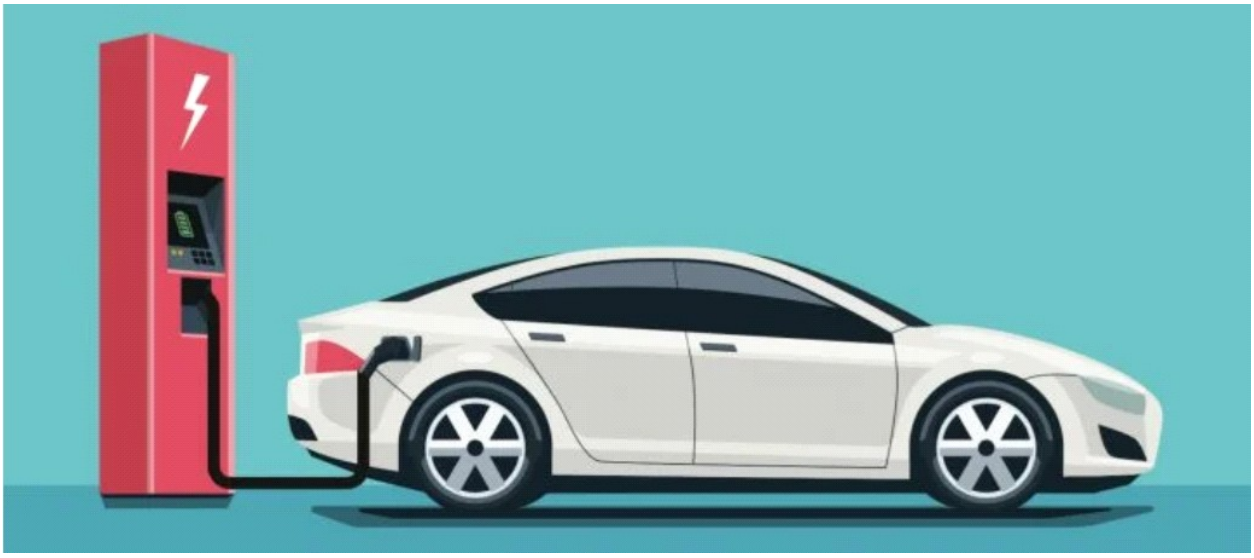
Selection of US Cities with the Highest Potential for Car Sales



Source: Greentech Media's website

by

Frantzdy Hervé

MSc Economics and Econometrics

August 2020

Table of Content

I. Introduction: Business Understanding

The car industry has been booming recently in the United States (Source: Investopedia (2020)). The tendency now favors more and more electric vehicles as they are viewed to be more environmentally-friendly, less expensive in the long run, and their design gives a feel of vehicles of the future. An automaker, namely Tesla, has had its market valuation grown at a staggering rate. It rose at a more than 700% yearly increase from Summer 2019 to Summer 2020 - in the middle of a pandemic (Bloomberg (2020)). Its sales are projected to only increase in the future (Source: Energy.gov: History of the electric vehicles (2020)). Consequently, the ambitious investor witnessing this tremendous shift in the car consumption patter might well be wondering: 'In what US cities this dynamic might see the highest movement?'

Obviously, it is needless to state that building the infrastructure for electric cars would require massive amounts of money. This is partly due to the fact that electric vehicles are almost completely different to the traditional combustion-engine vehicles when it comes to their engine outlook, their transmission system, their source of power, and the highly advanced technology needed for their computing wiring. Additionnally, we can argue that investing in this sector requires serious considerations on the demand since these cars have on average a higher Manufacturer's Suggested Retail Price. However, since this segment of the car industry in the US seems to be only promising, this might be exactly what any swift investor willing to find a niche for highly profitable investments would be interested in.

I.a. Business Problem

Electric vehicles represent a niche that requires massive investment to build the infrastructure to support its development. (Source: Estimates for building a charging station). This obstacle could be largely offset, though, by the financial gain any smart investment in the electric car industry might yield. As a matter of fact, based on the most pessimistic predictions, the traditional combustion engine will go extinct in about ten years while the increase in consumer's demand for electric cars and auto parts will continue at a rate as high as 21% (Source: Booming demand for electric vehicles). The bottleneck is that the demand is not uniform in the US. That is, all the states are not equal when it comes to growth in any sector. And, specifically, because of disparity in terms of the growth in the population by state and by urban area, income per capita, the cautious investor would gain much by being carefully selective. They would pick the locations for their investment in the said sector based on these precise considerations. In other words, the need to identify the states and the cities where the demand for reliable transportation

would be the most promising for the next 5 to 10 years. However, two questions arise from this cogitation: What states would see these demands manifested the highest, and what cities should be selected from the said states to explore potential investment ventures?

I.b. Analytic Approach

The following lines will provide a comprehensive approach to answer the aforementioned questions. A successful answer to the questions would be a group of 10 states where the demand for cars and auto parts will likely continue to grow. A satisfactory answer would also identify the urban areas where the population seems to be the most active so to integrate a key aspect of the answer into any strategic infrastructure planning. The proposed solution is made possible by the data provided from the US Bureau of Economic Analysis (BEA) and the platform designed by Foursquare to locate areas where there is the highest number of venues. We have then leveraged data using the BEA's API guidance and the well-structured platform designed by Foursquare.

I.c. Data Requirements

Our approach is two-fold:

1. First, we pull data from the BEA's website related to the Gross Domestic Product (GDP) level in each US state, their population, their income per capita and the consumer's spending per capita in motor vehicle and parts for the past 20 years. This would allow us to use enough data that would validate our assumptions in terms of the selection of the 10 states with the highest potential for demand in cars and auto parts in the years to come; and,

2. Finally, we use the Foursquare platform to locate 5 urban areas where the population seems to be the most active from 5 of the 10 states previously identified.

I.d. Objectives

Our main objective is to determine a list of 5 US cities from a list of 15 cities (3 from each of the 5 states selected) that present the highest potential for investment in the car industry.

This objective is contingent upon:

1. Obtaining all the necessary data for key variables that affect consumption in motor vehicle and parts (GDP, Income Per Capita, and Population) ;

2. Selecting and sorting a list of 10 states based on the average values of the target variable;

3. Selecting and sorting a list of 3 cities from the top 5 states with the highest potential for car sales; finally;

4. Selecting a city from each one of the 5 selected states that present the highest opportunity.

II. Methodology

The methodology adopted in this analysis is based on the use of descriptive statistics to select the cities with the highest potential for car sales. It can be summarized as a multi-level selection technique where each group of geographical units (states or cities) is sorted with respect to the key variable considered for a given level. The primary operation carried out is, thus, groups sorting leading to the selection of the top elements based on the rank of each element accordingly.

The starting point is the group of 51 us states (including the district of Columbia) on which two sorting criteria are applied. The first group is obtained after sorting the said states based on the variable that has the highest correlation with the consumer's per capita spending in motor vehicle and parts. Similarly, the second group is obtained after sorting the first group based on the second highest correlated variable with the target variable.

After obtaining a list of 10 states, further refinements are applied by pulling a list of 3 cities from each one of the 5 first states. The last group of 5 cities is then obtained from sorting the list of 3 cities according to the number of venues recorded from each one of them within 1 km of their most populous city center.

II.a. Data Collection

Access to the data was made via API calls to the BEA's and the Foursquare's websites. The data obtained from the BEA's website was organized into CSV files for each key variable that is correlated to the consumer's per capita spending in the segment of motor vehicle and parts. The quantitative variables are namely: GDP, Population, Income per capita, and timespan starts from 1999 to 2018. The data obtained from the Foursquare's websites was a list of venues which were counted in order to select the 5 states having the

highest consumer's spending per capita for motor vehicle and parts.

II.b. Data Analysis

The Data Analysis consists of:

- First, the visualization of data to gain insights about any outstanding features it might present;
- Secondly, the extraction of key statistics from the distribution of each variable across the whole timeframe and the display of scatter plots to define the magnitude and direction of the correlation our key variable has with each one of the three other variables. This latter step will ultimately lead us to the final list of selected states. From the said list, we would then analyze the repartitions of venues per city to finally obtain a list of 5 cities with the highest potential for the car industry.

II.c. Limitations

The methodology applied throughout this analysis could be qualified as naive to the extent that it does not build a predictive model based on a set of all possible factors that affect a consumer's decision to buy a car. We selected the variables on which sorting occurs on the assumption of the level of correlation they maintain with the consumer's spending in the said segment. The bottom line is that each one of the two factors on which the groups are sorted has at least a correlation value of 0.4. That is, the GDP, for instance, having a correlation value of 0.48 could explain almost 50% of the variation observed in the consumer's per capita spending taken separately all things being equal.

III. Results and Discussions

In this section, we will display the main results obtained from our analysis of the dataset put together for the sake of this endeavor. Each portion of the result presented will be accompanied with a brief discussion on any insight we might gain from them.

III.a. Description of the dataset

Key statistics on the dataset are presented below:

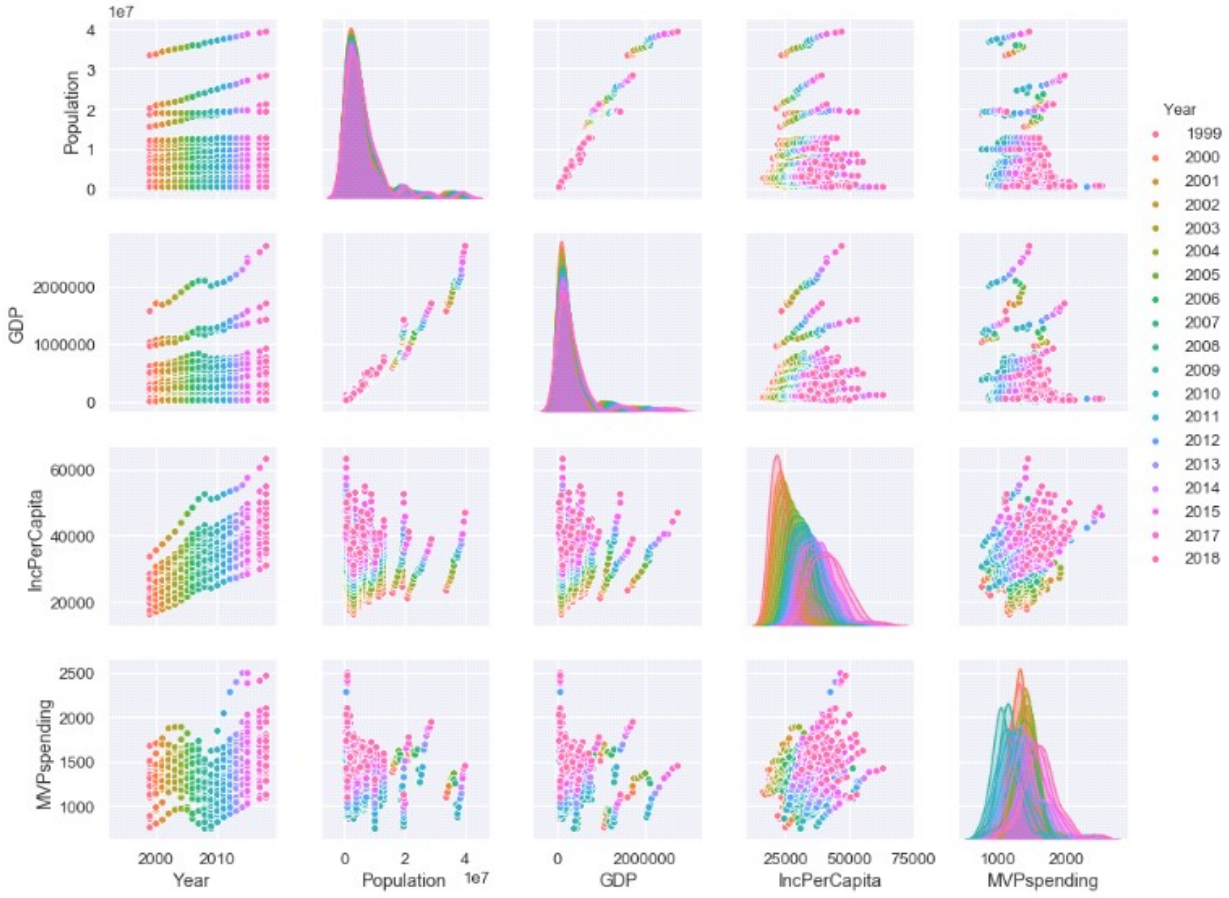
Descriptive Statistics on the dataset

	Year	State	Population	GDP	IncPerCapita	MVPspending
count	1020.00000	1020	1.020000e+03	1.020000e+03	1020.000000	1020.000000
unique	NaN	51	NaN	NaN	NaN	NaN
top	NaN	Florida	NaN	NaN	NaN	NaN
freq	NaN	20	NaN	NaN	NaN	NaN
mean	2008.45000	NaN	5.965344e+06	3.037542e+05	32155.465686	1374.028431
std	5.70784	NaN	6.689837e+06	3.755205e+05	7634.563168	239.721023
min	1999.00000	NaN	4.917800e+05	2.184340e+04	16184.000000	755.000000
25%	2003.75000	NaN	1.626136e+06	7.380845e+04	26504.250000	1222.750000
50%	2008.50000	NaN	4.166092e+06	1.816235e+05	31549.500000	1364.000000
75%	2013.25000	NaN	6.794228e+06	3.777569e+05	36812.750000	1509.000000
max	2018.00000	NaN	3.946159e+07	2.721651e+06	63151.000000	2499.000000

Source: Author

A brief glance at the table above shows that the dataset contained an observation for each year considered in this analysis. The magnitude of the standard deviation for most of the variables seems not to be high which might indicate a normal distribution where most observations gravitate around the mean.

Distribution of the variables



Source: Author

As mentioned earlier, the distribution of all of the variables seems to be mildly normal. Nevertheless, we observe some skewness from the population, the GDP, and the income per capita while the consumer's spending per capita in motor vehicle and parts seems to be normally distributed. The skewness observed in the former strongly indicates that for a few states the average level of the population, the GDP, and income per capita are well above the national average. This, thus, fairly validates the point that not all states are equal when it comes to their economic standing.

Another feature from the distribution is that all of the variables are increasing with time except for a couple of years around the year 2008. This indicates the impact of 2008 global financial crisis has had on said indicators but the population which inexorably continued to grow.

As per the relation between the consumer's per capita spending in motor vehicle and parts with the other variables, we could fit a line between the scatter points displaying such relation. The table below will give a more precise idea of the extent of such relations.

III.b. The list of 10 States

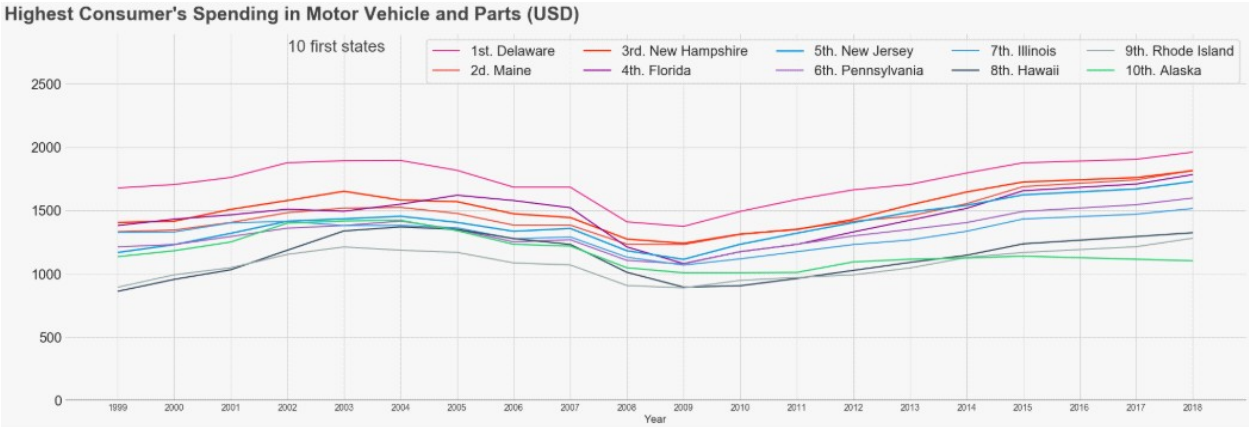
The selection of the 10 states with the highest potential is based on the correlation values populated from the table below.

Correlation table

	Population	GDP	IncPerCapita	MVPspending
Population	1.000000	0.973093	0.991718	0.354513
GDP	0.973093	1.000000	0.993228	0.483039
IncPerCapita	0.991718	0.993228	1.000000	0.411647
MVPspending	0.354513	0.483039	0.411647	1.000000

Source: Author

The correlation table presents the pairwise correlation using the Pearson method. It indicates that the consumer's per capita spending in motor vehicle and parts (MVPspending in the table) has the highest correlation with the GDP. The second highest correlation is from the income per capita variable. Therefore the two variables will be used to sort the list of 51 states described earlier. Although the population has the lowest correlation with the target variable, it is considered at the ultimate stage of selection of the cities where we select the city having the highest numbers of venues recorded on the Foursquare app.



Source: Author

The figure above shows the variation across 20 years of the 10 states obtained after three consecutive sortings of the dataset based on first the GDP, then the income per capita, and finally the consumer's per capita spending in motor vehicle and parts. We can also see that the first 5 states with the highest potential have held the top-5 position more or less consistently for the past 20 years. This will likely be the case for the next couple of years,

everything being equal.

We can see from our naive selection method, the first 10 states with the highest potential in car sales are Delaware, Maine, New Hampshire, Florida, New Jersey, Pennsylvania, Illinois, Hawaii, Rhode Island, and Alaska. For the sake of simplicity, we will select 1 of the 3 most populous cities where it would be ideal to make an investment in such an industry from each one of the first 5 states.

According to the online encyclopedia, Wikipedia, the following are the 3 most populous (number of inhabitants in parenthesis) for the first 5 states selected:

1.The State of Delaware (DE)

a) Wilmington (70,635)

b) Dover (38,079)

c) Newark (33,673)

2.The State of Maine (ME)

a) Portland (66,700)

b) Lewiston (36,299)

c) Bangor (32,568)

3.The State of New Hampshire (NH)

a) Manchester (109,565)

b) Nashua (86,494)

C) Concord (42,695)

4.The State of Florida (FL)

a) Jacksonville (903,889)

b) Miami (470,914)

c) Tampa (392,890)

5.The State of New Jersey (NJ)

a) Newark (285,154)

b) Jersey (270,753)

c) Paterson (148,678)

III.c. Leveraging data from Foursquare

API calls made to the Foursquare app made it possible to obtain a list of venues recorded on the app for each one of the 15 cities mentioned in the section above. The following is the result of their response from our requests related to the matter.

Number of venues per selected city

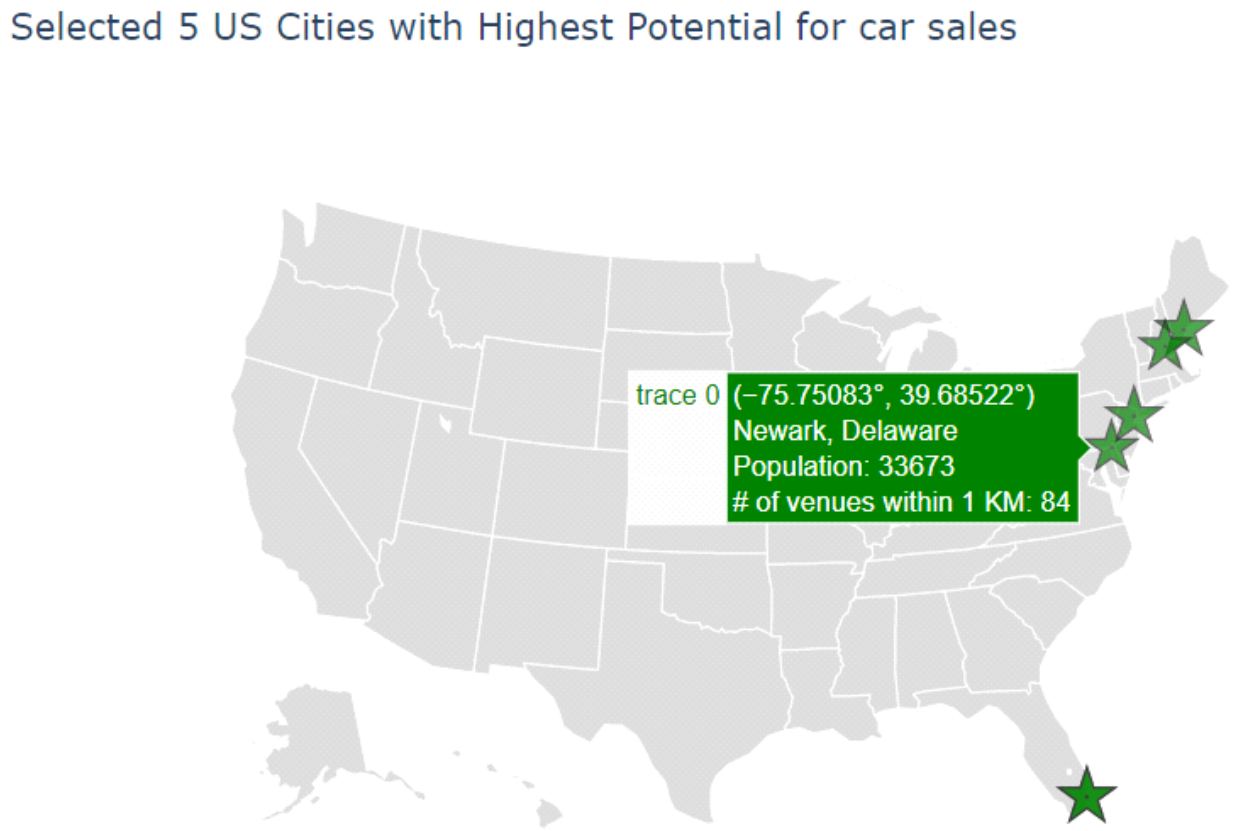
```
Within 1 KM of the given city, there are:
['67 venues in Wilmington, DE', '30 venues in Dover, DE', '84 venues in Newark, DE']
['100 venues in Portland, ME', '39 venues in Lewiston, ME', '82 venues in Bangor, ME']
['61 venues in Manchester, NH', '62 venues in Nashua, NH', '80 venues in Concord, NH']
['89 venues in Jacksonville, FL', '100 venues in Miami, FL', '93 venues in Tampa, FL']
['100 venues in Newark, NJ', '4 venues in Jersey, NJ', '63 venues in Paterson, NJ']
```

Source: Foursquare

As stated in the main objective in the introduction section of this document, we aimed at selecting the 5 cities with the highest number of recorded venues from the top-5 states with the highest potential for growth in the car industry. The list of cities above helps us select one city from each state where the highest number of venues are recorded according to the Foursquare app.

III.d. The list of 5

This section is a graphical representation of the 5 cities with the highest number of venues recorded on the Foursquare app. These cities are namely: Newark from Delaware, Portland from Maine, Concord from New Hampshire, Miami from Florida, and Newark from New Jersey.



Source: Author

IV. Conclusion

Throughout the analysis presented within this article, we realize that the fundamental questions that an investor might ask in order to determine the locations in the US with the highest potential in car sales have found their answers. The first list of states provided earlier was extracted from the total list of 51 US states (including the District of Columbia). The 10 states in question were found from sorting the total list three consecutive times with respect to the value of GDP, income per capita, and consumer's per capita spending in motor vehicle and parts from highest to smallest.

As a result of counting the number of venues recorded from the Foursquare app and selecting the cities with the highest number of venues, we obtained the list of 5 US cities with the highest potential for car sales in the future. These cities are namely: Newark from Delaware, Portland from Maine, Concord from New Hampshire, Miami from Florida, and Newark from New Jersey. It is our understanding that the concentration of the population coupled with their high standing of living should then indicate their highest likelihood to acquire electric vehicles and affect the growth in the sector more than any other city in the United States.

Future analysis might be offered in relation to this subject to study the stability of the car industry market in the said cities and how electric vehicles are viewed by the average consumer. By all means, regardless of how a selective investor might be, the cautious one has a place to start with this analysis.

V. References

1. Investopedia (2020). The car industry and the impact of electric vehicles. Consulted from: https://www.investopedia.com/articles/pf/12/auto-industry.asp#:~:text=The%20U.S.%20economy%20was%20booming,million%20new%20cars%20were%20sold.&text=Among%20the%20more%20notable%20early,difficult%20recession%20of%202007%2D2008

2. Energy.gov. History of the electric car. consulted from: https://www.energy.gov/articles/history-electric-car#:~:text=First%20Crude%20Electric%20Vehicle%20Is,an%20English%20inventor%20in%201884

3. Access to the Bureau of Economic Analysis platform via: https://www.commerce.gov/bureaus-and-offices/bea#:~:text=Today%2C%20the%20Department%20of%20Commerce's,annual%20rate%20of%204.8%20percent.

4. Access to Foursquare developper's corner via: https://fr.foursquare.com/developers/apps

5. All the codes used for this document can be accessed via the following link on github: https://nbviewer.jupyter.org/github/hertzy2/Coursera_Capstone/blob/master/Capstone%20Project_Selection%20of%20US%20cities.ipynb