

## Applied Data Science Capstone

# A New Laundromat in Walla, Walla, WA

Nathan Rathge

## Table of Contents

- Introduction/Business Problem
- Data
- Methodology
- Analysis
- Results and Discussion
- Conclusion

## Introduction/Business Problem

A client would like to know the best location for a laundromat in Walla Walla, WA.

A new laundromat would need to be located where there are no other laundromats nearby, but there are other factors. A successful laundromat needs customers. Customers for laundromats may not have a washer and dryer at home, or may be traveling, or may want to do multiple loads at a time. The reasons why a customer may use a laundromat may be varied or mysterious. With the assumption that successful laundromats would have similar items nearby,

data will first be collected about existing laundromats, and then applied to find similar areas of the city that do not have a laundromat.

## Data

Using the Foursquare API, laundromats in Walla Walla will be found. The locations near these laundromats will be analyzed, in the hopes of finding similar venues nearby, or finding patterns of venues. Using that information, a search of the entire Walla Walla metro area will hopefully reveal an area which has similar characteristics, yet does not have a laundromat in the area and may have a need for one.

Some characteristics to look for are:

- similar venues
- distance to travel-related locations

## Methodology

In order to determine the characteristics of an area with a laundromat, the first step is to determine where the laundromats are in Walla Walla. Foursquare has an API method for this, which is used by searching for the desired "location type" code.

The data visualization of the locations on a map by way of Folium are used to see where these locations are. Locations that are in the the same area can be clustered together by using K-means method. This will reduce the number of locations to be searched and also eliminate the possibility that nearby locations are counted more than once.

The venues that are found by the Foursquare search at each location can be counted and sorted, to see if there are any nearby locations that are common to the other laundromat locations. Since some locations may have more nearby venues, the count is "normalized" (or made to be relative to a total number of 1) to make sure this has some meaning.

The next step in the process is to make a grid of points over the entire city. These are equally spaced so there is no bias in their placement.

Foursquare is then searched for venues at all of these points. The locations are counted, sorted and normalized just like the "laundromats" were sorted above. These locations can then be compared to the above locations to see if there are any indicators of venues that are near laundromats.

The final step of the analysis is to use the K-means method, but in a more general way than the spatial analysis used above. When finding clusters of locations for spatial analysis, it is easy to choose a value for the number of clusters. However, when using it generally, or in a multi-dimensional way, several values of clusters are used, and analyzed to find the best value for number of clusters. The locations that get clustered with the laundromat locations are those that are selected as those which may be good for a new laundromat.

## Analysis

The "laundromat" locations are found in Foursquare by using the page:

<https://developer.foursquare.com/docs/build-with-foursquare/categories/>

This page lists all the categories of venues that are used to classify the locations in Foursquare. The types are given a category code used in the search. The page was searched and Laundromat (52f2ab2ebcb57f1066b8b33) was found and directly below it, Laundry Service (4bf58dd8d48988d1fc941735). Seven locations were found. These locations are shown in Figure 1 (two locations are very close to each other and appear to be the same dot).

It is clear from the map that there are four areas. The centroids of these clusters are found using a k-means algorithm with number of clusters equal to 4. The following four GPS coordinates are returned as the "laundromat" locations.

46.0674, -118.3352  
46.0396, -118.3870  
46.0755, -118.3077  
46.0552, -118.3319

Using an internet mapping program, it was determined that the two closest clusters were approximately 1 mile apart. This led to a determination that a radius of half a mile, or approximately 804 meters, is a good number to use as the

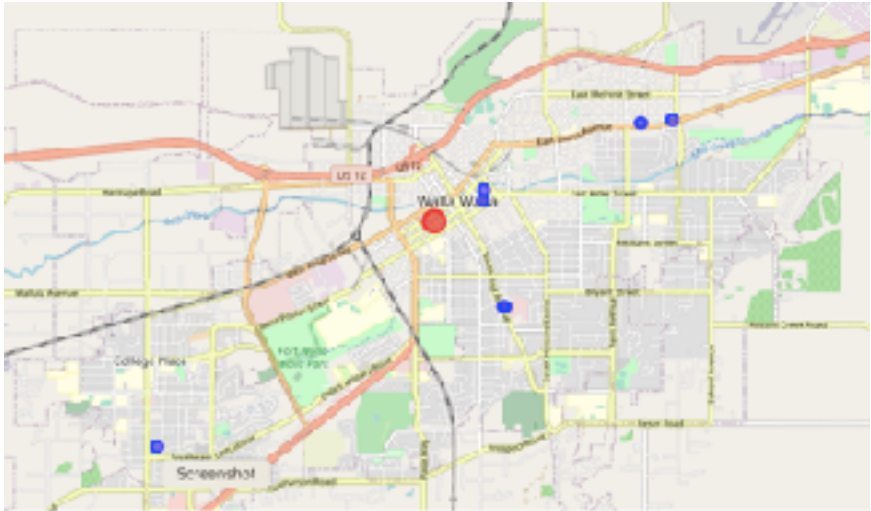


Figure 1: Locations of Laundromats in Walla Walla

radius parameter in the Foursquare search. For each one of these four locations, Foursquare was searched to find all the locations within a half mile radius. Ninety-five venues were found in total. The number of venues near each location is shown in Figure 2.

Neighborhood Coords	
46.03961569112464, -118.38701038135303	11
46.05518179797383, -118.33192727514648	5
46.06742819762553, -118.33520471888028	48
46.076538, -118.3077025	31

Figure 2: Number of Venues Found at Each Location

With each location having a drastically different number of different venues, it is difficult to make a good comparison to see if there are a lot of a certain type of venue near the laundromats. Therefore, it is necessary to collect all the venue types together, and then normalize the data, so it is set to a basis of 1.

The results of this exercise are shown in Figure 3. By analyzing this figure, there does not appear to be any pattern that stands out. There are a large number of places that serve food near each location, but there is a large variety and no clear indication of what type. One location has a post office & big box store, another has a credit union & convenience store, and yet another has a hotel & wine shop.

```

----46.03961569112464, -118.38701039135303----
      venue  freq
0  Pizza Place 0.09
1  Big Box Store 0.09
2  Coffee Shop 0.09
3  Post Office 0.09
4  Sandwich Place 0.09

----46.05516179797363, -118.33192727514648----
      venue  freq
0  Bakery 0.2
1  Grocery Store 0.2
2  Mexican Restaurant 0.2
3  Credit Union 0.2
4  Convenience Store 0.2

----46.06742819762553, -118.33520471888028----
      venue  freq
0  American Restaurant 0.08
1  Hotel 0.08
2  Lounge 0.04
3  Winery 0.04
4  Wine Shop 0.04

----46.075538, -118.3077025----
      venue  freq
0  Fast Food Restaurant 0.13
1  Coffee Shop 0.06
2  Chinese Restaurant 0.06
3  Mexican Restaurant 0.06
4  Sandwich Place 0.03

```

Figure 3: Top Five Venues Near Each Laundromat

The next step is to create a grid of points across the entire city. Spacing of approximately a mile is used in order to integrate well with the existing grid points. The grid is shown in Figure 4.

The Foursquare API was used to search for all the venues near these 40 locations. The data was counted and collected as they were at the four clusters above. Venues were only found at 36 of the 40 locations. The number of venues and type of venue were collected and normalized. Again, there was no apparent pattern of venues when the top five venues were collected together.

K-Means analysis is used to uncover which points are similar to the four laundromat points. The K-means method requires a number of clusters to determine similarity. K-means was ran 29 times (each value from 2 to 30) to determine the best cluster value. Each time, clusters were generated, and then plotted on the map. Similar clusters were plotted with the same color (see Figure 5). The best value was found to be 19, which is where the smallest number of

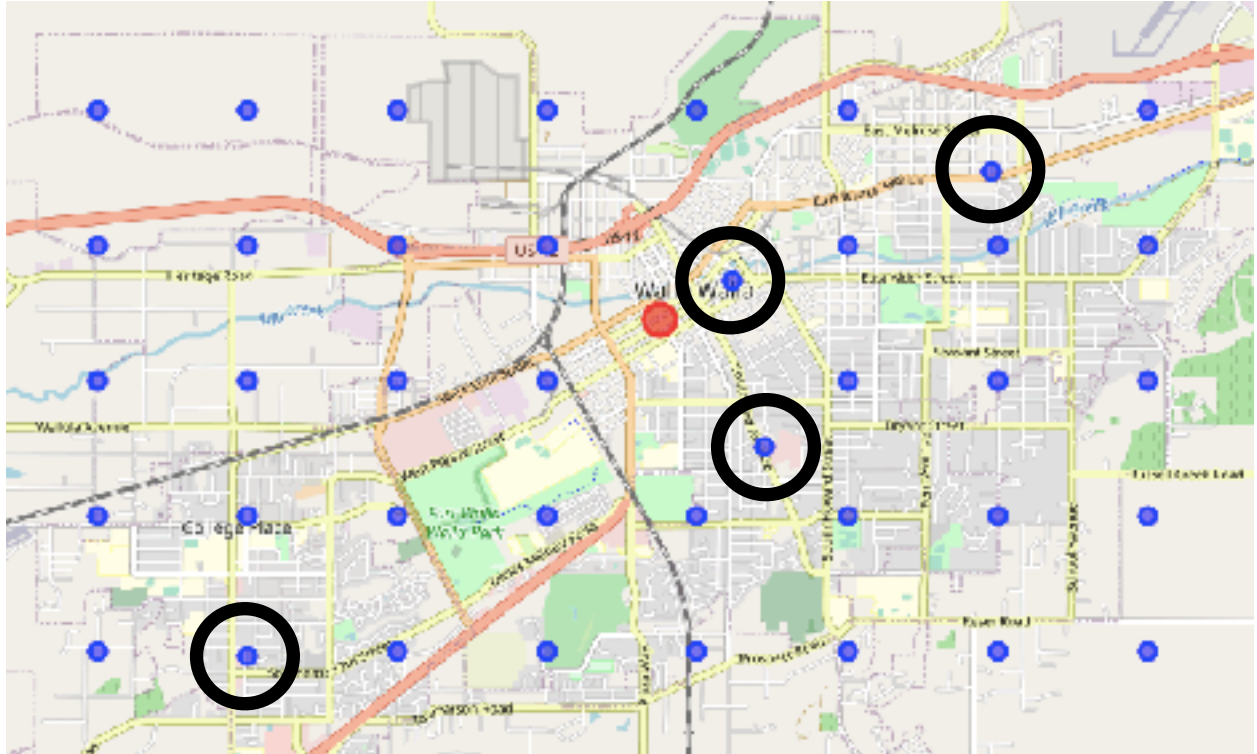


Figure 4: Grid of Points Used for Foursquare Search  
(Laundromat clusters circled)

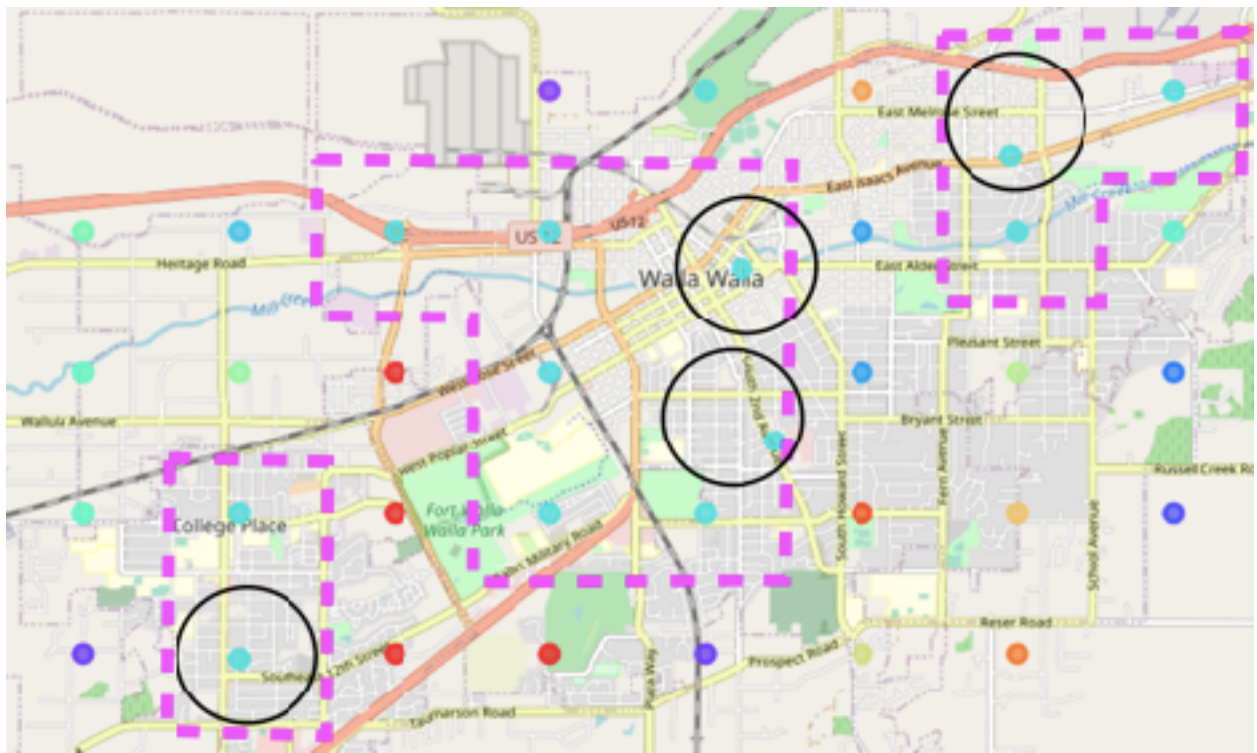


Figure 5: Colored Grid of Clusters  
(num\_clusters = 19)

new clusters were clustered with all of the laundromat clusters.

For k-means, it is also important to use a method to measure error. Two methods were used and graphed, which is supposed to find an "elbow" point where the error does not reduce much, but does not create too many clusters. However, no elbow point was found in either graph.

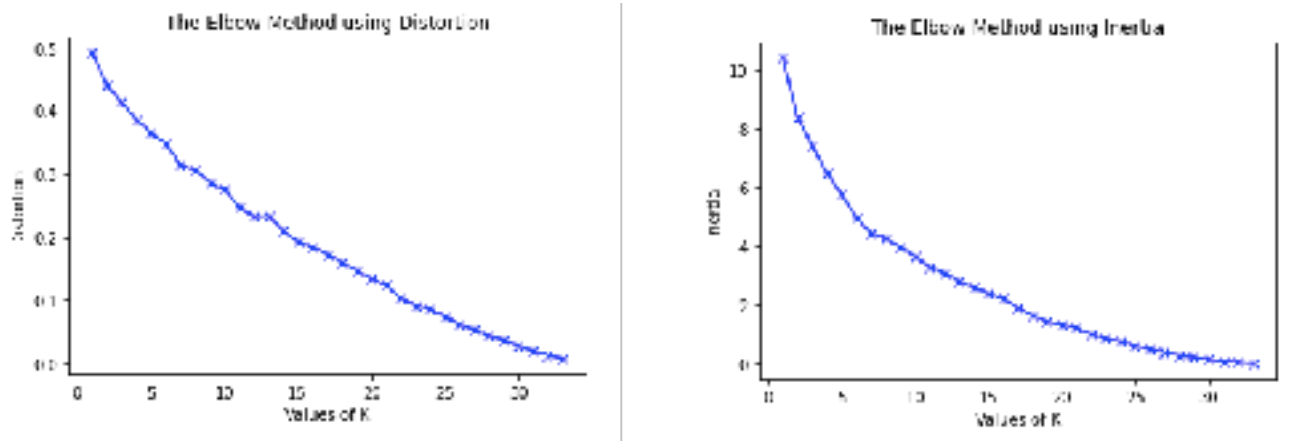


Figure 6: K-Means Error vs Number of Clusters

## Results and Discussion

A total of eight clusters were found that were similar to the four laundromat clusters. By looking at the map, three of the eight clusters were located near Highway 12, and one was located near a large park. The remaining four clusters were chosen as they are located in areas with small side streets. Each cluster is sufficiently far enough away from the other laundromats and have been determined to be sufficiently similar to the locations with laundromats.

The four locations are College Park (46.05, -118.387), Downtown West (46.06, -118.355), Jefferson Park (46.05, -118.339) and Mill Creek (46.07, -118.307). The client can choose from one of these four locations to determine a good location for the laundromat based on price or other options.

## Conclusion

This analysis of Walla Walla to find the best locations for a new laundromat turned out to be a good example of how data analysis can uncover hidden relationships. Doing simple things like collecting the types of venues and counting their frequency does not uncover anything evident. The k-means analysis uses many features to analyze the error and can determine similarity even when the similarity cannot be easily visualized.