Active Living Condition of Madison, WI

Sarat Sagaram

The Active Living Condition Index is an indicator of the level of activity in a certain region. In general, it is important that regions have a high active living condition, as it proves to be a good indicator of the overall health of that region. This is influenced by many factors, including access to bikes, transit, population density, and others. In my report, I intend to analyze which factors contribute highly to active living conditions. In addition, I will use machine learning to predict the active living condition based on the factors present. I hope this report can help places understand and improve aspects that can increase their active living index scores.

The dataset was gathered from the ArcGIS Hub, which contains many data hubs. The one I used is from the City of Madison, found here: https://data-cityofmadison.opendata.arcgis.com/datasets/active-living-index-composite-score?geometry=-89.913%2C42.893%2C-88.933%2C43.245&page=8
The information I utilized was Bicycle Facilities (1), Transit Access to Jobs (2), Intersection density (3), Population Density (4), Transit Service (5), which I believed could be the best features to predict Active Living Index Score(ALC Score). The first bucket of rows had many missing values, as well as many 0 values that spanned across several columns, which I removed to avoid many null errors.

In order to understand the relationship between a few variables, I started the analysis by plotting a line graph for some of the features, including Bike Score and Transit Service versus ALC Scores. In **Figure 1**, we see that as both the Bike Score and Transit Scores increase, their respective ALC Scores increase as well, spanning throughout the dataset. This means we can conclude that, in general, as Bike Scores and Transit Scores increase, ALC Scores increase as well, suggesting a positive correlation.

Next, **Figure 2** is a KMeans Analysis plot, attempting to understand the relationship between the number of clusters and the inertia of our dataset. Moreover, it is beneficial to determine how many distinct clusters there are in the data, grouping them into categories for further analysis. In the graph, we can see that after around 5 clusters, the inertia seems to level off and plateau. This plot is a good indicator that our data can be categorized into five main clusters, further classifying the dataset.

Finally, **Figure 3** visualizes the coefficients of a linear regression model between each of the features listed above and the ALC Score. The purpose of this conclusive graph is to use supervised machine learning to predict ALC Scores based on the features present. The full model is a pipeline containing a Polynomial Features component on Bike Score, Transit Score, and Inner Score, and a linear regressor component. We create a training/test split on our data (75/25 split), and fit the model around the training data frame. We then score the model around the test data frame, giving us a score of around 82%.

The coefficients from the model above display which features make the biggest impact on the ALC Score. As we can see, the features of Transit Access to Jobs and Population Density make the most significant impact on ALC Score compared to the rest of the features. This is a very important finding, as the data suggests that regions with higher transit access to jobs scores and population densities tend to have a higher active living condition index score. It is important to note that transit access to jobs and population density seem to go together, meaning that not all regions with high population densities will have a high ALC score.

In conclusion, our analysis ultimately suggests that active living conditions are affected by many factors, most notably transit access to jobs and population density. In addition, features like bike score and transit score only aid the active living condition score, as they are positively correlated. With these suggestions, regions can understand what affects the active living conditions of their area and make strides to improve them, which can better the overall health of the community. More research needs to be done, though, as we continue to search for a healthier, safer planet.

Figure 1: Bike and Transit Scores versus their ALC Score

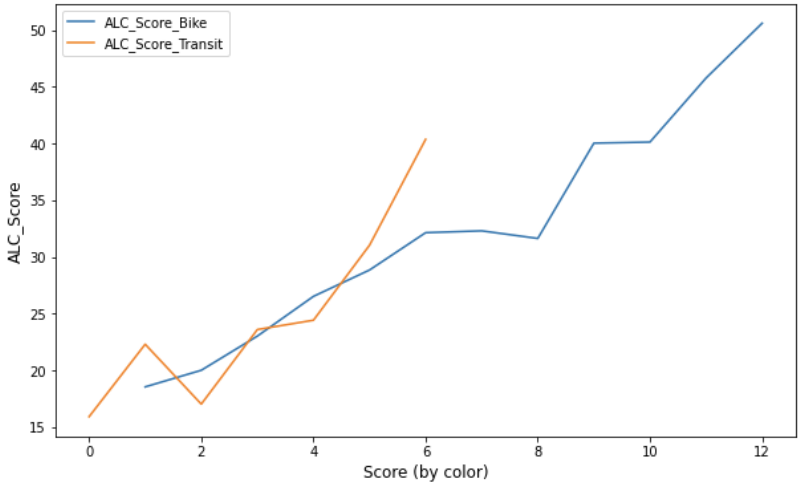

Figure 2: Cumulative Inertia through KMeans Analysis

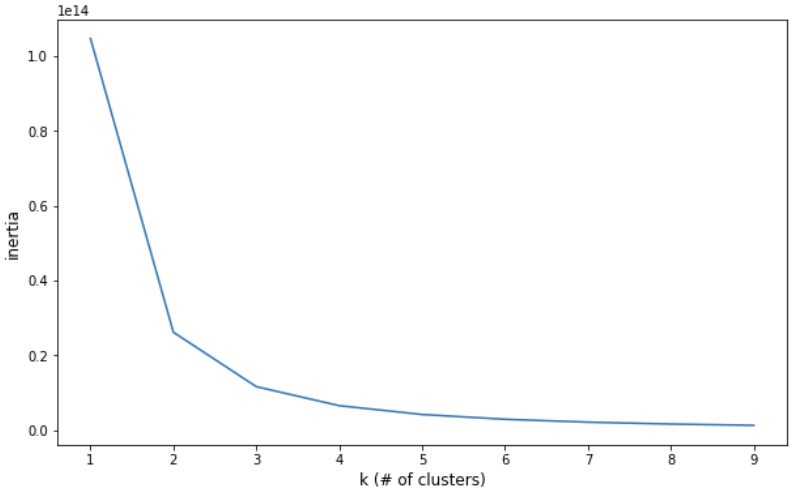

Figure 3: Coefficients from Supervised Pipeline Model