**Title**

Ryan Eudy

March 16, 2020

1. **Introduction**

Over the course of the last decade food trucks have become increasingly popular across the United States, including in Arlington County, Virginia a suburb of Washington D.C., where the number of food trucks has seemingly increase greatly in recent years. The mobility of food trucks is one of the reasons they are such a popular alternative to brick and mortar restaurants, as they are able to move location with ease to reach more patrons in different markets. However, as the number of food trucks on the road has greatly increase it has become not too uncommon to see many food trucks crowding a single location, battling over prime real estate. Some popular neighborhoods in Arlington experience this as food trucks gravitate to them. For food trucks profitable locations are critical to their success and it is often not easy to branch out and discover new viable locations making known locations more desirable.

This project is aimed towards food truck owners or managers in Arlington County, Virginia in order to explore what features successful neighborhoods for food trucks have and if there are any potential neighborhoods with similar feature and without any current food truck venues that would be make a viable option to explore business in.

1. **Data** 
   1. **Data sources**

Arlington County’s website <https://arlingtonva.us> has a projects section that includes data and research for topics such as Arlington’s demographics by geography. Within the provide data Arlington County breaks down demographic information from the 2010 U.S. Census by neighborhood or civic association. There are 61 pdfs documents containing the demographic break down, one for each of the neighborhoods determined by Arlington’s projects and planning. Within the pdf documents the demographic data includes total population of the neighborhood, population broken down on race, age and sex, total households, households broken down by type and size and lastly the total housing units and rather they are rented or owned and occupied or not. Overall from this data source a list Arlington neighborhood as well as their demographic makeup can be complied.

Geopy <https://geopy.readthedocs.io/en/stable/> is a python library which enables geographic coordinates to be attained from addresses, cities, countries, and landmarks. This library will be utilized in order to coordinates of each of the Arlington neighborhoods.

Four Square <https://foursquare.com> is a social networking service which provides an API with a feature that allows developers to find information about venues listed in their data that are nearby geographic coordinates. The API will be used to find venues that are nearby each neighborhood in Arlington in order provide data on the venue composition of them.

* 1. **Data cleaning**

Pdf files downloaded from the Arlington County’s website were manually converted into csv files by importing them into excel and saving them into a data directory in order to be ingested by pandas.

A demographics by neighborhood data frame was then created by reading each csv file in the directory, extracting the neighborhood percentage column and converting it into a single row with the neighborhood name as the index and demographics as columns.

Columns involving race, sex, single parent household, and person per household were dropped from the demographic by neighborhood data frame. Additionally, all percentages and totals were converted to floats and integers respectively.

A neighborhood location data frame was created using the geopy library however, some locations could not be found by the library or were found to be incorrect. The missing or incorrect data was manually patch using google maps with the expectation of several location that did not have clear coordinates in google and therefore were dropped from the data frame.

The location data was utilized to create a data frame of venues in Four Square nearby each of the neighborhoods. The venue category was found to be too unique for the project purposes with a single occurrence of many of the venue categories near any of the neighborhoods. In order to generalize the venue categories, the Four Square documentation was scrape for the general grouping of each category which would become the venue category. Miscellaneous categories that were not grouped into a more general category in the Four Square documentation were dropped from the data frame.

The venue data frame was then one-hot encoded by the new general venue categories. Once one-hot-encoded the data was then grouped by neighborhood taking the frequency of each venue category occurrences.

Food truck venues data was pulled from the nearby venue data frame and group by neighborhood in order to sum up the total nearby food trucks for each neighborhood.

Lastly the four data frames demographics by neighborhood, neighborhood location, frequency of venue categories by neighborhood and total food truck by neighborhood were inner joined on neighborhood, meaning if a neighborhood was missing form any data frame it was dropped.

1. **Methodology**

The main effort of this project is to identify new markets or neighborhoods to operate a food truck venue in Arlington. We can define a new market as a neighborhood without any nearby food trucks according to the Four Square data.

For the purpose of this project we will consider all known food truck venue locations in Four Square as successful locations. Therefore, we will want to determine what attributes successful locations have and if there are other neighborhoods within Arlington with similar attributes that lack any nearby food trucks.

Utilizing the collected data, the neighborhoods will be cluster with the Kmeans algorithm into common groups base on their attributes. The number of clusters will be determined by the elbow method of plotting the number of clusters against the inertia of each clustering.

The cluster of neighborhoods that has the majority of the known food truck venues nearby can be determined to have the most promising attributes of a successful food truck market. This cluster can then be explored in order to give insight on what may make a good food truck location.

Finally, neighborhoods within the identified cluster without any nearby food trucks can be determined to be potential successful new markets.

1. **Results & Discussion**

According to the Arlington County website there are 61 demographically distinct neighborhoods within Arlington as of the 2010 Census.

However, four of the neighborhoods were not consider in the analysis due to the lack of coordinate locations, Cherry Hill Nature Area, John M Longston, Leeway Overlee, and Tara Leeway Heights.

Additionally, another five neighborhoods were ruled out for having no nearby venues according to the Four Square data, Alcova heights, Ashston Heights, Dover Crystal, Old Glebe and Donaldson Run.

Of the remaining 52 neighborhoods analyzed 6 were found to have nearby food trucks venues leaving only a possible 46 new locations for food trucks.

Utilizing the elbow method, it was concluded that 5 clusters would be an appropriate number in order to cluster the 52 neighborhoods into common groups.

Upon examining the resulting cluster, we can observer that 7 out of the 12 food truck venues are located nearby cluster 5. While this is just over half of the food truck venues found by Four Square the next highest nearby count is cluster 0 and 1 with only 2 nearby food trucks venues. Therefore, it can be concluded cluster 5 contains preferable attributes for food truck venues.

Cluster number 4 contains 5 out of the 52 neighborhoods clustered, Bluemont, Buckingham, Colonial Village, Lyon Park and North Rosslyn. However out of these 5 neighborhoods Colonial Village and Lyon Park already have nearby food truck venues leaving us with 3 possible new locations that hold similar attributes to preferred known food truck locations, Bluemont, Buckingham and North Rosslyn.

Upon further examination of Cluster 5’s attributes we can see that the cluster has a high mean of percentage of young adult ages 18 to 34 living within the neighborhoods as well as the majority of housing being non-family households. Lastly, we can observer that the cluster has a high frequency of shopping and food venues.

1. **Conclusion**

The purpose of this project was to explore Arlington County neighborhoods in order to find potential new food truck locations with favorable markets. By taking known food truck venue locations from Four Square we were able to take neighborhoods with nearby food truck venues and find other neighborhoods with similar demographics and venues attributes. The results of this was a narrowed list of neighborhoods with many of the same features of neighborhoods with nearby food trucks. Additionally, we were able to observer that food trucks tended to be located near neighborhoods where the majority of population comprised of young adults, living in non-family housing.

The Final decision was that Bluemont, Buckingham and North Rosslyn would provide the most preferable conditions for expanding food truck operations into those neighborhoods.