Know the Signs: *Housing Choice and the Lived Environment*

Empowering consumers to make informed housing decisions by making environmental data centralized and accessible
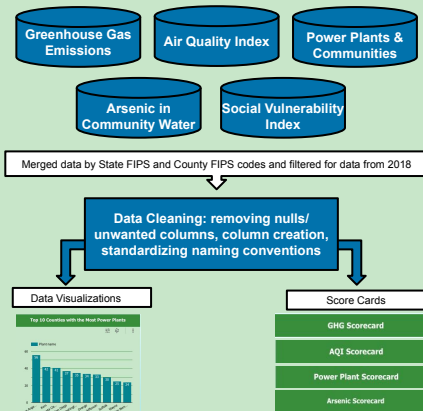
Team 61: Taylor Brown, Tinika McIntosh-Amouzouvi, Luis Esparza, Ashley Aviles, Chioma Dunkley

Motivation

When most think of the environment, they think of climate change and while that plays a role in the quality of life, more environmental factors can significantly influence our physical, social, and mental health.

Compiling pertinent environmental data and making it accessible for everyday consumers would allow them to make more informed decisions around housing.

Background

There is an absence of knowledge and transparency around environmental factors that have the ability to impact the health and well-being of homeowners and renters. This project aims to compile environmental data in order to create an environmental score for any given address in the United States and make it available on platforms for locating housing.

Data & Preprocessing

By looking into data sources from the Environmental Protection Agency and the Centers for Disease Control, this project examines environmental factors by state, county, and social vulnerability index.

Merged data by State FIPS and County FIPS codes and filtered for data from 2018

Data Cleaning: removing nulls/ unwanted columns, column creation, standardizing naming conventions

Data Visualizations

Score Cards

GHG Scorecard

AQI Scorecard

Power Plant Scorecard

Arsenic Scorecard

Data Insights

By leveraging our data sets from the EPA and CDC we created a clearer picture of the state of multiple environmental factors across the United States in the year 2018. With insights from our mentors and TAs, we looked to find data at the county level. We were able to find data on the greenhouse gas emissions, air quality index, Social Vulnerability Index, and Arsenic levels in community water. To understand our data better, we did a series of analysis including plotting to see which counties had the highest cancer prevalence and seeing if there were any correlations between our environmental factors and our health factors. While we could not find a correlation between our environmental factors and our health factor using the statistic models we learned in class. We believe that with more advanced statistical modelling we could start to identify key trends. Nevertheless, we were able to hone our data visualization skills in both python, Google Data Studio, and Tableau to visualize some geographicals trends within our dataset. We have showcased two of them down below:

Cancer Prevalence by County
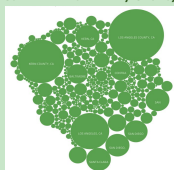
Total Greenhouse Gas Emissions by County
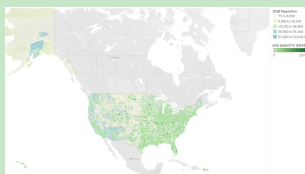
Additional Notes/ Recommendations:

Assumptions/Limitations to scorecard

The duration of the program was not sufficient for a team of novice data scientists to collect, curate, and analyze the data that was needed for the creation of the intended scorecard. However, we were able to do a great deal of important foundational work as well as made many discoveries about the data that would serve as a great starting point for a future project.

Although we limited the scope of this project to five main features impacting the prevalence of cancer, we realized that so many other factors can contribute to cancer outcomes outside of these features and outside of environmental factors alone. Therefore, it would be most ideal to create a score based on the solely on the impact of environmental factors on the quality of life. While it would be difficult to isolate the impact of these features, this score would be the most beneficial to those in the housing market as this score would allow a better understanding of the environmental around the residence that they are considering.

Proposed Solution

Scorecard

With insights from our mentors and TAs, we wanted to make a environmental scorecard where a user could input a county and see how it ranks with a score from 1 to 5. We cycled through different options to create the ranking system but ultimately decided to rank each county using the calculated quintile for each of our values. For our current model, we weighted each one equally to create an average overall score. However, we would like to note that the model can be modified weight values with a greater impact to a housing decision higher than others. Below you will see our proposed dashboard, which will display our given score for each our parameters and an overall score for the county.

We are currently still working to see if we can progress our model and create a predictive scorecard that could correlate a specific environmental factor to the likelihood of cancer prevalence.