5 areas with the highest mean income per household. The figure appears to depict food deserts in wealthy areas, but upon closer inspection it is observed that:
   - Zip 60614 is an area with restaurants, bars, cafes, etc. and in fact, there are several grocery stores within a 1.5-mile radius
   - Zip 60603 is in the downtown area with many overlapping inspection areas that indeed do have multiple grocery stores in the proximity

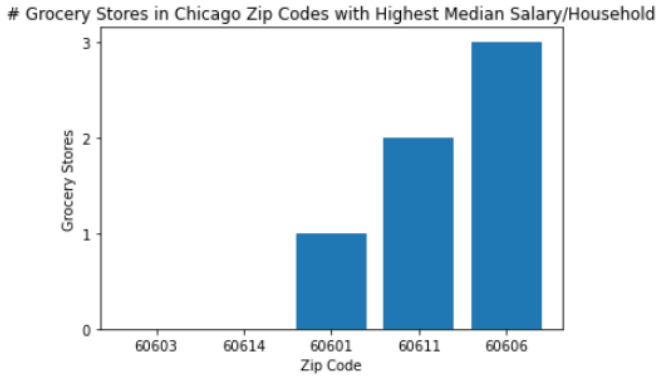

Figure 5: # Grocery Stores per Zip Code in Areas with Highest Medium Income

VI. DISCUSSION

The results obtained seem to depict that Chicago is not impacted by food deserts – per the availability of at least one grocery store per 1-mile radius, however, there is a stark difference between the areas with the highest median income per household vs those with the lowest median income per household. A more complete study would also look at other factors such as:

- Transportation Analysis: How feasible is it to reach these venues (grocery stores)?
  - Walkability - Is it safe (and practical) to walk 1 mile to purchase grocery stores?
  - Availability of public transit
  - Availability of ride share drivers
- Analysis of Grocery Stores:
  - Are the grocery stores indeed provide healthy items at an affordable price?
  - Quality of available healthy items

VII. CONCLUSION

It is good to see that the areas under study did not qualify as "food deserts" – suggesting the availability of healthy food to residents of Chicago, IL. The city has recently made efforts to reduce some of the food desert deltas such as:

- Opening of Whole Foods Market in Chicago's Englewood Community - More Info here
- Organizations such as the NFL Chicago Bears investing in healthy kitchen resources in Chicago's south side - More info here

There is still plenty of work needed to close the gap between Chicago's poorest and wealthiest neighborhoods. This study clearly presents opportunities for development from civic leaders, as well as opportunities for investment in private industry through the opening of additional grocery stores that can:

- Provide healthy food items at an affordable price
- Are easily accessible
- Cater to the culinary diversity of the local neighborhoods