==================================================================================

**Project report Obesity and Yoga**

==================================================================================

Table of Contents

[Introduction 2](#_Toc46645820)

[Data required 2](#_Toc46645821)

[Methodology 2](#_Toc46645822)

[Discussion 4](#_Toc46645823)

[Conclusion 4](#_Toc46645824)

# Introduction

We know that (child) obesity in the US is a large-scale problem. There are numerous factors that lead to obesity including unhealthy diet, lifestyle, smoking, drinking etc. etc.

Researchers also have found that there are a number of ways to reduce and fight this "epidemic", as it was labelled by "Centers for Disease Control and Prevention" in back in 1999. I believe that Yoga is a good way to fight obesity. Apart from good health, Yoga has many health benefits and is an enjoyable activity for the whole family

It would be good to have Yoga studio(s) at places where it would be most effective. In this project I want to show how we can use data and datatools to I want to use Foursquare API to see where we can start a Yoga studio.

# Data required

We do not want to start opening a Yoga studio just randomly. To combat obesity, we should target a city that has the highest level of obesity. A search on the Internet resulted in a dataset from <https://www.cdc.gov/500cities/index.htm>. The 500 Cities project is a collaboration between CDC, the Robert Wood Johnson Foundation, and the CDC Foundation. The purpose of the 500 Cities Project is to provide city- and census tract-level small area estimates for chronic disease risk factors, health outcomes, and clinical preventive service use for the largest 500 cities in the United States.

We downloaded the health dataset

<https://chronicdata.cdc.gov/500-Cities/500-Cities-Local-Data-for-Better-Health-2019-relea/6vp6-wxuq>. The original file is slightly more than 200 mb in size. We need only the BMI of the 500 cities.

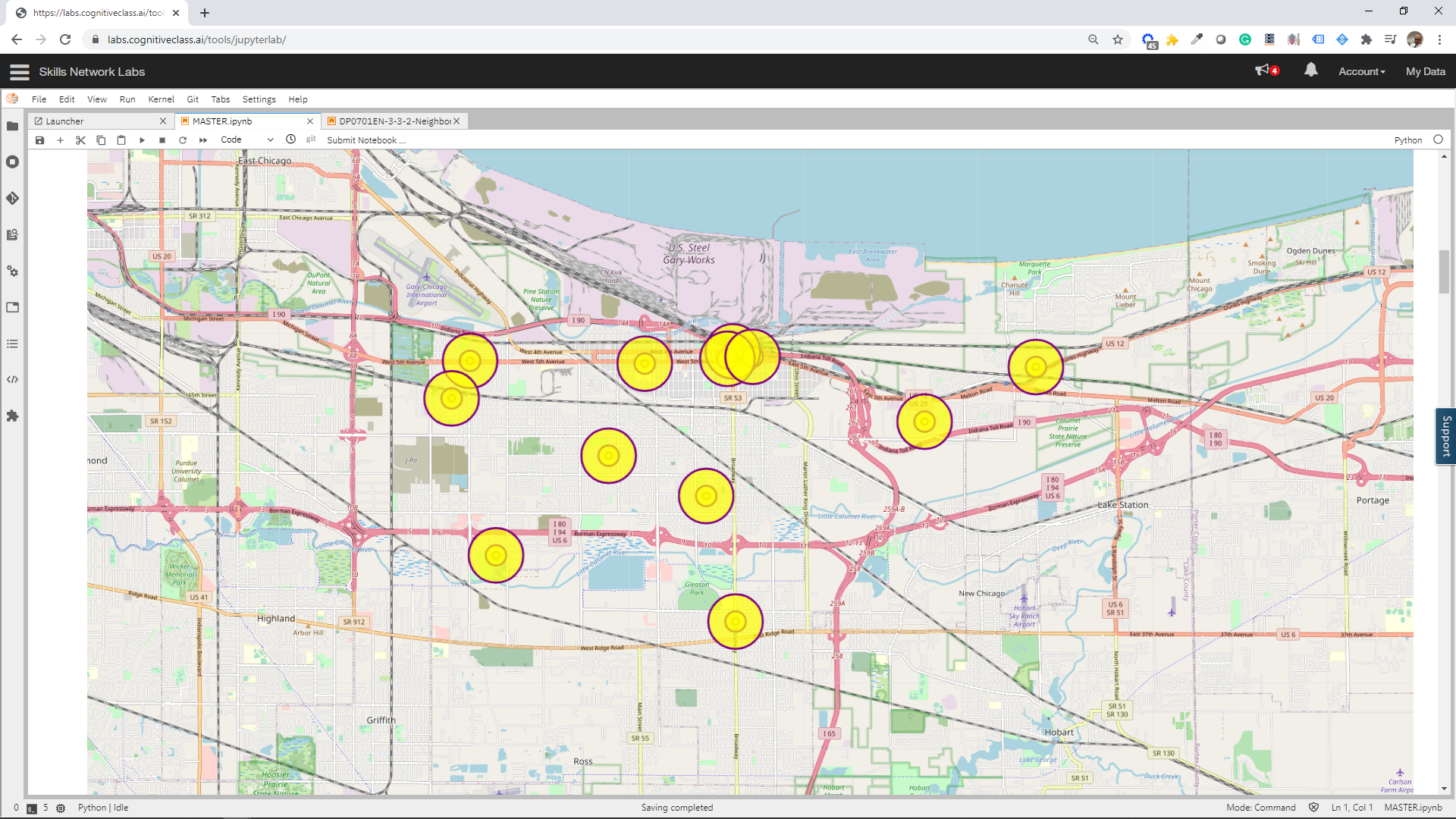
The resulting file can be found here:

<https://raw.githubusercontent.com/naveen1973/coursera_foursquare/master/obesity.csv>

# Methodology

1. We sort by BMI level and found that Gary, Indian has the highest levels of BMI. We will use the Foursquare data to find a where we the best place would be to open a Yoga studio.
2. Then we find the several geo-coordinates of the neighbourhoods of Gary, Indiana.  Please note that we will not use ALL the neighbourhoods as two of them (Pulaski and Morningside Historic District) do not have ANY venues as we will see when using Foursquare API. This gives us an error message.
3. Now we create a map of Gary to see where all the neighbourhoods are (see map - 1).

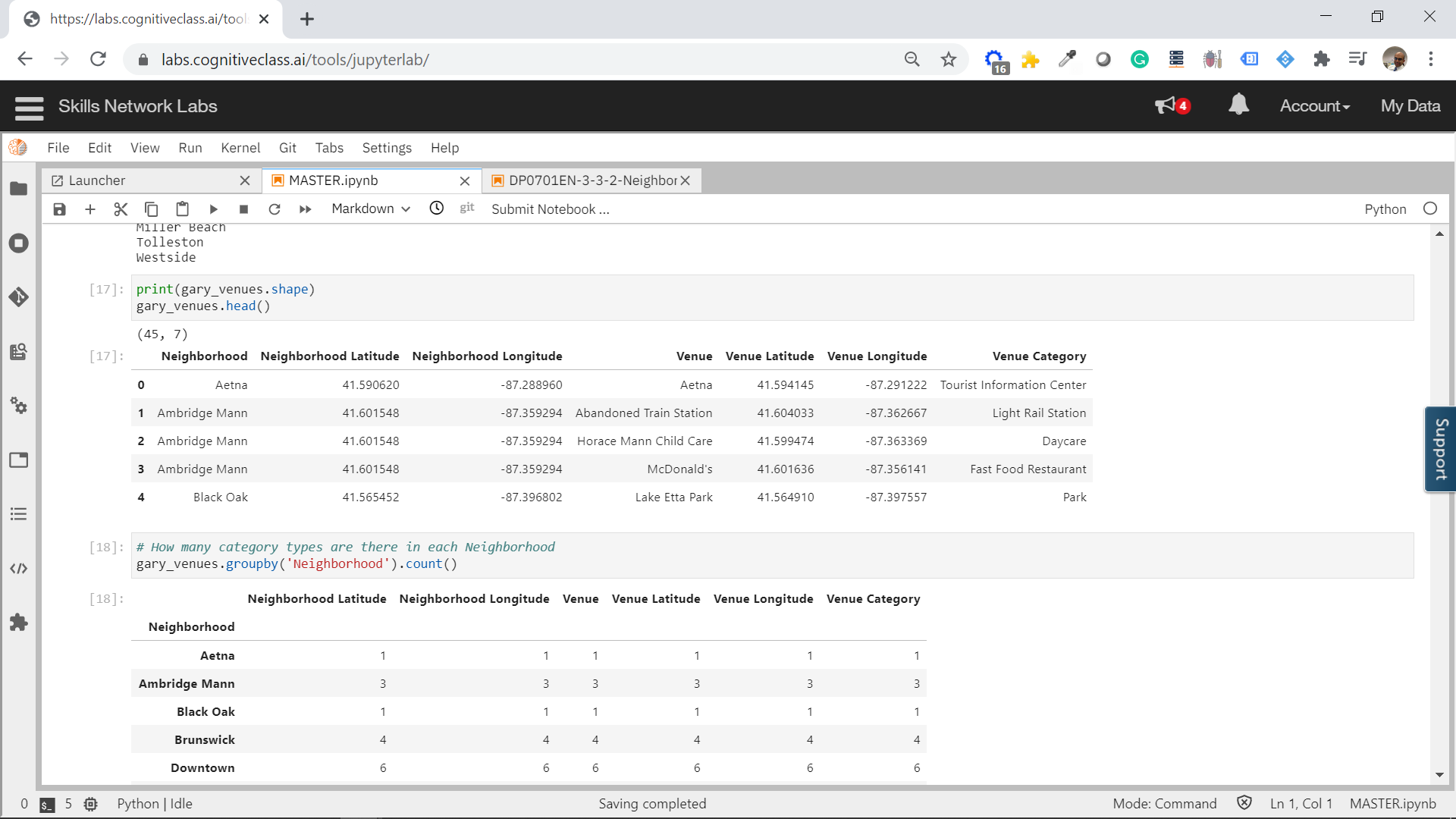
Map 1



1. Explore Neighbourhoods in Gary (Indiana)

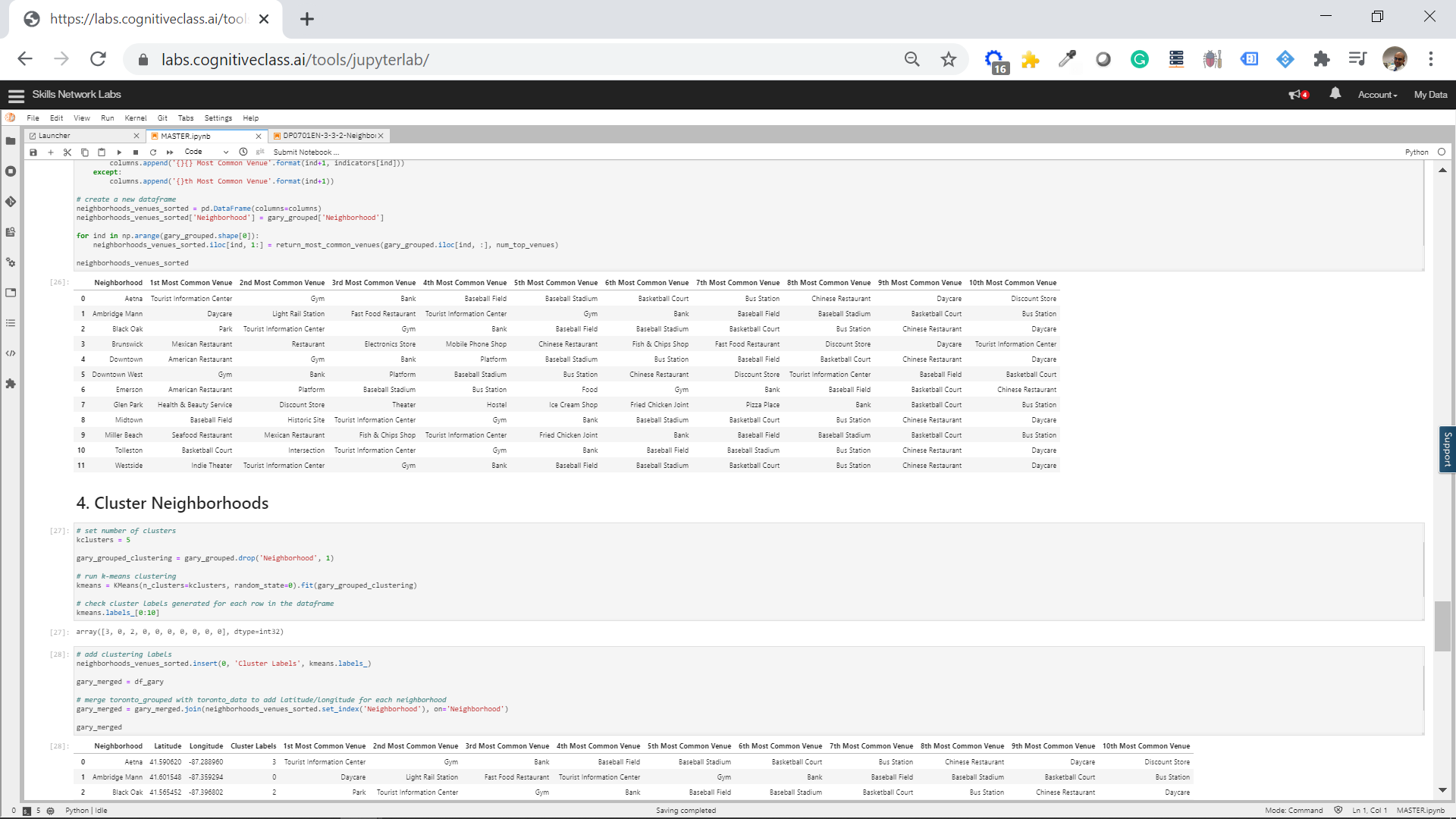
Using the Foursquare API we now explore the neighbourhoods of Gary. Overview 1 displays a sample of the overview of neighbourhoods. It shows the venue in a neighbourhood, its latitude and longitude, and to what category the venue belongs.

Overview -1

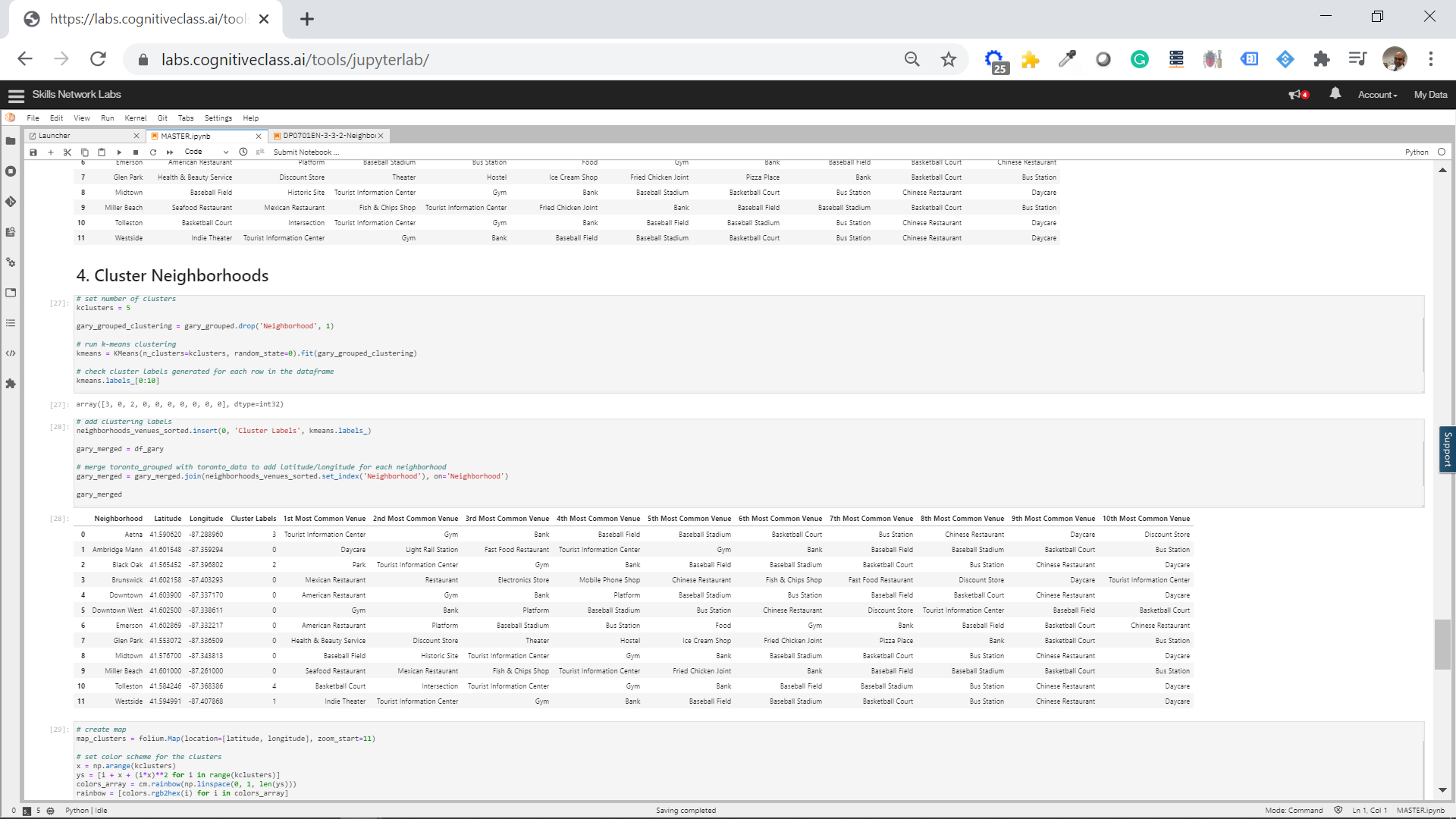


1. Analyse Each Neighbourhood

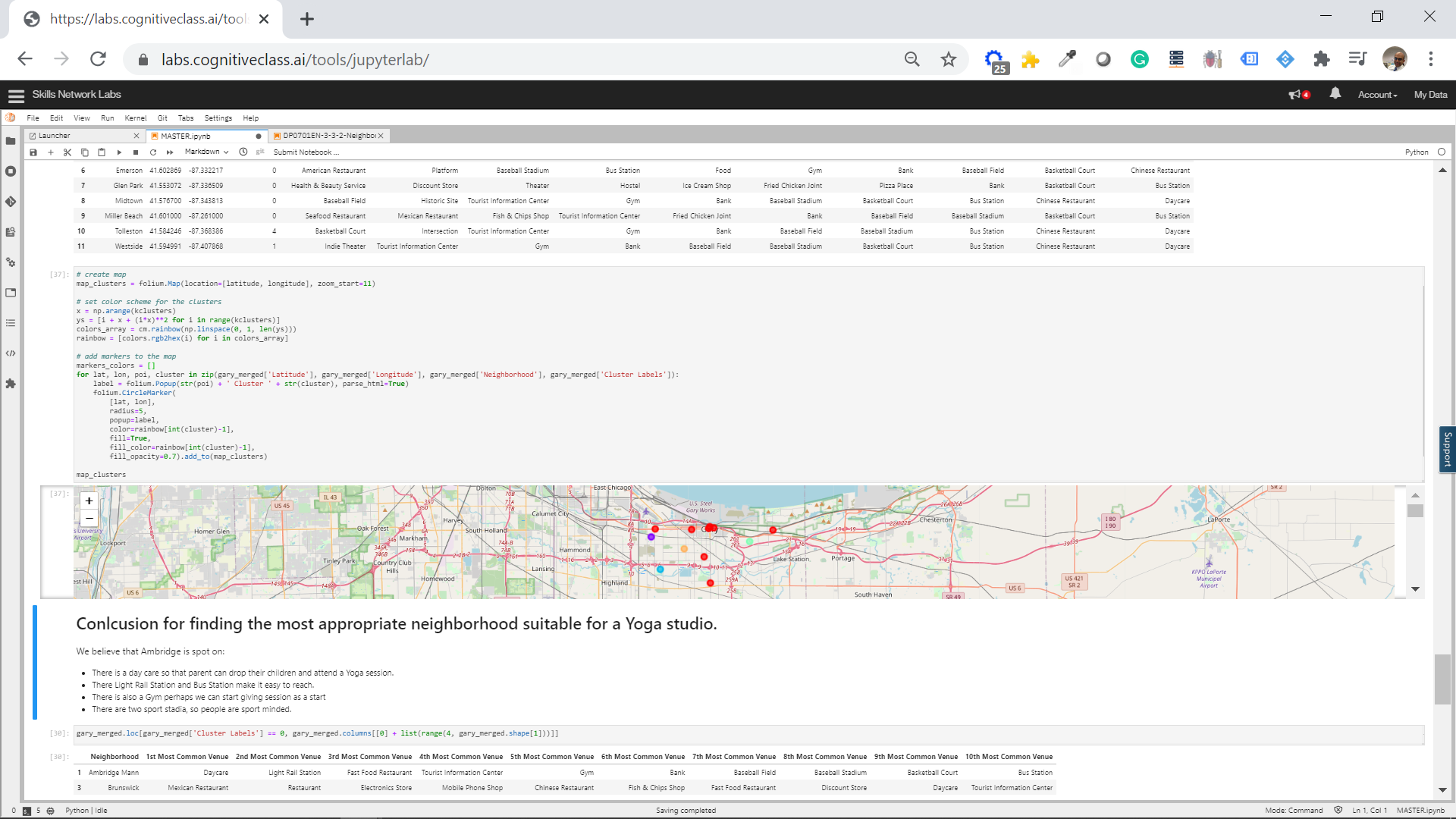
* We want to know how many times a venue type is present in neighbourhood. Using the one-hot method, we make an overview of venues.
* We now group the venues and see what portion a venue represents of all venues.



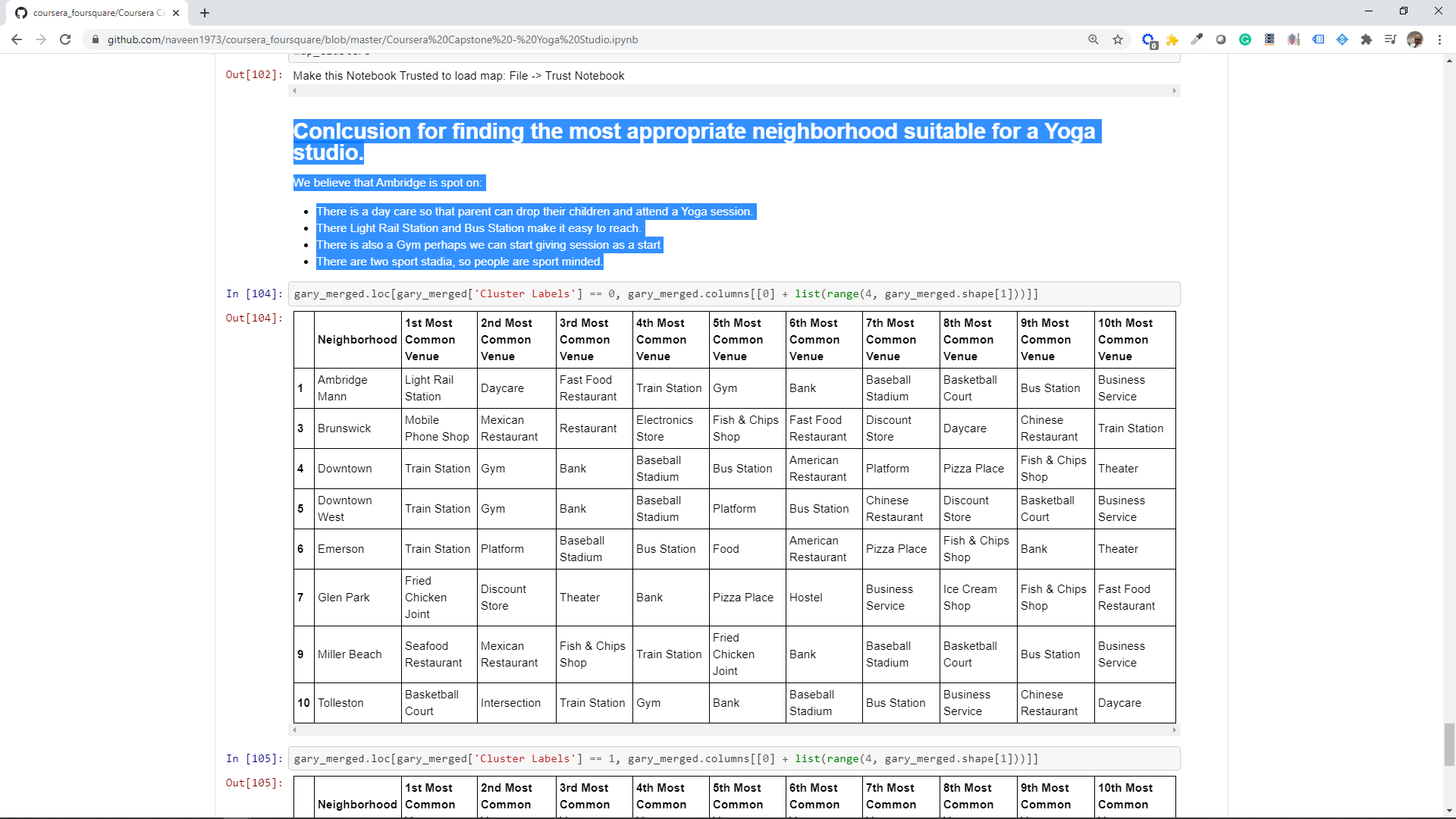
* Now we cluster the groups



…. which we map again



# Discussion



Based on our analysis above, we believe that Ambridge Mann is where we should open a Yoga studio:

* There is a day care so that parent can drop their children and attend a Yoga session.
* There Light Rail Station and Bus Station make it easy to reach.
* There is also a Gym perhaps we can give intro/taster sessions to being with.
* There are two sport stadia, so people are sport minded.

# Conclusion

This is a capstone project and my conclusion is based on our analysis using a specific set of data tools and a limited set of data. While we conclude that Ambridge Mann is a good neighbourhood to start a yoga studio. However, in real-life, we should investigating this further and the conclusion should be only a recommendation to assessing other factors such as:

* Safe for people to travel during the evening.
* (Health) Subsidies available from certain neighbourhoods could make it financially more attractive to start a yoga studio there.
* Real estate prices might not make it economically viable to start a yoga studio there. The yoga classes could maybe be offered as corporate classes as a start.

This project has been very interesting and I have also learned a lot from the reviewing projects of other participants. I think I could make similar analysis for other sectors perhaps using other machine learning methodologies.

I would like to thank Coursera, IBM and my fellow participants for this enjoyable learning journey.