Coursera Capstone – "The Battle of the Neighborhoods"

Comparative Analysis using location data via Foursquare

Problem Statement

Loudoun County is seated roughly 25 miles due west of Washington DC. It is the fastest growing county as well as the richest county per capita household income in the United States at $130,000. It is also where 80% of all internet traffic in the world flows. It seems like every week a new data center is being added to Loudoun County. By 2020, Apple, and Amazon will have second headquarters here in the area. It is the third most populous county in the state of Virginia and is a bedroom community for business moguls, diplomats, and politicians. It is also known for its rich history and farmlands which have been slowly converting into vineyards and breweries that draw crowds from all over the National Capital Area.

The objective of this Capstone project is to use Foursquare location data and regional clustering of venue information to determine what might be the "best" place in Loudoun County to open a restaurant. What makes this so challenging is that there is no shortage of restaurants offering every type of the worlds culinary cuisine. From Mexican to Chinese, and Russian to Middle Eastern, the offerings are seemingly endless, thus competition, is high. Also, Loudoun County commands huge rent costs due to the overwhelmingly expensive cost of land and development. Therefore, the objective is to find a location that is in a reasonable rent area and a population dense area, with high disposable income.

Data Overview

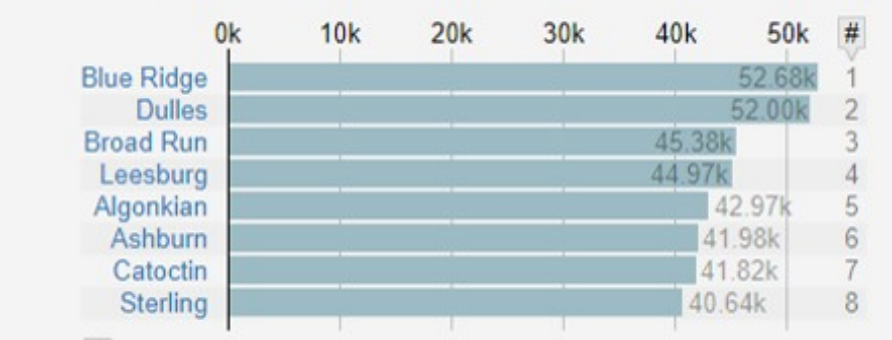
I will utilize data from Wikipedia regarding Loudoun County as well as Foursquare for this analysis. I will need to scrape the data, clean the data, and run analysis on the data. First I will have to determine the most likely neighborhoods based on mean rent prices and population density, as well as the disposable income. I will then consider the total number of venues and additional criteria to make a final determination. In addition, I am using data directly from the county office of planning and land management including Loudoun_Zoning.csv, Loudoun_population.csv, and Loudoun_income.csv, seen below on the next page.

Data Exploration and Methodology

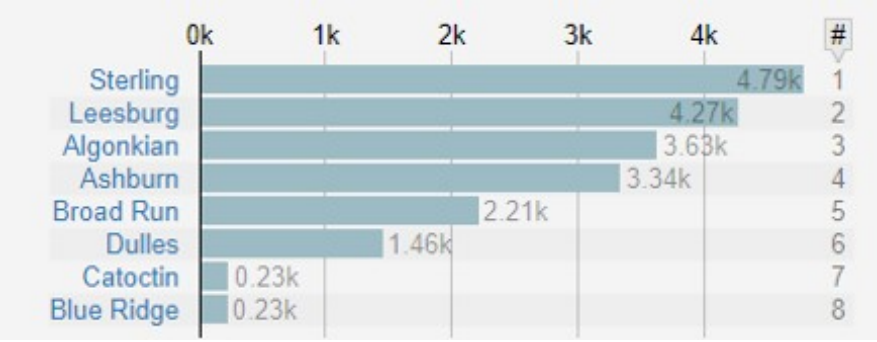
In first determining the most likely neighborhood for a restaurant based on rent and population density I analyzed Loudoun's subdivisions {Blue Ridge, Broad Run, Dulles, Leesburg, Algonquin, Ashburn, Catoctin, Sterling.}

This subdivision is important because it helps with the data analyses regarding both population density and median income distribution. These 8 areas are representative of both the most populous areas east of Leesburg, and the least populous areas west of Leesburg composed of farm land, wineries and undeveloped land. Further breaking down the data

into charts you can see the population density in each of the 8 subdivisions. In the following graphs below the population was broken out in the counties 8 subdivisions both in total and by number of people per square mile per.

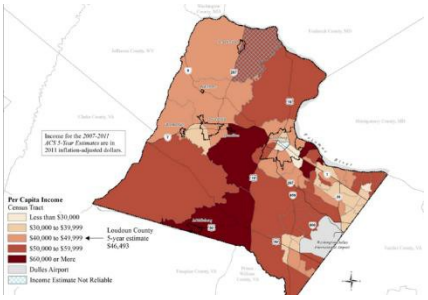


Total Population in each of the counties 8 subdivisions.

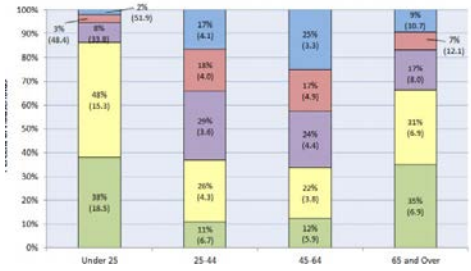


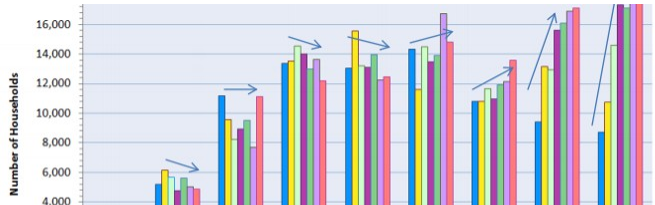
Total Population of County – Further reduced to People per square mile.

Next it was important to look at the income distribution across the county. While it is well founded that Loudoun County has the highest per capita median income in the entire country this does not mean that there are areas of the county where there are significantly lower wage earners and therefore a segment of the population who does not have the disposable income to be eating out at restaurants. Below is a choropleth map showing income distribution in the County. This map illustrates a large swath



area in both the Catoctin District and the Blue Ridge District where the average income per capita is greater than $60,000.00 per annum. In a household where the median income is $125,000 or greater there is an argument that there is disposable income for eating at a restaurant. I have also visualized the data for income in a different way using age as a coefficient of income as younger families tend to eat out more per week than older people. This graphical representation shows the percentage population broken down in ranges of income to help analyze and identify disposable income opportunities and predict whether or not these residents will utilize the restaurants in and around the area on a frequent basis. In addition, it was important to look at the income trends in the aggregate per household over a number of years to see if the potential for profit would exist in the future. The graph below represents hypothetical and projected income growth in the county for the years to come based on inflation and data pulled from the Bureau of Labor and Statistics.



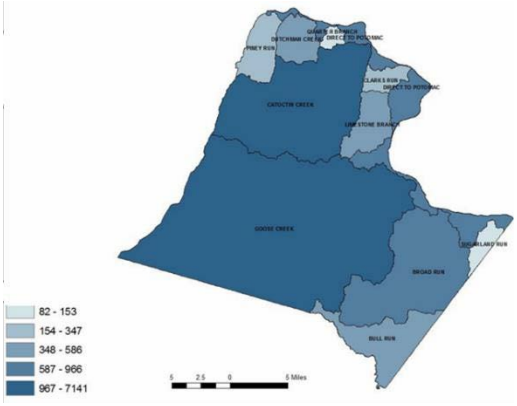


This graph shows that the trends of

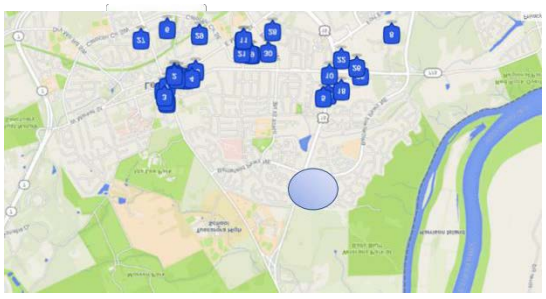
households making more than six-figure income is only increasing in the aggregate in the County and the forecast is that will continue into the future making a restaurant sustainable. As an aside this is also why this is the wealthiest County in the country. The amount of taxes collected is tremendous.

We also have to look at the costs of rent represented below by another Folium created choropleth map seen here in blue.

This map shows, not surprisingly, that the higher income areas of the Catoctin District and the Blue Ridge district correspond with higher rent prices per square foot and therefore in terms cost benefit this may be a wash. However there is a small swath of land within these two districts called the Limestone Branch where rent values are significantly lower than the surrounding areas for commercial purposes. As an aside part of the significantly high cost of land, and therefore, the associative high cost of rent has to do with the strong demand for secure data centers in the County.



Using clustering techniques developed in prior exercises during the capstone project I have examined the most common venues using the metrics stated above. The most populous neighborhood in the Limestone branch is Leesburg, which also happens to be the seat of Loudoun County. Leesburg proper has a significant amount of restaurants already and therefore competition would be great. However just on the outskirts there is a lot of development as well as income potential and population. Seen here in this API developed map there are a lot of options to choose from.



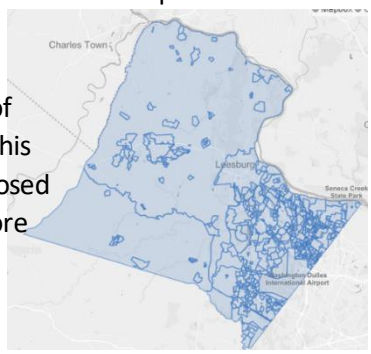
Result and Final Determination

As stated before, at first glance, Leesburg or Purcellville seem like the prime locations to place a restaurant. However, in the end after all the analysis is complete, the lime stone branch along the Rt 15 corridor, in and around, Lucketts, would be the optimal place to put a new restaurant. In consideration of all the factors medium rental costs, high population density, proximity to Leesburg and Lovettsville, as well as to many new brewery's and vineyards make it an optimal choice. In addition, the income levels are

higher in this area and the competition is sparse on the outskirts of the Le

	OBJECTID	ZO_ZONE	ZO_ORDINANCE	ZO_SPEC_CODE	ZO_PROJ_NUM	ZO_UPDATE_NUM	ZO_ZONE_DATE	ZO_ZONE_ORD	ZO_UPD_DATE
1	1			NV	NO VALUE	NO VALUE	1899-12-30T00:00:00.000Z	NO VALUE	1899-12-30T00:00:00.000
2	2	A10	1993 R		ZMAP-1998-0009	Z2000-0001	1900-01-01T00:00:00.000Z	A10 1993	2012-04-16T00:00:00.000
3	3	A10	1993 R		ZMAP-1997-0011	Z2000-0002	1998-06-03T00:00:00.000Z	A10 1993	2012-04-16T00:00:00.000
4	4	A10	1993 R		ZMAP-0000-0293		1980-05-19T00:00:00.000Z	A10 1993	2012-04-16T00:00:00.000
5	5	A10	1993 R		ZMAP-1995-0011		1996-09-18T00:00:00.000Z	A10 1993	2015-10-08T00:00:00.000
6	6	A10	1993 R		ZMAP-2000-0004	Z2002-0002	2001-10-15T00:00:00.000Z	A10 1993	2012-04-16T00:00:00.000
7	7	A10	1993 R		ZMAP-1999-0013	Z2001-0005	2000-12-18T00:00:00.000Z	A10 1993	2012-04-16T00:00:00.000
8	8	A10	1993 R		ZMAP-1987-0025		1988-06-20T00:00:00.000Z	A10 1993	2012-04-16T00:00:00.000
9	9	A10	1993 R		ZMAP-1985-0010		1986-04-21T00:00:00.000Z	A10 1993	2012-04-16T00:00:00.000
10	10	A10	1993 R		ZMAP-1987-0037		1988-08-01T00:00:00.000Z	A10 1993	2012-04-16T00:00:00.000

In addition I used the counties LOGIS mapping system and Folium to create choropleth maps of income and population as well as rent cost distributions to help in the identification of where to put the restaurant. As well as JSON files to identify county and city boundaries. In identifying and creating this map I realized that it was actually easier and more cogent to use the 8 districts as my first layer of data visualization as I peeled back the layers on the county data. I did this because the County itself is organized into 8 major subdivisions as opposed to using city, town, or ordinance boundaries. This keeps the datum more clean and easier to follow.



esburg town proper.