**Exploring Walkability and the Need for a Car in Select American Suburbs**

**Introduction:**

A couple wants to move outside of city limits. There are many advantages of urban living that the couple enjoys; one of them being the lack of need for a car. Walkability and public transit availability are important factors when determining where to live along with having a large number of recreational venues nearby. The couple has several children and would like to live in a more suburban environment but still would rather not have a car. What American suburbs would be best for the couple to move to?

**Data:**

Foursquare API: Venue specific types around top 108 cities calculate the average distance of city center

* recreational venues (bars, restaurants, coffee shops) – distance in m from city center
* grocery stores (groceries, market etc.) – distance in m from city center
* job venues (business buildings, offices etc.) – distance in m from city center

Data.gov:

* top 108 cities
  + walkability score (1-100)
  + bikeability score (1-100)
  + population

Walkability, bikeability, and public transit scores for American cities are available for free on the top 108 cities but are not available for American suburbs. Using data from Foursquare API and the walkability data available online at data.gov we can develop a model that accurately predicts whether a resident would need a car or no car to live in that city. Foursquare data of specific venue types: recreational venues (bars, restaurants, coffee shops), grocery stores (groceries, market etc.), and job venues (business buildings, offices etc.) will be used as features to predict a no car binary classification. The no car binary classification is the target variable and is calculated by using a city's walkability and bikeability score and averaging them. Then taking only the top tier and classifying them as being cities that you do not need a car. A list of America's top suburbs is listed and the Foursquare API data will be leveraged to predict whether or not these suburbs are walkable or bikeable enough to warrant not needing a car to live in.

**Methodology:**

To determine whether it is possible to live in a city/town without a car it must be walkable and/or bikeable. Data from data.gov regarding walkability and bikeability scores on a scale of 1-100 is only available for large cities. In order to obtain whether or not a car was necessary to have for the particular city I first averaged bikeability and walkability scores and plotted them in a histogram. There is a clear separation of the upper tier at a value of 65. This is used as a cutoff of determining whether a car is necessary or not necessary to live in a city.

This problem is a classification problem and can be solved by leveraging foursquare venue data of the specific suburbs in America. In order to live in a city the venues that must be available nearby are grocery stores, jobs, as well as recreational or entertainment venues. I looped through the top 108 cities that had walkability and bikeability scores and used Foursquare API data for each of the cities to calculate the average distance and count of specific venues. The specific venue ID’s were grouped into recreational venues (bars, restaurants, coffee shops), grocery stores (groceries, market etc.), and job venues (business buildings, offices etc.). The classification theory is that the smaller the average distance was for each of these the more walkable or bikeable the city may be. Using the job venue average distance as features and the no\_car binary classification as the target variable I explored various types of classification algorithms.

**Results:**

The data from 108 American cities were split to testing and training data to determine the most accurate algorithm in the problem. The algorithms explored were the following: K-nearest neighbors, decision tree, support vector machine, and logistic regression. Accuracy was calculated using the f1 score statistic and was highest in the K-nearest neighbors with a testing algorithm of 83%. A list of 15 American suburbs were given and the average distance of grocery venues, work venues, recreational venues, and the number of recreational venues were calculated using the Foursquare API. The trained k-nearest neighbors model was applied to the suburb data and only two suburbs were deemed live-able without a car: Naperville, IL and Arlington, VA.

**Discussion**:

Out of the 15 suburbs selected it is not too surprising that only 2 of them were deemed live-able without a car. Most American suburbs are by nature more “spread out” than cities. Though the calculation of average distances of specific venues has the correct idea, I noticed that not all of the venue ID’s fell within each of the selected features. Venues included in the calculation are only those venues that are on Foursquare and may not be representative of the entire suburb or city as Foursquare registration may be more skewed towards specific types of venues that are more advanced in terms of social media presence and tech presence in general. It is also not surprising that the top two suburbs are not far from extremely large cities Chicago, IL and Washington, D.C.

**Conclusion:**

The couple that is searching for a city that they may not need a car is limited to Naperville, IL and Arlington, VA. Naperville has significantly more recreational venues than Arlington and because of this should be the choice for the couple to begin exploring. A map of the Foursquare venues are shown in the notebook and can help the couple decide where to begin looking for a house. There are a large amount of additional variables that will need to be looked at before the couple may be satisfied, but they are well on their way after discovering which suburbs they can live in without a car. 