Aerotropolis Aero-City Land Use Design Using Machine Learning

Bill Huang

CP6552 Design of Smart Urban Systems

Prof. Perry Yang

Spring 2023

Contents

Motivation

The motivation for using data to determine the land use for Airport City started for asking the question "How should Atlanta's Airport City be designed and can we use a data-driven methodology to assist us in doing so?" Airport's City design can be thought of in urban form and infrastructure. The urban form and infrastructure of a development is informed by its zoning and land use. Those two codes in Atlanta define the physical and activity constraints on the design of a building, infrastructure, or landscape. The idea behind this algorithm would be to suggest a land use design for the locale of Airport City based upon how other developments in Atlanta have been built while accounting for the specific details of the Airport City region.

Methodology Road Map

To make the algorithm work, it needs to be trained to the data of other developments around the Atlanta area. A diagram of the process to arrive at a suggested land use for Airport City by the data is detailed below in Figure 1. After collecting data the model first part is trained on the data and outputs an ideal scenario of land uses for Airport City. However, these ideal scenario land use outputs need to be applied to the real world, so real world constraints are applied by using an optimization model on the regression output.

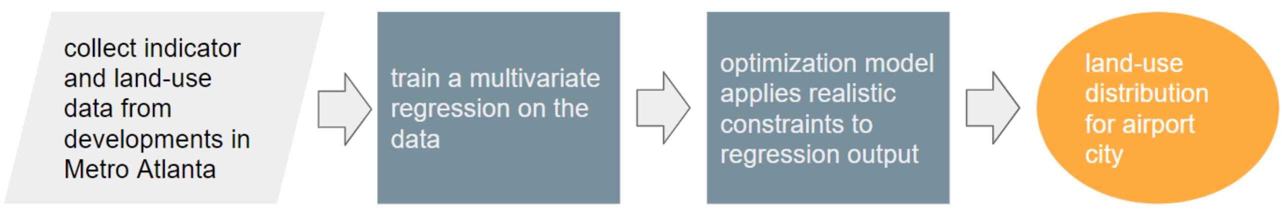


Figure 1: Methodology Map for Airport City Land Use Determination

Data

The data collected took the form of finding data for each independent and dependent variable of each development in the dataset. The data included 70 different developments within the Atlanta Metro. The developments were chosen based on local knowledge of the region. The hope was that these developments would be the most popular ones in the Atlanta region and that they would have diversity. The developments categorized by type are shown in Table 1. The Suburban development type was purposefully chosen to be the most frequent as the airport is in a suburban area. The urban, tourist, and university data points will inform the model to account for the non-local population of the airport as many transplants live and visit the urban, tourist, and university areas of Atlanta. A map of where the development data points are by type is displayed in Figure 2.

Table 1: Development Data Point Type Breakdown

Dev. Type	Dev. Count	Percent of Data
Suburban	42	60%
Urban	16	23%
Tourist	8	11%
Univ.	4	6%

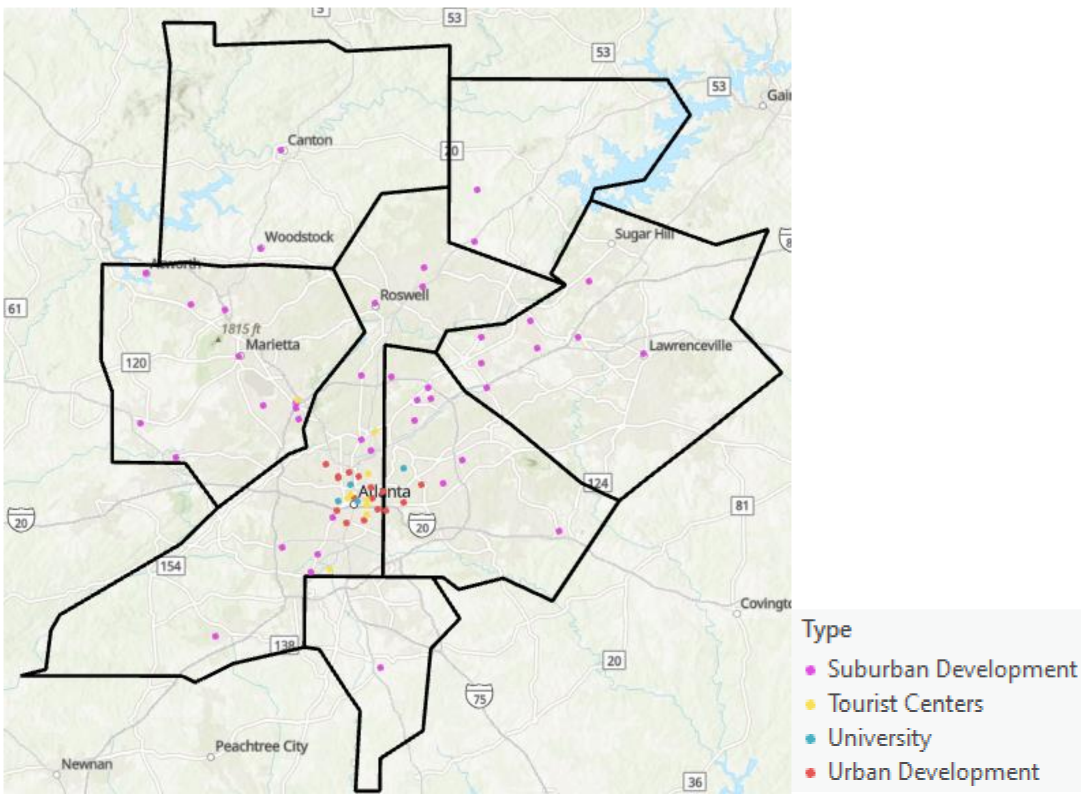


Figure 2: Development Data Points to Train the Model on by Type (U.S. Census Bureau, 2021)

For each of the developments, 14 pieces of information must be collected. These pieces of information are the variables for the model and are shown in Table 2. If the data collection could be automated in the future, two reasons would make it difficult. The first being that some learning model or human must decide which developments to include in the dataset. The second being that the land use percentages of each development do not come from uniform data source. Each city has its own land use and/or zoning map with different codes. More than 30 different land use data sources were accessed to create the dataset for the dependent variables in this project. The 7 land use categories chosen here as the dependent variables are based on the most popular and wide coverage categories in the Atlanta Metro. The areas represented by the collected data are all census tracts within a 1-mile radius of a development center.

Table 2: Variables in Model

Variable Name	Variable Type	Data Source
Latitude	Independent	Google Maps (Google Maps, 2023)
Traffic Congestion	Independent	Google Maps (Google Maps, 2023)
Highway Nearness	Independent	US Census TIGERLine (U.S. Census Bureau, 2021)
Median Age	Independent	US Census ACS 5-Year (U.S. Census Bureau, 2021)
Total Population	Independent	US Census ACS 5-Year (U.S. Census Bureau, 2021)
White Population Percentage	Independent	US Census ACS 5-Year (U.S. Census Bureau, 2021)
Area Median Income	Independent	US Census ACS 5-Year (U.S. Census Bureau, 2021)
Open Space LU %	Dependent	City Land Use/Zoning Map (City of, n.d.)
Low Density Residential LU %	Dependent	City Land Use/Zoning Map (City of, n.d.)
Medium High Density Residential LU %	Dependent	City Land Use/Zoning Map (City of, n.d.)
Commercial LU %	Dependent	City Land Use/Zoning Map (City of, n.d.)
Industrial LU %	Dependent	City Land Use/Zoning Map (City of, n.d.)
Mixed-Use LU %	Dependent	City Land Use/Zoning Map (City of, n.d.)
Institutional LU %	Dependent	City Land Use/Zoning Map (City of, n.d.)

The input data for the site merged data from the local area of the airport. It was assumed that the airport passenger's data reflected trends in airport data for the whole industry. This assumption is made more valid by the fact that Hartsfield Jackson has the highest passenger traffic of any airport in the world. Table 3 displays the data. The maps of the data are included in the appendix.

Table 3: Data Point Difference Between Airport Passenger Traffic Inclusion and Local Residents (Airlines for America, 2023) (Statista, 2015) (United States Department of Transportation, 2023) (Atlanta Airport, 2023)

Data Coverage	Median Age(Years)	Total Population	White Population (%)	Area Median Income(USD)
Airport + Local	39.2375	234167	40.73625	65158
Local Only	36.475	12257	29.4725	56276

Multivariate Multiple Regression Model

While the Multivariate Multiple Regression Model is referred to as one model in the paper often, it is actually seven different multivariate regression models. Each model has one output. The output of each model is the predicted ideal land use of a development based on the development's indicators. The model was built and trained in Python using linear regression in the sklearn package (Pedregosa et al., n.d.) and resulted in Equation Set 1. Table 4 displays the evaluation of each model. The adjusted R squared metric shows how much of the result of the regression model is explained by all the indicator variables in the regression. An adjusted R squared of around 0.3 is very good for socio-temporal-space variables. Generally, the trained models are just fine. Another evaluation metric used was cross validation. In machine learning, cross validation tells us the accuracy of the model for a new next data point. A seven-fold cross validation was performed on this model. It seems that the model accuracy hovers just below 50%. This is average performance for this problem space.

$$y_{open\ space} = -0.144x_{latitude} + 0.023x_{traffic} + 0.007x_{highway} - 0.0065x_{age} + 0.0037x_{population} + 0.0037x_{white} - 0.0027x_{income} + c$$

$$y_{low\ res} = -0.071x_{latitude} - 0.00013x_{traffic} + 0.00046x_{highway} - 0.001x_{age} - 0.0007x_{population} + 0.00048x_{white} - 0.000036x_{income} + c$$

$$y_{med\ high\ res} = -0.097x_{latitude} + 0.00039x_{traffic} - 0.001x_{highway} + 0.0023x_{age} + 0.00063x_{population} + 0.0021x_{white} - 0.00079x_{income} + c$$

$$y_{commercial} = -1.05x_{latitude} - 0.056x_{traffic} - 0.0058x_{highway} - 0.0023x_{age} - 0.0035x_{population} - 0.00082x_{white} - 0.0022x_{income} + c$$

$$y_{industrial} = -0.22x_{latitude} + 0.01x_{traffic} - 0.025x_{highway} - 0.013x_{age} - 0.0035x_{population} - 0.00082x_{white} + 0.0022x_{income} + c$$

$$y_{mixed\ use} = 1.65x_{latitude} + 0.02x_{traffic} + 0.014x_{highway} + 0.011x_{age} + 0.007x_{population} - 0.0062x_{white} + 0.0024x_{income} + c$$

$$y_{institutional} = -0.036x_{latitude} + 0.0035x_{traffic} + 0.01x_{highway} - 0.005x_{age} + 0.0014x_{population} + 0.00052x_{white} - 0.000059x_{income} + c$$

Equation Set 1: Multivariate Linear Regression Models for Each Land Use

Table 4: Evaluation Metrics of Linear Regression Models

Regression Model	**Adjusted R Squared(0-1)**	**Cross-Validation Accuracy(%)**
Open Space	0.11	72
Low Density Residential	0.07	10
Medium High Density Residential	0.12	20
Commercial	0.14	46
Industrial	0.13	80
Mixed-Use	0.24	13
Institutional	0.10	31

With the regression models being trained and validated, the Airport City specific information can be entered into the model to produce a suggested land use for the new development. Table 5 displays the input data for Airport City and Table 6 shows the model's output. The model output suggests values that are somewhat sensible, but also some do not make any sense. Some values such as Medium High Residential being 135% makes no sense, as it is over 100%. Thus, an additional model is required to interpret these results into a usable state.

Table 5: Airport City Specific Site Information Model Input

Independent Variable	**Airport City Values**
Latitude	33.658975
Traffic LOS (ratio)	0.32
Highway Proximity (mi)	0.274115158
Median Age (years)	39.2375
Total Population	234167
White Population Percentage	40.73625
Area Median Income (USD)	65158

Table 6: Airport City Site Specific Regression Model Output for Land Use

Dependent Variable/Model Output	**Airport City Values (%)**
Open Space LU %	94
Low Density Residential LU %	19
Medium High Density Residential LU %	135
Commercial LU %	70
Industrial LU %	28
Mixed-Use LU %	160
Institutional LU %	13

Optimization Model

The optimization model was developed in python using the PuLP package (Department of Engineering Science The University of Auckland September, 2011). The PuLP package only performs linear optimization, so that is also a limitation that only allows the functions and constraints for the model to be linear. Equation 1 shows the basic objective function to minimize. The goal of this objective function is to minimize the difference between the ideal values of land use that were found in Table 6 and the realistic design values we want to find in qi.

$$\min_{q_i} \sum_i |p_i - q_i|$$

i= one of seven land use categories in this project (such as industrial or open space)

q_i= the percentage of a development that should be the land use of category i

p_i= the percentage of a development that is ideally of land use category i(taken from Table 6)

Equation 1: Objective Function that attempts to get designed land use close to ideal land use

However, a large barrier that determines realistic land use is often political or administrative. A penalty term is added to our objective function to account for this. Equation 2 shows this penalty term and Equation 3 shows an alterative objective function combining Equations 1 and 2. The idea of the penalty term is to say that if much of the land use is changing in airport city then it is less likely to be a practical design because it would require much rezoning, political resistance, and more stakeholders.

$$a \sum_i |c_i - q_i|$$

c_i= percentage of development that exists on Airport City plot of land use category i

a= penalty term weight

Equation 2: Penalty term that attempts to get designed land use close to current land use

$$\min_{q_i} \sum_i |p_i - q_i| + a \sum_i |c_i - q_i|$$

Equation 3: Objective function with the penalty term leads to a more realistic design projection

Lastly, the realistic contraints are applied to the model in Equation 4 and 5. Equation 4 states that all land uses must take a meaningful percentage of the land in the airport city develop. This means that all land use percentages should add up to 100% unlike the results from the regression model in Table 6. Additionally, all designed land use percentages must be non-negative as that also makes no logical sense.

$$\sum_i q_i = 1$$

q_i= the percentage of a development that should be the land use of category i

Equation 4: This is the first constraint that says all land use percentages summed must be 100%

$$q_i \geq 0$$

q_i= the percentage of a development that should be the land use of category i

Equation 5: Second constraint that says all land use percentages must be non-negative

Modeling Results

Figure 3 below the model output from the regression model visualized. The existing land use is on the right-hand side of the chart. The no constraints output shows the optimization model output without a penalty term where the penalty bar shows the output with the penalty term. The no constraints model output suggests slightly lowering the mixed-use percentage, increasing the medium high density residential percentage and adding open space. The intuitive interpretability of this result is difficult between the regression model is a seven-dimensional space to think in. In addition, there is also a optimization model applying high dimensional calculus on that seven dimensional space to arrive at this result. However, some interpretability can be gleaned from the penalty bar result. The mixed use and low residential proportions are kept the same as existing, but the model turns the commercial to high residential as well. Here, we can see the model applying smaller changes, but keeping large difficult to change portions of the existing land use. The no institutional bar shows what the model thinks the zoning should be if no anchor institution like a large company or university exists in the development. The no industrial bar shows the suggested land use proportions when no industrial land use is permitted in airport city. These two scenarios occur quite often in developments, so they were included here. Finally, the regression scaling bar shows what the regression model outputted, but just scaled down to fit 100% land use limit rather than passing the results through an optimization model to apply realistic constraints on the design.

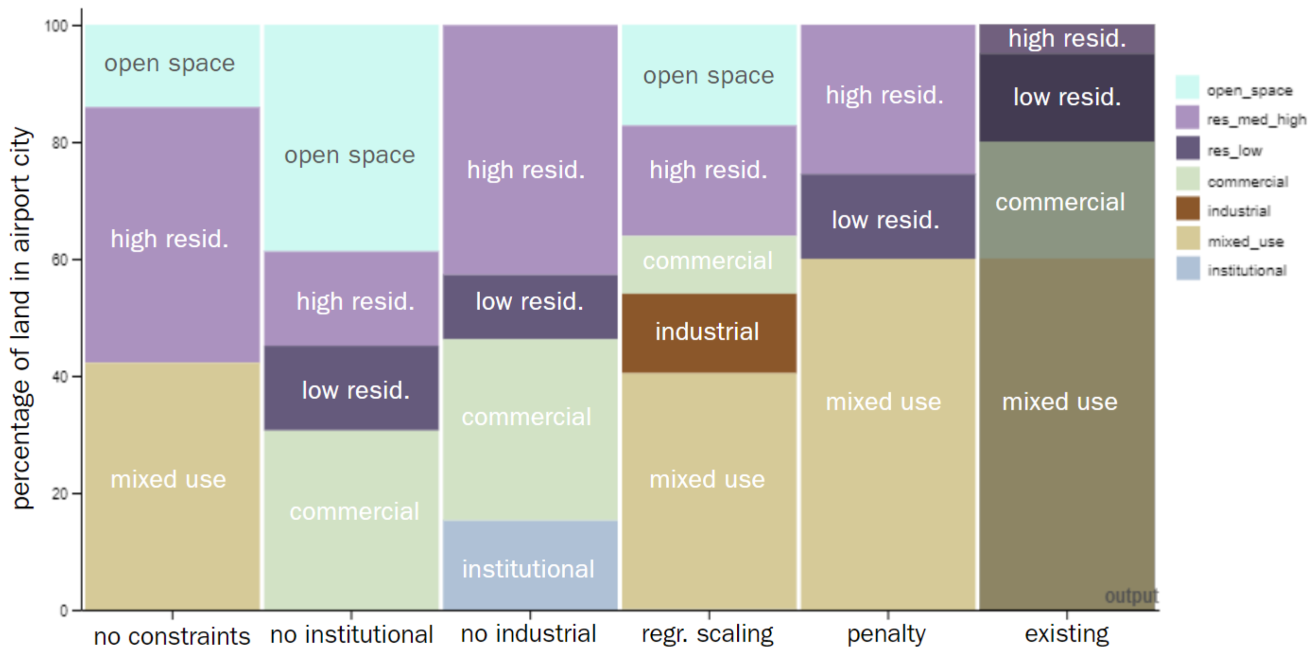


Figure 3: Suggested Land Use Proportions from the Machine Learning Model

While this model can only determine land use proportions for the site area of a development, at minimum this would assist development designers and developers to have guidelines to operate by to make their job easier and more successful. Examples are shown below of how urban designers could choose to use the proportions for Airport City. Figure 4 shows an example of the no constraints model suggestion where is suggests turning the commercial site into a park that connects to the high density residential and mixed use development. Figure 5 is the existing land use layout. Figure 6 is the penalty model suggestion where the commercial and some low density residential is turned into medium high density residential.

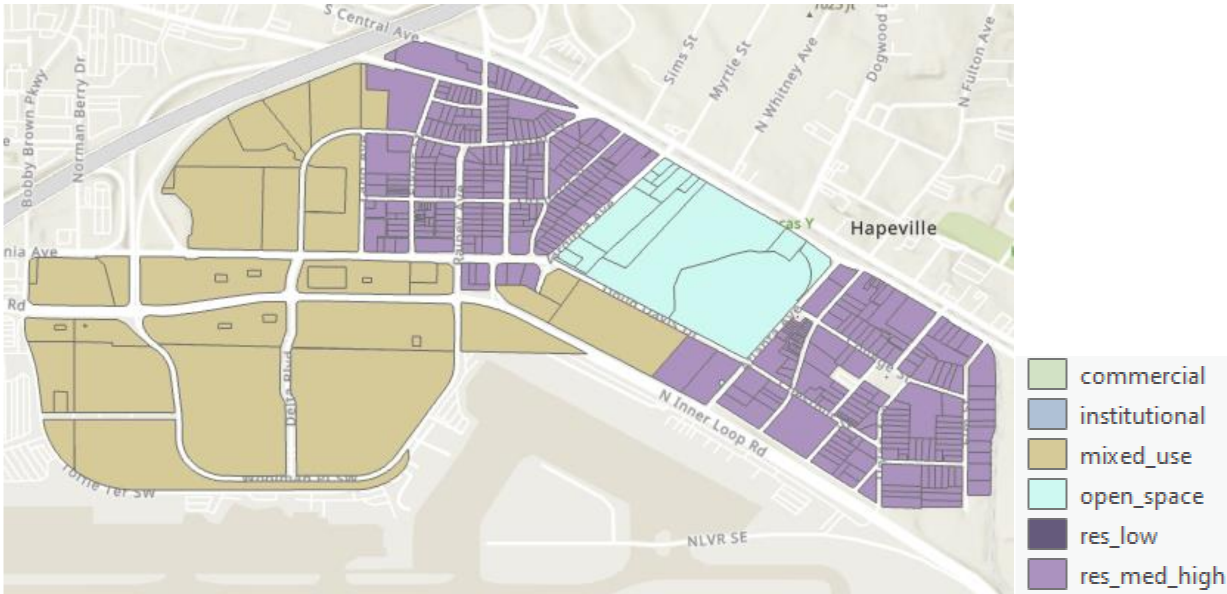


Figure 4: No Contraints Model Output Example Land Use Design (City of Hapeville, 2022)

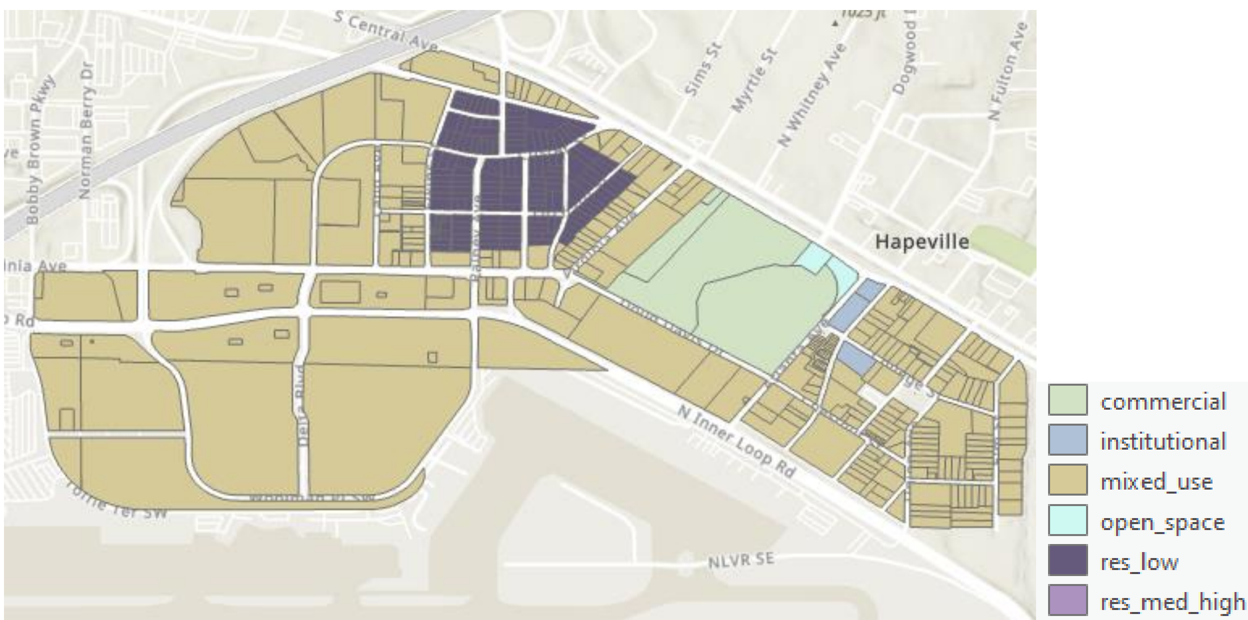


Figure 5: Existing Land Use Design of Airport City (City of Hapeville, 2022)

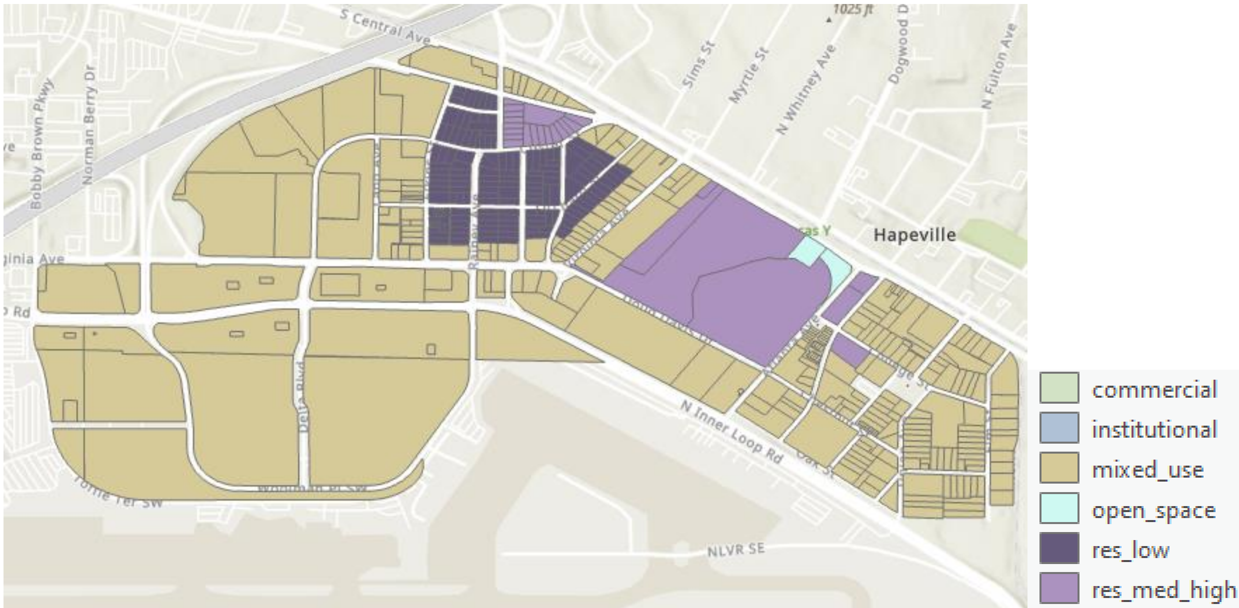


Figure 6: Penalty Term Model Output Example Land Use Design (City of Hapeville, 2022)

Future Work

The outcomes of this project were exciting as this outcome was a proof of concept for a machine learning based urban design methodology that has not been thoroughly explored yet in research. This current process gives helpful guidelines to urban designers and developers on what a successful development for the locale may look like. It would be interesting to interface this development orientated scope with the parcel resolution land use methodologies to determine what parcel within a development should be a specific land use. Methodology changes that could account for more hyper local data from the cities by the airport would increase accuracy. Methodology changes to optimize around carbon emissions, development contraints around the airport region and other such ideal factors would be helpful to explore as well. Lastly, a step further than suggesting land uses for a development that would be successful is seeing what would happen to the locale if the suggested land use was adopted. This would be an exciting area to explore, but the methodology would look vastly different and even more detailed data would have to be collected.

References

Airlines for America. (2023). *Air Travelers in America Key Findings of a Survey Conducted by Ipsos*. Airlines for America. https://www.airlines.org/wp-content/uploads/2023/03/A4A-Air-Travel-Survey-2023-Key-Findings.pdf

Atlanta Airport. (2023, January). *STATISTICS FOR ATLANTA AIRPORT*. Atlanta Airport. https://atlairport.net/statistics/

City of Acworth. (n.d.). *Official Zoning Map*. City of Acworth. Retrieved April 13, 2023, from https://cms5.revize.com/revize/acworthga/document_center/Applications%20-%20Permits%20-%20Ordinances/Zoning%20Map.pdf

City of Alpharetta. (n.d.). *AlphaGIS City of Alpharetta GIS Mapping Application*. City of Alpharetta. Retrieved April 13, 2023, from https://alphagis.alpharetta.ga.us/alphagis/

City of Atlanta. (n.d.). *Maps and GIS*. City of Atlanta. Retrieved April 13, 2023, from https://www.atlantaga.gov/government/departments/city-planning/maps-and-gis

City of Austell. (n.d.). *Official Zoning Map*. City of Austell. Retrieved April 13, 2023, from https://www.austellga.gov/Assets/Files/Government/Forms/AustellZoningMap2016Adopted.pdf

City of Avondale Estates. (n.d.). *Official Zoning Map*. City of Avondale Estates. Retrieved April 13, 2023, from https://www.avondaleestates.org/DocumentCenter/View/2186/Zoning-Map---Updated-July-28-2021

City of Canton. (n.d.). *Canton GIS Open Data*. City of Canton. Retrieved April 13, 2023, from https://map-cantonga.opendata.arcgis.com/

City of Chamblee. (n.d.). *Official Zoning Map*. City of Chamblee. Retrieved April 13, 2023, from https://www.chambleega.com/DocumentCenter/View/3239/Official-Zoning-Map-01182022?bidId=

City of Clarkston. (n.d.). *COMMUNITY VIEW MAPPING*. City of Clarkston. Retrieved April 13, 2023, from https://www.clarkstonga.gov/community-view-mapping

City of College Park. (n.d.). *Zoning*. City of College Park. Retrieved April 13, 2023, from https://www.collegeparkga.com/departments/planning_growth___management/zoning

City of Cumming. (n.d.). *Zoning Map*. City of Cumming. Retrieved April 13, 2023, from http://www.cityofcumming.net/zoning-map

City of Decatur. (n.d.). *City Maps and GIS*. City of Decatur. Retrieved April 13, 2023, from https://www.decaturga.com/planning/page/city-maps-gis

City of Doraville. (n.d.). *Maps*. City of Doraville. Retrieved April 13, 2023, from https://www.doravillega.us/departments/community_development_department/mappdfs.php

City of Duluth. (n.d.). *Maps and GIS*. City of Duluth. Retrieved April 13, 2023, from https://www.duluthga.net/services/planning___development/maps___gis.php

City of Dunwoody. (n.d.). *Open Data Dunwoody*. City of Dunwoody. Retrieved April 13, 2023, from https://get-dunwoody.opendata.arcgis.com/

City of East Point. (n.d.). *Planning and Community Development*. City of East Point. Retrieved April 13, 2023, from https://www.eastpointcity.org/planning-community-development/

City of Fairburn. (n.d.). *Official Zoning Map*. City of Fairburn. Retrieved April 13, 2023, from https://www.fairburn.com/sites/default/files/uploads/departments/planningandzoning/zoning_map_pdf.pdf

City of Hapeville. (2022). *City of Hapeville Official Zoning Map* [Map]. City of Hapeville. https://www.hapeville.org/DocumentCenter/View/4870/Zoning-Map-02-01-2022

City of Jonesboro. (n.d.). *Official Zoning Map*. City of Jonesboro. Retrieved April 13, 2023, from https://www.jonesboroga.com/Assets/Files/Zoning/JonesboroZoningMap%20January%2013%20 2020.pdf

City of Kennesaw. (n.d.). *Planning and Zoning*. City of Kennesaw. Retrieved April 13, 2023, from https://www.kennesaw-ga.gov/planning/

City of Lawrenceville. (n.d.). *Official Zoning Map*. City of Lawrenceville. Retrieved April 13, 2023, from https://www.lawrencevillega.org/DocumentCenter/View/3164/Official-Lawrenceville-Zoning-Map-Adopted-May-20-2020-PDF

City of Lithonia. (n.d.). *Official Zoning Map*. City of Lithonia. Retrieved April 13, 2023, from City of Jonesboro

City of Marietta. (n.d.). *Zoning*. City of Marietta. Retrieved April 13, 2023, from https://www.mariettaga.gov/815/Zoning

City of Norcross. (n.d.). *Official Zoning Map*. City of Norcross. Retrieved April 13, 2023, from https://www.norcrossga.net/DocumentCenter/View/3559/Norcross-UDO-Official-Zoning-Map-2020-12-07?bidId=

City of Peachtree Corners. (n.d.). *Maps*. City of Peachtree Corners. Retrieved April 13, 2023, from https://peachtreecornersga.gov/178/Maps

City of Powder Springs. (n.d.). *Official Zoning map*. City of Powder Springs. Retrieved April 13, 2023, from https://www.cityofpowdersprings.org/DocumentCenter/View/1001/Zoning-Map-PDF?bidId=

City of Roswell. (n.d.). *Planning & Zoning Development Map*. City of Roswell. Retrieved April 13, 2023, from https://www.roswellgov.com/government/departments/community-development/planning-zoning/planning-zoning-development-map

City of Sandy Springs. (n.d.). *Development Code and Zoning Maps*. City of Sandy Springs. Retrieved April 13, 2023, from https://www.sandyspringsga.gov/development-code-and-zoning-maps

City of Smyrna. (n.d.). *Mapping*. City of Smyrna. Retrieved April 13, 2023, from https://www.smyrnaga.gov/departments/mapping

City of Suwanee. (n.d.). *Maps*. City of Suwanee. Retrieved April 13, 2023, from https://www.suwanee.com/explore-suwanee/maps

City of Woodstock. (n.d.). *Zoning Map*. City of Woodstock. Retrieved April 13, 2023, from https://files4.1.revize.com/woodstockga/Document%20Center/GIS%20Services/Maps/Zoning_20200225_18x24_202003031252127670.pdf

Cobb County Government. (n.d.). *Zoning Maps*. Cobb County Government. Retrieved April 13, 2023, from https://www.cobbcounty.org/community-development/zoning/maps

Department of Engineering Science The University of Auckland September. (2011). *PuLP: A Linear Programming Toolkit for Python*. https://optimization-online.org/2011/09/3178/

Forsyth County. (n.d.). *Official Zoning Map of Forsyth County*. Forsyth County. Retrieved April 13, 2023, from https://www.forsythco.com/Portals/0/Documents/GIS/OpenData/SubFolder/ZONING_34x44_frames.pdf

Fulton County. (n.d.). *Property Map Viewer*. Fulton County. Retrieved April 13, 2023, from https://gis.fultoncountyga.gov/Apps/PropertyMapViewer/

Google Maps. (2023). *Google Map data* [Map]. Google. https://www.google.com/maps/@33.7695028,-84.385734,15z

Gwinnett County. (n.d.). *Land Use/Development/Zoning*. Gwinnett County. Retrieved April 13, 2023, from https://www.gwinnettcounty.com/web/gwinnett/services/homeownersassociations/landusedevelopmentzoning

Pedregosa et al. (n.d.). *Scikit-learn: Machine Learning in Python* (JMLR 12). Retrieved April 13, 2023, from https://scikit-learn.org/stable/about.html

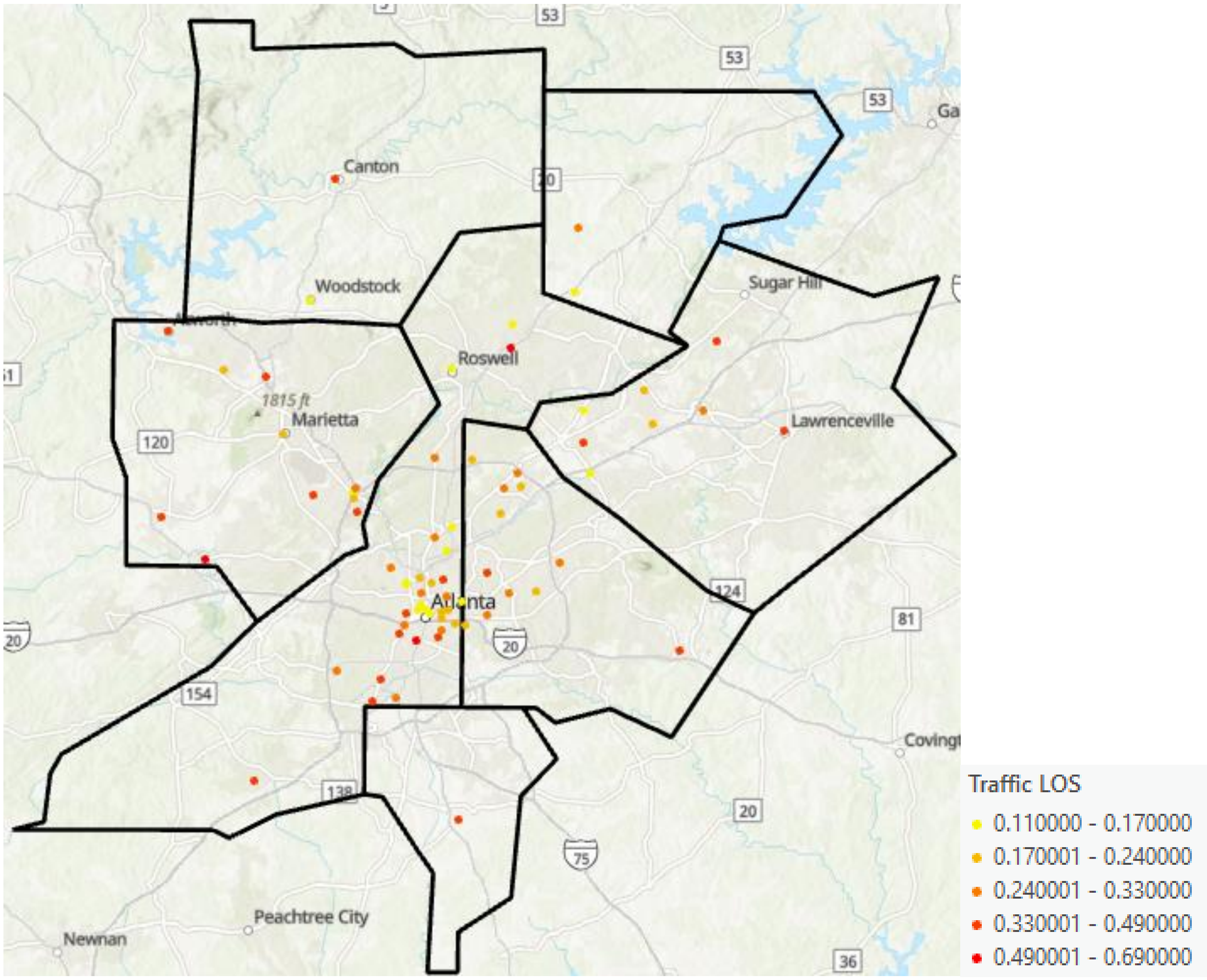
Statista. (2015, June). *Air travel frequency in the United States as of June 2015, by income*. Statista. https://www.statista.com/statistics/316376/air-travel-frequency-us-by-income/

United States Department of Transportation. (2023, April 11). *Bureau of Transportation Statistics*. United States Department of Transportation. https://www.transtats.bts.gov/airports.asp?pn=1&Airport=ATL&carrier=FACTS

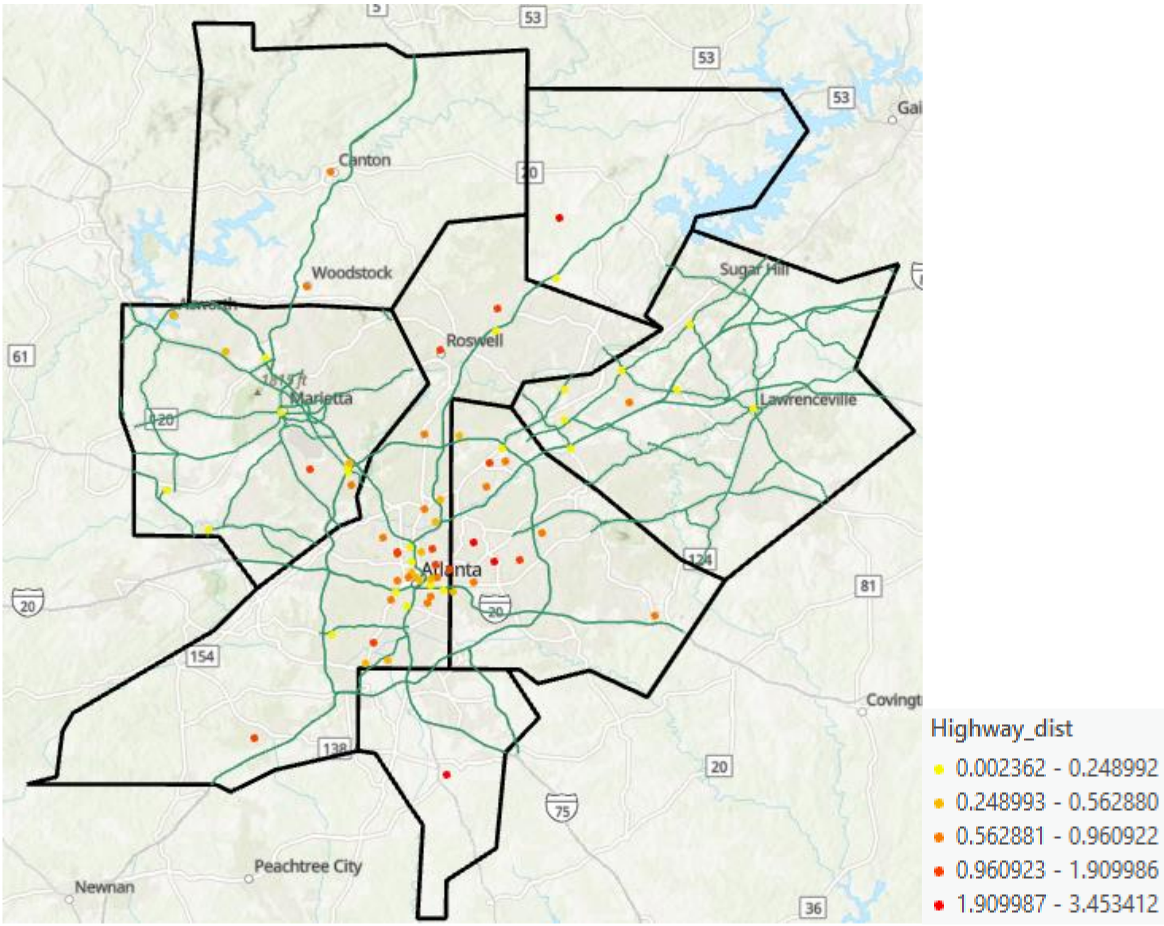
U.S. Census Bureau. (2021a). *ACS 5-Year Suvery 2021 by Census Tract*. U.S. Department of Commerce. data.census.gov

U.S. Census Bureau. (2021b). *TIGERLine Highway*. U.S. Department of Commerce. data.census.gov

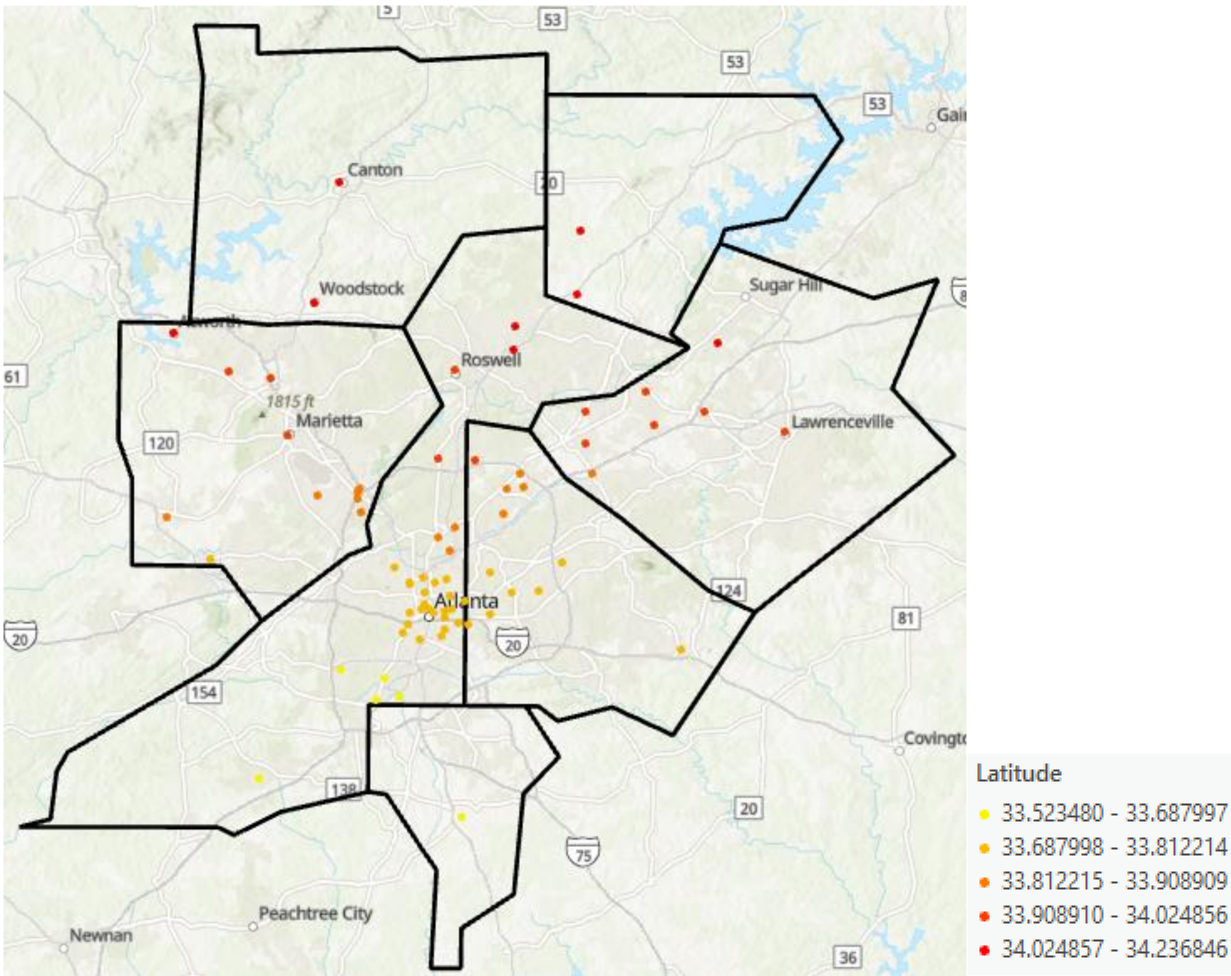
Appendix



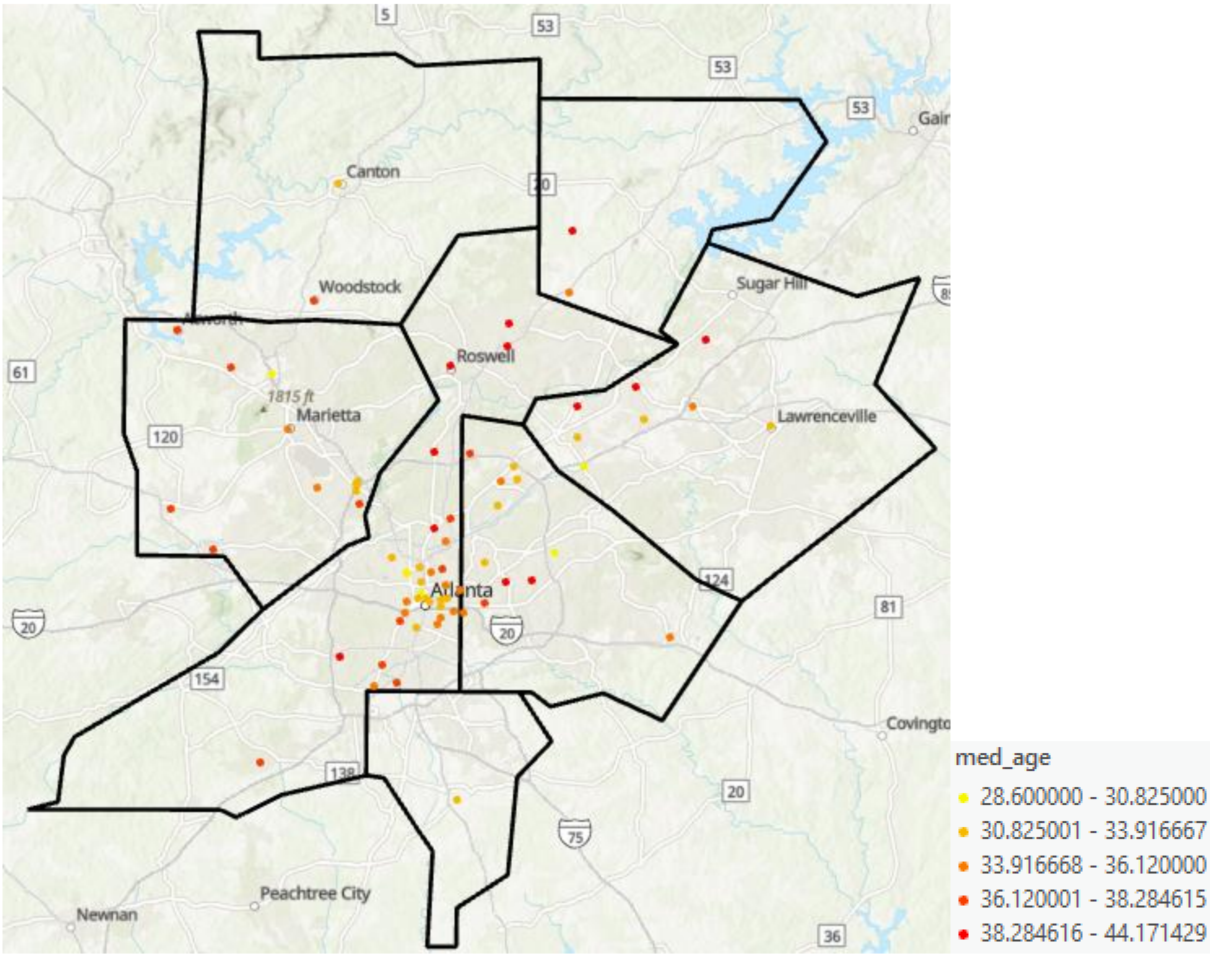
Appendix 1: Traffic LOS for Dataset (U.S. Census Bureau, 2021)



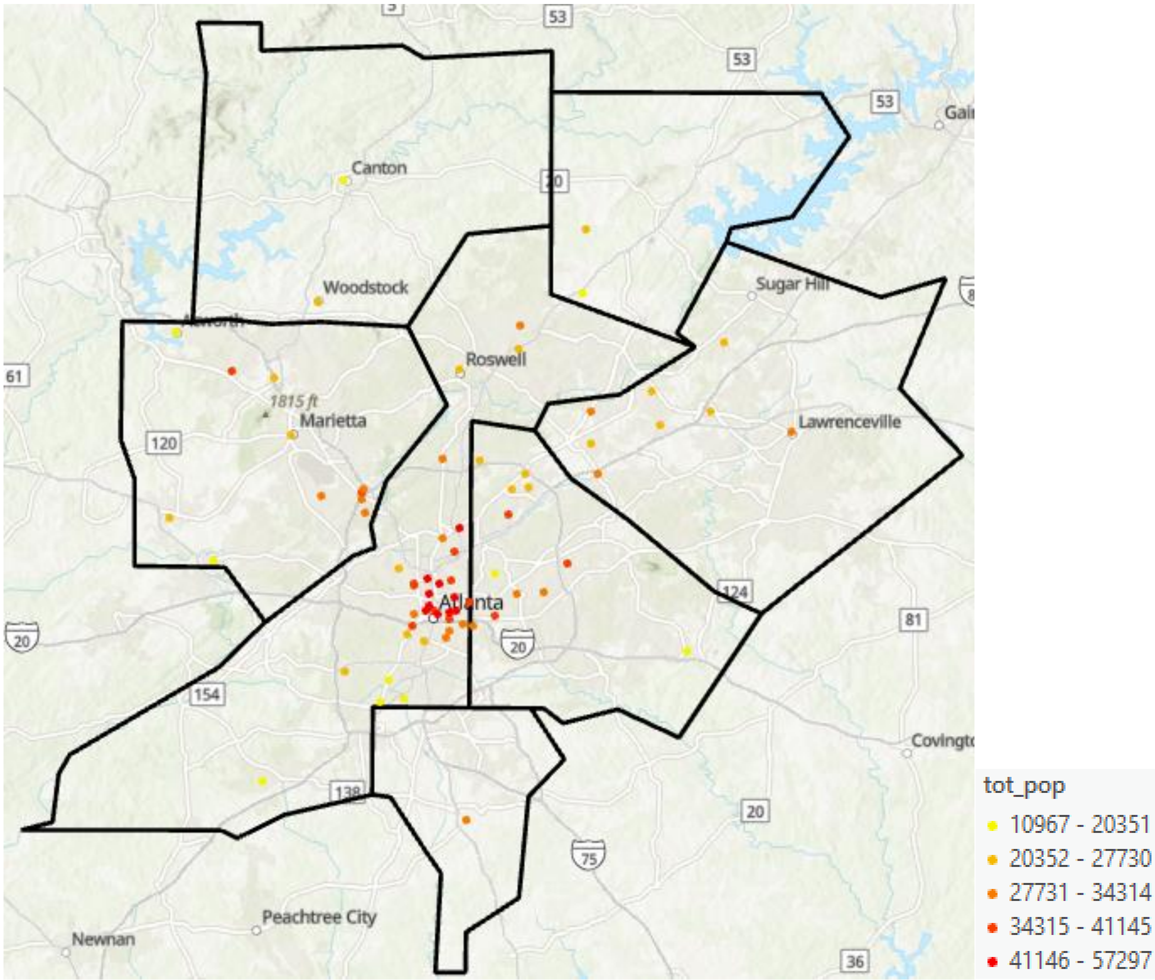
Appendix 2: Highway Distance for Dataset (U.S. Census Bureau, 2021)



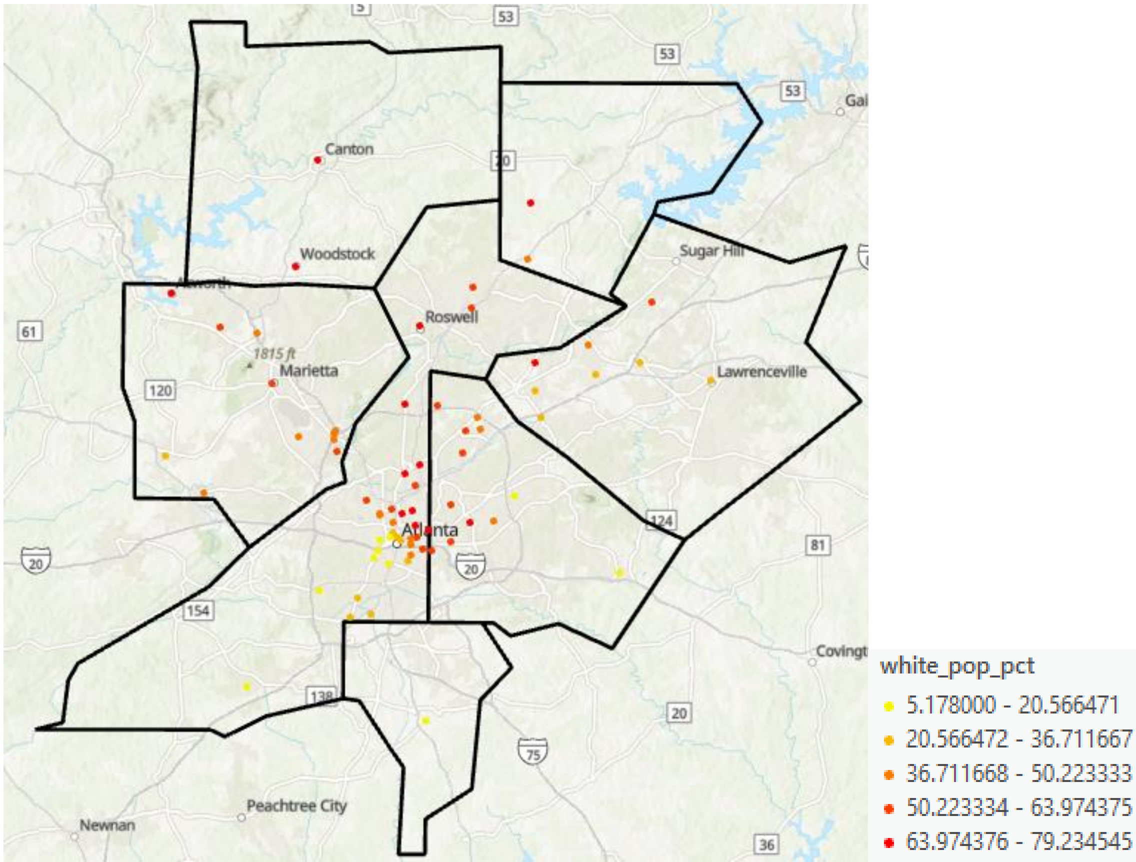
Appendix 3: Latitude for Dataset (U.S. Census Bureau, 2021)



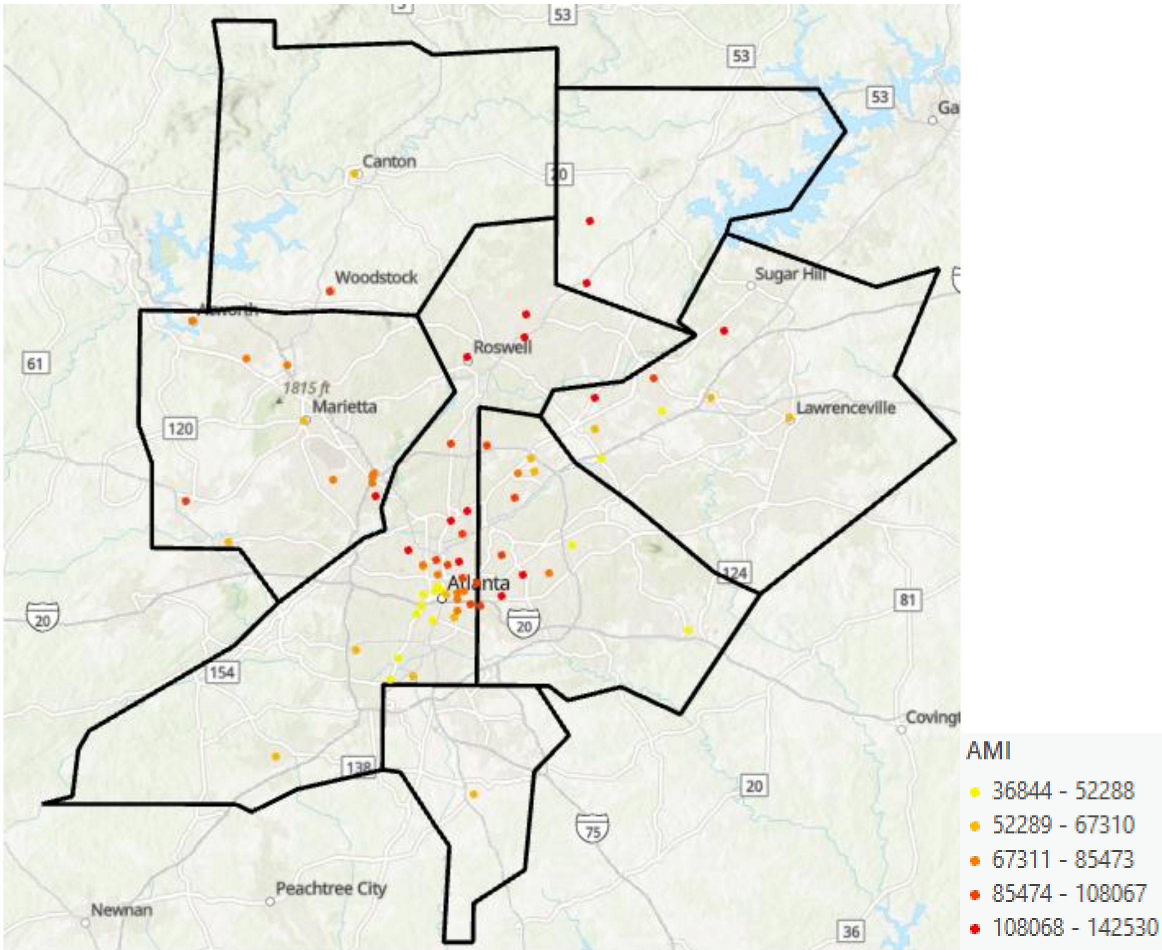
Appendix 4: Median Age for Dataset (U.S. Census Bureau, 2021)



Appendix 5: Total Population for Dataset (U.S. Census Bureau, 2021)



<u>Appendix 6: White Population Percentage for Dataset (U.S. Census Bureau, 2021)</u>



Appendix 7: Area Median Income for Dataset (U.S. Census Bureau, 2021)