Finding the Location for a New Whole Foods in Washington, DC

Greg Babbitt

24 August 2020

Introduction:

Grocery stores need to be located as near as possible to their customers. This convenience allows easy access for the customers, who then end up spending more at the stores. Choosing where exactly to place a new grocery store is incredibly important. Building a new store is expensive and requires investment in labor, construction costs, and marketing. The audience or stakeholders for this project are grocery store management, who needs to know they can maximize revenue to justify the investment.

Business Problem:

Choose the best location for a new Whole Foods Market in my home city, Washington, DC. Whole Foods is a high-end grocery store chain owned and operated by Amazon. Whole Foods attracts clients with a high relative income and there are already several stores throughout the city. The goal is to choose a location where incomes have risen recently, because that population has many potential Whole Foods customers, and where there are no other grocery stores nearby.

Data:

The DC city government makes income and census tract data available to the public. The data contain median income by census tract for the years 2005 and 2018, which should give an accurate picture for how incomes have grown in certain neighborhoods.

Here is an example of one row of data in the census/income file:

A	B	C	D	E	F	G	H	I	J	K	L	M
OBJECTID	GIS_ID	FEDTRACT	TRACTNO	AREASQM	POPDENSI	TOTAL	NAME	FULLTOTA	AGE0_17	AGE0_17P	AGE18PLU	AGE18PLU
1	Tract00_0	22.01	22.1	0.16027	21869.35	3505	2000 Tract	1135	824	0.23509	2681	0.7649

AV	AW	AX	AY	AZ	BA	BB	BC	BD	BE	BF	BG
SINGLECO	MULTICOS	MULTIPER	SINGLEPEF	FAGI_TOT	FAGI_MEA	FAGI_MED	FAGI_TOT	FAGI_MED	AREA	LEN	SHAPE_Le
0	0	61	272	53943913	36179.69	28487	49226475	32345	0	0	2575.694

The income data have a federal tract numbers and geographic locations, so it will be possible to compare that data to the neighborhood data in Foursquare. The data also come in a Shapefile format, a Geographic Information Science datatype that includes a geography variable. This would allow me to see the exact shape and locations of the tracts.

You can find a link to the DC income data here:
https://opendata.dc.gov/datasets/census-tracts-by-median-income-2000 and
https://opendata.dc.gov/datasets/acs-2018-median-household-income-variables-

tract/data?geometry=-77.153%2C38.891%2C-76.892%2C38.938&orderBy=B19049_001E&orderByAsc=false

The Foursquare venue data has the grocery store category, so we can filter the queries for to discover competitor grocery stores and other Whole Foods Markets in Washington, DC.

Methodology:

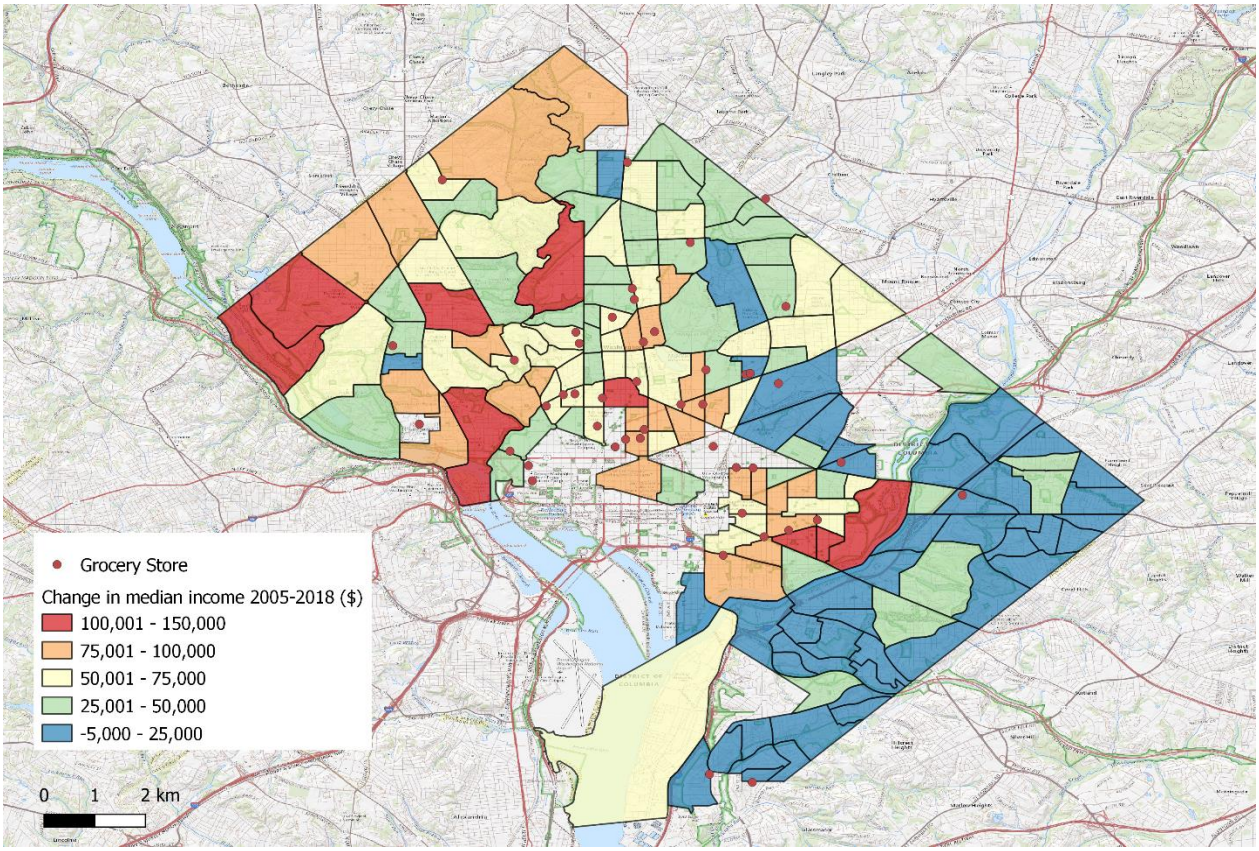
This project took three main steps. The first step involved downloading, cleaning, aggregating, then analyzing the Washington, DC Census Tract data. I downloaded data from 2005 and 2018 that included median income for 161 census-delineated tracts (neighborhoods). Each tract has a unique ID, which allowed me to join the two datasets using Python Pandas. I needed to trim the data to include only the required columns and to alter the format of the data in the columns so Pandas would recognize the unique IDs as string and the incomes as real numbers. I could then use Pandas to calculate the change in median income between 2018 and 2005, which gave me the top candidate tracts that experienced significant income growth.

The second step was to use the FourSquare API to query for venues in Washington, DC. I used GIS Software called QGIS to calculate the center point for each census tract, then used those coordinates to query FourSquare for all venues nearby. I then used Pandas to extract only the venues with the category “grocery stores” because these include Whole Foods Markets and other competitors.

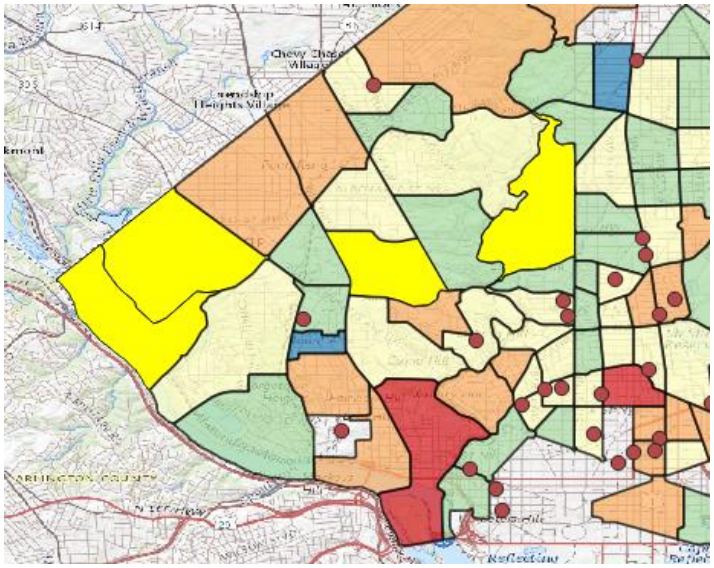
Now that I had the location of all grocery stores in Washington, DC, in the final step I used QGIS to calculate in which neighborhood each grocery store sits. This allowed me to simultaneously compare the grocery store locations with the income change data to determine the best candidate spots for a new Whole Foods.

Results:

This graphic shows the locations of existing grocery stores in Washington, DC and the



census tracts that experienced income growth between 2005 and 2018. You can see that there are nine tracts that median income grew between $100,000 and $150,000. However, only four of those tracts (6, 9.01, 9.02, and 26), seen below highlighted in yellow, are not near an existing grocery store. These are the best candidate locations for a new Whole Foods Market. They offer both a population with increased income and limited competition from other grocery store chains.



Discussion:

This analysis provides an excellent start in choosing a new location for a Whole Foods Market by determining the most suitable neighborhoods based off change in median income and the potential presence of competitors. There are several ways to further refine our search for a new Whole Foods Market by including additional datasets. Other important factors for the location of a grocery store include access to transportation, traffic patterns, zoning restrictions, price/availability of real estate, etc… Additional analysis could include these variables to try and focus in on specific, address level targets for a new store.

Conclusion:

This project used two disparate datasets and fused them based on geography and income changes. The result were neighborhood level recommendations for a new Whole Foods Market that could attract wealthy customers to ensure high sales and justify the cost of building a new grocery store. Further analysis will narrow down the options within those neighborhoods by incorporating additional data sources.