

# Applied Data Science Capstone S2

Peer-graded Assignment:  
Capstone Project – The  
Battle of Neighborhoods  
(Week 1)

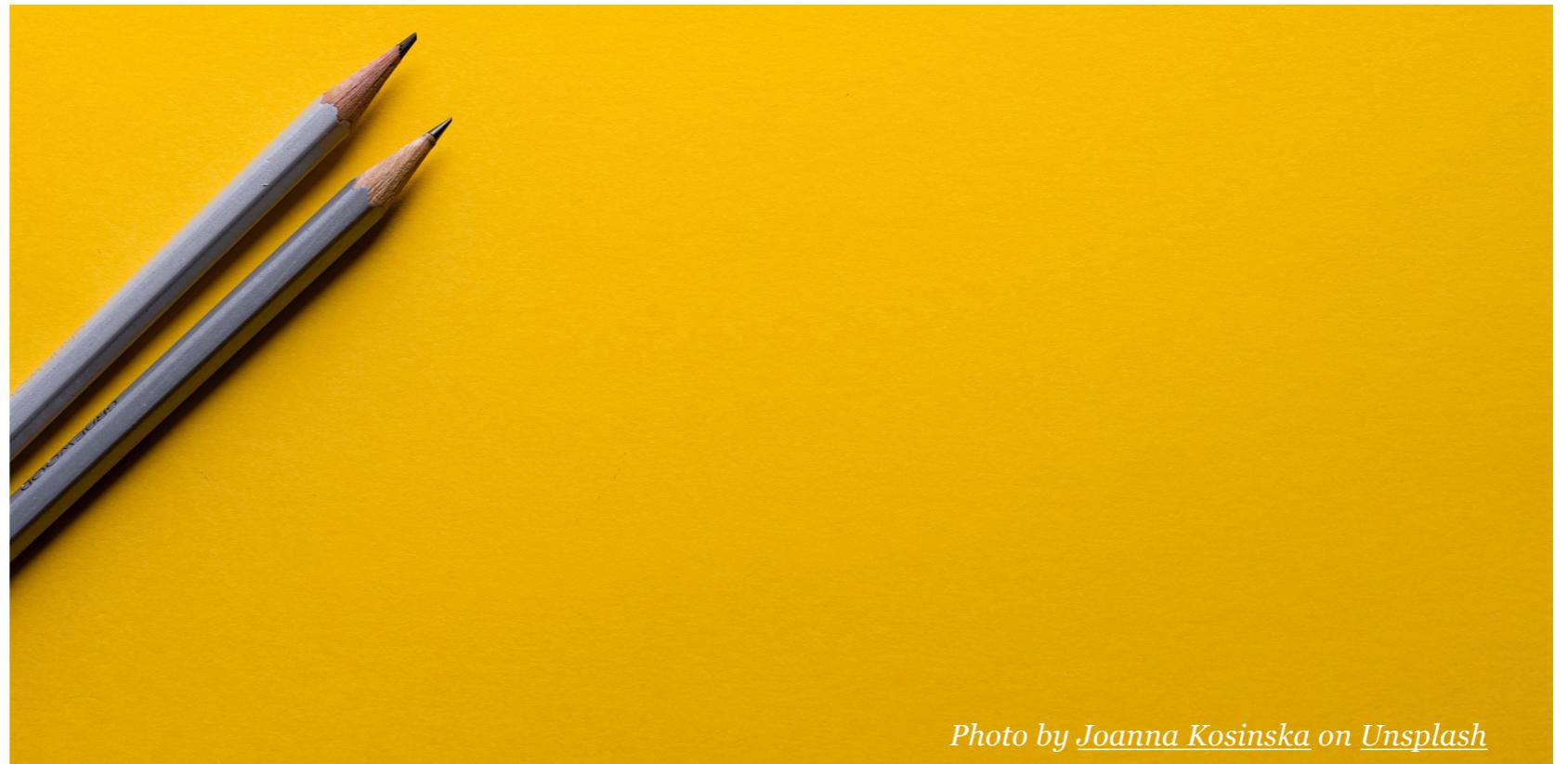


## Section 2 Instructions

Describe the data that you will be using to solve the problem or execute your idea. Remember that you will need to use the Foursquare location data to solve the problem or execute your idea. You can absolutely use other datasets in combination with the Foursquare location data. So make sure that you provide adequate explanation

and discussion, with examples, of the data that you will be using, even if it is only Foursquare location data.

This submission will eventually become your **Data** section in your final report. So I recommend that you push the report (having your **Data** section) to your Github repository and submit a link to it.



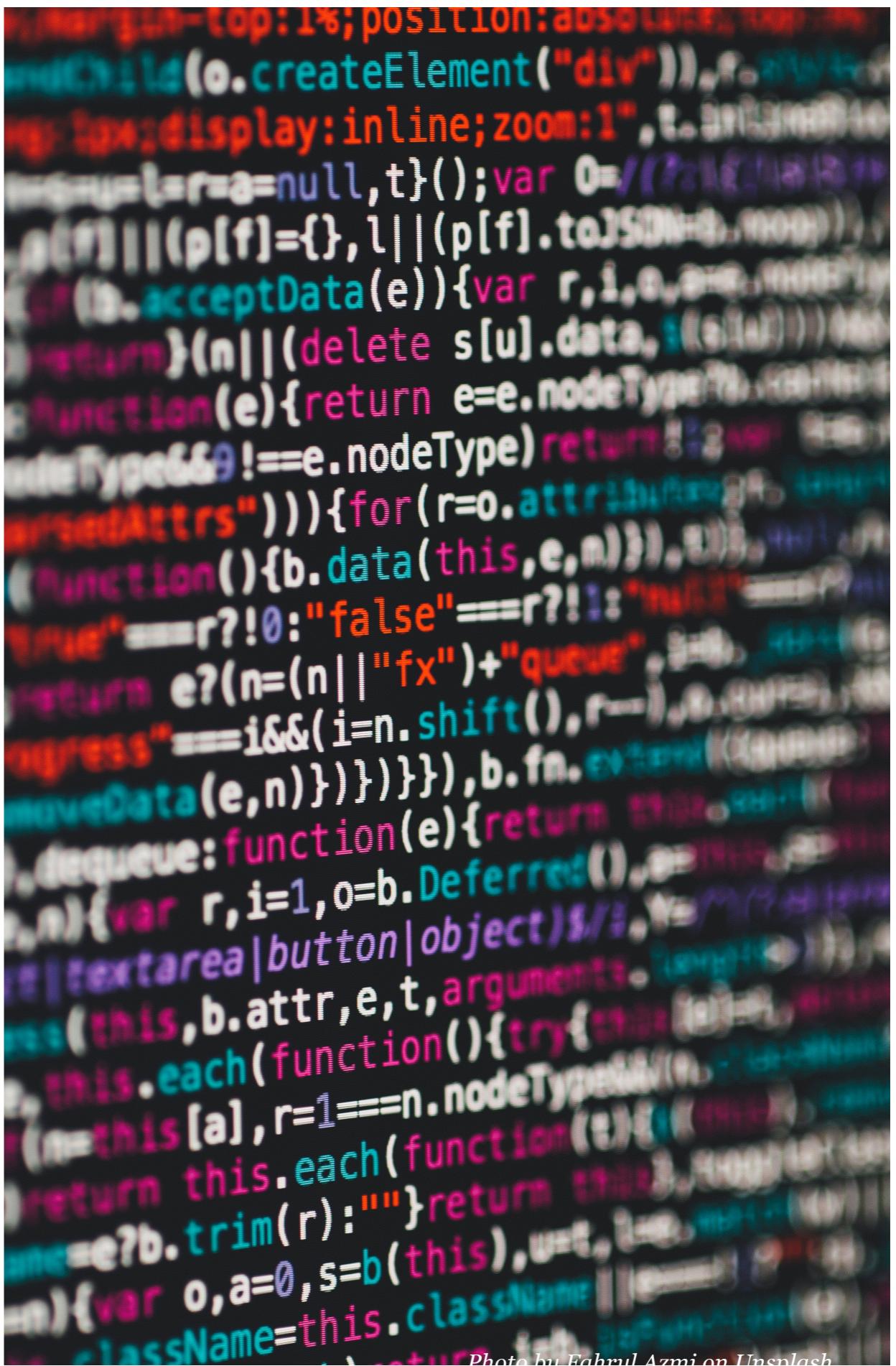
*Photo by Joanna Kosinska on Unsplash*

# The Data

A description of the data and how it will be used to solve the problem. (15 marks)

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 teams 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 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 retail shops in Detroit, MI.



## **Step 1 of 6: Framing the problem.**

### What is the specific business problem?

- Leaderships of both organization are not clear on the Economics, Diversity, Health & Safety, Education or Housing & Living status in the city of Keego Harbor, MI. It's also understood that the primary goal is building an Organic Cannabis farm in Keego Harbor, MI. The next phase is to build retail shops in both Keego Harbor & Detroit, MI. Solving the ambiguity concerns linked to the communities is contingent on understanding the key factors listed above.
- Dr. Luther Jr. has decided that the problem is not an investment problem per-say but one of inadequate human assets. Or a better way of stating it is not “what” is in the city of Keego or Detroit, MI but “who” is located in these areas.

## **Step 2 of 6: Collecting the right data.**

### What data is available?

- Keego Harbor, MI has a very small population and historical data is scarce. [DataUSA.io](#) provides

exceptional unstructured and structured data to solve leaderships questions.

- Dr. Luther will also utilize a few features on [Foursquare](#). For example, ratings, lists, local searches, recommendations, and etc. to align the idea of building retail shops in both cities.

### How should you extract the data?

- Download the raw data via Data USA's website.

### What format to store and perform data analysis?

- Both Excel and CSV files to be stored on the IBM Cloud for compilation.

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

### Apply the following principles to the data:

- Identify any missing values.
- Format and normalize the data.
- Identify if the key factor are represented in the extracted data.
- Decide if additional resources needed for story telling.
- Remove any relevant or irrelevant data.



Photo by [Markus Spiske](#) on [Unsplash](#)

## Step 4 of 6: Explore your data

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

Using the “Descriptive” and “GroupBy” Statistical methods to support recommendations. Wrangle any outliers that confirms and aligns with leaderships vested interest.

## Step 5 of 6: Analyze Your Data In Depth

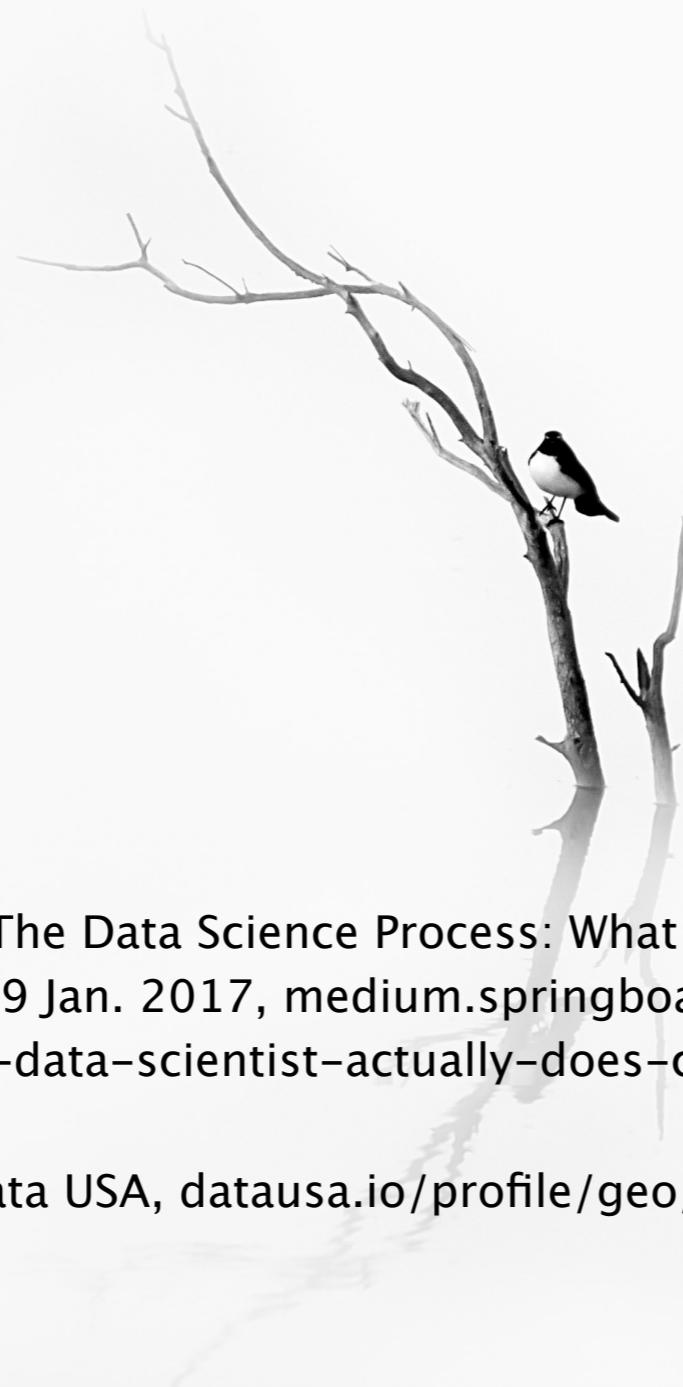
Can you draw on the data, create a data story, and resolve the ambiguity associated to the inadequate human capital?

- Dr Luther will apply some Data Science specialization and analytical skill of Segmenting and Clustering Neighborhoods based Foursquare database. Next he will apply Basic to advanced Data Visualization with Python, and etc.

## Step 6 of 6: Visualize & Communicate findings

- Dr Luther plans to communicate the data in a visual presentation with a focus of exercising the art of Data Storytelling to provide recommendations based on the results.

# 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/).