Walkability Analysis

Simon Kaminski

Mesa Community College

CIS 480

Dr. Kazuaki Kashihara

November 21, 2025

Introduction

Project Purpose

The primary objective of this project was to gather both conventional and geospatial data regarding the topic of car dependency indicators, such as cars per household, average commute time to work, location affordability, etc, and determine the level of dependence a certain area has on vehicles. Questions asked through this analysis are focused on answering which neighborhood-level factors, such as job accessibility, walkability, transit availability, and housing affordability, are the strongest predictors of car dependency. Answering these questions will help to understand and predict future car dependency using current measures of transportation infrastructure, public transit access, and quality-of-life indicators. These questions will also help to answer what neighborhoods are projected to experience the greatest increases in car dependency if housing development continues without corresponding investment in transit infrastructure.

Initial Hypothesis

My initial hypothesis for this analysis was that improved quality of life factors in a given area and improved public transportation would reduce the need for vehicular dependence in a city. I also believed that an increased population density may contribute to lower car dependency as more tightly packed areas would possibly mean people would walk or bike to places more frequently. This is something I took into consideration given cities such as NYC from personal experience as I found it was easier to walk in more densely populated areas than it would be to drive. I also found that certain areas have better public transit setups than other areas which I also believed would cause a decrease in car dependency

Project Scope

The scope of this project varied greatly from the initial planning phase. Originally, the plan was to collect data for each state based on 'Census Tract' geospatial data. The final analysis was limited to a set of states at the block level. The reasoning behind why this project was limited to a certain number of states is addressed later on in this report, a simple overview as to why was regarding the method of collection and the limitations that I ran into with PolicyMap. The reason for switching from Census Track to block-level was that it integrated better with certain data points that were only collectable at the block level. I did not lose any granularity in my analysis as a result of this change and feel that this is a better choice over Census Tract.

Data Collection and preparation

For the data collection portion of my project, I focused on using PolicyMap as my main source for data. PolicyMap was my main choice as it would allow me to collect many points of data for quality of life factors, general road infrastructure, public transit metrics, etc. PolicyMap also gave me geospatial data from a block-level scope for the year 2021. A limitation that I overlooked by only using PolicyMap was that many points of data that I would use in this analysis were limited to 2021, which meant that I was not able to perform a time series analysis as I originally planned on.

Each point of data that I collected from PolicyMap was provided to me in a CSV format. Each file that I collected would have to be combined into a singular dataset, which I would be able to use for further analysis and visualization. This step of data collection and cleaning took a long time for me, as it was much more complex that I originally planned it out to be. I started by loading each CSV file into a python script which would perform various cleaning operations such as converting numeric types, removing or imputing missing values, depending on how I saw fit, and removing or renaming columns. I did lose some information through the process of cleaning, but I believe that the idea of having data of higher quality outweighs having a large amount of unusable data. I created a function whose purpose was to reshape the data using Pandas 'melt' method. After that, I removed rows with missing values and pivoted the melted DataFrame to allow each category to become its own column. These operations would give me a cleaned DataFrame in a wide format, in which each row contains columns with separate categories.

Model Selection

The model I chose to train in my analysis was a Linear Regression model and a RandomForest Regressor. The reason for training a Linear Regression model was that I thought it would do a good job when dealing with continuous outcomes. The other reasons why I chose this model were due to its simplicity and available metrics libraries that I can use to determine and improve the model's accuracy. I made use of the RandomForestRegressor model to understand the importance of different features as it relates to walkability. This is important as some of the factors, such as distance to work or transit, do not have a linear relationship, which I could capture using this model.

Analysis Findings and Implications

The main purpose behind completing this analysis is to highlight and analyze factors that have an affect on walkability in certain areas. These factors have a significant practical implication, particularly in supporting healthier, more connected and economically vibrant cities. This analysis provides actionable insights into how changes in the built environment, such as sidewalk connectivity, street design, and pedestrian amenities, directly influence people's daily walking, public transportation habits. There are many findings that show that an improved walkability in an area can boost general physical activity and reduce reliance on cars, which in turn helps reduce air pollution. Research shows that many people do not participate in enough physical activity, "An increasing portion of the population, including many children, lack regular physical activity. Although there are many ways to be physically active, walking is one of the most practical ways to increase physical activity among a broad population." (Litman, 2024) One great way to improve the amount people are able to achieve their daily exercise goal would be to improve road infrastructures to promote walking, or at the very least improve public transportation access. The following sections define and introduce the connections between this analysis and different societal issues which it covers

Transportation

Improving walkability creates alternatives to car travel which can mean an improvement in cost savings for many individuals, reduced traffic congestion, and better access to public transit. Well connected and maintained pedestrian infrastructure allows for much safer travel and increases the reach of transit systems

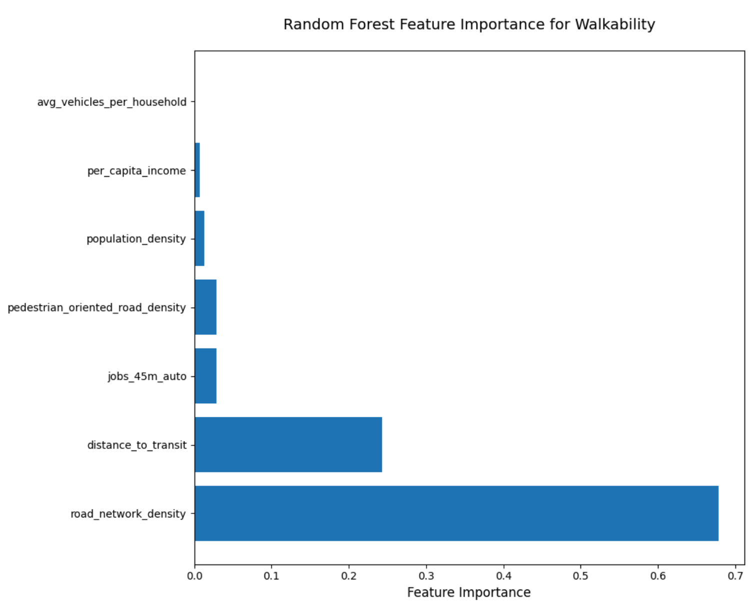
Public Policy

From the findings in this analysis, policymakers and urban planners are better equipped to advocate for pedestrian-friendly investments. Cities can use this research to help justify policies that promote compact, mixed-use development and measure the success of interventions, such as sidewalk improvements.

Health and Environment

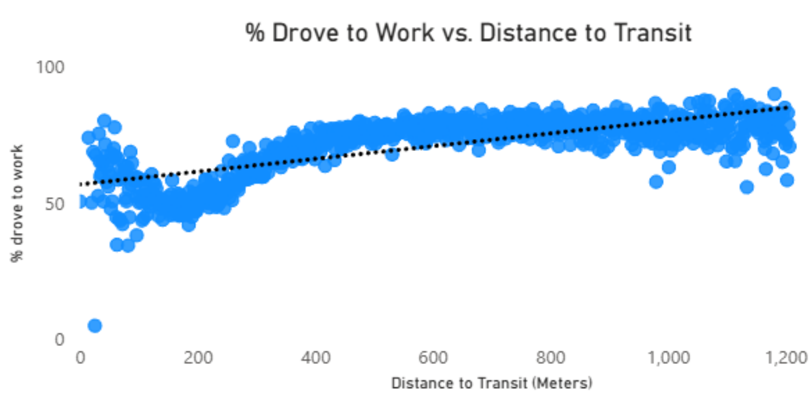
More walkable communities lead to an increase in physical activity, which means generally better public health. This also comes with the benefit of reducing pollution. This is also supportive of small businesses as more people can access local businesses by means of walking a short distance

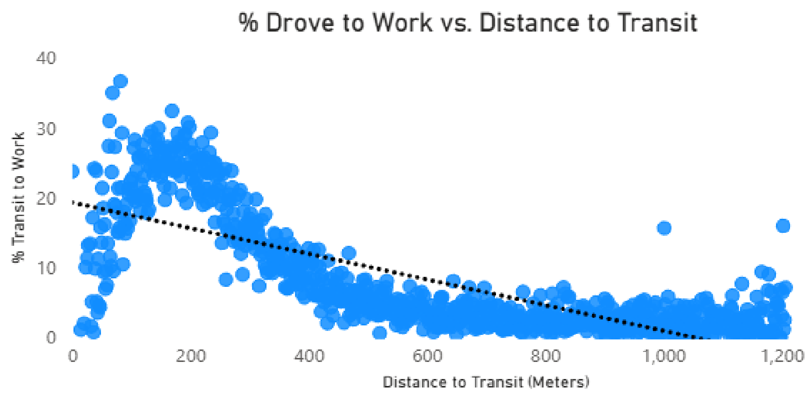
This analysis brought to attention four main factors that had an effect on the walkability of a given block group. These factors, from most important to least, include: road network density, number of jobs within 45 minutes by car, population density, and pedestrian-oriented road density. The most interesting factor to look at was road network density, as it held the most importance out of each factor.



I initially assumed that population density and per capita income would have a larger effect on the walkability of an area. In hindsight, this outcome makes sense as the walkability of an area is affected by the urban planning and design, such as street connectivity, sidewalks, and general distance to goods. Higher-density areas can still be unwalkable if they have wide streets, few intersections, and a lack of sidewalks. Income is also a factor I assumed would have a higher impact on walkability, as I assumed there would be more contributions made to improve general road infrastructure to possibly support walkability. This assumption was incorrect, as it assumes that everyone prioritizes

Something else interesting that I found from my visualizations was the relationship between the percentage of people who drove to work compared to the distance to public transit stations. The two graphs attached below show the trends found from the data. It seems that when a public transit station is around 100-300 meters away from a block group, people are more incentivised to commute to work using public transit rather than driving. This does not necessarily mean that distance to public transit is the main factor for this, as things like the average distance to work, affordability of transportation, and various other quality of life factors.





Research Questions

My analysis attempted to answer the following research questions, each of which attempt to understand different factors that contribute to walkability and public transportation access in areas as it relates to car dependency

Does longer commute time increase car dependency?

It was found that longer commute time does lead to an increase in car dependency. Every state that was covered in this analysis saw an over 50% reliance on cars to get to work regardless of commute time, but a general trend showed that as there are more jobs with a longer commute time, people generally tend to commute using their own cars. This result supports my hypothesis in which I believe that longer commute times will lead to an increase in car dependency

Does better public transit access reduce car dependency?

To answer this question I looked into the distance to transit compared to the percent of people who commute to work by car. I found that if a transit stop is near 200-300 meters of a block group, the amount of people who commute to work by transit increases. Larger distances to transit stops show a much higher dependency on cars. This supports my hypothesis as I believed that providing an improved infrastructure for public transit will lead to a decrease in car dependency

Does proximity to jobs reduce car dependency?

Similar to my first research question, it was found that the proximity to jobs (i.e. jobs greater than 45 minutes away) has an increase on car dependency for a given block group. This supports my partially supported hypothesis. Although I did see that being in further proximity to jobs had an increase on car dependency, I believe that the impact I saw came from other factors such as distance to public transit stations and population density

Does walkability decrease car dependency?

The research shows that as an area's walkability index increases closer to 20, the percentage of people who drive to work decreases. Other factors such as population density which can affect walkability also play a role in this calculation but, in general, it was found that quality of life factors that support walkability in a given area will decrease the dependency on cars.

Does greater road infrastructure increase car dependency?

Improved road infrastructure showed to have a direct impact on car dependency. The road infrastructure, represented by road network density, which is calculated from PolicyMap by dividing the number of roadway links from intersection to intersection by the land area of the block group. The results from the analysis show that as road network density increases, the percentage of people who drive to work shows a decline, whereas more people choose to commute using public transportation.

Do quality-of-life factors (walkability, affordability, transit) interact to influence car dependency?

It was found that quality of life factors such as walkability, affordability and transit accessibility were main factors that interacted with car dependency. It was found that as an area's walkability index approached 20, the highest score, the percentage of people who drove to work decreased. Other factors such as population density which can affect walkability also play a role in this calculation but, in general, it was found that quality of life factors that support walkability in a given area will decrease the dependency on cars.

Project Improvements

To continue working on this project, I would like to collect more data points from each state rather than just three. Doing so would allow me to get more information regarding geographic factors as to how different areas of the U.S provide more walkable areas. An issue that I had during the development of this analysis was understanding the way I would have liked my dataset to be structured. With the way that I was allowed to collect data from PolicyMap, I ran into an issue in which the process of collecting points for each layer would be slow. Getting and preparing the 13 layers per state took me a long time, as I had to navigate through the web interface and download them one by one. Repeating this process for all 50 states would have taken much more time than I had planned for the data collection phase of this project, so I decided to stick with the three states I have in the analysis. My method for combining the data is very easily reproducible for any new states for which I download the data, which would be my goal in the future. If I had spent more time on the data collection phase, I would have liked to understand how I could use the PolicyMap API to possibly download all of the data in an automated way. This improvement in data collection would allow me to get all of the data I need for each of the states, which, for this project, would be very beneficial as I have more points of data to create an analysis from.

References

Complete Streets Case Studies. (n.d.). United States Department of Transportation. https://highways.dot.gov/complete-streets/complete-streets-case-studies

Litman, T. (2024, December 16). *Economic Value of Walkability*. Victoria Transport Policy Institute. https://www.vtpi.org/walkability.pdf

Randall, S. (n.d.). *Walkability Analysis for Improvement* [Prioritizing Potential Walkability Projects in a High-Vacancy Environment].

Zhu, X., Lu, Z., Yu, C.-Y., Lee, C., & Mann, G. (2020, Janurary 20). *Walkable communities: Impacts on residents' physical and social health: Researchers from Texas A&M University studied residents in a newly developed 'walkable community' in Austin, Texas to see how it changed their habits for physical activity and ...* PubMed Central. https://pmc.ncbi.nlm.nih.gov/articles/PMC8776244/