**Introduction**

**Business Problem**

Atlanta is the capital and most populous city of the U.S. state of Georgia. The city serves as the cultural and economic center of the Atlanta metropolitan area, home to more than 6 million people and the ninth-largest metropolitan area in the nation. It ranks in the top twenty among world cities and 10th in the nation with a gross domestic product (GDP) of $385 billion. Atlanta's economy is considered diverse, with dominant sectors that include aerospace, transportation, logistics, professional and business services, media operations, medical services, and information technology.

Every year, people from all over the country move to Atlanta, GA. It’s on the list as one of the top-ranked inbound states in the country for several years. With diverse, vibrant communities, a booming economy, and plenty of green spaces everyone from millennials to retirees are moving to Atlanta from all over the country. The city is home to promising startups and business opportunities, excellent rates of job growth, and the most sought-after neighborhoods for families. According to researchers, the city is expected to see significant population growth over the next two decades. Atlanta is poised to be the next center of the digital age with the city ranked #4 among Venture Capitalists and is on pace to become the nation’s hottest startup venue. All of these would make Atlanta to be a sought-after destination for job seekers.

The affordable housing cost combined with promising career paths has people flocking to the metro area. These factors lead to an increase in demand for real estate properties. With over 29 counties and more than 140 cities in Metropolitan area, it makes it very difficult for anybody looking to move to Atlanta and choose a city to invest for buying a house. When people are trying to buy houses, they need to consider lot of factors right from location of the neighborhood to amenities such as grocery stores, shops, and restaurants. Schools are another important amenity as the quality of local schools is an important factor in determining the resale value of the house. These problems led me to analyze on the housing prices and help numerous people planning to move and explore venues in Atlanta.

The target audience and stakeholders for my project would be anybody wanting to buy a house in Atlanta's metropolitan cities. This could be individual families or real estate agents who could use my project analysis results for initial analysis and to better connect with prospective home buyers. This could also be used by people visiting Atlanta to get familiar with the different venues to visit. I would leverage foursquare to get location neighborhood data around required cities targeted by the audience. My analysis would segment the cities into different clusters based on price, neighborhood amenities and venues. Prospective buyers could narrow their search and focus on a selected few cluster of cities and make their house buying research simpler and more effective.

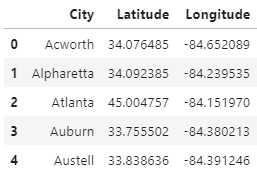
**Data Collection, Cleaning and Preparation**

I will need the following different types of data for my analysis.

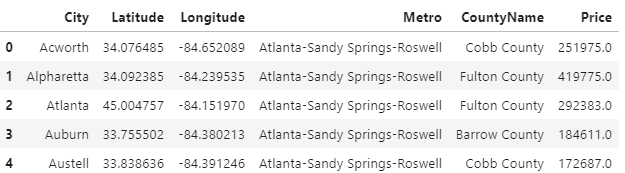
1. I will need to get a list of cities around metro Atlanta. For this, I will use BeautifulSoup to Scrape data from the web page <http://atlanta.metroguide.com/list_of_cities.htm>. Here is a screenshot of the first 5 cities.



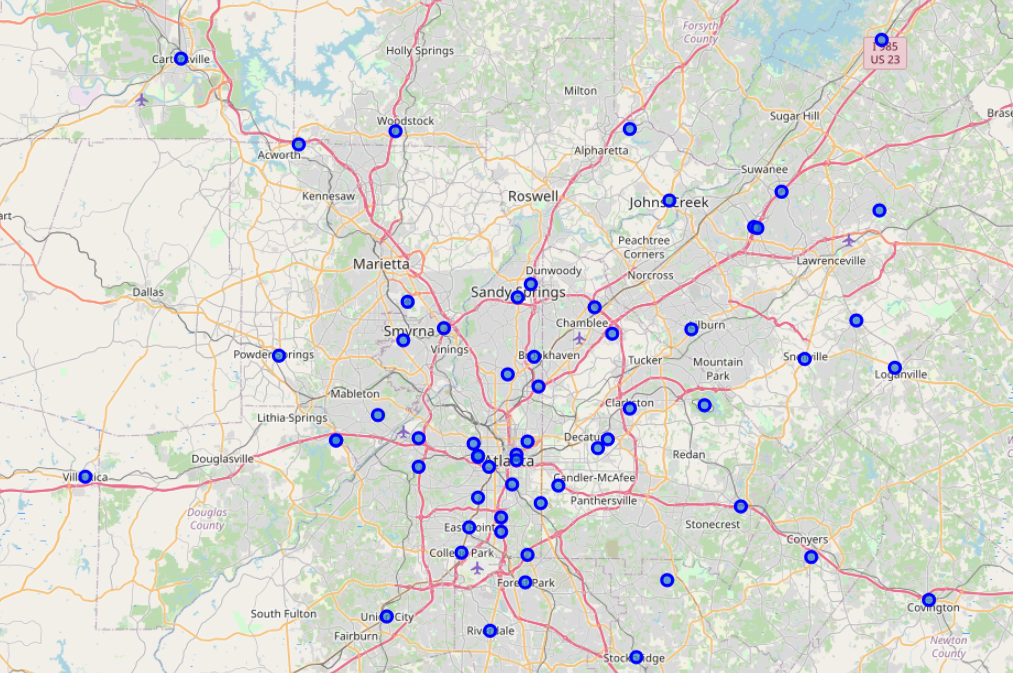
1. I will need to fetch the latitude and longitude information for the cities fetched above. I got the latitude and longitude information using geopy.geocoders. I was not able to get information for some of the cities. I used geopy.geocoders again with a zip code tied to the location and I was able to successfully fetch the geocentric values for all the required cities.



1. I will need the current data on house prices for different houses in Atlanta metropolitan city limits. I will get the house prices from <https://www.zillow.com>. I was not able to get most current information from the site. Had values only until March 2020. I used the aggregate price value of house prices from Jan 2019 to Mar 2020 for my analysis.



1. I will use foursquare to get the neighborhood venues close to different cities based on latitude and longitude of cities.



1. I will also need the current school ratings data to segment the cities into different clusters. I will get school rating values for different local public schools from <https://schoolgrades.georgia.gov/dataset/school-level-data/resource/3ab1377d-1f02-4c3c-99c5-e876ae766329#{}>. This site has information only from 2019 as there were no end of grade tests for many schools for 2020.

Getting school information for the specific cities was tricky as there are multiple type of schools and each city had multiple schools. To get it into a format that could help with house price analysis, I got high level information on different schools. This logic could show at a high level what percentage of A Grade schools a city has. This was done for all 3 levels as this would segment the prospective buyers.



This analysis can be updated with the latest data when it becomes available.