**A New Shopping Center in Raleigh, NC**

**Introduction Problem Description**

Raleigh is one of the fastest-growing cities in the United States with a growth rate of about 3.4% per year. A Forbes study last year found the Raleigh metropolitan area was the fastest growing in the United States, increasing 47% from 2000 to 2012. Between 2000 and 2009, the Raleigh-Durham-Cary metropolitan area grew 40%, leading the nation. The three-year change in the median household income is 12.37%. It is predicted that Raleigh will remain the fastest-growing metro area through 2025.

The increasing population and income boost the need for more shopping centers. The objective of this project is using advanced data science methodology and machine learning techniques to find the best location in the city of Raleigh to open a new shopping center.

**Data**

To complete this project, we will need the datasets below:

* list of neighborhoods in the city of Raleigh
* Geographical coordinates (latitude and longitude) of these neighborhoods
* FourSquare API that provides the surrounding venues of given coordinates of these neighborhoods

**Methodology**

* Get a list of neighborhoods in Raleigh, NC from Wikipedia (<https://en.wikipedia.org/wiki/Raleigh,_North_Carolina_neighborhoods>)
* Web scraping using Python requests and beautifulsoup packages to extract the data
* Get the geographical coordinates (latitude and longitude) using geocoder package
* Use pandas and folium package to visualize the neighborhoods in a map
* With the geographical coordinates, use Foursquare API to get the top 100 venues that are with in a radius of 2km, and venue data such as name, category, and coordinates will be gathered.
* Group venues by neighborhoods and count the frequency of each venue category.
* Analyze data using K-means clustering to analyze the distribution of shopping centers to help us decide where to open a new shopping center

**Results**

**Discussion**

**Conclusion**