**SIADS 591 (FALL 2020)**

**FINDING AFFORDABLE HOUSING SITES IN MICHIGAN**

**Alan Fehsenfeld, Koon Leong Ho, George Thio**

**MOTIVATION**   
Our project will aggregate data from multiple APIs and databases to enable users to more easily identify potential locations for affordable housing developments in the state of Michigan. Potential sites for affordable housing will be identified using data such as proximity to transportation, [Walk Score®](https://www.walkscore.com/), lot size, crime statistics, etc. For a complete list of criteria, see Appendix A. Using this data, we will create interactive visualizations, with heat maps, markers, and filters that will identify areas amenable to affordable housing.

**OBJECTIVES**Less sophisticated affordable housing developers identify potential development sites through word of mouth or by driving by potential sites. More sophisticated developers use single databases. We propose to go one step further by aggregating information from multiple databases into an integrated data set. By providing a robust set of reporting features that will allow developers to easily customize data views, we will maximize both their chances of project success by ensuring all relevant criteria have been captured and the efficiency of their search process.  
  
Given caps on maximum rental rates, the ability to secure adequate funding is an important determinant of success and financial viability for affordable housing projects. In Michigan, the portion of the annual state budget earmarked to support new affordable housing development is managed by the [Michigan State Housing Development Authority](https://www.michigan.gov/mshda/) (MSHDA) which uses an objective points system to score project proposals and determine the distribution of government grants. The MSHDA scoring criteria are shown in items #1-7 of **Appendix A - Commercial Property Criteria** which includes an explicit calculation of Affordability (#5). While points are given for Historic Rehabilitation Potential (#7), sites with existing buildings and sites with raw land are eligible for grants and accordingly, our project will allow users to identify both types of locations. We will also consider all properties whether they are currently offered for sale or not.

**DATA SOURCES**For a full list of data sources, see Appendix A. Two primary sources will be the [Google Maps API](https://cloud.google.com/maps-platform/?utm_source=google&utm_medium=cpc&utm_campaign=FY18-Q2-global-demandgen-paidsearchonnetworkhouseads-cs-maps_contactsal_saf&utm_content=text-ad-none-none-DEV_c-CRE_460848633508-ADGP_Hybrid%20%7C%20AW%20SEM%20%7C%20BKWS%20~%20Google%20Maps%20API-KWID_43700033921822021-kwd-335425467-userloc_9060815&utm_term=KW_google%20maps%20api-ST_google%20maps%20api&gclid=Cj0KCQiAw_H-BRD-ARIsALQE_2Mb26ql5HxkookcoUUDKMCiUdR2CSv7n1ZIzdqqjSQxORLJ66t--MYaAgNgEALw_wcB) and the [Walk Score API](https://www.walkscore.com/professional/api.php). To identify public transportation stops, we will use Google Maps which is estimated to contain petabytes of data. Information about the transit dataset can be found here: <https://developers.google.com/transit/gtfs#how-do-i-start>

Google sourced files will be comma delimited text files with the following schema:  
<https://developers.google.com/transit/gtfs/reference#stopstxt>

There is also a Python module called [transitfeed.py](https://pypi.org/project/transitfeed/) that can be used to access Google transit data.

[Walkscore.com](https://www.walkscore.com/) will be used to identify the Walk Score of properties for potential affordable housing developments based on the premise that greater proximity to amenities is a desirable characteristic for affordable housing dwellers from the perspective of both cost and convenience. The Walk Score is a measure of how easy it is to walk to nearby amenities and ranges from 0 (a car is needed to easily reach local amenities) to 100 (many amenities are within easy walking distance). The Walk Score API contains records for most addresses in the United States, Canada, and Australia. According to the [United States Postal Service](https://www.usps.com/) (USPS), there are 160 million delivery addresses in the United States. We estimate that there are at least this many Walk Score API records. The API records can be accessed through this URL: <https://api.walkscore.com/score>.

The Walk Score schema can be found here: <https://www.walkscore.com/professional/api.php>. The data is returned in a JSON file.

To keep data costs manageable while still establishing clear proof of concept, we may complete the project based on a single county in Michigan, before extending the scope state-wide or nation-wide as time and cost permit.

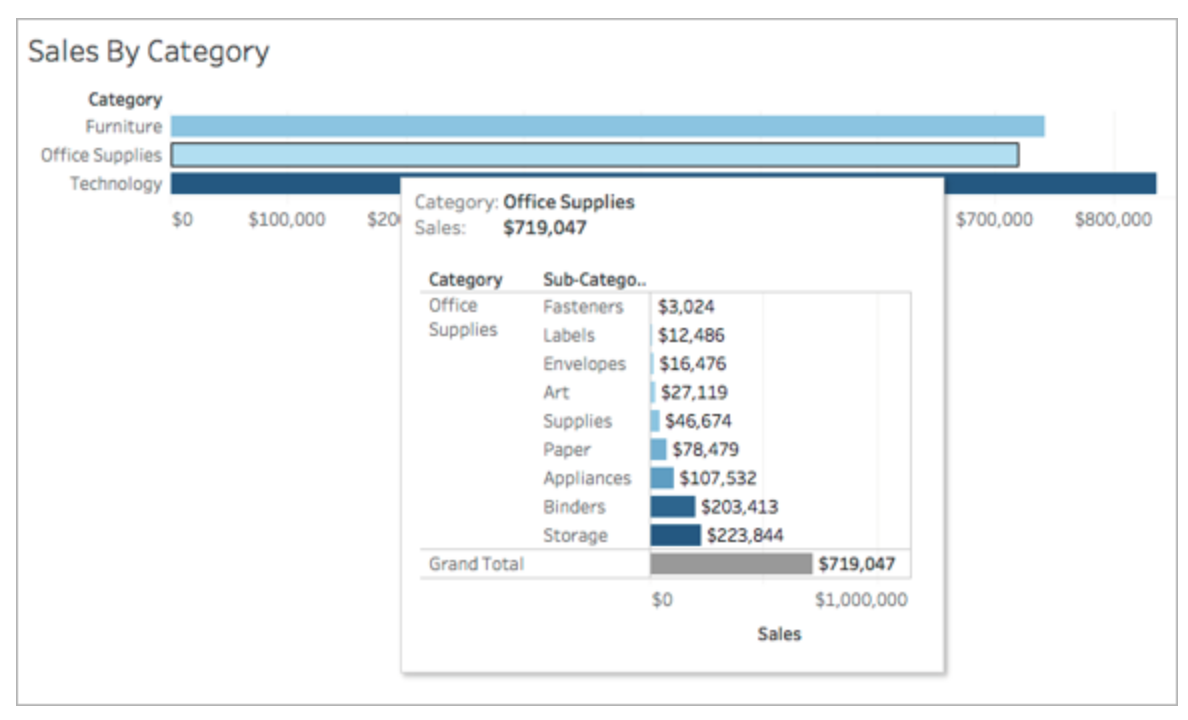
**DATA PROCESSING**  
For each record in our datasets, we will identify a USPS address, geocode, or other unique geographic identifier ([GEOID](https://www.census.gov/programs-surveys/geography/guidance/geo-identifiers.html)) for each record. We will convert all point locations to latitude and longitude and will encode county level information with 5 digit [FIPS codes](https://www.nrcs.usda.gov/wps/portal/nrcs/detail/national/home/?cid=nrcs143_013697). For the Walk Score API and the Google transit API, we are primarily interested in the latitude and longitude records

The Walk Score API returns a latitude and longitude record with four decimal places.

The Google transit data contains data sets for each public transportation system. We will have to merge data files for all Michigan transit systems. Next, we will have to make sure that each stop has a regular route. Finally, we will return a latitude and longitude record for each stop.

**DATA AGGREGATION AND MANIPULATION**Some datasets will contain records for specific properties. For instance, one dataset might have data related to the size in acres of a property. Other datasets will contain records that will be used to describe the area around a specific property. As an example, the Google transit dataset will be used to determine the proximity of the nearest transit stop to each property.

For all property specific data, we will need to link the records by latitude (lat) and longitude (lng). For data describing the area around a specific property, we will need to calculate the distance to each property and calculate a value for each property.  
  
Some of the Python libraries and packages we plan to use include numpy, pandas, requests, and datetime. Where available, appropriate and affordable, we will write functions that take a data frame of raw property listings (i.e. based on tax and census block data) as arguments and aggregate the Commercial Property Criteria using lat/long coordinates which will be done using API calls where available, appropriate and affordable. Based on our preliminary EDA, these functions can be written using basic data manipulation and data conversion techniques.

**DATA STORAGE**The data will be stored in a Postgresql database after it has been processed. The visualization will use the data from the database.  
  
**VISUALIZATIONS**Good sites for affordable housing developments should have multiple favorable characteristics. But in addition to producing a multi-factor composite score, we will also allow users to filter and display multiple criteria based on their particular characteristics of interest -- e.g. by use of a tooltip as shown in the Tableau example below.  


Source: [Tableau: Create Views in Tooltips (Viz in Tooltip)](https://help.tableau.com/current/pro/desktop/en-us/viz_in_tooltip.htm#:~:text=A%20Viz%20in%20Tooltip%20is,that%20mark%E2%80%94in%20its%20tooltip)  
.  
For example, some developers may be more sensitive to crime rates and be willing to sacrifice proximity to public transportation in favor of maximum neighborhood safety. The final product will be an interactive map with heat maps, markers and filters that allows users to compare properties based on multiple characteristics.

**PROJECT RESPONSIBILITIES**See Appendix B for a list of responsibilities and a timeline.

**APPENDICES  
  
Appendix A - Commercial Property Criteria**( #1-7 are MSHDA geographic related criteria)

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Criteria** | **MSHDA Criteria?** | **Data Source** |
| 1 | Proximity to Transportation (5 pts) | Y | Google Maps  <https://developers.google.com/places/web-service/overview> |
| 2 | Proximity to Amenities (4 pts)   * Grocery stores * Pharmacy * Doctor/Clinic * Parks (greenspace) * Schools * Senior Centres | Y | Google Places  <https://developers.google.com/places/web-service/overview> Google? |
| 3 | Education, Health and Well-Being, etc (4 pts)  (from “Opportunity 360)   * State -> County -> Census Tract (size based on pop density) * Go by Census Tract or if ⅓ mile next to another one, can use that one | Y | MSHDA  <https://developers.google.com/places/web-service/overview>  <https://www.michigan.gov/documents/mshda/mshda_li_ca_34_tab_y_opportunity360_points_679968_7.pdf> |
| 4 | Developments Near Downtowns/Corridors (10 pts) | Y | Walk Score®  <https://www.walkscore.com/> |

**Appendix A - Commercial Property Criteria (cont’d)**( #1-7 are MSHDA geographic related criteria)

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Criteria** | **MSHDA Criteria?** | **Data Source** |
| 5 | Affordable/Market Rent Differential (0 or 5 pts)   * Is there an apartment building within ¼ mile that is >= 40% above the affordable rate (i.e. median annual income x 30% / 12 x 60%) * Affordable = 100%; Commercial > 140% * Have to match similar properties (e.g. 1br vs 1br) * Trying to establish need for affordable housing in areas with high housing costs (probably should look > ¼ mile) | Y | Apartments.com [apartments.com](https://www.apartments.com/)  National Multifamily Housing Council (NMHC) <https://www.nmhc.org/research-insight/data-sources/data-sources-apartment-rents/>  Realty Mole <https://www.realtymole.com/api>  Rentometer <https://www.rentometer.com/> |
| 6 | Developments in Opportunity Zones (HUD)/Rising Tide Communities (MHDA) (3 pts)   * Special Census Track designations * Merge data into our own data | Y | Housing and Urban Development (HUD) <https://www.hud.gov/>  MSHDA <https://www.michigan.gov/mshda/> |
| 7 | Historic Rehabilitation Potential | Y | Core Logic1 |
| 8 | Crime Data | N | Michigan.gov <https://www.michigan.gov/documents/msp/Violent_Property_Crimes_CityTwp_2019_697037_7.pdf> |
| 9 | Lot Size | N | Core Logic |

**Appendix A - Commercial Property Criteria (cont’d)**( #1-7 are MSHDA geographic related criteria)

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Criteria** | **MSHDA Criteria?** | **Data Source** |
| 10 | SEV2 | N | Core Logic |
| 11 | Currently Zoned as Multi-Family Housing | N | Core Logic |
| 12 | Housing Vacancies | N | Core Logic |
| 13 | Neighborhood Type (Commercial, Residential, Mixed) | N | Core Logic |
| 14 | Unemployment Rate | N | U.S. Bureau of Labor Statistics (BLS) <https://www.bls.gov/> |

1[Core Logic](https://www.corelogic.com/about-us/data.aspx) is a leading supplier of U.S. real estate, mortgage, consumer and specialized business data

2[SEV](https://www.michigan.gov/taxtrib/0,4677,7-187-25923-126336--,00.html#:~:text=State%20Equalized%20Value%20%2D%20One%20half,each%20tax%20year%20at%20issue.): State Equalized Value is one half (1/2) of a property's true cash value  
  
**Appendix B - Responsibilities and Timeline**

|  |  |  |
| --- | --- | --- |
|  | **Timeline** | **Due Date** |
| 1 | Decide on Commercial Property Criteria | 11/30/2020 |
| 2 | Identify Data Sources | 12/3/2020 |
| 3 | Assign Responsibilities to Examine Data Sources | 12/7/2020 |
| 4 | Submit Draft Proposal | 12/9/2020 |
| 5 | Pull, Verify, and Clean Data | 12/16/2020 |
| 6 | Data Integration | 12/23/2020 |
| 7 | Develop Interactive Map | 12/30/2020 |
| 8 | Written Report - First Draft | 1/6/2020 |
| 9 | Written Report - Final Draft | 1/13/2020 |
| 10 | Submit Project | 1/17/2020 |