The Urban Heat Island Effect in Phoenix, Arizona – A Machine Learning Approach

Charlie Sturr[1]

Columbia University – Masters Candidate in Data Science

December, 2021

Abstract

The Urban Heat Island effect has been studied and analyzed for well over two decades [1]. This phenomenon is classified as occurring: "when cities replace natural land cover with dense concentrations of pavement, buildings, and other surfaces that absorb and retain heat. This effect increases energy costs (e.g., for air conditioning), air pollution levels, and heat-related illness and mortality" [2]. Traditionally, the UHI effect has been analyzed using classification analysis or simple regressions, which creates an opportunity to apply machine learning techniques to better classify and understand the phenomenon. This paper explores 2 machine learning techniques from regression trees: Random Forests and XGBoost. Factors (independent variables) are considered geographically and demographically, as the UHI is a uniquely urban phenomenon, and therefore has implications for future development of cities.

[1] hcs2150@columbia.edu

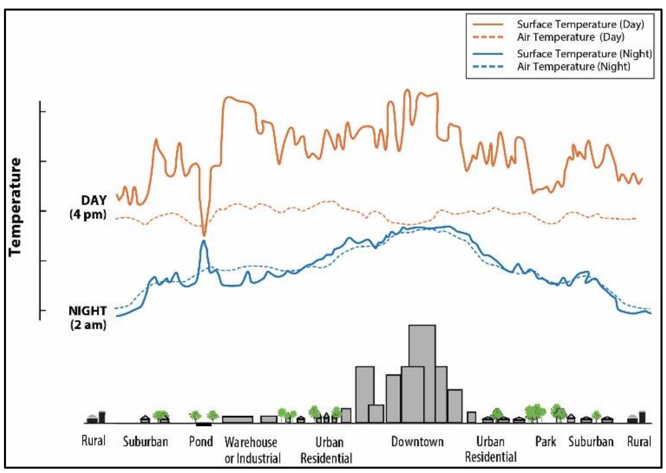
1. Introduction

Phoenix was recently named the fastest growing city of 2021 [3], while also setting records for the hottest recorded months ever observed [4]. As the city continues to grow, rising temperatures will have significant implications for future development, power demand, population flow, agriculture, and more. Furthermore, Phoenix has experienced significant variability in its annual summer monsoon season [5], which typically generates over half of the city's rainfall each year [6].

Unlike other major metropolitan areas in the United States, Phoenix is uniquely sparse in terms of population density. Despite ranking 5th in overall population, it ranks 378th in population density (for cities with greater than 50,000 residents) [7]. This, combined with Phoenix's severe heatwaves and 85% sunshine per year, creates a unique geography to study a phenomenon like the UHI.

This paper aims to better classify, and understand the main drivers of the Urban Heat Island effect in Phoenix using a machine learning approach. Using both a Random Forest, and XGBoost regression, the analysis will help drive predictive ability and factor importance for the UHI phenomenon. The UHI is categorized by increased surface temperatures in more dense, urban areas, compared to outlying suburban / rural regions.

Figure 1: The Urban Heat Island Effect



The study of the Urban Heat Island effect is especially important today, as the severity of increasing temperatures in urban centers across the United States has increased since the 1990s [1]. As temperatures in cities increases, the externalities related to the UHI, including increased mortality rates and health complications [8], will continue to rise. As such, it is crucial to better understand the main drivers of the UHI effect so cities can better address these causes in future development. Because the Urban Heat Island effect has implications for future urban planning and zoning, the data collected for this analysis incorporates both geographic features (such as Land Surface Temperature, Normalized Difference Vegetation Index, Percent Impervious Land Cover), and socioeconomic / demographic factors (Population Density, Median Household Income, Percent Rental Occupancy), to try and establish a relationship between both the built environment, and population dynamics contained within. As proposed by Huang et al, in many urban centers the impact of the UHI is more disproportionately felt by denser, poorer areas [9], so including these socioeconomic factors is a key step to better understanding not only the drivers of the UHI effect, but also who is most impacted by it.

2. Study Area

Maricopa County is the largest county in Arizona, the 4th largest county in the United States (by population). For this paper, analysis was conducted across all 2,505 census block groups within Maricopa County. Each CBG has a unique identifier, and different characteristics collected from the US Census.

Figure 2: Maricopa County in Arizona

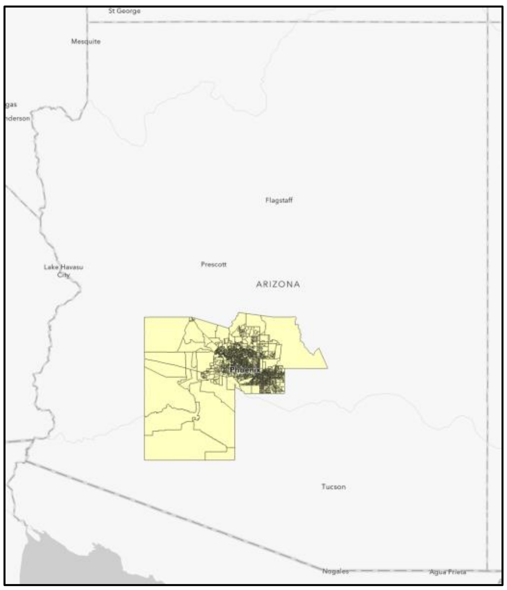
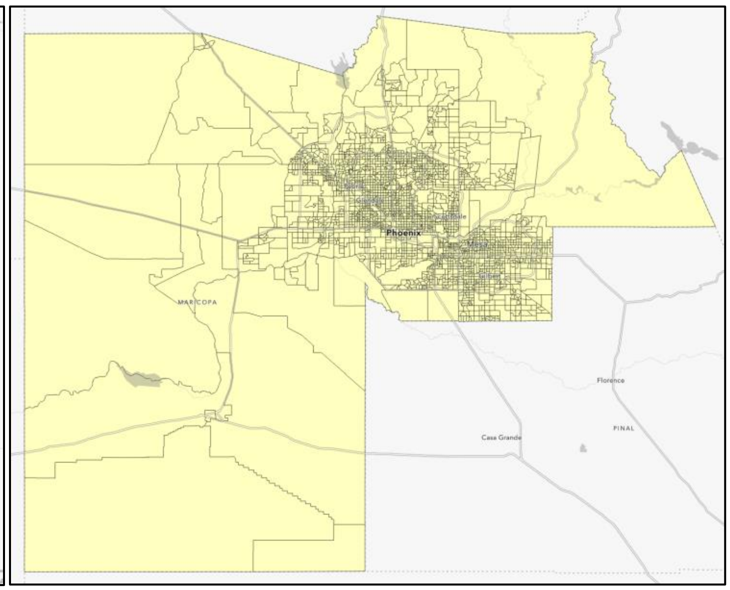


Figure 3: Maricopa County Census Block Groups



3. Data

The data collected was collected in 2016 due to consistency of inputs. One of the most important factors in understanding the UHI effect is the built, urban environment, and the most comprehensive collection of data that spans buildings, impervious / land use cover, and satellite data comes from 2014-2016. Data collection is split between physical / geographic inputs, and socioeconomic / demographic inputs. Once data was collected, it was aggregated on the CBG level. There are 2,505 census block groups in Maricopa County, and the following data summary represents the data included in this analysis:

3.1 Physical / Geographic Inputs

Variable	Definition	Unit	Measurement	Source
Dependent Variable				
LST	Land Surface Temperature	ºC	Pixel (30m x 30m)	LANDSAT8
Independent Variables				
Dist_center	Distance from CBG centroid to urban center	M	Distance	Google Earth Engine

CBG_Area	Area of the CBG	M^2	Area	Google Earth Engine
NDVI	Normalized Difference Vegetation Index	N/A	Pixel (30m x 30m)	LANDSAT8
Building_Area	Total building footprint in CBG	M^2	LIDAR	ASU / NGA – 2014
Building Height (Avg.)	Average Building Height in CBG	M	LIDAR	ASU / NGA - 2014
Building Height (Max)	Max Height of Buildings in CBG	M	LIDAR	ASU / NGA - 2014
Roof type Max Height	Type of Roof of Max Building	Flat / Complex / Pitched	LIDAR	ASU / NGA - 2014
Roof type Largest Footprint	Type of Roof of Biggest Building	Flat / Complex / Pitched	LIDAR	ASU / NGA - 2014
High Density Land Usage %	Percent of CBG	%	Pixel (30m x 30m)	NLCD 2016
Medium Density Land Usage %	Percent of CBG	%	Pixel (30m x 30m)	NLCD 2016
Low Density Land Usage %	Percent of CBG	%	Pixel (30m x 30m)	NLCD 2016
Percent Shrub Landcover	Percent of CBG	%	Pixel (30m x 30m)	NLCD 2016
Percent Impervious Land	Index 0-100; impervious land cover	%	Pixel (30m x 30m)	NLCD 2016

Table 1

3.2 Socioeconomic / Demographic Inputs

Variable	**Definition**	**Unit**	**Measurement**	**Source**
Independent Variables				
Median HH Income	Median HH income by CBG	$	Per HH	US Census 2016
Population Density	Total # of People per Square Mile	Person / Sq. Mile	Per CBG	US Census 2016
Percent Rental Home	Total Percentage of HHs renting	%	Per CBG	US Census 2016

Table 2

4. Data Collection & Cleaning

Collecting and assembling the data for this project was quite time intensive and took a significant amount of manipulation and pulling. Given the range of sources and the density of some of the files, it required using Python, QGIS, ArcGIS, and Microsoft Excel. Below is an overview of data sources and the required manipulation.

LANDSAT8 2016 Data

In order to aggregate and clean the LANDSAT data, I used the Google Earth Engine Python package. From here, we can calculate NDVI, and LST. Once I pulled the required LANDSAT image collection, I filtered on a given date range, and then selected the least cloudy image. For this study, we compare summer months (~June – September), and winter months (~December – March). Typically, UHI analysis is focused on temperatures between 4-5pm, but for Arizona, because of the satellite's orbit, all LANDSAT images for the Phoenix region are collected at ~5-6pm.

From here, we generate a LANDSAT8 composite layer that is the least cloudy image from within our timeframe. The summer image was collected on 2016-07-12 18:03:56, and the winter image was collected on 2016-12-03 18:04:20. In order to calculate NDVI, I used the built in Google Earth Engine function `normalizedDifferece`() . NDVI is an important feature in understanding the UHI effect due to its ability to predict the abundance of actively photosynthesizing vegetation [10]. From here, we can calculate LST using the following procedure, which was published in the International Journal of Applied Engineering Research in 2017 [11]. Because land surface temperature (LST) uses NDVI as an input, this will create implications in future analysis given the high degree of correlation between LST and NDVI. Because of this, we will adjust our analysis accordingly, to account for outcomes that both include, and exclude NDVI given that LST is our target factor we are trying to predict.

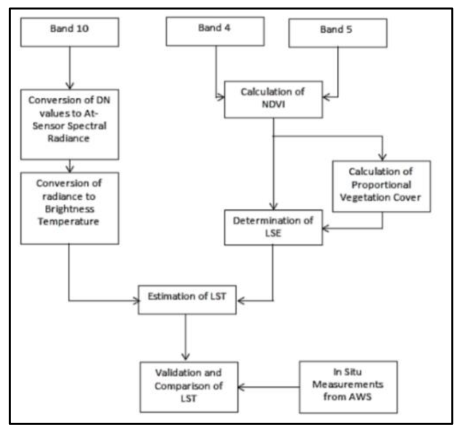


Figure 4: LST Calculations

1) **Top of Atmosphere Radiance (TOA)**: $TOA = M_L * Q_{CAL} + A_L$
   a. M_L = band specific rescaling factor, Q_{CAL} = band 10, A_L = band specific additive factor

2) **TOA to Brightness Temperature**: $BT = \frac{K_2}{(Ln(\frac{K_1}{TOA})+1)} - 273.15$
   a. K_1, K_2 are band specific thermal conversion constants

3) **NDVI**: $\frac{Band5-Band4}{Band5+Band4}$

4) **Proportion of Vegetation**: $P_v = (\frac{NDVI-NDVI_{MIN}}{NDVI_{MAX}-NDVI_{MIN}})^2$

5) **Emissivity**: $\varepsilon = 0.004 * P_v + 0.986$

6) **LST**: $\frac{BT}{(1+\left(0.00115*\frac{BT}{1.4388}\right)*Ln(\varepsilon))}$

ASU / NGA LIDAR 2014 Data

The majority of the time in data collection and cleaning was spent on the ASU / NGA LIDAR 2014 data. The dataset contained over 800,000 entries for buildings in the Phoenix metropolitan area, and therefore merging these required use of ArcGIS. The data set include a range of attributes associated with each building, including max height, average height, footprint, roof type, and more. At the CBG level, I aggregated the total building area (per CBG) to generate building footprint, the max height of a building within a CBG, the average height of buildings within a CBG, and also catalogued the type of roof for the building with the max height, and the largest area.

NLCD 2016 Data

The National Land Cover Data came from the United States Geological Survey. Each year, the USGS classifies surface types based on satellite imagery and produces comprehensive data sets at the pixel level for the entire United States. In order to use in our UHI analysis appropriately, I pulled the following:

- High Density Land Usage (as a percentage of area in the CBG)
- Medium Density Land Usage (as a percentage of area in the CBG)
- Low Density Land Usage (as a percentage of area in the CBG)
- Shrub Cover (as a percentage of area in the CBG)[2]
- Percent Impervious Land (as a percentage of area in the CBG)
    - Impervious land cover is classified as being highly impervious (e.g. a parking structure or asphalt road, or pervious (e.g. a golf course, grass, etc.) and is classified on a scale of 0-100.

5. Methodology

To run this analysis, I used a random forest algorithm, as well as an XGBoost algorithm to see which performed better and which features were more important. The random forest classifier uses improved bagging and boosting [12], and is less sensitive to noise and overtraining [13]. On the other hand, XGBoost, which is also a decision tree based algorithm, relies on extreme gradient boosting (hence the name), and also has widespread applications in regression type analysis because of its speed and flexibility across data types [14].

Furthermore, each model was run four times:

1. Summer Months Including NDVI
2. Summer Months Excluding NDVI
3. Winter Months Including NDVI
4. Winter Months Excluding NDVI

The reason for excluding NDVI is because it is highly correlated with LST in the summer months. When running a factor importance analysis, NDVI was the most important factor by ~60%, so I wanted to see if we could pick apart the data and see what

[2] Note: because the regional vegetation in Phoenix is that of an arid desert climate, shrub cover was used instead of tree canopy cover

other drivers exist. As we can see, the NDVI is highly negatively correlated with LST in the summer months:

Figure 5: Summer Months Correlation

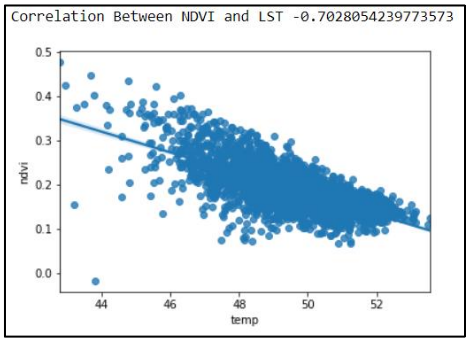
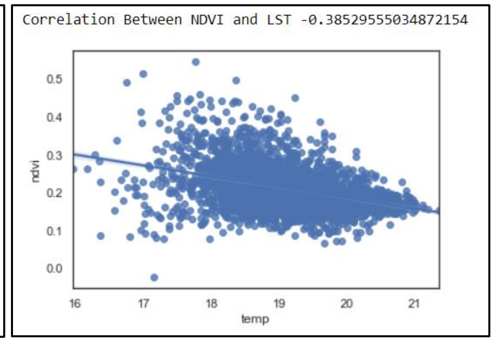


Figure 6: Winter Months Correlation



When we examine the feature importance of an analysis run on default parameters, we can see that NDVI drives a significant portion of our results in the summer months (note these were generated with default hyperparameters set for our random forest model):

Figure 7: Winter Feature Importance; Including and Excluding NDVI

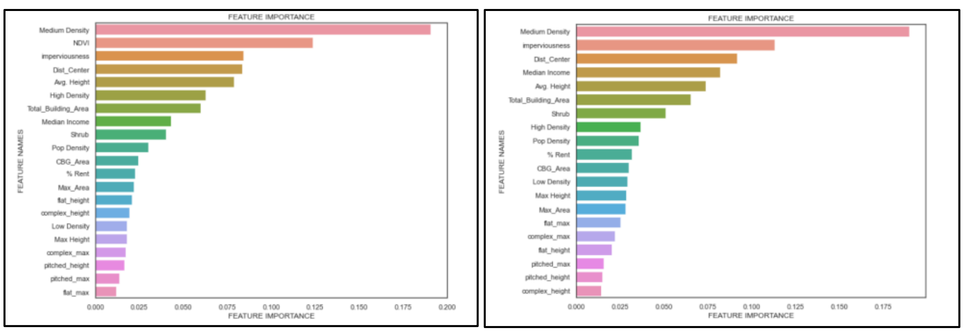
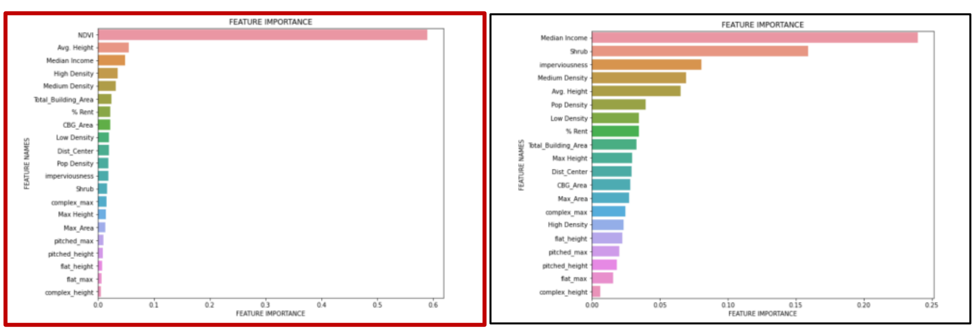


Figure 8: Summer Feature Importance; Including and Excluding NDVI



Therefore, in order to best classify our analysis, the plan is to run 4 random forest models, and 4 XGBoost models – two in winter and two in summer, with NDVI included and NDVI excluded.

6. Tuning Hyperparameters

In order to optimize our hyperparameters, I used Scikitlearn's built in Randomized Search Cross validation to tune our Random Forest:

Summer Excluding NDVI

```
rf_random.best_params_
```

```
{'n_estimators': 1600,
 'min_samples_split': 2,
 'min_samples_leaf': 2,
 'max_features': 'auto',
 'max_depth': 70,
 'bootstrap': True}
```

Summer Including NDVI

```
rf_random.best_params_
```

```
{'n_estimators': 2000,
 'min_samples_split': 2,
 'min_samples_leaf': 2,
 'max_features': 'auto',
 'max_depth': 70,
 'bootstrap': True}
```

Winter Excluding NDVI

```
{'n_estimators': 2000,
 'min_samples_split': 5,
 'min_samples_leaf': 2,
 'max_features': 'auto',
 'max_depth': 110,
 'bootstrap': True}
```

Winter Including NDVI

```
{'n_estimators': 2000,
 'min_samples_split': 5,
 'min_samples_leaf': 2,
 'max_features': 'auto',
 'max_depth': 110,
 'bootstrap': True}
```

For the XGBoost, I used the package in Python XGBTune:

Summer Excluding NDVI

```
{'eval_metric': 'rmsle', 'max_depth': 8, 'min_child_weight': 1, 'gamma': 0.0, 'subsample': 1.0, 'colsample_bytree': 0.55, 'a
lpha': 0, 'lambda': 1, 'seed': 0}
```

Summer Including NDVI

```
{'eval_metric': 'rmsle', 'max_depth': 8, 'min_child_weight': 1, 'gamma': 0.0, 'subsample': 0.95, 'colsample_bytree': 0.95,
'alpha': 0, 'lambda': 1, 'seed': 0}
```

Winter Excluding NDVI

```
{'eval_metric': 'rmsle', 'max_depth': 8, 'min_child_weight': 1, 'gamma': 0.0, 'subsample': 1.0, 'colsample_bytree': 0.9, 'al
pha': 0, 'lambda': 1, 'seed': 42}
```

Winter Including NDVI

```
{'eval_metric': 'rmsle', 'max_depth': 8, 'min_child_weight': 1, 'gamma': 0.0, 'subsample': 1.0, 'colsample_bytree': 1.0, 'al
pha': 0.1, 'lambda': 1.1, 'seed': 0}
```

7. Results

Before we analyze our results from the Random Forest and XGBoost analysis, first, let's take a look at the distribution of temperatures from both seasons to get a better understanding of the range of values we are seeing, and to better interpret the MSE.

Figure 9: Winter Temp Distribution

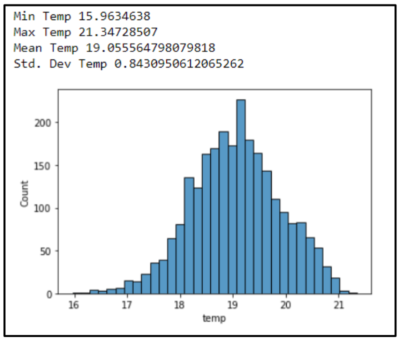
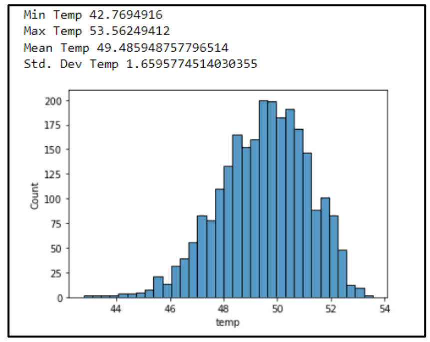


Figure 10: Summer Temp Distribution



From the above histograms, the distribution of temperatures in the summer months has a higher standard deviation compared to winter months, and is slightly more left skewed indicating a higher concentration of LST in higher temperature ranges.

Additionally, we can examine the correlation for summer and winter:

Figure 11: Summer Variable Correlation

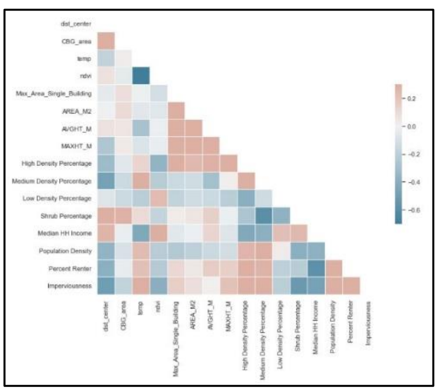
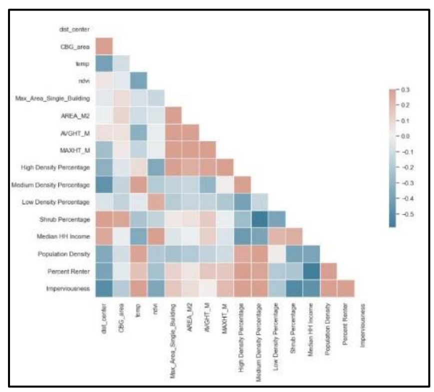


Figure 12: Winter Variable Correlation



As the plots above exhibit, the temperature is highly correlated with NDVI in the summer months. In the winter months, medium density percentage appears to be more highly correlated compared to summer. Furthermore, the distance to the urban center is much more negatively correlated in the winter months than it is in the summer months. For both summer and winter, medium household income is negatively correlated with land surface temperature.

Now, we can examine the results of the Random Forest and XGBoost, using the parameters specified above. We include the MAPE (Mean Absolute Percentage Error) because just looking at the MSE will create lower values for the winter months due to the lower standard deviation.

Table 3: Error, R^2, and MAPE (Mean Absolute Percentage Error) Output

	MSE	**R^2**	**MAPE**
Random Forest			
Summer (Incl. NDVI)	0.81195	0.69074	**1.410%**
Summer (Excl. NDVI)	1.29188	0.50794	1.793%
Winter (Incl. NDVI)	0.28311	0.59482	**2.170%**
Winter (Excl. NDVI)	0.32055	0.54124	2.327%
XGBoost			
Summer (Incl. NDVI)	0.77036	0.70658	**1.349%**
Summer (Excl. NDVI)	1.24832	0.52445	1.754%
Winter (Incl. NDVI)	0.29168	0.58255	**2.166%**
Winter (Excl. NDVI)	0.321475	0.53992	2.322%

As we can see from the table above, across all analysis schema, the ones that include NDVI are the most accurate (in terms of MAPE). When we compare the differences between Random Forests and XGBoost, the improvements from XGBoost are only marginal, especially in the winter. What is promising, though, is that both the R^2 value and the MAPE is relatively low, which gives us confidence that our model is accurately describing the most important features that drive land surface temperature within our test area. From here, we can begin to do a deeper dive on the actual results from the tests in graphical form. Furthermore, it is important that we evaluate, and compare the most important features from our eight different tests. This will help us take a look "under the hood" of the model, as we can start to see what the major drivers are for the UHI effect in Phoenix.

Random Forest Analysis – Error

Figure 11: Summer Excluding NDVI:

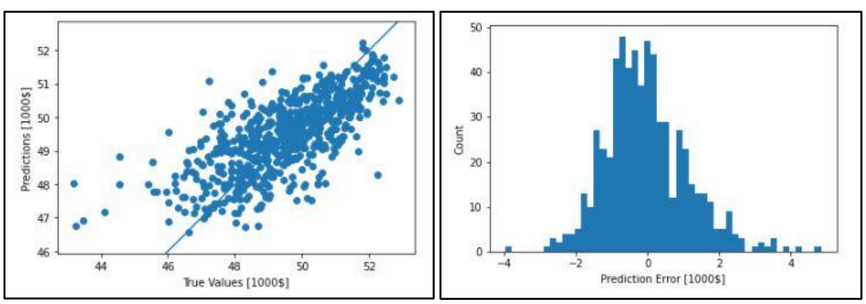


Figure 12: Summer Including NDVI:

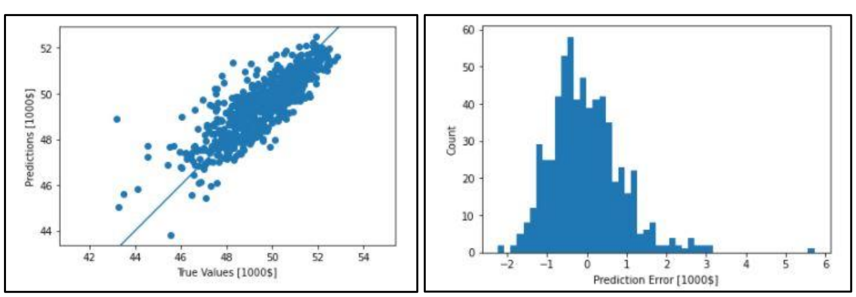


Figure 13: Winter Excluding NDVI:

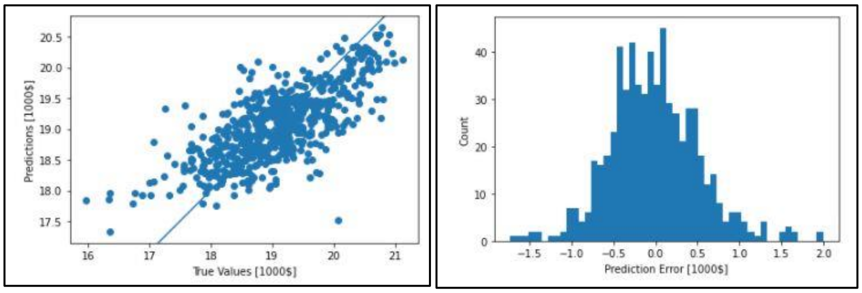
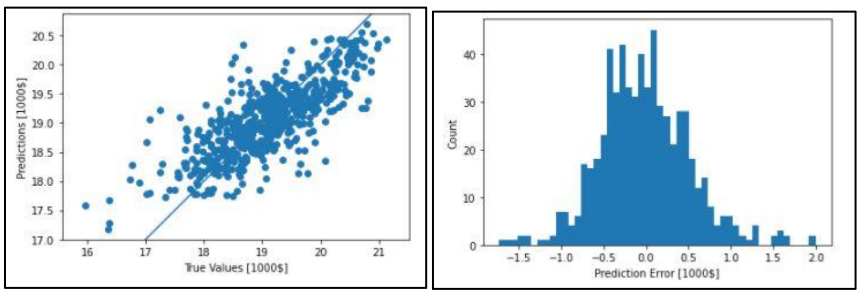


Figure 14: Winter Including NDVI:



When we examine the output from the Random Forest model, the MSE and MAPE start to take shape in graphical form. As we can see, in the summer when we include NDVI, the results are highly clustered around the best fit line, compared to the winter when the results are somewhat more dispersed. Next, we can turn to the feature importance to understand what the major drivers are across the different models.

Random Forest Analysis – Feature Importance

Figure 15: Summer Including NDVI

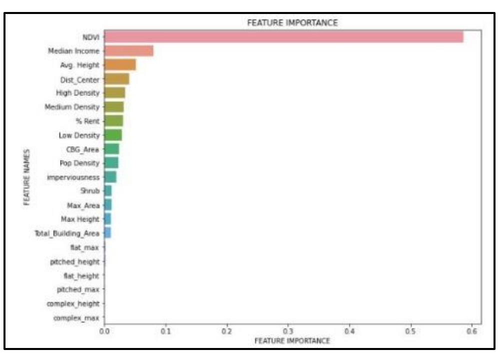


Figure 16: Summer Excluding NDVI

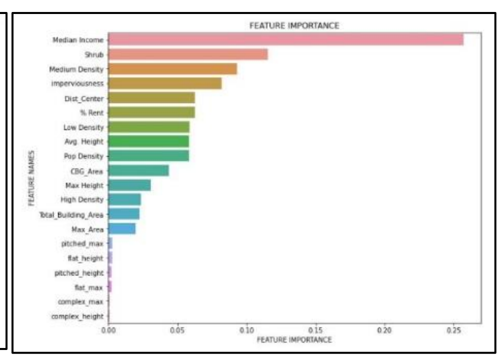


Figure 17: Winter Including NDVI

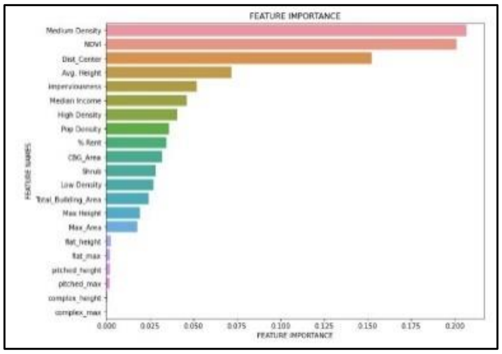
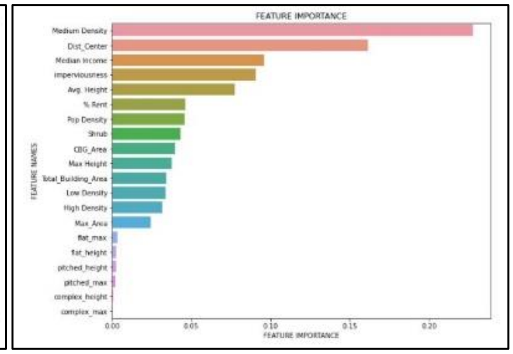


Figure 18: Winter Excluding NDVI



The feature importance comparison between winter and summer, and including and excluding NDVI creates some of the most interesting and compelling output from our analysis. As we can see, the pattern of NDVI driving a significant portion of the analysis in the summer months holds true, but once we remove this factor, some of the other features gain new importance. For instance, when we examine the importance of shrub cover in summer months, its importance when we include NDVI is roughly 1%. Compared to excluding NDVI, the shrub cover importance jumps to ~12%. What's also interesting is the fact that a socioeconomic factor plays such a significant role in the feature importance for the summer analysis without NDVI. Median HH income is the most important feature when we exclude NDVI. Given that wealthier areas are traditionally less dense, and have higher vegetation indices [15], this impact makes logical sense.

Additionally, when we consider the difference between the summer analysis and winter analysis, across both inclusive and exclusive of NDVI, we see different features with different levels of importance. For both the winter analyses, medium density land cover is the most important feature, whereas in the summer, it's the 6th and 3rd most important feature. Furthermore, in the winter months, the distance to the urban center is a top 3 most important feature. This is especially interesting when we consider the definition of the UHI effect in relation to urban densities, given traditional UHI models expect lower temperatures the farther away from the urban center.

XGBoost Analysis – Error

Figure 19: Summer Excluding NDVI:

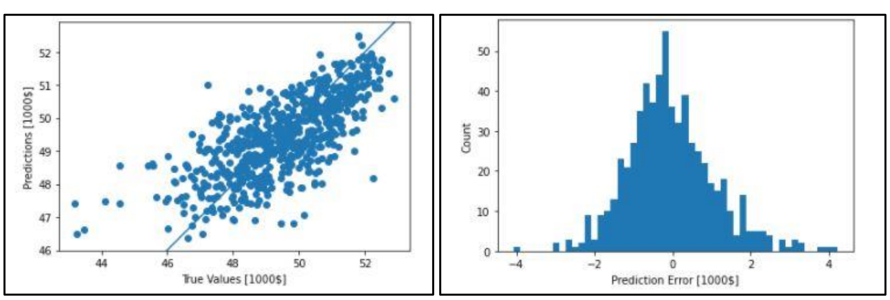


Figure 20: Summer Including NDVI:

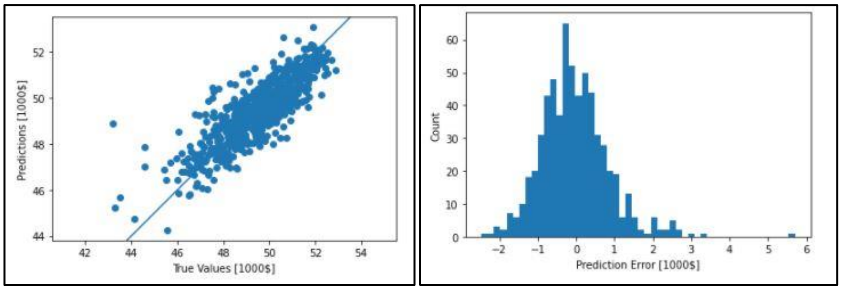


Figure 21: Winter Excluding NDVI:

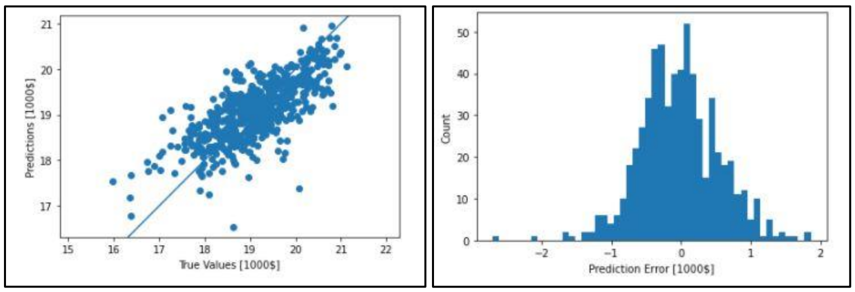
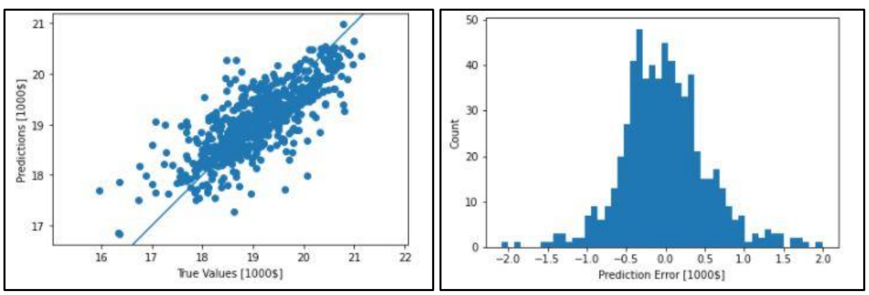


Figure 22: Winter Including NDVI:



Similar patterns are again exhibited in the XGBoost model, and the results from the previous table are once again reinforced graphically. What's interesting is that there is a higher incidence of outlier values in the summer including NDVI analysis compared to the Random Forest approach.

XGBoost Analysis – Feature Importance

Figure 23: Summer Including NDVI

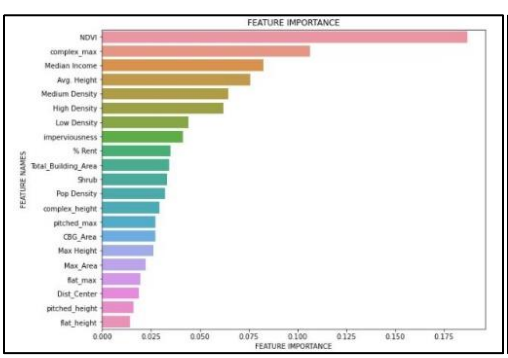


Figure 24: Summer Excluding NDVI

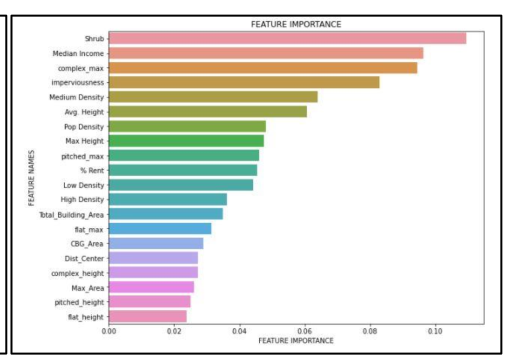


Figure 25: Winter Including NDVI

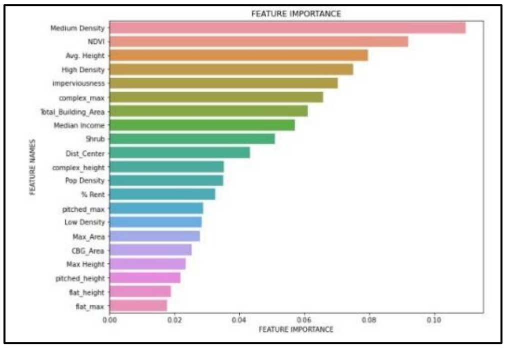
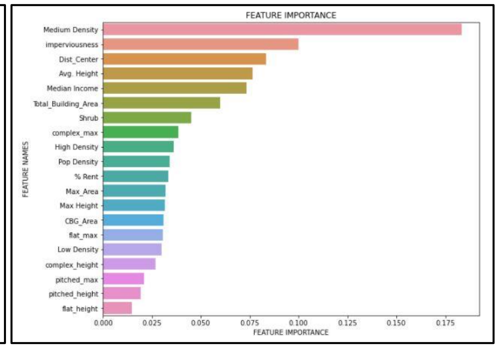


Figure 16: Winter Excluding NDVI



When we examine the feature importance from the XGBoost, particularly compared to what we saw from the Random Forest approach, we see some interesting dynamics. For both of the summer analyses, the roof type of the building with the max height in the CBG is a top 3 feature. When we contrast this with the random forest approach, all roof type variables were some of the lowest importance features. According to Deng, et al, roof type plays a significant role in the radiance and heat generation of a given building [16], so it's interesting to see that in the XGBoost these types of factors drive significant levels of importance. Looking at the winter analysis, the most important driver for both NDVI inclusive and exclusive is percent medium density land development. In the context of a city like phoenix, where there are lower levels of high density development, and it is essentially one large suburb, it makes sense that medium density development drives a significant level of importance given that there are lower incidences of high density developmental areas.

8. Conclusions

Overall, the results of both the XGBoost model and the Random Forest model show promising output in terms of reducing error. Having all errors be less than 2.5% across all testing schemes suggests a high degree of accuracy, and this is reinforced by our corresponding R^2 values. We achieved this by randomized optimization schemas for our hyperparameters, and by selecting variables that have historically held significant importance in physical modelling processes for the UHI effect.

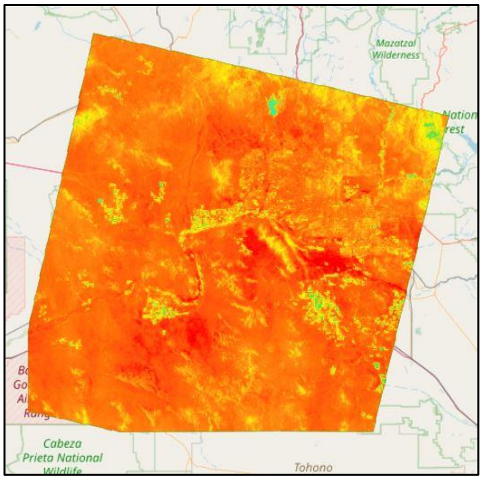
When we consider the feature importance across all testing schema, we start to see greater levels of nuance, which have more significant implications for future urban development. For the summer months, wealthier, less dense communities with higher degrees of shrub coverage saw lower temperatures compared to denser, more urban areas. This is not surprising when we consider the root drivers of the UHI effect previously explored in physical modelling, but it does help us understand which areas of the city will be impacted more severely as the UHI effect becomes more severe. What's interesting, is that particularly in the Phoenix metropolitan area, more affluent communities have a higher concentration of golf courses, back yard swimming pools, and larger property sizes. Therefore, seeing these results is not surprising when we consider higher density areas trap more heat.

When we think about the winter months, it's interesting to see that one of the most important features is distance to city center (especially considering this was a variable with lower importance in the summer months). One hypothesis that could be explored further is that because Phoenix is such an arid, hot place to begin with, the UHI effect behaves differently in the summer months when temperatures reach their extremes. Because the density in Phoenix is lower, maybe instead of being an urban heat island effect, it should be a subclass called the ‘suburban heat island effect’.

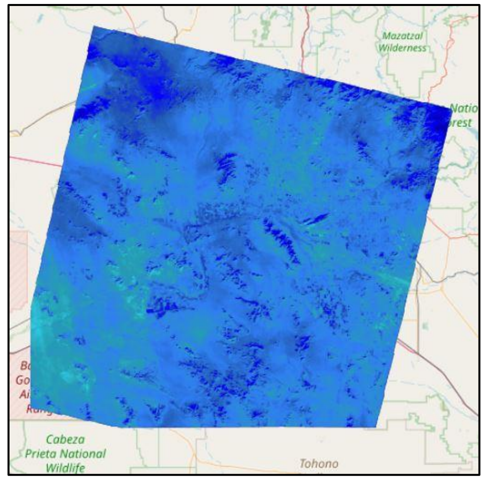
As we look to future implications and continuations of this research, it will be interesting to revisit our calculations when contemporary data is more readily available. As the city has changed in the last five years, what new dynamics exist? Furthermore, an extension of this project could be examining how the UHI effect impacts the seasonal monsoons that typically pass through Phoenix during the June / July months.

Appendix of Images:

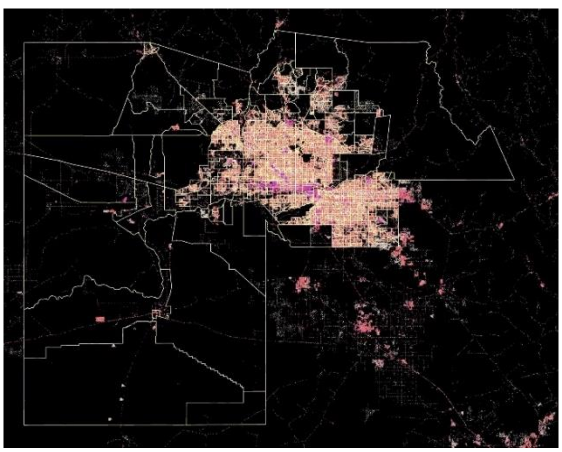
LANDSAT8 LST – Summer



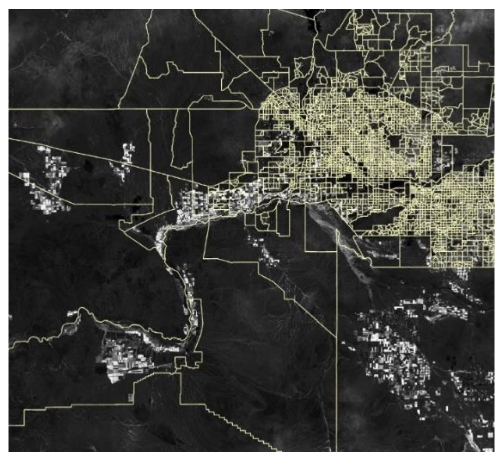
LANDSAT8 LST – Winter



Impervious Land Cover



NDVI



References:

1. Habeeb, D., Vargo, J., & Stone, B. (2015). Rising heat wave trends in large US cities. *Natural Hazards*, *76*(3), 1651–1665. https://doi.org/10.1007/s11069-014-1563-z
2. Hibbard, K. A., Hoffman, F. M., Huntzinger, D., & West, T. O. (2017). Changes in land cover and terrestrial biogeochemistry. climate science special report: Fourth national climate assessment, volume I. https://doi.org/10.7930/j0416v6x
3. Healy, J. (2021, August 12). No Large City Grew Faster than Phoenix. *New York Times*. Retrieved December 23, 2021, from https://www.nytimes.com/2021/08/12/us/phoenix-census-fastest-growing-city.html.
4. Munson, O. (2021, July 1). June 2021 breaks record for hottest June on record in Phoenix, and the trend could continue. *AZ Central*. Retrieved December 23, 2021, from https://www.azcentral.com/story/news/local/phoenix-weather/2021/07/01/june-2021-breaks-record-hottest-june-record-phoenix/7829100002/.
5. National Weather Service . (2010-2021). NOAA Online Weather Data. *National Weather Service NOAA Data*. chart.
6. United States, CLIMAS. *Southwestern Monsoon*, NOAA. https://climas.arizona.edu/sw-climate/monsoon. Accessed 23 Dec. 2021.
7. United States Census, 2021
8. Tan, Jianguo, et al. "The Urban Heat Island and Its Impact on Heat Waves and Human Health in Shanghai." *International Journal of Biometeorology*, vol. 54, no. 1, 2009, pp. 75–84., https://doi.org/10.1007/s00484-009-0256-x.
9. Huang, Ganlin, et al. "Is Everyone Hot in the City? Spatial Pattern of Land Surface Temperatures, Land Cover and Neighborhood Socioeconomic Characteristics in Baltimore, MD." *Journal of Environmental Management*, vol. 92, no. 7, 2011, pp. 1753–1759., https://doi.org/10.1016/j.jenvman.2011.02.006.
10. Johnson, D. P., & Wilson, J. S. (2009). The socio-spatial dynamics of extreme urban heat events: The case of heat-related deaths in Philadelphia. Applied Geography, 29(3), 419–434. doi:10.1016/j.apgeog.2008.11.004
11. Digavinti, Jeevalakshmi & Reddy, S. & Manikiam, Balakrishnan. (2017). Land surface temperature retrieval from LANDSAT data using emissivity estimation. International Journal of Applied Engineering Research. 12. 9679-9687.
12. Breiman, Leo. *Machine Learning*, vol. 45, no. 1, 2001, pp. 5–32., https://doi.org/10.1023/a:1010933404324.
13. Pal, M. "Random Forest Classifier for Remote Sensing Classification." *International Journal of Remote Sensing*, vol. 26, no. 1, 2005, pp. 217–222., https://doi.org/10.1080/01431160412331269698.
14. McCarty, Dakota, et al. "Machine Learning Simulation of Land Cover Impact on Surface Urban Heat Island Surrounding Park Areas." *Sustainability*, vol. 13, no. 22, 2021, p. 12678., https://doi.org/10.3390/su132212678.
15. Ravulaparthy, Srinath K., et al. "Spatial Firm Demographic Microsimulator: Development and Validation for Phoenix and Tucson, Arizona, Megaregion." *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2671, no. 1, 2017, pp. 59–70., https://doi.org/10.3141/2671-07.

16. Deng, Y., Chen, R., Xie, Y., Xu, J., Yang, J., & Liao, W. (2021). Exploring the impacts and temporal variations of different building roof types on surface urban heat island. *Remote Sensing*, *13*(14), 2840.