**Segmenting and Clustering Neighborhoods in Novi, MI**

**Applied Data Science Capstone Project Report**

By Yunjiao Cai July 1, 2019

# 1. Introduction to the Opportunity in Novi

### 1.1 Background

Novi is a city in Oakland County in the U.S. state of Michigan. The city is located approximately 25 miles northwest of the center of Detroit. Novi was originally a township organized in 1832, but it continues to be one of the fastest growing cities in Michigan today.

### 1.2 Problem

As the city grows and develops, it becomes increasingly important to examine and understand it quantitiatively. The city of Novi provides open data for everyone and encourages entrepreneurial use to develop services for the benefit of its citizens.

Developers, investors, policy makers and city planners have an interest in answering the following questions as the need for additional services:

* ***Whether the population increases continuously since Novi is the one of the fastest growing cities in Michigan?***
* ***Is population increasing correlated to economy?***
* ***How is crime incident today in Novi?***
* ***Where is the best place to open another coffee shop?***

# 2. Data

### 2.1 Data Sources

Most data and information about the City of Novi can be found in the City of Novi Open Data Center:

* [https://www.opendatanetwork.com/entity/1600000US2659440/Novi\_MI/](https://www.opendatanetwork.com/entity/1600000US2659440/Novi_MI/demographics.population.count?ref=search-entity&year=2017)
* Novi Developers Access to venue data: <https://foursquare.com/>

### 2.2 Data Selection

Normally, we can use the following data sets to indicate the growing status and economy healthy situatin of the city:

1. Population changing information
2. Residents’ earning and education degree
3. The debt and revenue of the city
4. Crime incident information of the city

In the case, we can select and focus on the following data sets from the city Open Data Center:

* The City demographic data will show us the population changing rate, it will also show the residents’ earning and education information.
* Economy and finance data will enable us to conclude economy statuation of Novi.
* The public safety data will let us to compare the population density and crime.
* Neighborhoods and location data can let us to capture venues information, so we can determine where to open another coffee shop.

# 3. Methodology

All steps are referenced below in the Appendix: Analysis Section.

### 3.1 Data Cleaning

Data scraped and downloaded from multiple sources were in all different type (ie. HTML, XML, xlsx), and it was in a mess and unreadable.

I used *Beautiful Soup* to pull data out of HTML and XML files, and parsed data in order to find ways of navigating and searching the useful information. Then I used *json* library to convert those parsed data into lists or dictionaries.

After that, I combined some related lists and dictionaries into different DataFrame or tables by using *Pandas* library. However, there were a lot of missing values in the DataFrame. I use *Pandas* to clean, fill those missing values and reorganize some DataFrame.

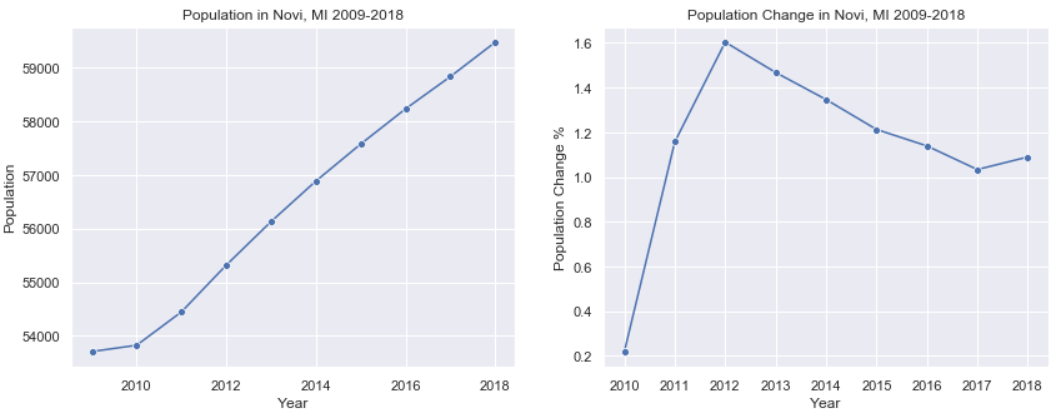
### 3.2 Exploratory Data Analysis

***In order to determine the living condition of Novi, I explore and analyzed the each data set to see how does data affect to the city. I also studied and understood the correlation between population, economy and crimes in the city. Then I performed statistical analysis on each data, created summary tables and figures to show how those data affects the city.***

#### 3.2.1 Relationship between population changing and city growing

In general, more population means more business opportunities and more tax income to the city. I captured the population changing information of Novi from 2009 – 2010, and also calculated and plot the rate change of the population.

The following figures show the population increases continuously while the city is growing from 2009-18



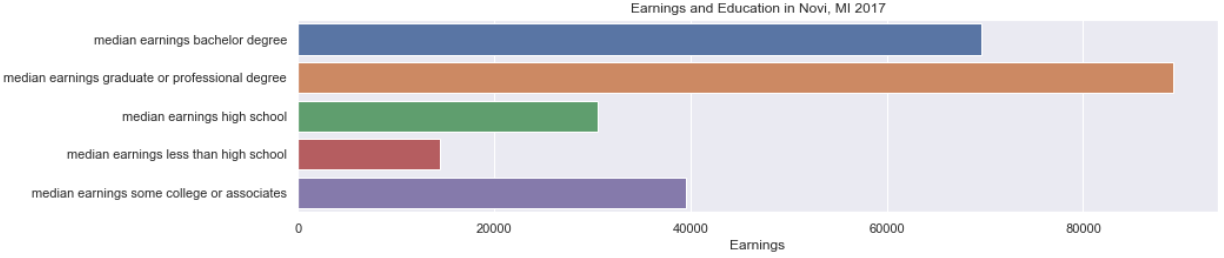
#### 3.2.2 What kind of people move to Novi

This will help the city to have more business opportunities if people who live in Novi have high income. More shopping mall, groceries, supermarkets, stores will occur and provide more tax to the city, then the city will have sufficient resource to develop.

Also, the people who have high education degree will make the city living condition and school district to be better.

The following table and figure show people who come to Novi have high earning and education degree.

|  |  |
| --- | --- |
|  |  |
|  |  |

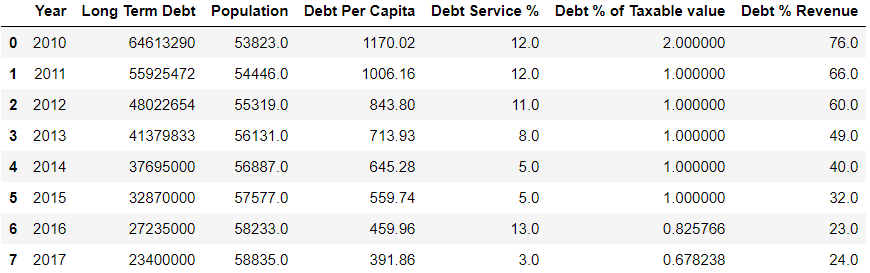
****

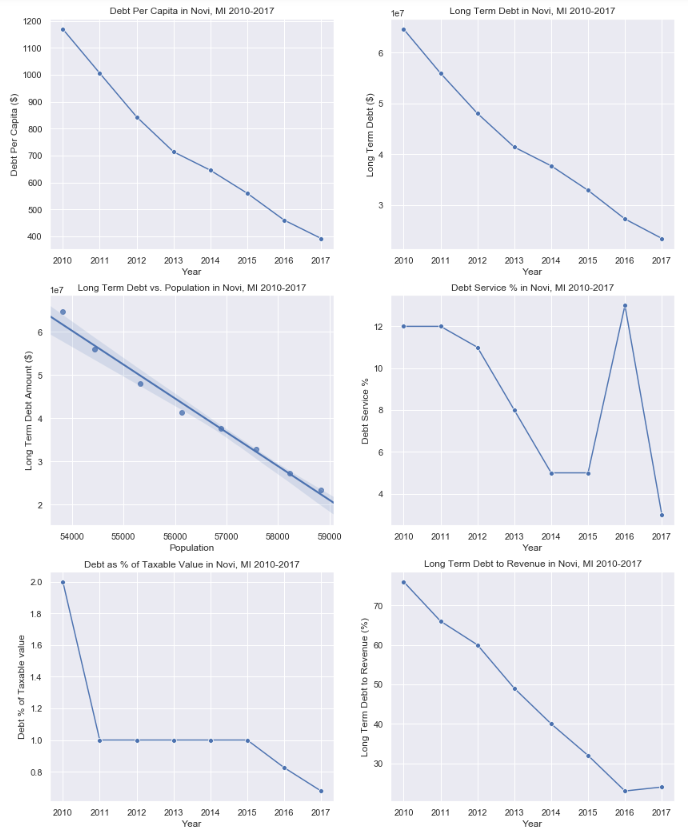
#### 3.2.3 Check the city debt and revenue information

Since Novi is the one of the fastest growing cities in Michigan. I assumed the revenue of the city should increase and the debt should decrease due to more and more high income people move to Novi.

The following table and figures show debt decreases and revenue increases in Novi from 2010 – 2017.

* Debt refers to the total amount of outstanding dollars owed by a government entity.
* Debt Per Capita is the amount of outstanding debt per resident.
* Long Term Debt is the unmatured principal of bonds, or other forms of noncurrent or long-term obligation indebtedness.
* Debt Service Ratio is the percentage of annual revenues committed to paying back debts.
* Debt as a Percentage of Taxable Value is the amount of long term debt compared to taxable value.
* Long Term Debt Revenue is the ratio of outstanding long term debts to revenues.

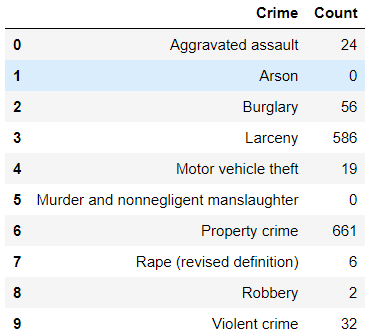
****

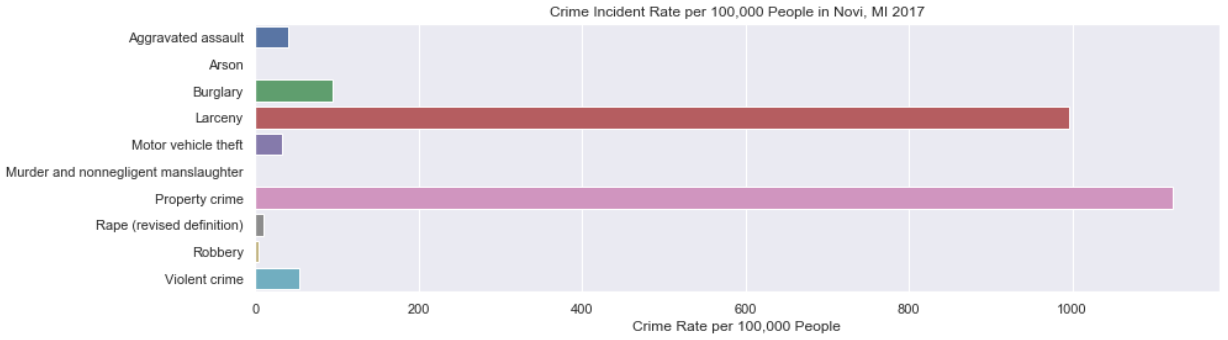
****

#### 3.2.4 How about the crime incident since more population come to Novi

I summaried all crime occurred in Novi in 2017, sorted and listed them with crime name and count number. In order to explore the relationship between population density and crime, I calculated and created a bar chart to display the crime incident rate per 100,000 People in Novi.

The following table shows the number of crime occurs in Novi in 2017.



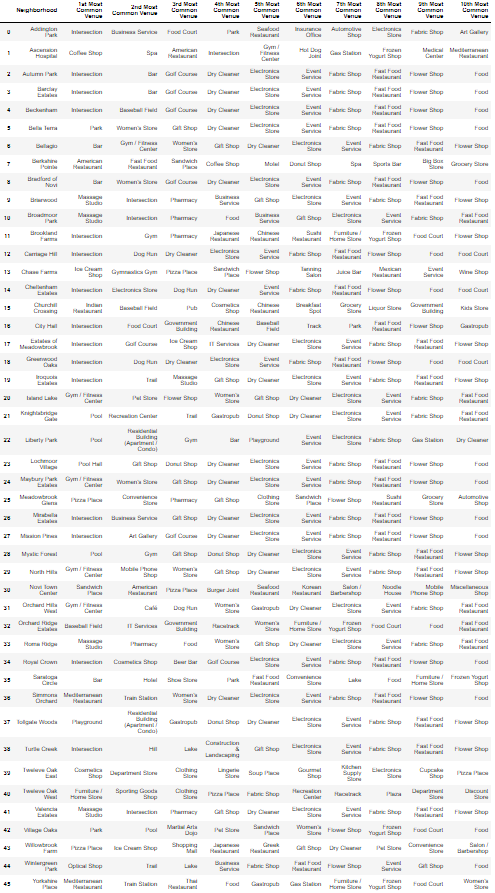
The figure shows the crime incident rate per 100,000 People in Novi.****

#### 3.2.5 Explore and Cluster the Neighborhoods in Novi

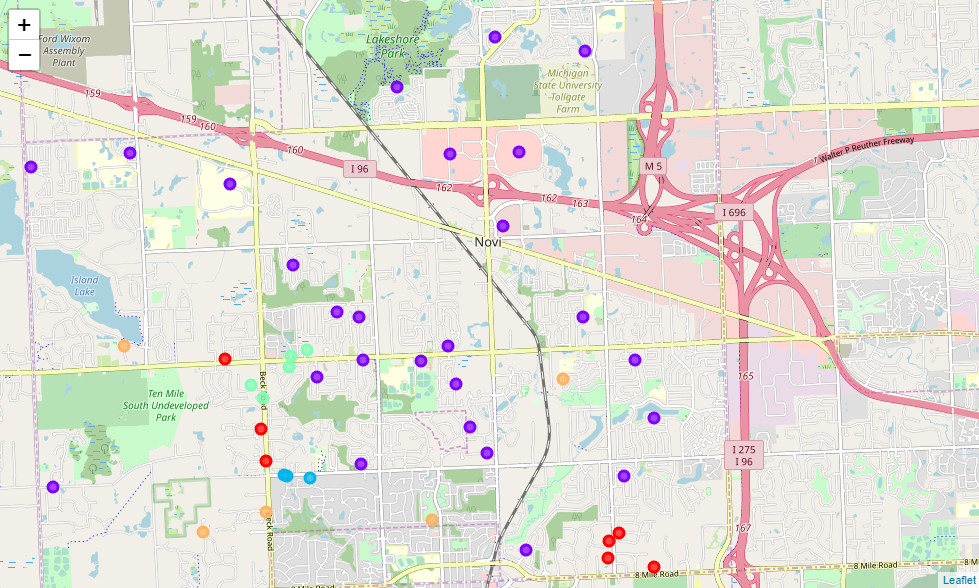
I created data includes all neighborhoods with their latitude and longitude values in Novi, then I used *Foursquare API* to explore those neighborhoods, used the *explore function* to get the most common venue categories in each neighborhood.

Finally, I used the ***k-means clustering algorithm*** to group the neighborhoods into clusters, and use the *Folium library* to visualize the neighborhoods in Novi and their emerging clusters.

The following table shows the top 10 most common venues in Novi:



The following figure shows the map of cluster neighborhoods in Novi after using k-means machine learning technology:

****

# 4. Result

In this study, I analyzed the relationship between Population, Residents’ earning and education, Cities’ Debt and Revenue and City Economy, Living condition.

***Data shows more and more high income and education population come and live in Novi in the past 10 years. At the same time, more and more business opportunities occurrs, and the city has more revenue and less debt while the city is growing. Also, the high population does not increase the crime indicent significatly. The crime indicent per population density is still very low.***

Overall, Novi is the one of the best city to live in Michigan and it is also the one of the fastest growing city. Better living condition, school district and safety will attract more high income and education people move to Novi, then those people will bring more benefit and opportunities to the city. It is a great virtuous cycle.

# 5. Discussion

There is more and more business opportunities in Novi. Based on the data analysis, machine learning and cluster visualization, we can see that maybe the best place to open a coffee shop will be at the section of Ten Mile Rd and Novi Rd.

The reason is : 1) The center of the city . 2) A lot of subdivisions locate around that section. 3) Close to the entry of the heavily used highway in Novi***.***