

Gregory Gray Jr.

# PEER-GRADED ASSIGNMENT:

# CAPSTONE PROJECT

The Battle of  
Neighborhoods  
(Week 2)



# CONTENTS

1. A full report consisting of all of the following components (15 marks):

## Introduction Section

- Discussion of the business problem and external stakeholders

## Data Section

- Description of the data to solve the problem and its the source.

## Methodology Section

- Main component of the report, exploratory data analysis, inferential statistical testing performed, machine learnings and the utilization.

## Results Section

- Observations and Recommendations

## Conclusion Section

- Reported Findings

2. Link to Notebook on the Github Repository (coding provided (15 marks):

## Link Section

- See provided link

3. Your choice of a presentation or blogpost. (10 marks)

## Presentation Section

- See provided link

# INTRODUCTION SECTION

## DISCUSSION OF THE BUSINESS PROBLEM AND EXTERNAL STAKEHOLDERS

In November of this year the legalized and regulated sale of Cannabis was approved in the state of Michigan. Currently, Ann Arbor, MI has some of the strictest laws against the use of cannabis (Wikipedia.org).

In contrast, Keego Harbor, MI voted to legalize marijuana for people 21 and older (up to 1 oz.) and only in non-public places (e.g. a house). With the newly found legislation many venture capitalists are trying to leverage the 12.16% of Michiganders who use some form of

cannabis (CBSNews) into a lucrative business. Recently, exclusive projections from the 2018 edition of the Marijuana Business Factbook suggested that retail sales of medical and recreational cannabis in the United States are expected to hit \$8 billion – \$10 billion this year – a nearly 50% increase from 2017 – and rise as high as \$22 billion by 2022 (MJBizDaily). One of the fastest growing venture capitalist firms, Nilla Bean Inc., believes that “seed and growth-stage” is best fit



Photo by [rawpixel](#) on [Unsplash](#)

for serving an emerging market such as the Cannabis industry (CanopyBoulder). Nila Bean's goal would be to invest at a discounted valuation to achieve outsized returns within the next 5 to 15 years.

## OBJECTIVES

- Build a leading organic Cannabis farm in Keego Harbor, MI and recruit employees who live within a 100 miles radius.  
(Leverage Structured and Unstructured Data)
- Although the law permits the sale only in non-public places (e.g. a house). If and when the public sale is approved the investment team would like to build retail shops specializing in the sale of cannabis products near heavily populated areas in Detroit, MI (Leverage Foursquare Location Data)

## PROBLEM

The executive team at Nila Bean Inc. are cautiously optimistic that with a minimal investment in Keego Harbor, MI makes perfect sense. Furthermore, they have found the entrepreneurs who have the skills to move

forward with the opportunity. Entrepreneurs of famed Black & Gold Leaf, a Cannabis farm is looking to relocate from California to Keego Harbor is ideal. Yet, the Black & Gold Leaf's leadership team has no data about Michigan let along the city of Keego Harbor as it relates to the Cannabis industry. If the Entrepreneurs at Black & Gold Leaf hopes to secure funding they need to capitalize on the financing that Nila Bean is offering. Black & Gold Leaf requires the expertise to extract the data and communicate the potential of success.

Black & Gold Leaf has decided to hire Dr. Luther Jr., a Data Scientist from a Grey Slate Inc. to provide data to aid in the final investment decision. Dr. Luther Jr. primary objective will be to learn and address the ambiguity concerns linked to the communities of Keego Harbor and Detroit Michigan through analysis of the following:

- Wrangling Structured and Unstructured data
- Understanding the relationships of the data
- Communicating the findings to stakeholders

# DATA SECTION

## DESCRIPTION OF THE DATA TO SOLVE THE PROBLEM AND ITS THE SOURCE.

After weeks of discussing the primary problem Dr. Luther Jr. has decided to sign a multi year contract, with the potential for an extension, to resolve both leaderships concerns."Data science as one's attempt to work with data, to find answers, to questions that they are exploring."Murtaza Haider, PhD Ted Rogers School of Management at Ryerson University

Dr. Luther Jr. is scheduled to present his approach in an all encompassing meeting and plans to outline the **Data Science Process** with the goal of properly leveraging the data to suggest if **(a)** the city of Keego, MI provides human assets to build an Organic Cannabis farm and **(b)** is there formidable infrastructure to build retails shops in Detroit, MI.

## Step 1 of 6: Framing The Problem.

### What is the specific business problem?

Leaderships of both organizations are not clear on the Economics, Diversity, Health & Safety, Education or Housing & Living status in Keego Harbor, MI. Furthermore, the primary goal of building an Organic Cannabis farm in Keego Harbor which would naturally lead to building retails shops in both Keego Harbor & Detroit, is really contingent on understanding the five key factors listed above.

Dr. Luther Jr. has decided that the problem is not an investment problem per-say but one of inadequate human capital. Or a better way of stating it is not "what" is in the city of Keego or Detroit but "who" is located in these areas. To address this problem Dr. Luther Jr. plans to capture data that aligns with both organization's strategies.

## **Step 2 of 6: Collecting The Right Data.**

### What data is available?

- Per Data USA, a website created in 2014, by Deloitte, Datawheel, and Cesar Hidalgo, Professor at the MIT Media Lab and Director of Collective Learning, has a wealth of public US Government statistical data that's perfect for this data analysis.
- Compared to other US cities, Keego Harbor, MI has a very small population and the historical data is scarce. To ensure the “right data” is collected DataUSA.io was utilized to provide a database of unstructured and structured data which is key to solving leadership’s investment concerns.
- Dr. Luther Jr. will also utilize a few API features provided by Foursquare to recommend if building retail shops in both cities are ideal. Some of the features are ratings, lists, local searches, and other

options that provides the best fit for adequate statistical analysis.

### How should the data be extracted?

The data was download from <https://datausa.io/profile/geo/keego-harbor-mi/> and transformed using Microsoft Excel and exported in CSV format.

### What format will be used and how will the data be stored for data analysis?

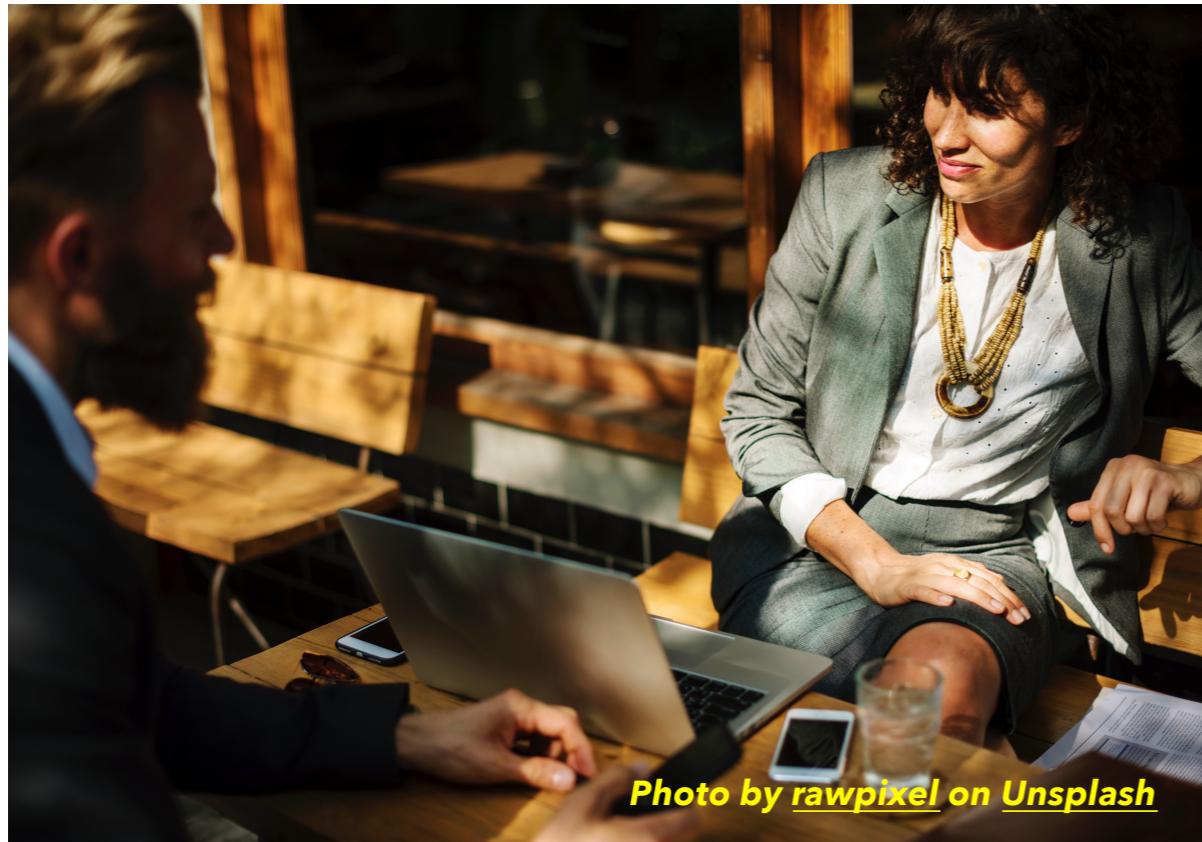
Both Excel and CSV files will be stored on the IBM Watson Studio Cloud for compilation.

## **Step 3 of 6: How to wrangle the data.**

### Principles applied during the data wrangling phase:

#### Identify any missing values.

- Data containing more than required e.g. unnecessary columns and rows were eliminated.



*Photo by rawpixel on Unsplash*

### Formalizing and normalizing the data

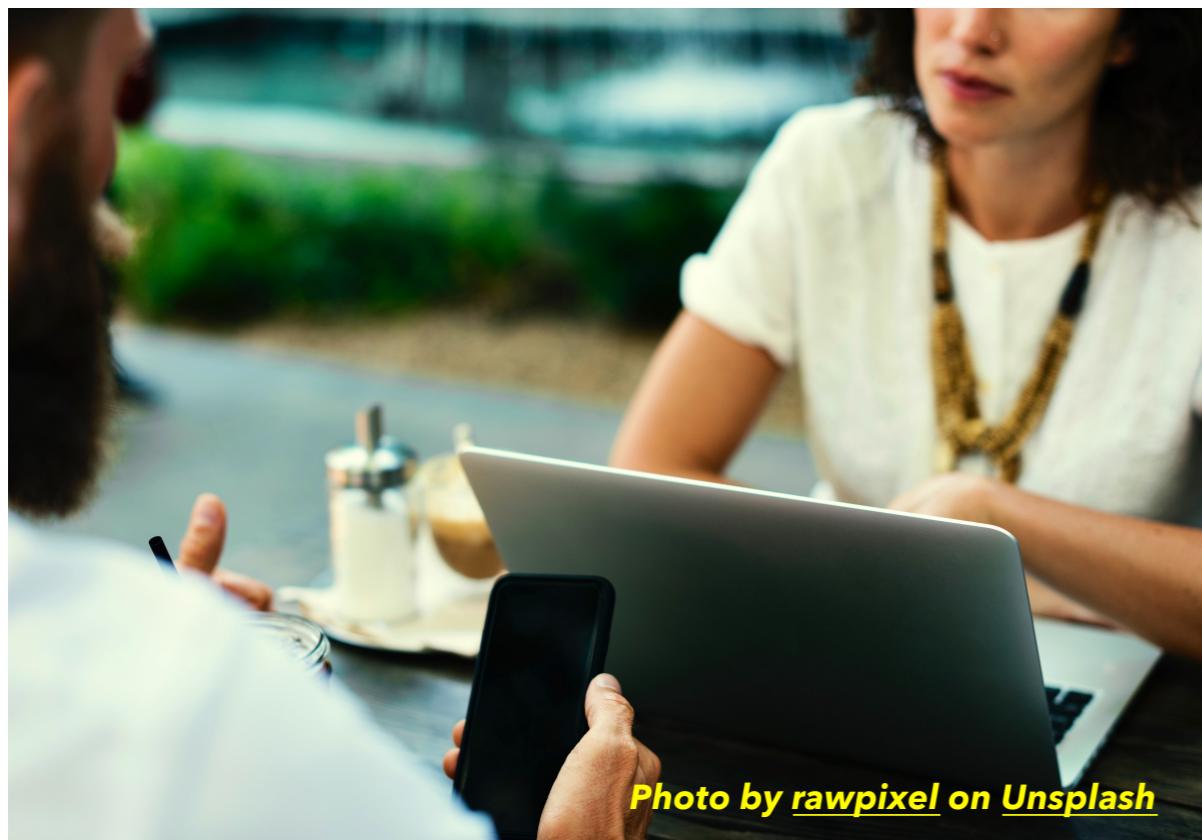
- IBM Watson Studio Cloud has an import feature that automatically codes in Python.
- To ensure easy translation the data was in CSV format (instead of Excel). CSV format eliminates the need to code which is traditionally required in Python.

### Identify the key factors that are represented in the extracted data.

- DataUSA.io data was dated from 2013 to 2017.

### Remove any irrelevant data.

- After further analysis the Educational data was removed. Educational data lacked any significant correlation link to either approving or denying investment legitimacy.
- Removed any comparison to the city of Keego Harbor with the entire United States.



*Photo by rawpixel on Unsplash*

## Decide if additional resources are needed for story telling

- No additional resources were required for the initial investment.

## **Step 4 of 6: Explore The Data**

### How does the data point back to the inadequate human capital concerns?

- Using the “Descriptive” and “GroupBy” statistical methods to support the recommendations is the best approach. It was important to choose charting methods that
  - (a) to show the best representative to data and
  - (b) aid in the efforts to deny or confirms leaderships vested interest.

## **Step 5 of 6: Analyze the Data In Depth**

### Create a data story to resolve the ambiguity associated with inadequate human capital?

- Dr. Luther Jr. will apply some Data Analysis and Visualization methods in Python to Segment and Cluster Neighborhoods information provided by the Foursquare API database.

### Eliminate unrealistic analysis tools.

- After careful analysis advance features of Data Science e.g. Machine Learning methods were not applicable

## **Step 6 of 6: Visualize & Communicate The Findings**

- Dr. Luther Jr. plans to communicate the data in a visual presentation through the use of charting showing vested correlation.

# METHODOLOGY SECTION

MAIN COMPONENT OF THE REPORT, EXPLORATORY DATA ANALYSIS,

INFERENTIAL STATISTICAL TESTING PERFORMED, MACHINE

LEARNINGS AND THE UTILIZATION.

## Exploratory Data

### Analysis

- Basic Visualization: tools utilized were Box Plot, Pie, Line and Bar Charts.

- Advanced Visualization: tools deployed were Geospatial Data to create a map of Michigan (Feego Harbor and Detroit) using Folium a marker was added to display the locations.

- Raw data from [DataUSA.io](#) allows for a generalized conclusion of the growth and human resource potential in Keego Harbor and Detroit, MI.

- Applying Excels Data Analysis tools generates the mean, median, mode, range, standard deviation, variance etc. which can be easily communicated in the data story.

## Inferential Statistical Testing Performed





## Machine Learning and Utilization

- Skilled Data Scientist understand that not all business decisions require advanced techniques. Applying Machine Learning to this particular business problem would be counter intuitive and would failed to address

the concerns identified during the Data Section phase.

- In retrospect, Dr Luther Jr. came to the conclusion that fitting Machine Learning methods would not be the best approach.

# RESULTS SECTION

## OBSERVATIONS AND RECOMMENDATIONS

### Observations

Citizens of Keego Harbor, MI has the demographics to fully staff an Organic Cannabis farm.

- i. Population, as of 2016, is 3,022 with a 0.2% growth rate.
- ii. Medium Age is 37 for Foreign and US born citizens.
- iii. Total number of employees is 1,699 with a growth rate of 8.77 %.

- iv. Largest industries are Retail at 288 per capita and Accommodation & Food Service at 214 per capita.
- v. Education was not a considered factor.



Foursquare API data suggests building retail shops in Keego Harbor or Detroit MI.

## Keego Harbor

- i. There is limited competition for building Cannabis retail shops.
- ii. The top rated businesses in Keego Harbor is Harbor Steakhouse.
- iii. Foursquare identified 30 businesses near the Top Rated business in Keego Harbor
  - None are linked to the manufacturing or distribution of Cannabis.
- iv. Of foursquare's provided list there is only one cross street.
  - Building Cannabis retail shops on both sides of the road at a reasonable distance maybe beneficial.

## Detroit

- i. Foursquare advised that the Central Market is one of the "Top Trending" shopping facilities.

- ii. There are a total of 15 related shop within the search parameters. The list is a very diverse community of businesses
  - None of business are related to the Cannabis industry.

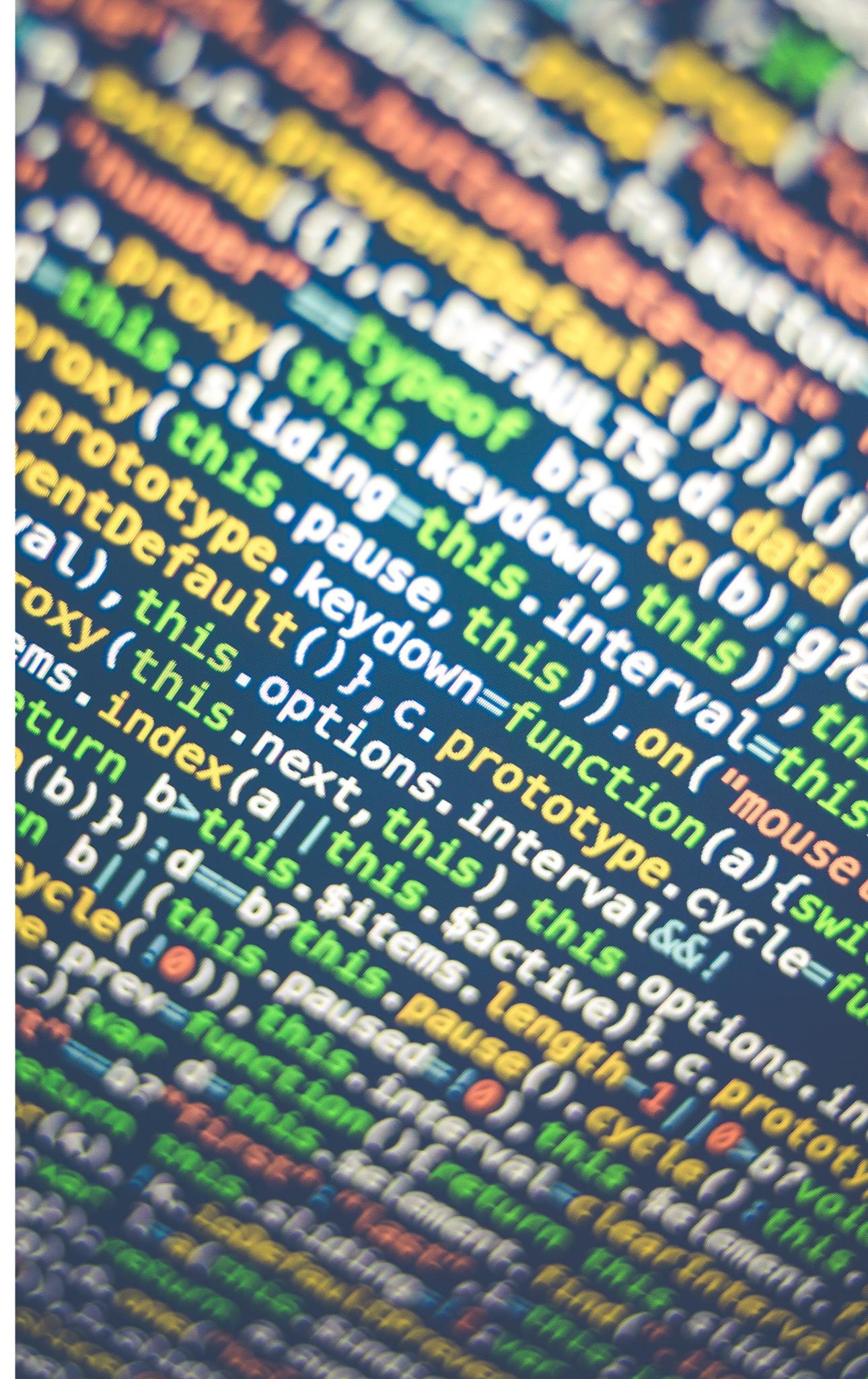
## Recommendations

- Entrepreneurs at Black & Gold Leaf should consider relocating to Keego Harbor, MI and building an Organic Cannabis Farm with the expectation that their staffing needs.
- Entrepreneurs at Black & Gold Leaf should consider the wage of retail and agricultural industries as a mode for hiring.
- Nila Bean Investment is well intended in building Cannabis retail shops near "Top Rated" businesses based on Foursquare's API research feature.
- Nila Bean Investment would benefit from the workforce growth potential provided by Keego Harbor, MI.

# CONCLUSION SECTION

## REPORTED FINDINGS

1. The business problem was an Exploratory endeavor primarily geared on presenting facts as they are.
2. Before the initial investment is made more effort should be taken into understanding the psychological or attitudes towards the use, manufacturing, and distribution of Cannabis in Keego Harbor and Detroit, MI.
3. After careful investigation it was determined that Machine Learning should not be used before the initial investment but it would be a worthy application during manufacturing and distribution.
4. As technology grows so will the skills needed to drive down cost and both organization should consider improving their training and development strategies.



Notebook on GitHub labeled (Battle of Neigh.ipynb):

<https://github.com/DailyDataSci/Battle-of-Neighborhoods.git>

Notebook on IBM Watson Cloud Direct Link (showing map and markers):

[https://dataplatform.cloud.ibm.com/analytics/notebooks/v2/23dc4eb6-95e8-4d50-9814-369191f3b98c/view?access\\_token=20a8a6b6576df012a8c9b1d8b9e77e66081c366a72f064f25f30564fe79c4d35](https://dataplatform.cloud.ibm.com/analytics/notebooks/v2/23dc4eb6-95e8-4d50-9814-369191f3b98c/view?access_token=20a8a6b6576df012a8c9b1d8b9e77e66081c366a72f064f25f30564fe79c4d35)

# Work Cited

Bandyopadhyay, Raj. "The Data Science Process: What a Data Scientist Actually Does Day-to-Day." Springboard, Springboard, 9 Jan. 2017, [medium.springboard.com/the-data-science-process-the-complete-laymans-guide-to-what-a-data-scientist-actually-does-ca3e166b7c67](https://medium.springboard.com/the-data-science-process-the-complete-laymans-guide-to-what-a-data-scientist-actually-does-ca3e166b7c67).

"Keego Harbor, MI." Data USA, [datausa.io/profile/geo/keego-harbor-mi/](https://datausa.io/profile/geo/keego-harbor-mi/).

CBS News. "17 Stoner States: Where's Marijuana Use Highest?" CBS News, CBS Interactive, 25 Oct. 2011, [www.cbsnews.com/pictures/17-stoner-states-wheres-marijuana-use-highest/4/](http://www.cbsnews.com/pictures/17-stoner-states-wheres-marijuana-use-highest/).

Daly, Celia. "News & Articles." California Proposition 64, Part III: Revenue – Who Benefits?, go.canopyboulder.com/news/who-is-investing-in-the-cannabis-industry.

McVey, Eli, et al. "Exclusive: US Marijuana Sales May Reach \$10 Billion This Year, \$22 Billion by 2022." Marijuana Business Daily, Marijuana Business Daily, 27 Sept. 2018, [mjbizdaily.com/exclusive-marijuana-sales-may-reach-10-billion-this-year-22-billion-by-2022/](https://mjbizdaily.com/exclusive-marijuana-sales-may-reach-10-billion-this-year-22-billion-by-2022/).