Remote Sensing Correlations Between Income and Land Cover to Analyze Urban Heat Islands in Austin, Texas

Authors:

Sophia Myers[1], Hubery Pai[2], Noah Peralez[3], Abhiram Raju[4], Anna Shifman[5], Grace Valdez[6#], Erika Podest[#], Peder Nelson[#]

[1]South Carolina Governor's School for the Arts and Humanities
[2]Rick Reedy High School
[3]STISD Science Academy
[4]Chirec International School
[5]Agoura High School
[6]A.N. McCallum High School
[#]Advisor

Contributions:

A.S. extracted all of the data from Collect Earth Online to create one large dataset of the points with impervious surfaces and greenness from 2022-2024. She also created visualizations for the impervious surfaces and greenness using SeaBorn and MatPlotLib. A.S. preprocessed and cleaned the dataset to be ready for the different models. She then developed, analysed, and visualised, an unsupervised machine learning model using K-means clustering. Lastly, she edited the abstract and gathered photos from GLOBE Observer for cross checking.

A.R initiated and developed advanced machine learning models using Hierarchical clustering and Anomaly detection techniques. He also extracted, cleaned, and mapped income data from the US Census using geodesic and the Vincenity algorithm, aligning it precisely with designated AOI locations. A.R also helped create these visualizations for income data using SeaBorn and MatPlotLib. In addition to preprocessing the data, Abhiram visualized the distributions of variables in relation to the generated clusters, finding the optimal number of clusters. Finally, he also helped create the abstract and organized the files and data.

S.M. performed data extraction and mapping for ECOSTRESS land surface temperature and part of the GLOBE Observer land cover dataset, handled and organized the data files, cleaned and filtered the datasets, and designed the project logo and presentation. Additionally, she handled the communications involved in the project, edited the final video for the SEES Virtual Science Symposium along with A.S., and organized, wrote, and archived the GitHub repository. She also helped create the abstract.

H.P. performed visual and statistical analysis. He mapped income data using SeaBorn and MatPlotLib, mapping it based on intensity at AOI locations. H.P. applied p-value analysis and Pearson correlation (r) to identify statistical significant correlations within the factors. He also implemented the DBSCAN algorithm. Additionally, he created the heat maps for the correlations between all 4 factors, highlighting key trends and relationships. Finally, he adjusted the filtering algorithms for the data to increase data quality and quantity, only removing data when comparing certain factors.

N.P. and A.R. collaborated on drafting the abstract and organizing files and data. N.P. also helped create heatmaps to illustrate correlations among the four factors, assisting H.P., which revealed key trends in the relationship. Additionally, N.P. reviewed and looked over the data files for land surface temperature as part of the GLOBE Observer cover dataset. Furthermore, he took part in creating a poster for the NASA SEES Symposium, designing each section of the poster.

G.V. served as a peer mentor and generated the original project idea as well as providing resources for study. E.P. and P.N. also served as mentors. E.P. assisted with assessing data quality, expanding the scope of the project, and writing the abstract. P.N. provided guidance on using GLOBE Observer, Collect Earth Online, and assisted in data extraction on ArcGIS Online.

Abstract

Urban Heat Islands (UHIs) significantly impact urban environments such as Austin, Texas, particularly among socioeconomically disadvantaged communities. Due to tightly packed housing, low-income areas typically have less vegetation and more impervious surfaces, leading to higher land surface temperatures. Understanding the factors worsening extreme heat events is crucial to protect those affected, especially as climate change exacerbates UHIs.

We focused on household income, percent greenness, and impervious surfaces as variables to answer these research questions:

1. What factors contributed to an increase in land surface temperature the most?
2. How do these variables correlate in the context of Austin, TX?

We analyzed land cover for twelve areas of interest (AOIs) created by NASA SEES interns from 2022-2024. Each AOI, arranged in a $9km^2$ grid, contained 37 coordinates. From the 444 resulting points, we derived greenness and impervious surface percentages (Collect Earth Online), income data (U.S. Census), and land surface temperature (ECOSTRESS via APPEEARS). We validated the impervious surface and greenness percentages using GLOBE Observer ground photos. Preliminary analysis revealed a significant correlation ($p=0.002$) between socioeconomic disparities and impervious surfaces, with lower-income areas having more impervious surfaces and higher UHI risk.

We applied unsupervised learning algorithms to identify trends in our data, fitting K-means, Hierarchical, and DBSCAN clustering models. These models grouped data points by similar land cover and income features. We analyzed cluster patterns and determined the optimal number of clusters using Silhouette and Davies-Bouldin indexes, resulting in two clusters for K-means, three for Hierarchical, and five for DBSCAN. K-Means corroborated that our hypothesis was accurate; communities at the highest risk for UHIs have 61% less vegetation and 48% lower income than those at low risk.

Next, we plan to combine our methods with Supervised Learning in California to analyze a wider range of data and create more accurate correlations.

Key Words: Urban Heat Islands, Remote Sensing, Socioeconomic Data, Unsupervised Learning, GLOBE Observer, Citizen Science

Research Questions

1. What are the primary factors contributing to increased risk of urban heat islands (UHIs) in socioeconomically disadvantaged areas of Austin, Texas?
    a. Identifying key contributors to higher land surface temperatures in lower-income areas addresses local environmental justice issues. This is important for developing targeted mitigation strategies.
    b. These environmental justice issues in Austin, Texas can be applied to other urban areas in order to analyze contributors to UHIs and climate change across the world.
2. To what extent do these primary variables correlate with each other?
    a. Identifying correlations between variables provides insight into which variables, if any, are most responsible for generating extreme heat.
3. Out of the twelve areas of interest (AOIs) in Austin, Texas we gathered data from, which ones are most likely to be affected by urban heat islands based on these correlations?
    a. Understanding which AOIs are at the highest risk for the UHI effect allows for local initiatives to be made providing targeted relief for those affected. Additionally, mapping out these high-risk areas allows city planners to mitigate the effects of extreme heat by making use of the variables described in the questions above (e.g. planting more vegetation, limiting impervious surface land cover).

Introduction

As global warming and climate change intensify, urban areas are increasingly pressured to lessen and mitigate their effects. Compared to rural areas, cities, characterized by dense populations and developed infrastructure, typically feature less vegetation and more impervious surfaces such as concrete and asphalt. These surfaces absorb and retain heat, exacerbating the urban heat island effect: a phenomenon in which certain areas of urban regions experience much higher temperatures than rural environments. This effect is caused by high concentrations of vehicles, machinery, and other sources of greenhouse gasses and fossil fuels, contributing significantly to atmospheric warming and the formation of urban heat islands in large cities. The urban heat island effect has the potential to become severe if left unchecked by cities. NASA Applied Sciences warns, “The rapid growth of urban populations, the urban heat island effect, and increase in the frequency and duration of heat waves due to climate change, raise a series of issues about the increased health risks of sensitive urban populations to extreme heat and the effective means of mitigating the impacts of heat waves.” (ARSET, 2022) In addition, increased land surface temperatures may affect unexpected areas of urban infrastructure: material and energy flow, water quality, soil quality, and biological habits could all be subject. (Li Yang et al, 2016) Extreme heat in cities not only harm the health of the environment, but also of the citizens living there, meaning that sustainable solutions must be addressed in order to maintain healthy communities.

In this research, our team will attempt to investigate the most significant driving factors behind the formation of urban heat islands (UHIs) in the urban city of Austin, Texas and analyze the correlations between these factors using unsupervised learning algorithms in order to inform further environmental justice applications. We decided to focus on four distinct variables in our analysis. Firstly, we investigated percentages of greenness and impervious surfaces across our area of study. These two factors are already known contributors to the UHI effect, and literature suggests that low amounts of vegetation and high amounts of impervious surfaces worsen the effects of extreme heat due to impervious surfaces’ absorption and retaining of the sun’s radiation (Climate). Therefore, we expect to see similar patterns in our analysis. The next variable we focused on is land surface temperature, which will allow us to analyze specific temperature changes over time and identify possible UHIs. Finally, we chose to use median household income as an additional variable in order to connect our study of urban heat islands to environmental justice and identify any correlations between wealth and susceptibility to extreme heat events. If there are any locations in our study area with low household income but a high risk for urban heat islands, the people living there may not be as well equipped to handle the effects of these high temperatures. In summary, by identifying areas with land cover features that correspond to high land surface temperatures, we can evaluate the risk that area has for the formation of urban heat islands. Then, we will be able to combine that risk factor with the area’s median household income, allowing us to inform possible environmental justice initiatives to better prepare the population of the area to withstand extreme urban heat or even work to minimize the effects of urban heat.

The increasingly prevalent challenge of urban heat islands calls for comprehensive research to inform effective mitigation strategies. By analyzing key factors contributing to UHIs in Austin, Texas, this study aims to provide actionable insights that can enhance urban populations’ resilience to climate change. Our findings will not only advance scientific understanding but also support cities’ and communities’ efforts to create healthier, more sustainable urban environments.

Methods and Materials

Choosing and Analyzing an Area of Study

Before choosing our area of study, we had to make sure there would be enough data for our regression analysis available to us. By looking at a map of all of the AOIs in the GLOBE Observer land cover database from 2022 to 2024, we noticed that there is an abundance of data in Austin, Texas from previous and current NASA SEES interns. This abundance of data combined with the fact that Texas - and by extension, the city of Austin - is an extremely ecologically and socioeconomically diverse state led to us choosing it as our primary area of study.

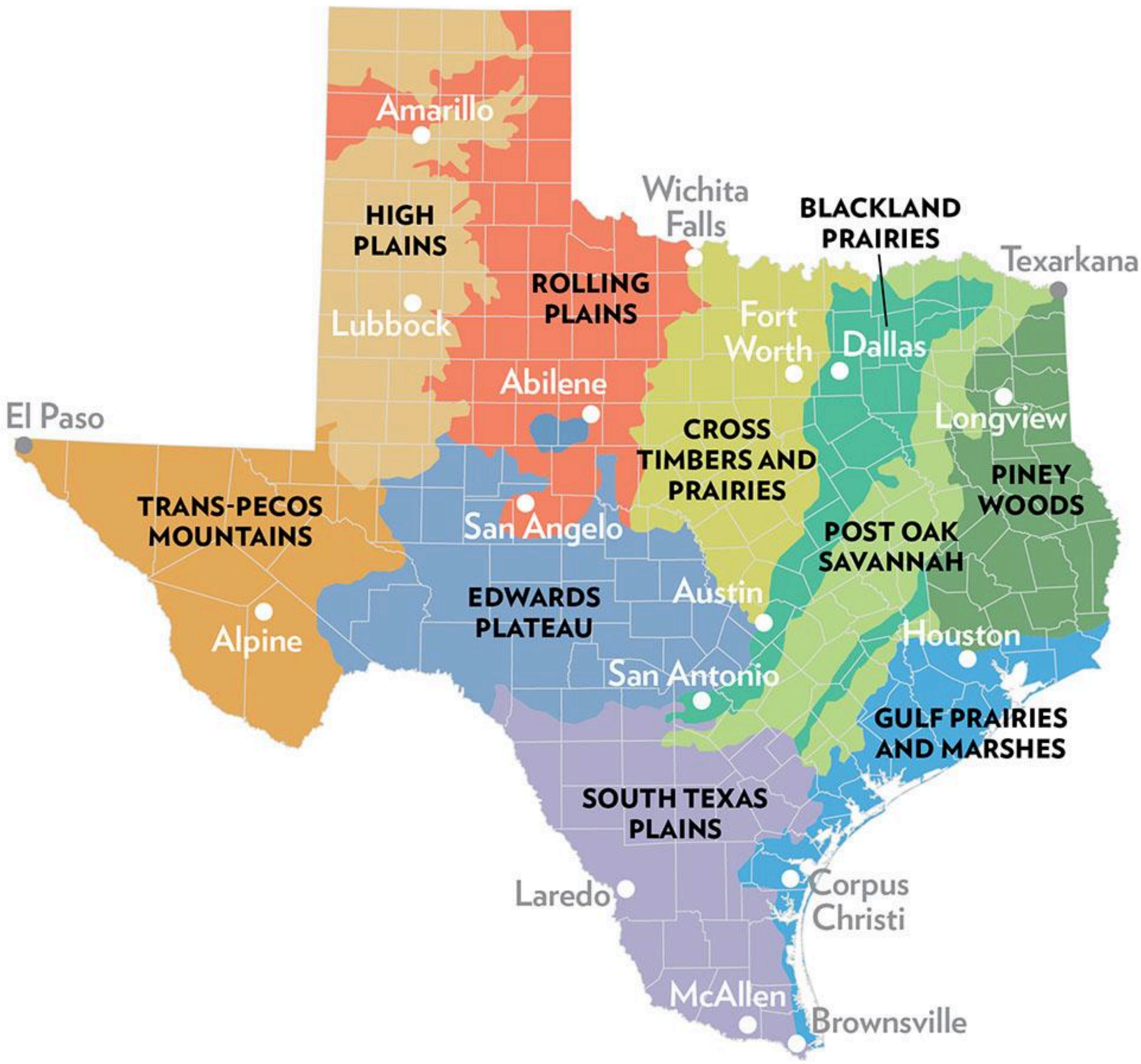


Fig. 1: A map of the ecoregions and major cities of the state of Texas, courtesy of Texas Highways via https://texashighways.com/outdoors/wildflowers/the-wildflower-regions-and-vegetational-areas-of-texas/ .

Austin, Texas is located between the natural regions of the Blackland Prairies and the Edwards Plateau. Because of its close proximity to multiple different ecoregions, the ecology and climate of Austin vary. However, Austin is also located in the area of central Texas known as the Texas Hill Country, which is characterized by shallow soil, hills and caves made primarily of limestone, grassland, mesquite savannah, and temperate juniper or oak woodlands (Texas, 2022). Additionally, the city of Austin contained a population of roughly 958,000 people in 2022 and had an average median yearly income of about $86,500 (Austin).

We found fourteen total AOIs in Austin between 2022 and 2024; however, two were deemed unreliable due to missing land cover data, leaving us with a total of 12 AOIs with 37 data points inside each to work with. Additionally, each of the 37 points in an AOI are further subdivided into a 10x10 grid, creating 100 secondary points, which were then labeled using land cover descriptors in Collect Earth Online (Fig. 2). Using these secondary points, we are able to get the exact percentages of impervious surfaces and greenness (tree canopy cover, grass, and bush added together) and coordinates for each AOI point.

Fig. 2: A screenshot of an AOI point in Austin, Texas being labeled using Collect Earth Online's land cover description system. The screen is made up of a 10x10 point grid overlaid on a satellite map with red points representing impervious surfaces and green points representing any tree cover or grass.

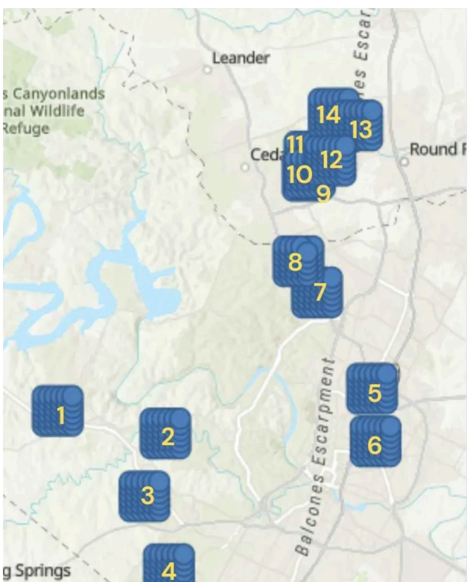


Fig. 3: The 14 total AOIs found in the Austin, Texas area visualized in an ArcGIS map and numbered. Note that two of these AOIs were not included in the final correlation analysis because of poor data quality from Collect Earth Online.

Data Extraction

As a part of our data analysis, we extracted a .csv file of the coordinates, dates and times of measurements, GLOBE Observer user IDs, and percentages of impervious surfaces, tree cover, grass, and shrubbery of all 37 points per each of the 14 AOIs that we observed in Austin from the GLOBE Observer database. After downloading this file, we uploaded it into a webmap in the geographic information system ArcGIS Online in order to layer all of our data files into one large, visual dataset. Next, we requested land surface temperature (LST) data and their corresponding quality checkers through NASA AppEEARS for the same group of points. We sourced our LST data from the ECOSTRESS instrument aboard the International Space Station and chose to collect data daily between 2022 and 2024. All of our LST data was also exported as one .csv file and imported into the ArcGIS web map. After importing the LST data, it was filtered using the quality checkers requested from AppEEARS to limit the dataset to only data points that were deemed good quality.

The income data was accessed from a 2011-2015 ACS 5-year documentation provided by U.S Census Reports from Kaggle. As shown, these incomes were converted to discrete variables over certain ranges:

Standard Income Table:

Low (L)	High (H)		Midpoint (I)	Range (R)
\$0	\$10,000		\$5,000	\$10,000
\$10,000	\$14,999		\$12,500	\$4,999
\$15,000	\$19,999		\$17,500	\$4,999
\$20,000	\$24,999		\$22,500	\$4,999
\$25,000	\$29,999		\$27,500	\$4,999
\$30,000	\$34,999		\$32,500	\$4,999
\$35,000	\$39,999		\$37,500	\$4,999
\$40,000	\$44,999		\$42,500	\$4,999
\$45,000	\$49,999		\$47,500	\$4,999
\$50,000	\$59,999		\$55,000	\$9,999
\$60,000	\$74,999		\$67,500	\$14,999
\$75,000	\$99,999		\$87,500	\$24,999
\$100,000	\$124,999		\$112,500	\$24,999
\$125,000	\$149,999		\$137,500	\$24,999
\$150,000	\$199,99		\$175,000	\$49,999
\$200,000	\$275,000		\$237,500	\$75,000

This data was filtered to only include areas within Austin. For each income location, the dataset included mean, median and standard deviations of income. Using each of the fourteen AOIs, we mapped each one to the nearest known household income. In order to do this, we utilized the geodisc library which calculates this using the Vincenity formula, which is included in the "Data Availability" section of this report.

To analyze the possibility of error when using Vincenity, we also calculated the distances between these locations and made sure that they matched what we saw on a satellite map. After applying this, we generated a heatmap showing all of the median incomes in our AOI locations, with yellow representing lower income and red representing higher income:

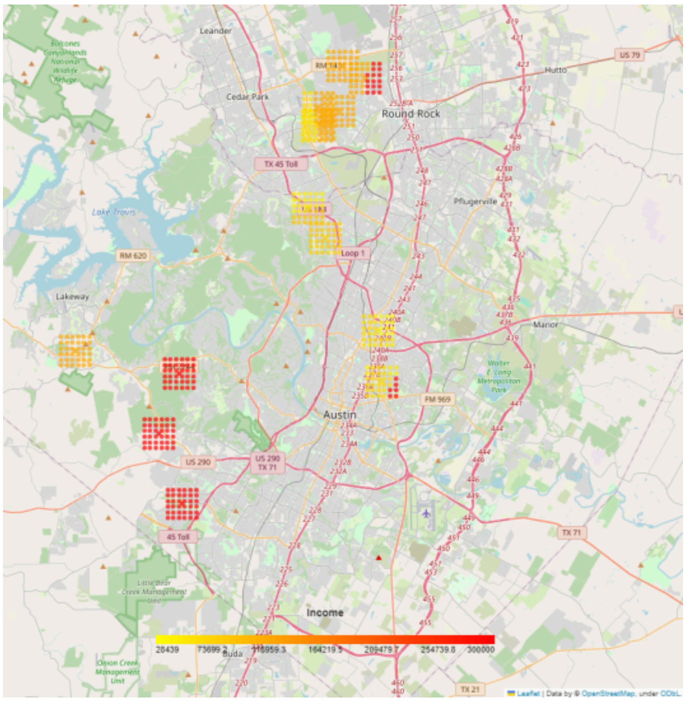


Fig. 4: a screenshot from an OpenStreetMap satellite map displaying a heatmap scale from yellow to red of median household incomes across fourteen AOIs in Austin, Texas.

Data Processing

After having the complete dataset, we started to preprocess the data. We selected the following columns to analyze within our data: longitude, latitude, grass, impervious surfaces (not including buildings), tree canopy cover, bush/scrub, median income, and average income. We then summed tree canopy cover, bush/scrub, and grass into one variable called vegetation to get the percentage of greenness within a city. We also dropped any columns that had missing data (with the exception of analysis for correlations, in which case columns would only be dropped if the columns being compared were missing data). After cleaning the data, we had a total of 444 rows of data representing 444 coordinates in Austin. We then preprocessed and normalized the data by importing StandardScaler from sklearn.preprocessing. Normalizing the data allows all of the values to scale so the machine learning algorithms can run effectively. We then visualized the frequency distributions of income as well as that of the vegetation and impervious surface data using SeaBorn and Matplotlib

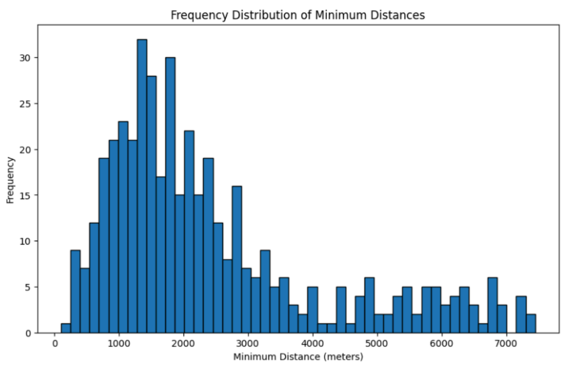


Fig. 5: A histogram generated in Matplotlib that visualizes the frequencies of minimum distances, in meters, between each of the 444 AOI points.

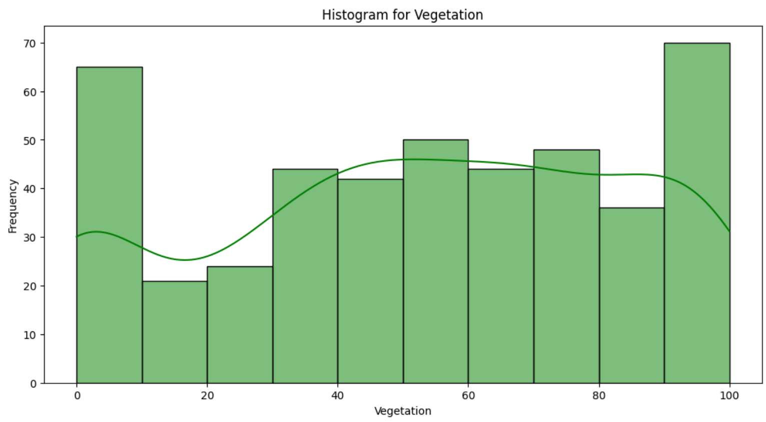


Fig. 6: A histogram generated in Matplotlib that visualizes the frequencies of vegetation (tree canopy cover, bush/scrub, and grass percentages added together) percentages in each of the 444 AOI points.

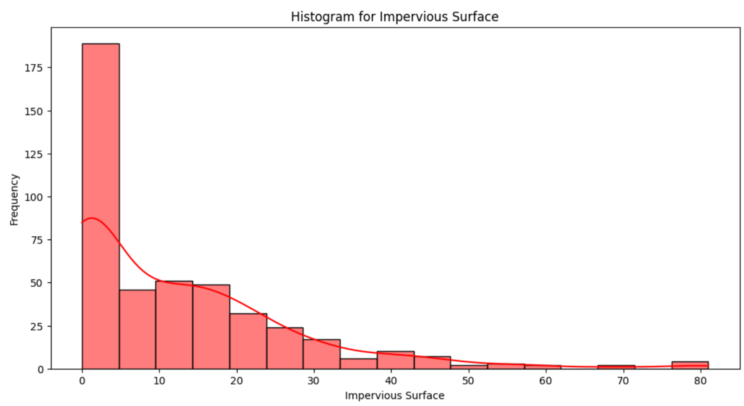


Fig. 7: A histogram generated in Matplotlib that visualizes the frequencies of impervious surface percentages in each of the 444 AOI points.

Data and Results

We performed Pearson correlation analysis to determine the relationships and statistical significance between the correlations of the variables. To do this, we calculated the p-value and r-value between all possible combinations of factors from scipy.stats. After calculating the results, we identified statistically significant correlations if $p < 0.05$, such as income v.s. total greeness ($p = 6e-18$), providing correlations and statistical significance with a 95% confidence for all factors.

Correlation and Statistical Significance Table

Compared Factors	Correlation	p-value	Statistical Significance
ECO_L2T_LSTE_002_LST with Median Income_LST	-0.01777787722432561	0.7109330360323557	Not significant
ECO_L2T_LSTE_002_LST with Total Greeness	-0.050761976472202534	0.2896911516107493	Not significant
ECO_L2T_LSTE_002_LST with Impervious Surface (no building)	-0.058487001652537764	0.2223939927527586	Not significant
Median Income_LST with Total Greeness	0.3937482404747515	6.451200348467655e-18	Statistically significant
Median Income_LST with Impervious Surface (no building)	-0.14329101055917032	0.002474440934546296	Statistically significant
Total Greeness with Impervious Surface (no building)	-0.28934510307828376	5.195441555043095e-10	Statistically significant

Fig 8: Table depicting correlation r and p-value between all factors. Note that median income and greenness, median income with impervious surfaces, and greenness with impervious surfaces are statistically significant and most likely did not occur by chance.

As noted from Fig. 8, we found 3 significant and expected correlations: median income and greenness, median income with impervious surfaces, and greenness with impervious surface. Median income and total greeness were weakly (r = 0.39) positively correlated, as the higher income an area has, more efficient and green infrastructure can be developed. Median income and impervious surfaces were very weakly (r = -0.14) negatively correlated, as with lower wealth, less consideration is placed on urban heat. Finally, total greenness with impervious surfaces was weakly (r = -0.28) negatively correlated, which aligns with most modern developments, as when impervious surface use increases, greenness in the area decreases.

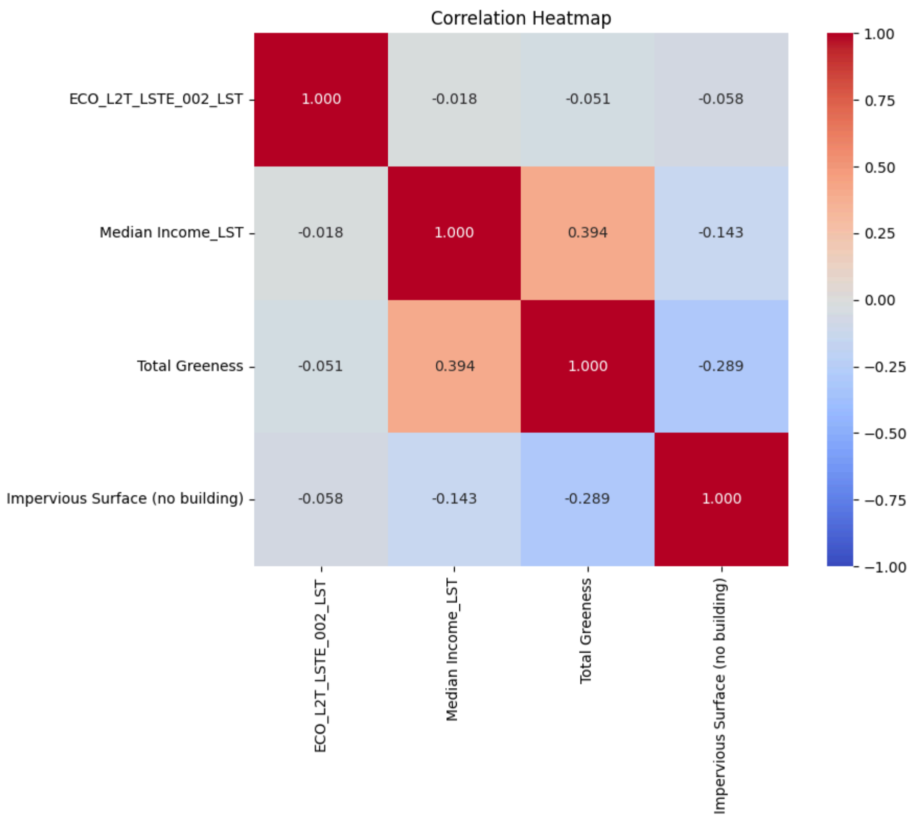


Fig 9: This heatmap displays the correlation r between all factors. This heatmap was created using matplotlib.pyplot.

We tested multiple unsupervised machine learning models in order to create similar clusters among all of our points of study, including K-means. To find the optimal number of clusters using this model, we calculated the Silhouette score, Davies-Bouldin score, and Calinski-Harabasz score from sklearn.metrics. We looped through a range from two to five clusters to find the amount most effective for our analysis that leads to the lowest possibility of error. After calculating and plotting the results, we found that two clusters would give the most accurate results.

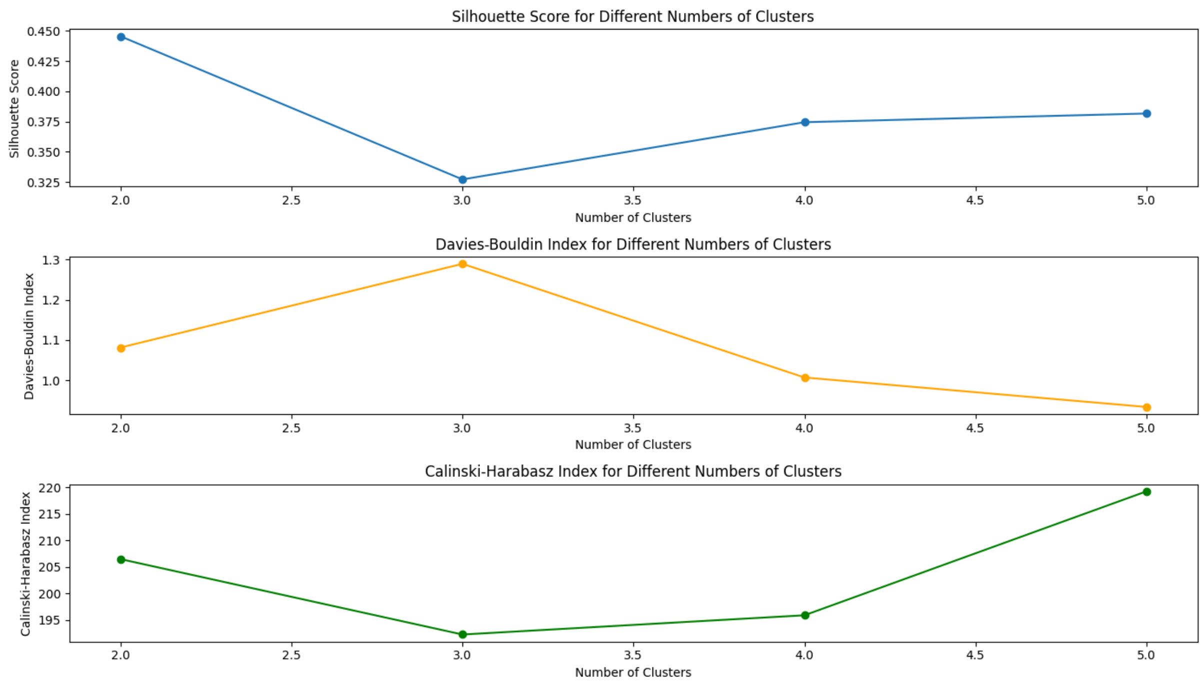


Fig. 10: Three charts showing three different index scores used to calculate the optimal number of clusters. From top to bottom, they are the Silhouette Score, the Davies-Bouldin Index, and the Calinski-Harabasz Index. These charts were created in Matplotlib.

Next, we ran a K-means algorithm on our data in order to begin determining which AOIs in our study are at the highest risk for the urban heat island effect. As seen through the visualization of our clustering data below, the locations of the clusters generated by K-means line up with the wealth disparity shown in Fig. 4; the AOIs with the highest income fall into the "low risk" cluster.

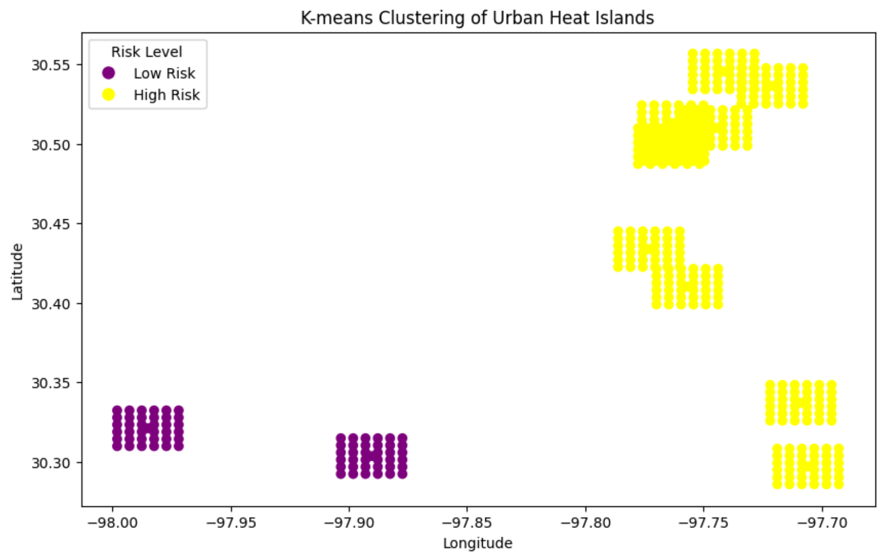


Fig. 11: A visualization generated in Matplotlib that graphs the twelve AOIs by longitude and latitude and separates them by cluster into either low risk (shown in purple) or high risk (shown in yellow).

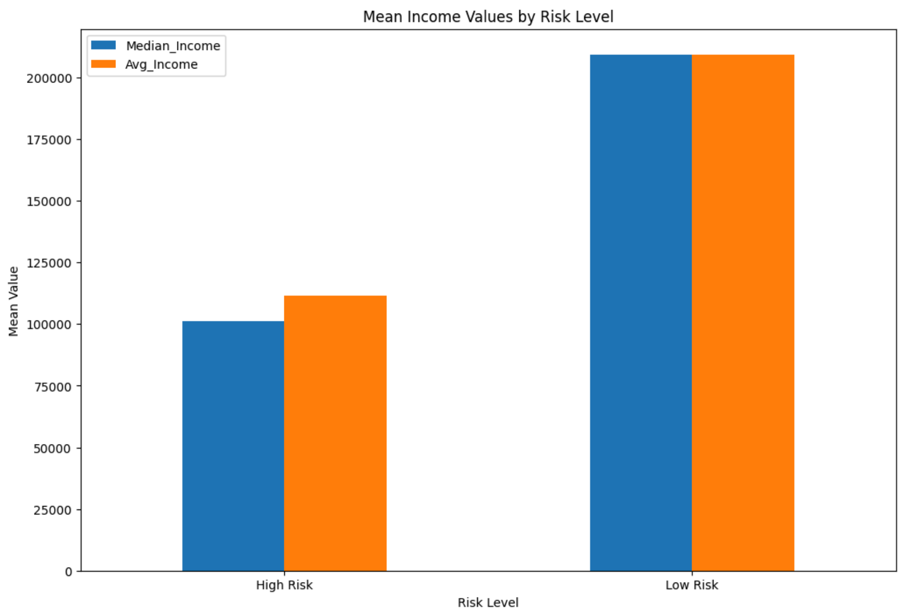


Fig. 12: A bar chart generated in Matplotlib that shows the disparities between median household income (shown in blue) and average household income (shown in orange), in US dollars, in the low risk versus high risk clusters. This chart displays that AOIs at the highest risk for the UHI effect have a median income roughly 44% lower and an average income roughly 50% lower than those at low UHI risk.

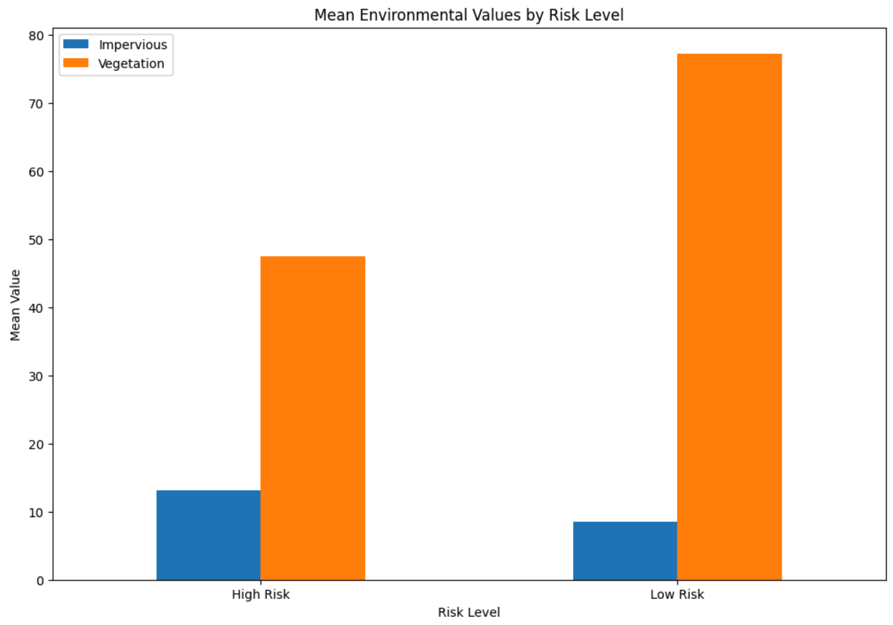


Fig. 13: A bar chart generated in Matplotlib that shows the disparities between the percentages of impervious surfaces (shown in blue) and vegetation (shown in orange) in the low risk versus high risk clusters. This chart displays that AOIs at the highest risk for the UHI effect contain roughly 30% less vegetation and 5% more impervious surfaces than those at low UHI risk.

From the K-means clustering results, it is evident that there is a significant income gap within Austin, Texas that puts low income communities at a greater risk for UHIs than those with comparatively higher incomes. This phenomenon is made worse by the evidence that lower income communities contain less vegetation and more impervious surfaces than those with high income, likely due to these communities' lack of excess wealth to build large green spaces like parks and community centers and high amounts of tightly-packed housing units. The results from the K-means clustering gave us a final silhouette score of 0.4456.

We also performed Density-Based Spatial Clustering of Applications with Noise, optimal for real world data. Using the Silhouette Score as the benchmark, we iterated through different combinations of minimum points required for each cluster and maximum distance (epsilon) between points of a cluster. After iterating through all combinations of epsilon (0.7-1.5, increasing by 0.01) and minimum points (1-5, increasing by 1), the highest silhouette score was 0.3215.

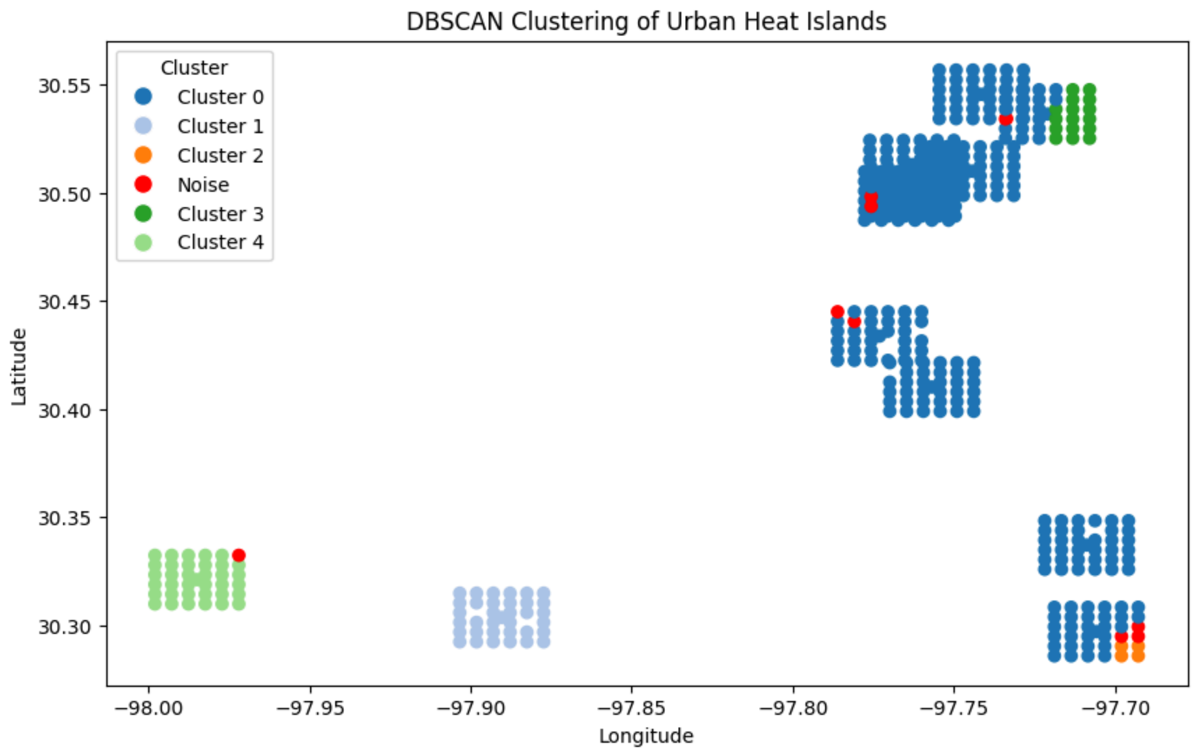


Fig 14: A visualization generated in Matplotlib that graphs the twelve AOIs by longitude and latitude and separates them by clusters with similar factors and UHI risk.

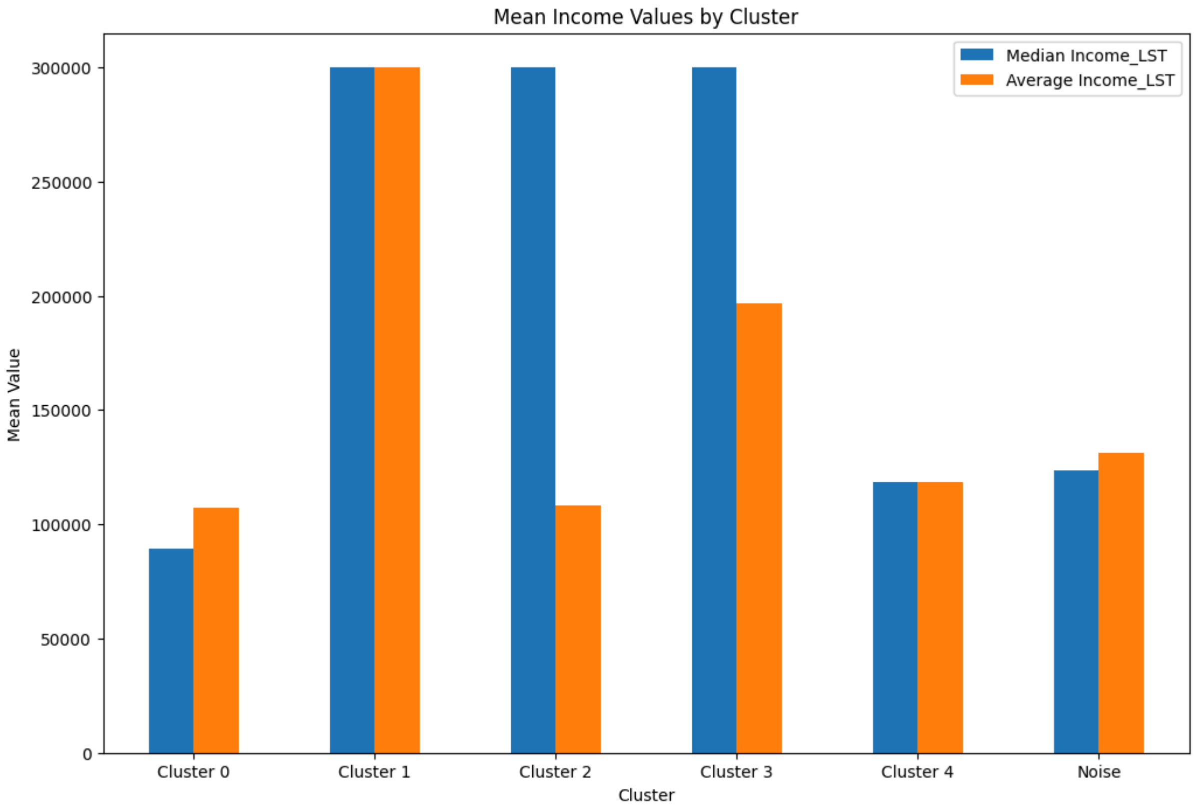


Fig 15: A bar chart generated in Matplotlib that graphs the mean income values of 5 clusters identified and noise.

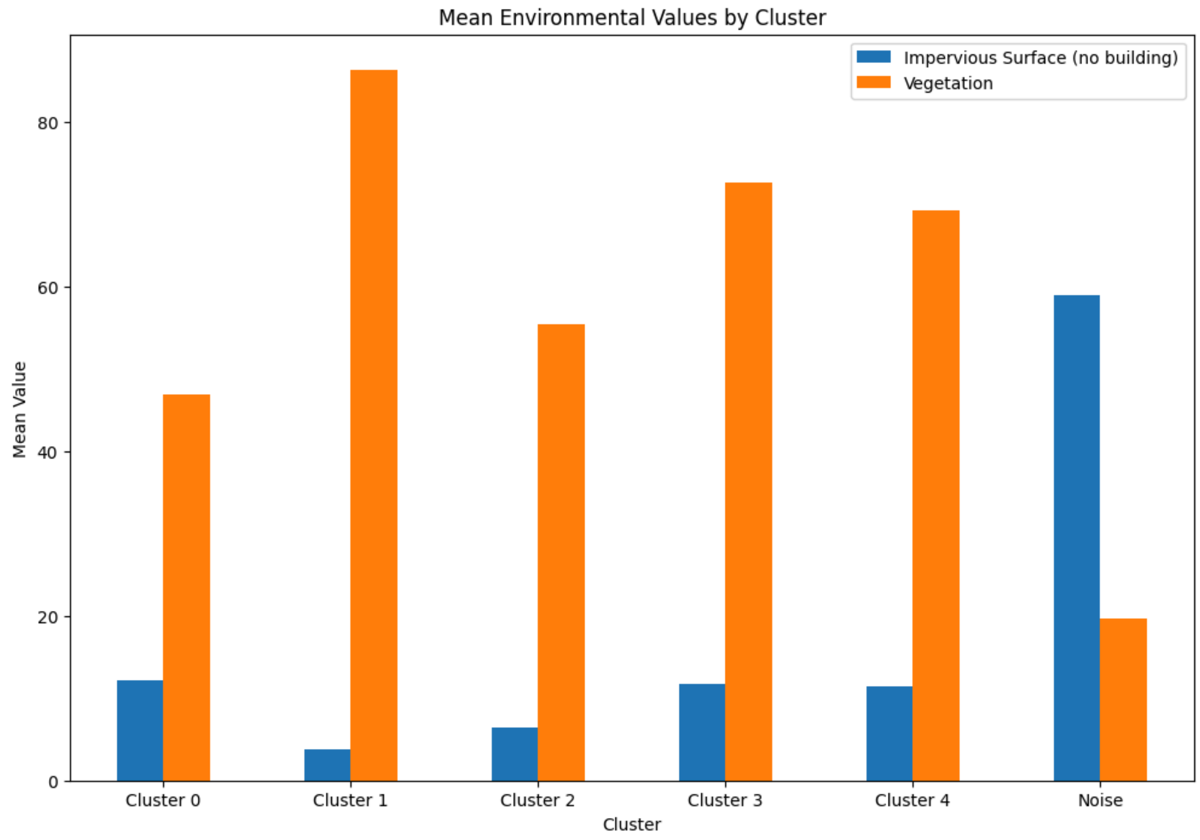


Fig 16: A bar chart generated in Matplotlib that graphs the vegetation and impervious surface values within the 5 clusters identified and noise.

In comparison to K-means clustering, it can be seen that DBSCAN differentiates within clusters, allowing identification of low risk areas surrounded by low income areas. By comparing this to the income data visualization in Fig 4, it can be seen that the clusters formed align with AOIs where there are mass differences in income. Additionally, cluster 0 has the lowest median income and average income in all of the clusters, and resulted in the highest mean value of impervious surfaces and lowest values of vegetation (ignoring noise), aligning with the trend described with Pearson analysis.

We also performed Hierarchical (Agglomerative) clustering algorithms to determine which AOI locations had the greatest risk of high land surface temperatures and urban heat islands. Like the previous algorithms, Hierarchical clustering was run on a range of different numbers of clusters to find the optimum levels and relations of different variables.

Figures 16-18 show the results on different numbers of clusters.

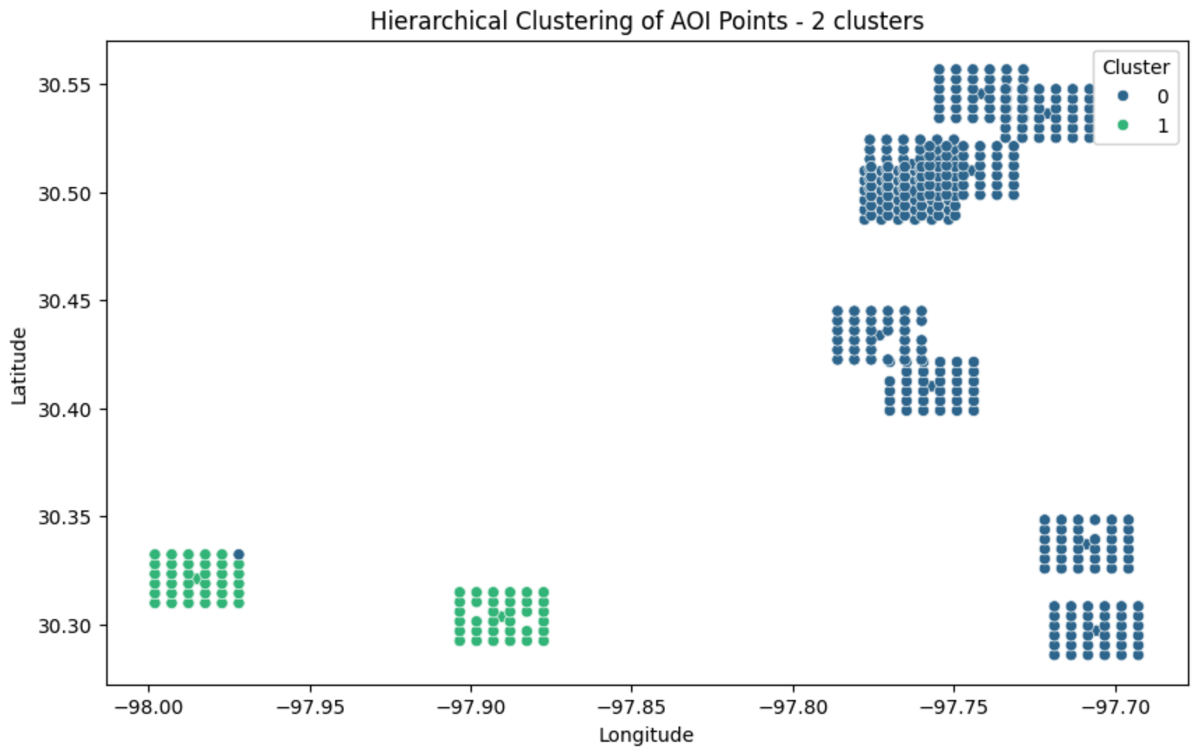


Figure 16- A bar chart generated in Matplotlib that graphs the AOI point coordinates into 2 clusters.

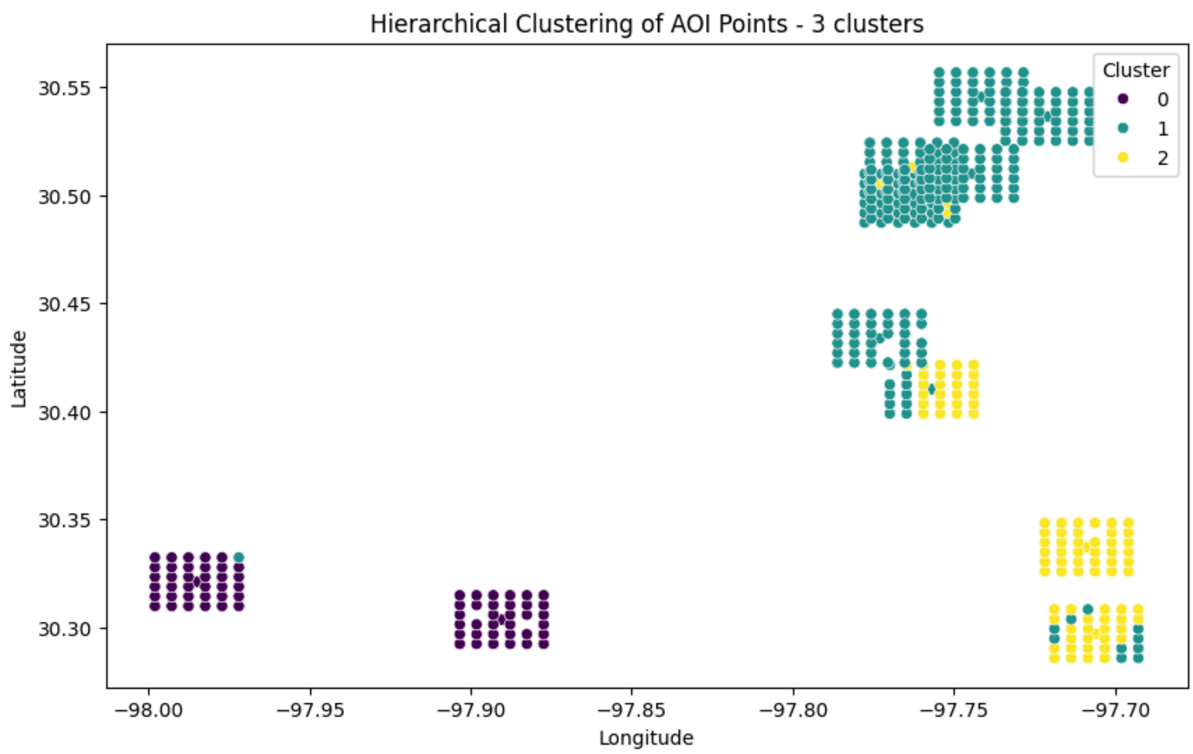


Figure 17- A bar chart generated in Matplotlib that graphs the AOI point coordinates into 3 clusters.

In figures 18 and 19, the results income levels in clusters are shown for 2 and 3 as the number of clusters:

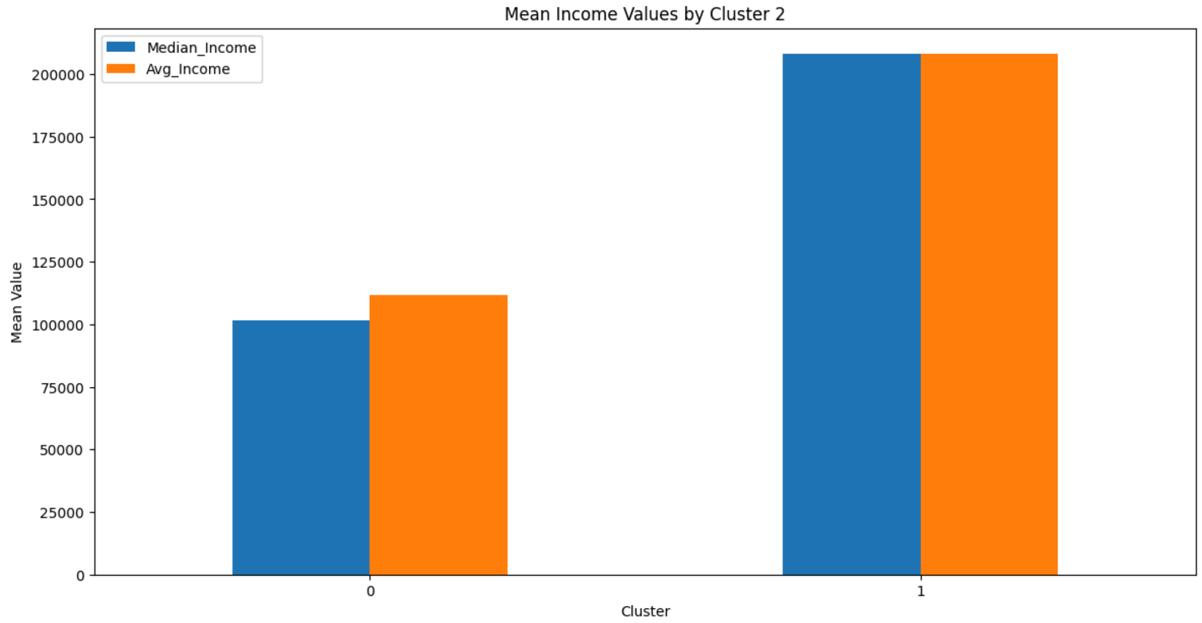


Fig 18: A bar chart generated in Matplotlib that graphs the median and average AOI income values within the 2 hierarchical clusters.

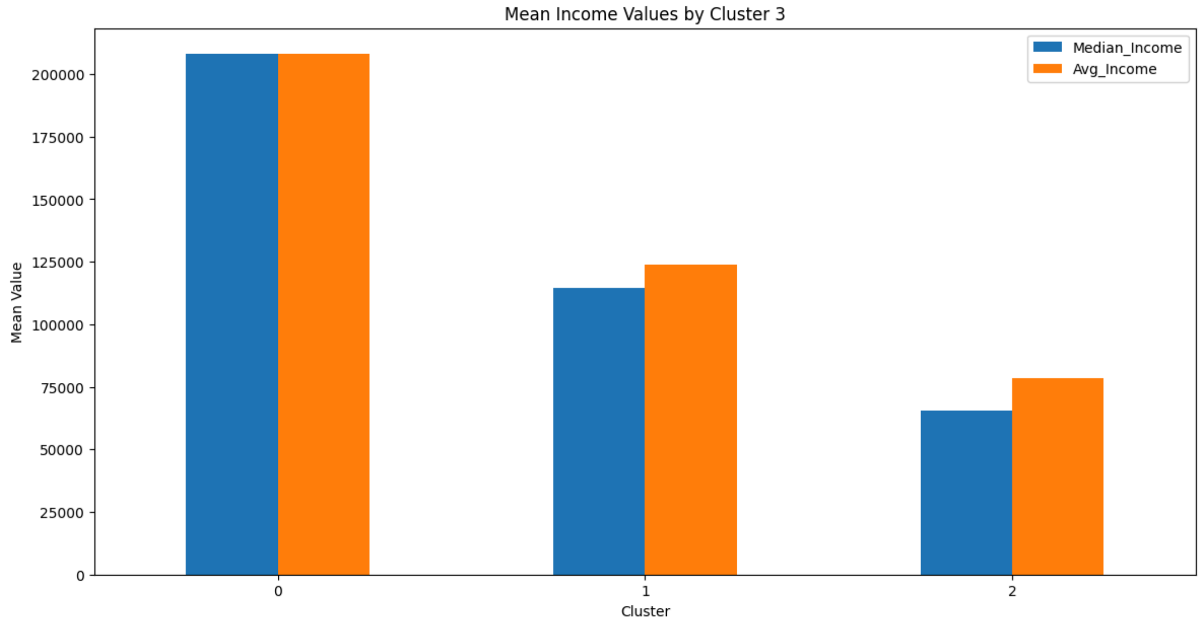


Fig 19- Fig 18: A bar chart generated in Matplotlib that graphs the median and average AOI income values within the 3 hierarchical clusters.

This follows from the trend from the other clustering algorithms in that there is a clear disparity between income levels of different clusters. The advantage of using 3 clusters is that there is an additional group for “medium risk”- Cluster 1 - which means that every AOI location doesn’t need to fall under a binary high or low risk of UHIs.

Figures 20 and 21 validate this wealth disparity as clusters with low incomes have high impervious surface percentages and low vegetation percentages. In poorer neighbourhoods, the lack of trees and more impervious surfaces is what causes this higher UHI risk.

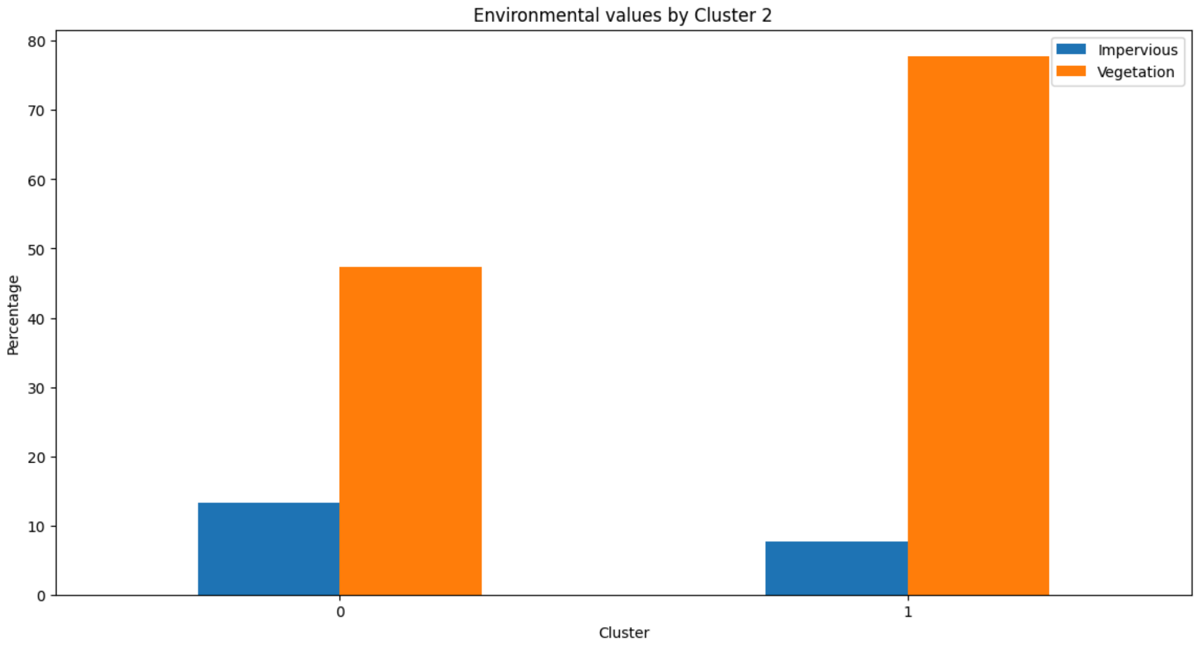


Fig 20- A bar chart generated in Matplotlib that graphs the Impervious surface and vegetations percentages within the 2 hierarchical clusters.

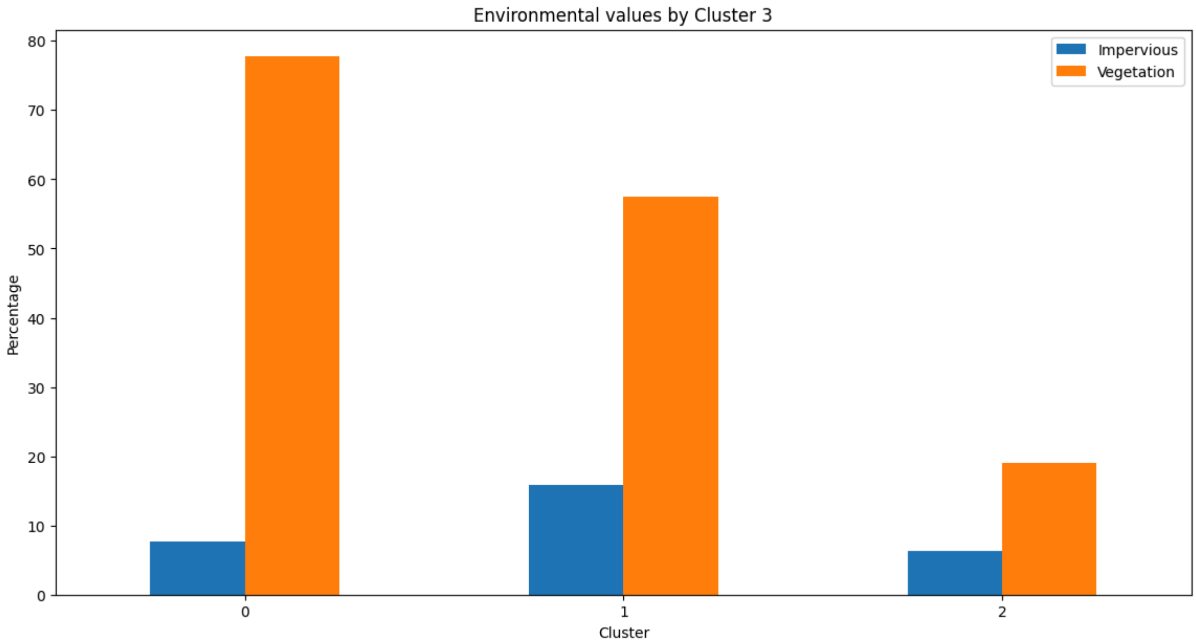


Fig 21- A bar chart generated in Matplotlib that graphs the impervious surface and vegetation percentages within the 3 hierarchical clusters.

To look at a spread of every variable's distribution with one another, we generated pyplots shown in figure 22.

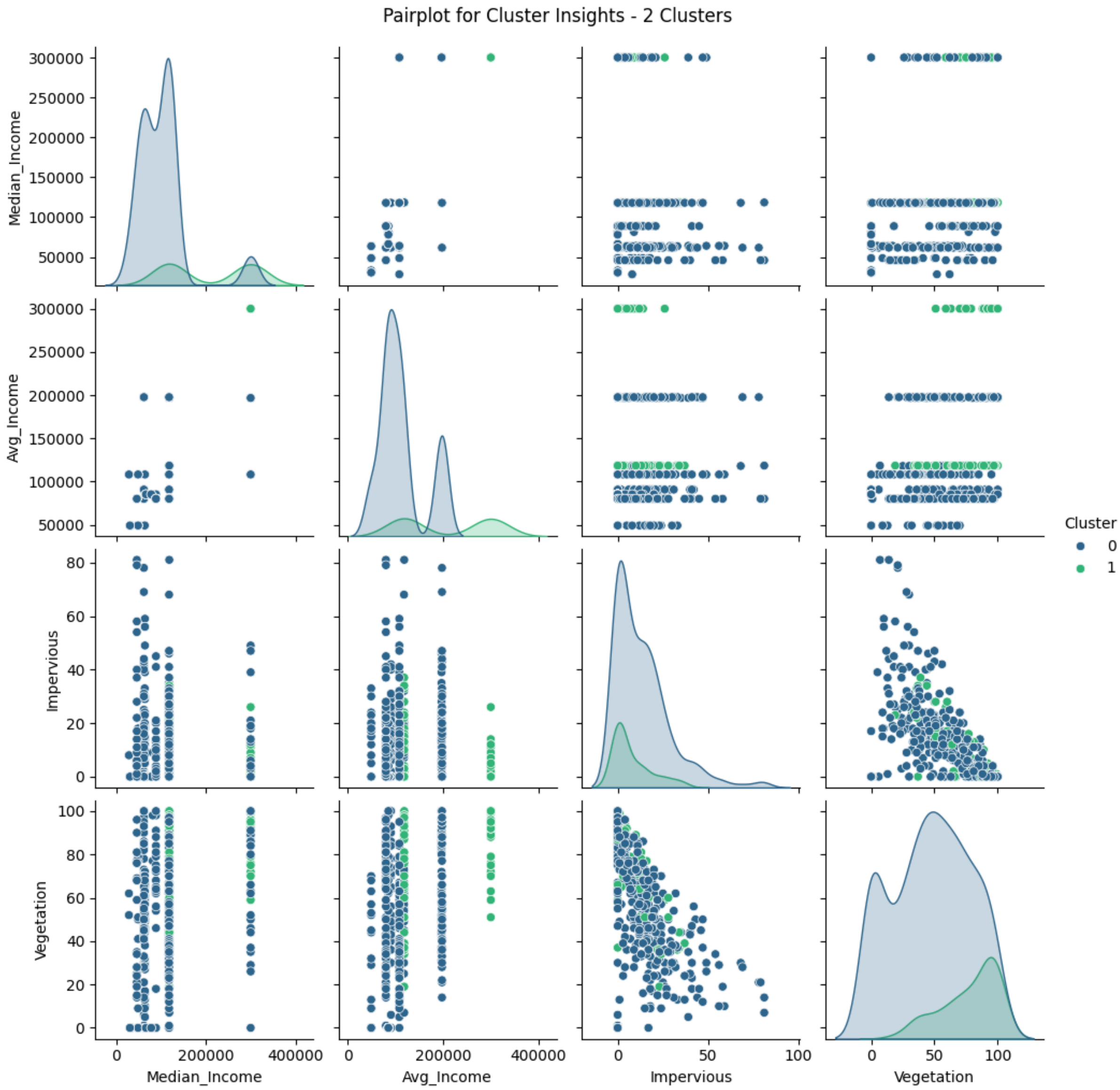


Fig 22: Pyplot distributions of our 4 features for 2 clusters.

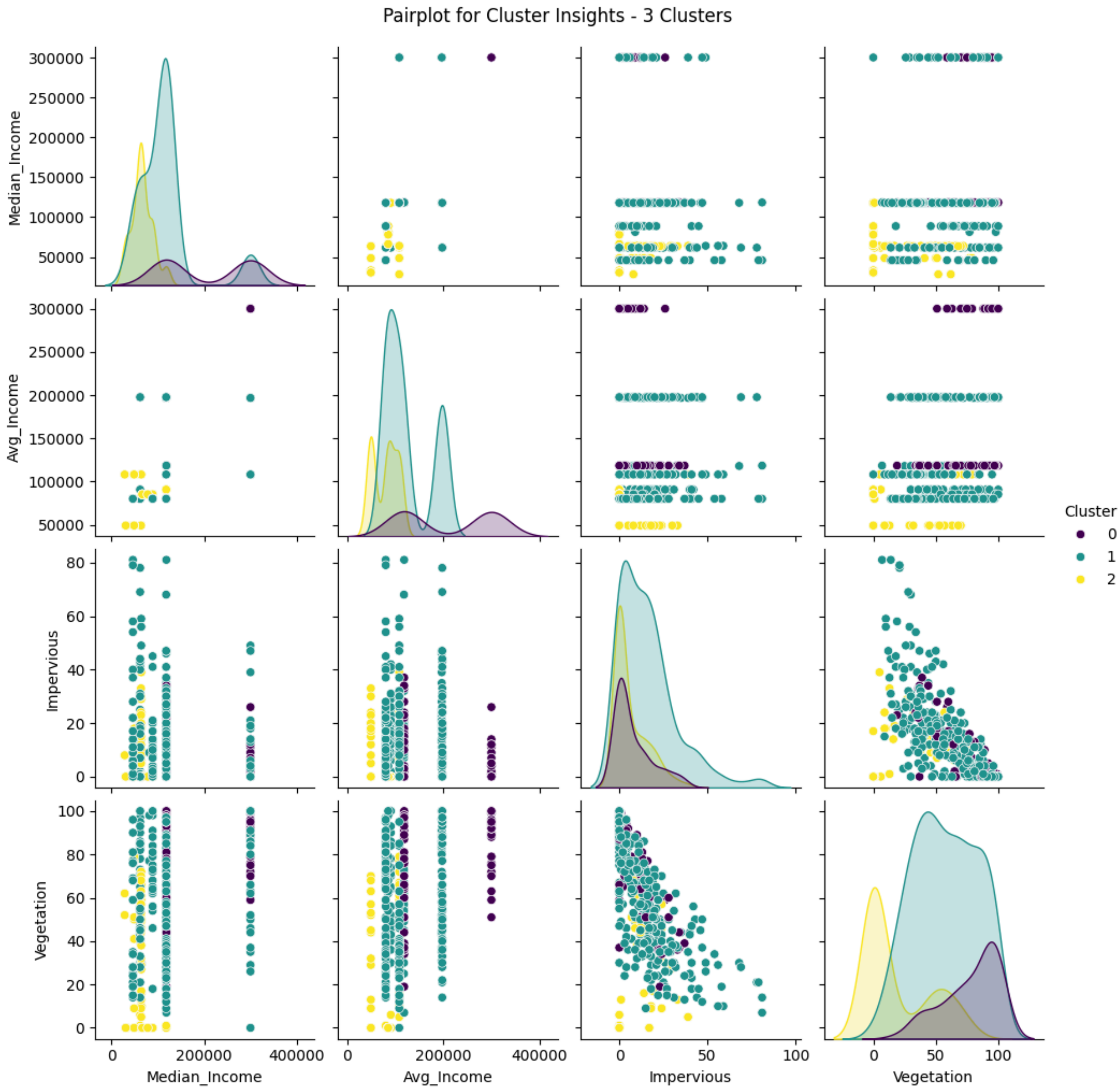


Fig 23: Pyplot distributions of our 4 features for 3 clusters.

The highest silhouette score that arose from hierarchical clustering was 0.386 for 2 clusters. Given this score, we carried out an Anomaly Detection Algorithm to filter out the anomalies and compare performance. This is shown in figure 24:

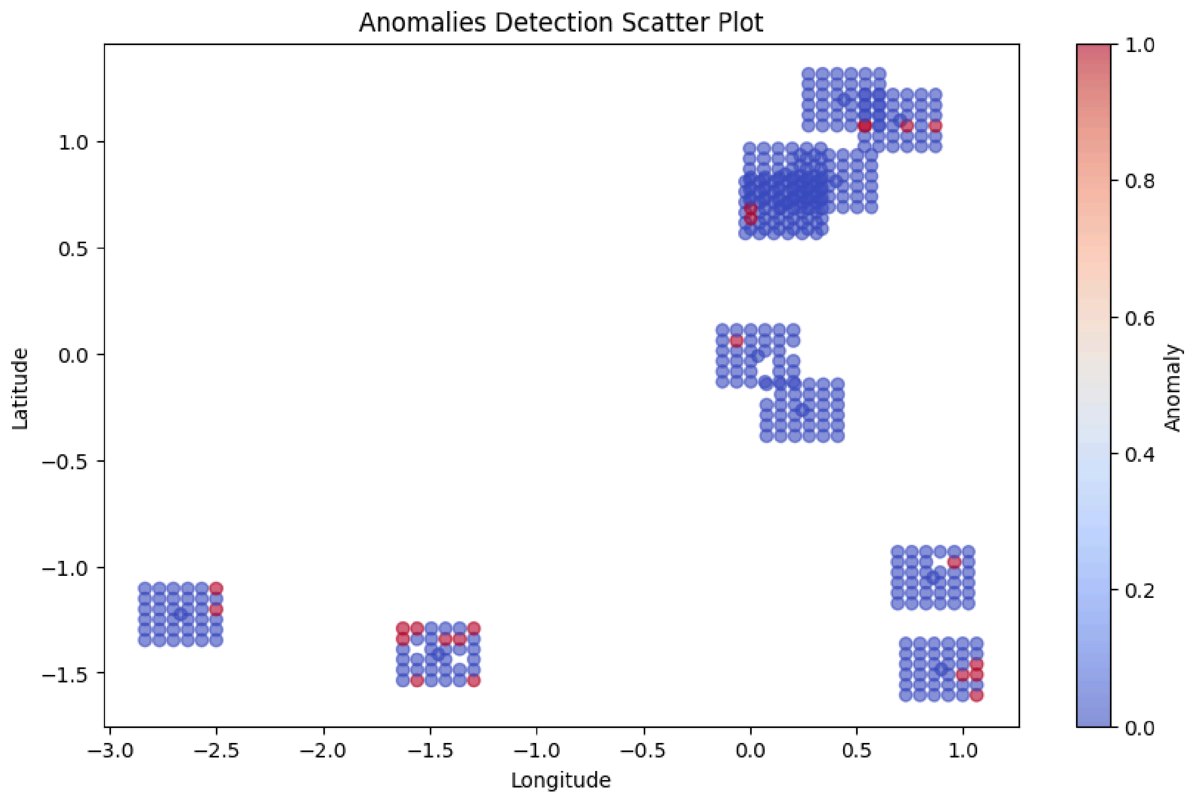


Figure 24: Anomaly detection algorithm generated in Matplotlib and applied to all twelve AOIs.

After filtering out these outliers, the performance of the anomaly detection algorithm improved to a silhouette score of 0.468 for the 2 clusters.

Discussion

Our results support our original hypothesis because we initially predicted that lower income and more impervious surfaces would lead to an increase in urban heat island risk.

Some errors may arise in our research due to not having all the GLOBE Observer photos to cross check the data points with. While there is a wealth of GLOBE Observer data available in Austin, not all of our AOIs could be matched with their corresponding ground photos, meaning that some of the impervious surface and greenness percentages used could not be validated. Additionally, the ECOSTRESS instrument can take LST measurements at different times of day, so the timing for every LST data point is not entirely consistent, and some LST data points are completely missing because of cloud cover or instrument malfunction when they were recorded. Lastly, once the LST data was included, we found that it had limited correlations with the other factors. This was an unexpected result that contradicted our hypothesis, in which we guessed that LST would strongly correlate with the other variables, which is why LST is not present in our graphs comparing our correlations.

In the future, we would like to expand this method to the entire state of California or select counties due to the area's vast differences in land surface temperature. LST measurements within Austin varied minimally, resulting in low correlations with other factors. The coastal areas and valleys of California and the different income brackets in Los Angeles may provide more diverse results, which would strengthen our correlations, benefit urban planning, and help contribute to future ideas on how to decrease the LST discrepancies between different areas of large cities. We would also want to use supervised learning and have a target variable in these future analyses.

Conclusion

Ultimately, the correlations that we discovered between land surface temperature, percentages of greenness and impervious surfaces, median household income, and UHI risk support the conclusion that socioeconomically disadvantaged areas, usually containing less vegetation and more concrete and asphalt, are on average at a much higher risk for extreme heat events resulting from urban heat islands than high-income areas of Austin. We found weak, yet statistically significant, negative correlations between median income, impervious surfaces, and land surface temperature, and weak positive correlations between median income and greenness (see Fig. 9). This means that as median income decreases in an area, the area typically has less vegetation, more concrete and asphalt, and higher temperatures than the rest of the city.

Our conclusion is somewhat contradictory in practice due to the fact that low-income neighborhoods struggle the most with urban heat islands but are the least equipped to protect themselves from their effects, leaving the most already vulnerable populations in Austin the most susceptible to heat-related hazards. By publishing this research, we hope to identify these disadvantaged areas so that urban planners can create targeted strategies on how to reduce urban heat island risk for the citizens living there.

Extreme heat can lead to the spread of diseases, cause heat stroke, and worsen chronic illnesses for those most affected (Temperature). It can also create drought and harm crops, vegetation, and local wildlife, creating a host of economic and ecological issues as well. As the planet warms exponentially, urban heat islands become even bigger threats to larger numbers of cities, meaning that their effects must be mitigated before the risk continues to spread. Our team calls upon local legislatures, urban planners, and other city officials to keep the variables described in this research in mind when making improvements to the city of Austin and other cities like it. Planting more vegetation that cools the surrounding area, keeping this vegetation healthy, and substituting concrete and asphalt for materials that better release the sun’s radiation could all be ways to keep people living in high-risk urban heat areas safe and improve the overall health of the city. Sustainability and climate-conscious initiatives aren’t just for the health of the planet; they are for the health of humans as well.

The experience of completing this research was extremely fulfilling and could not have been done without the help of our project mentors Grace Valdez, Peder Nelson, Erika Podest, Cassie Soeffing, Rusty Low, and Andrew Clark. There are many improvements that can be made on our methods (see “Discussion” for information on possible sources of error), including using more complete datasets, employing a wider range of satellites, recording more precise data about data collection times, and gathering more GLOBE photos, and we hope to see our work expanded in the near future based on the foundations we have set here.

Data Availability

All datasets used in this project can be found in our GitHub repository located at https://github.com/abhiramraju7/-Predicting-UHIs-with-ML .

Additionally, a copy of our video presentation briefly explaining our process shown at the SEES 2024 Virtual Science Symposium can be found at https://drive.google.com/file/d/1bxMNVqxsKrOWAQOOjjTbuoS7zNeVFGpI/view?usp=drivesdk .

View an explanation of the math process behind the Vincenity formula (see "Data Extraction" for further details) used in this project at https://drive.google.com/file/d/1biy8OAmMbeIUrtGraRH21Yi82tAtAx2n/view?usp=drivesdk .

All land cover data referenced in this paper is made publicly available by and courtesy of NASA, Collect Earth Online, and the GLOBE Observer program. The GLOBE Observer database can be found via https://observer.globe.gov/get-data/land-cover-data and the resources provided by Collect Earth Online are available at https://www.collect.earth/ .

Income data used in this project is courtesy of the United States Census Bureau and can be located at: https://data.census.gov/, https://www2.census.gov/programs-surveys/acs/summary_file/2015/data/5_year_by_state/, or https://www.kaggle.com/datasets/goldenoakresearch/us-household-income-stats-geo-locations/data.

Land surface temperature data from the ECOSTRESS instrument is sourced from NASA and requested through the AppEEARS app via NASA EarthData. These data can be accessed at https://urs.earthdata.nasa.gov .

Acknowledgements

The authors would like to acknowledge the support of the 2024 Earth System Explorers (ESE) Team, NASA Science Mentors, and ESE peer mentors. The authors would also like to personally acknowledge and thank Cassie Soeffing, Rusty Low, and Andrew Clark for organising the NASA SEES Earth System Explorers team and providing guidance throughout the duration of the research. NASA STEM Enhancement in the Earth Sciences (SEES) Virtual High School Internship program. The NASA Earth Science Education Collaborative leads Earth Explorers through an award to the Institute for Global Environmental Strategies, Arlington, VA (NASA Award NNX6AE28A). The SEES High School Summer Intern Program is led by the Texas Space Grant Consortium at the University of Texas at Austin (NASA Award NNX16AB89A0)

Contact

Anna Shifman
anna.shifman@yahoo.com
https://orcid.org/0009-0004-4292-8170

Abhiram Raju
abhiramraju7@gmail.com
https://orcid.org/0009-0002-2181-4743

Sophia Myers
sophiamyersdesigns@gmail.com
https://orcid.org/0009-0008-7134-1730

Hubery Pai
hubobopai@gmail.com
https://orcid.org/0009-0002-1409-1444

Noah Peralez
Noahperalez2025@gmail.com
https://orcid.org/0009-0003-2685-5322

Works Cited

ArcGIS online. arcgis.com. (n.d.). https://www.arcgis.com/index.html.

ArcGIS. (n.d.). LST data. Retrieved from https://igestrategies.maps.arcgis.com/home/item.html?id=8de7c6f3909d42bf901adae6894f6f5d.

ARSET - Satellite Remote Sensing for Measuring Urban Heat Islands and Constructing Heat Vulnerability Indices | NASA Applied Sciences. (2022, August 2). Appliedsciences.nasa.gov. https://appliedsciences.nasa.gov/get-involved/training/english/arset-satellite-remote-sensing-measuring-urban-heat-islands-and.

Austin, TX. Data USA. (n.d.). https://datausa.io/profile/geo/austin-tx/#:~:text=In%202022%2C%20Austin%2C%20TX%20had,%2486%2C556%2C%20a%209.61%25%20increase.

Bird, T. (2024, March 21). *The wildflower regions and vegetational areas of Texas*. Texas Highways. https://texashighways.com/outdoors/wildflowers/the-wildflower-regions-and-vegetational-areas-of-texas/.

CLIMATE AND EQUITY IMPERVIOUS SURFACES (%) IMPERVIOUS SURFACES (%) AND HEAT. (n.d.). https://affordablehousing.miami.edu/_assets/pdf/impervious.pdf.

Collect Earth Online - Satellite Image Viewing & Interpretation System. (2023, August 30). https://www.collect.earth/.

GLOBE Land Cover Data. GLOBE.gov. (n.d.). https://observer.globe.gov/get-data/land-cover-data.

Li Yang, Feng Qian, De-Xuan Song, Ke-Jia Zheng. Research on Urban Heat-Island Effect. Procedia Engineering, Volume 169, 2016, Pages 11-18, ISSN 1877-7058, https://www.sciencedirect.com/science/article/pii/S1877705816332039

NASA LP DAAC. (n.d.). Application for Extracting and Exploring Analysis Ready Samples (AppEEARS). Retrieved from https://appeears.earthdatacloud.nasa.gov/

SERVIR – a joint NASA and USAID program. (n.d.). Collect Earth Online. Retrieved from https://www.collect.earth/

Temperature-related Death and Illness. (n.d.). National Institute of Environmental Health Sciences. https://www.niehs.nih.gov/research/programs/climatechange/health_impacts/heat#:~:text=Loss%20of%20internal%20temperature%20control

(*Texas ecoregions*. Texas Ecoregions - Texas Parks & Wildlife Department. (2022, December 16). https://tpwd.texas.gov/education/hunter-education/online-course/wildlife-conservation/texas-ecoregions.

United States Census Bureau. (2022). Explore Census Data. United States Census Bureau. https://data.census.gov/.

U.S. Census Bureau. (n.d.). Retrieved from https://data.census.gov/

IVSS Badges

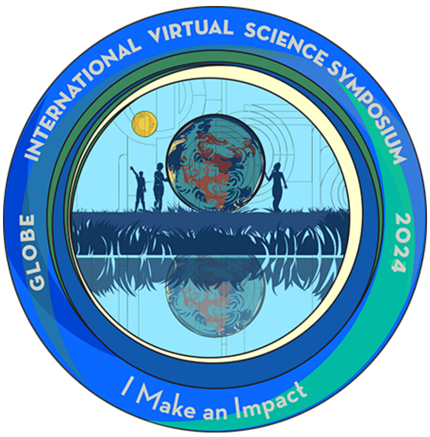
I WORK WITH A STEM PROFESSIONAL

In this research, we worked with scientist Dr. Erika Podest from NASA's Jet Propulsion Laboratory in Pasadena, California in order to make our research as accurate as possible. Dr. Podest is well versed in machine learning and data extraction from satellites - two subjects that we worked closely with in our project - and was able to give us valuable tips on how to make our data accessible and how to navigate the data having holes or being low quality. She gave us more information about the resolution and performance of ECOSTRESS and gave us a better perspective on how to properly prepare our research to present to the scientific community.

I AM A COLLABORATOR

We all collaborated as a team from different schools around the world. Our team consisted of Sophia Myers, Hubery Pai, Noah Peralez, Abhiram Raju, and Anna Shifman. The report goes into details of all the contributions of each member. Working with students from other schools helped our communication and teamwork skills as we communicated online and had to face many hurdles with time differences. We used platforms like Zoom to collaborate online. Additionally, using all of our diverse skills we have previously learned helped create a strong team as we all collectively brought something to the table.

I MAKE AN IMPACT



Our peer mentor, Grace Valdez, is an Austin, Texas native. She originally came up with our project idea when she observed the discrepancies between wealthy and poor neighbourhoods in her community and wanted to conduct research on how to bridge that gap effectively and improve the health of the environment. She noticed that low-income neighbourhoods were much less green and had more tightly packed housing than other areas of the city, so in the summer, the heat became extreme. As our team conducted research on the factors that contribute to urban heat island formation, we identified four variables - percent greenness, percent of impervious surfaces, land surface temperature, and median household income - that could be used to make recommendations on how to mitigate the effects of extreme heat in Austin (for example, planting more trees in low-income areas to provide more shade, especially in areas with a lot of housing).