Coursera Capstone Final Report: Siting a new coffeeshop in Denver

Manuel Schnabel

September 2nd, 2020

Introduction

The objective of this study is to make an informed decision on where best to site a new coffeeshop within the city of Denver. This could be of interest both to entrepreneurs looking to start a new, local business, as well as large chains such as Starbucks or Caribou Coffee who are looking for additional opportunities to increase revenue. Several factors affect where locating a coffeeshop makes sense. For example, a new coffeeshop should not be located right next to competing coffeeshops, especially if those are already popular. However, the absence of competitors is not sufficient to establish that a location is viable because some areas may be devoid of existing coffeeshops for a reason (for example, they may be industrial areas, military bases, vacant land, etc). The objective of this study is to find a location with few competitors but which is generally viable for opening a coffeeshop.

Data

The question of where in Denver to site a new coffeeshop will be tackled using Foursquare data, Folium maps, and clustering. Denver is a large city, exceeding the typical search radius of Foursquare queries. Furthermore, it is an urban sprawl, branching into suburbs with no clear divisions between towns. For simplicity, the area to be studied is defined as a rectangle between 39.68 and 39.78° latitude, and -104.94 and -105.05° longitude. This approximately lines up with the boundaries of the city of Denver and represents effective limits between city and suburbs in the recollection of the author who used to live there.

This area will be rastered with calls to the Foursquare API. Three types of data will be collected: coffeeshops (venues/search? call, search_query='COFFEE'), trending coffeeshops (venues/trending? call, search_query='COFFEE'), and venues (venues/search?, no search_query). The purpose of the last call is to verify whether areas with no coffeeshops contain other businesses, or if they are simply devoid of businesses entirely and therefore a bad location for a new business.

The data will be mapped using Folium maps. Regions with few coffeeshops, and far from trending coffeeshops, will be identified by eye, and my K-Means clustering. The city will be divided up into ~1km2 squares, the number of coffeeshops and total businesses in each will be summed, and mapped to find areas with many businesses but few coffeeshops. The Folium map itself will be used to check for the type of area (residential, commercial, industrial, military, empty). Based on these analyses, suitable locations for a coffeeshop will be identified.

Methodology section

The location data for coffeeshops, trending coffeeshops, and all businesses was called using the Foursquare API (all of these calls are regular calls, ratings are not called as those are premium calls). The data is received as a dictionary, which contains the results in a list nested within it. Concatenation of lists A and B can be performed as A=A+B, and this procedure is used to build up 3 concatenated lists: coffeeshops, trending coffeeshops, and all businesses. Then each list is converted to Pandas Dataframe and duplicates are dropped. Finally, only the necessary columns are kept: latitude, longitude, name, and id (to serve as a key or index).
The commands .head, .shape, and .size are used to get a feel for the size of the datasets. These revealed that there were only 1-2 trending coffeeshops in Denver, making competition from trending coffeeshops irrelevant in most parts of the city. On the other hand, the size of the dataset with all businesses was >10'000, making it futile to plot all of them on a map individually. Instead, a double 'for' loop over latitude and longitude was used to cluster cafes and total businesses by the 1km2 they were in. This was visualized by plotting a circle with radius proportional to the number of cafes or businesses at each grid point.
Two maps were generated: one showing all cafes individually, and one showing coffeeshops number and total business number per gridpoint. Furthermore, in order to find locations with few coffeeshops that might have escaped the naked eye, the coffeeshops were clustered using K-Means clustering, and mapped color coded by cluster with together with a large circle centered on the cluster center and with radius commensurate with the cluster size. This third map revealed additional potential locations for a coffeeshop within the city of Denver.

Results section

Figure 1 shows a Folium map with the coffeeshops within the city of Denver. The rastered area is also drawn in as a rectangle. It can be seen that there is only one trending coffeeshop (purple), so whether coffeeshops are trending or not can largely be neglected. However, it is immediately obvious that there is a large region in the south-west of Denver with no coffeeshops. It can also be seen from the Folium map that the south-west of Denver is residential, making it in principle suitable for a coffeeshop.

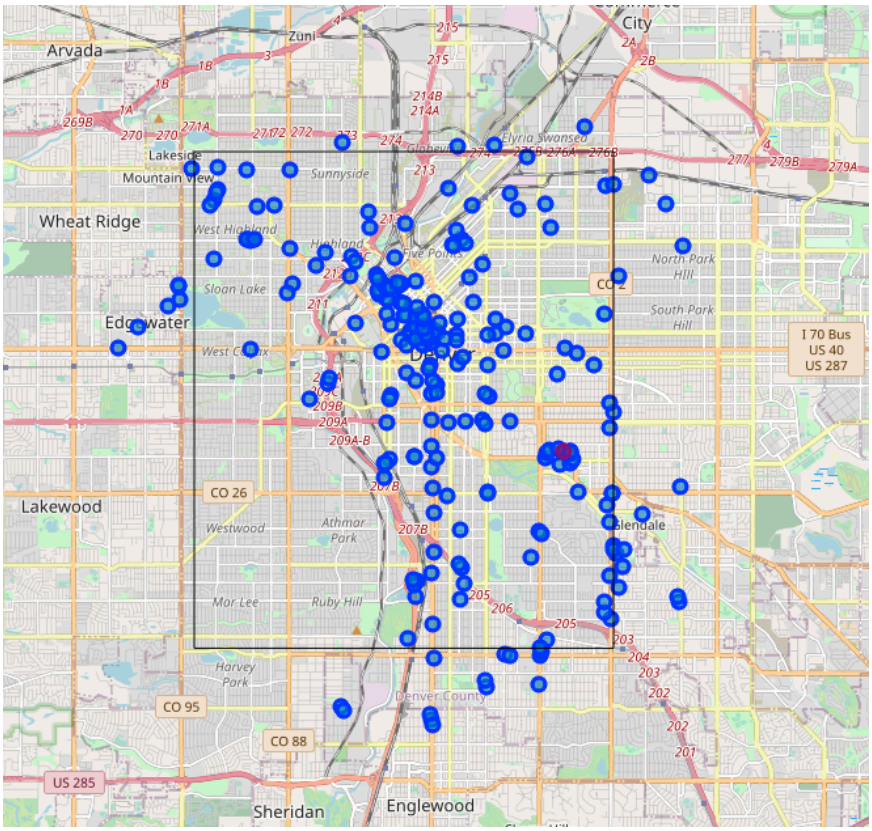

Figure 1. Folium map of coffeeshops within the city of Denver (black rectangle). Coffeeshops are shown in blue, trending coffeeshop is shown in purple.

However, it is possible that there are other reasons not to site a business there – perhaps residents prefer to drive to a large mall outside the city limits for all their needs, or perhaps it is a dangerous neighborhood where it is unviable to open any kind of business. To examine whether this is an issue, the number of coffeeshops and number of all businesses were aggregated by ~1km2 square (strictly speaking, a 0.01° by 0.01° square).
Statistics were collected on number of coffeeshops and venues using dataframe.describe(), and it was found that the number of coffeeshops varies much more across grid squares (standard deviation > mean, min=0) than the number of total businesses (standard deviation < mean, min>0). Since there are some businesses everywhere, but some places have no coffeeshops, we expect that there will be locations with no existing coffeeshops but with existing other businesses that might draw customers to a new coffeeshop. To visualize this, number of coffeeshops and total businesses aggregated by grid square are displayed in Figure 2, which shows a map of Denver with circles whose size is proportional to the number of coffeeshops or businesses.

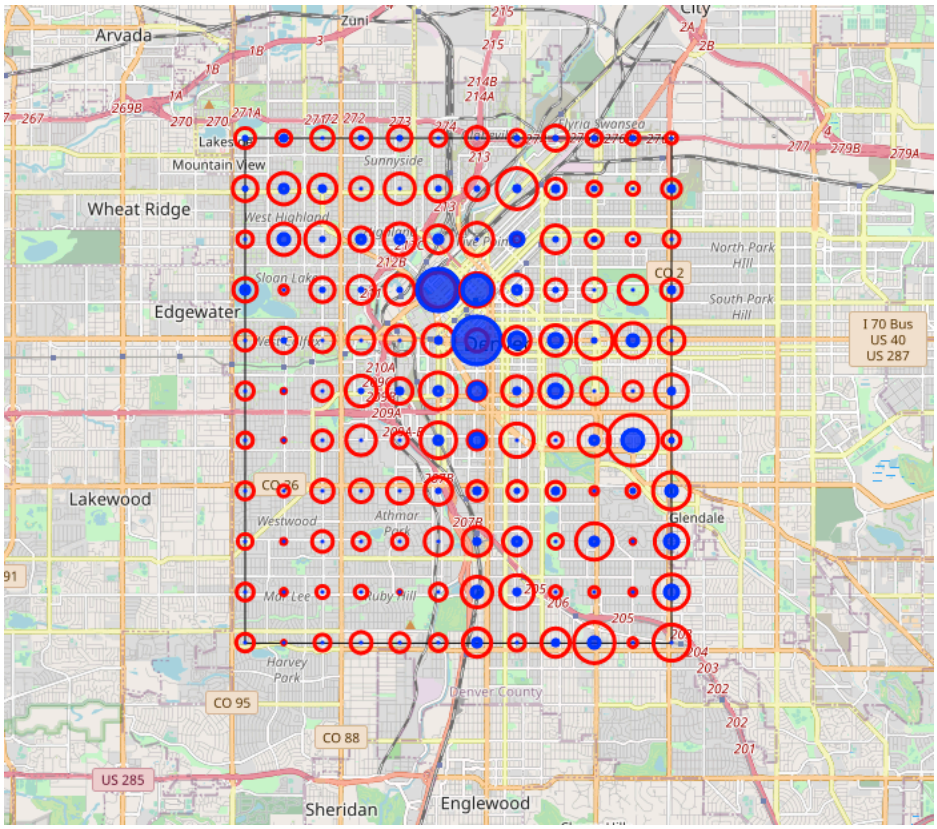

Figure 2. Folium map of aggregated number of coffeeshops (blue) and businesses (red) within the city of Denver (black rectangle). Number of coffeeshops and businesses is aggregated by ~1km2 square, diameter represents number of each. A given diameter represents 10x more businesses than coffeeshops.

It can be seen in Figure 2 that the highest coffeeshop density is in a few downtown grid squares, as we might expect. In the south-west, we see that one line at the very west seems to have very few businesses in total, whereas other grid squares in the southwest have quite a few businesses (large red circles) but few coffeeshops (small blue circles). On the other hand, the north-east has few total businesses but actually does have a few coffeeshops, suggesting the north-east would be a terrible place for a new coffeeshop.

In order to gain further insight that is less prone to outliers, K-Means clustering of the coffeeshops in Figure 1 was employed in order to find areas that lie outside of a coffeeshop cluster. K-Means optimization was run for various numbers of clusters. The .inertia_ attribute was called for each value of cluster number as a measure for quality of fit, and is shown in Fig. 3. It can be seen that while the fit continues to improve for more clusters, there is a “knee” at a cluster number of 5, which is generally taken to be the optimum number for fitting without overfitting. K-Means clustering was performed once more with 5 clusters, the clusters were color-coded by cluster, and are shown in Figure 4. The coffeeshop data was also grouped by cluster (cafe.groupby('ClusterID').count()), indicating that 4 roughly equally large clusters with 25-40 coffeeshops each, and one large “downtown” cluster with 99 coffeeshops was formed.

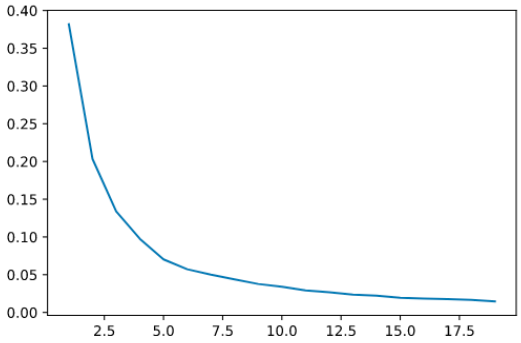

Figure 3. Inertia value of K-Means clustering of coffeeshops (vertical axis) as a function of number of clusters (horizontal axis).

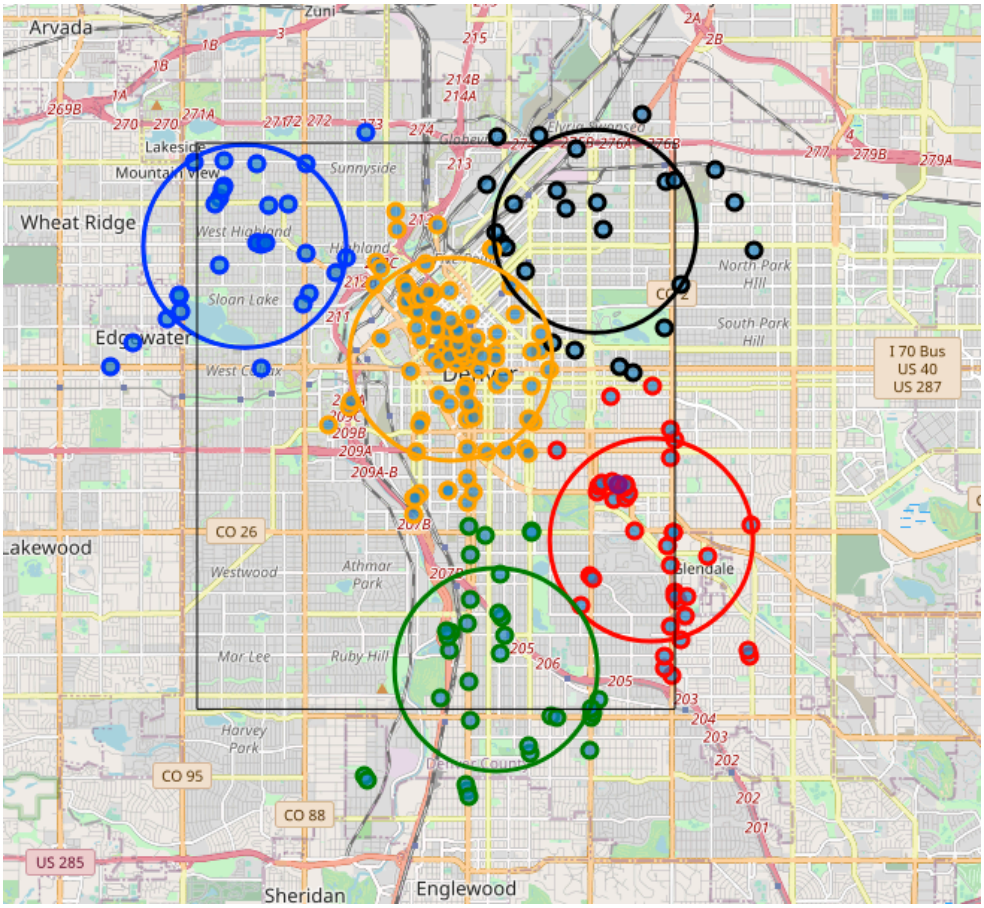

Figure 4. Folium map of coffeeshops from Fig. 1 after K-Means clustering, color-coded by cluster. A large circle indicating the approximate extent of each cluster is also shown for each cluster.

Discussion section

Examining the locations of existing coffeeshops and existing trending coffeeshops (Fig. 1) revealed two things: there are too few trending coffeeshops to use them as a guide to where not to site a new coffeeshop, and there is a total absence of coffeeshops in the southwest of Denver, even though inspection of the Folium map itself revealed that it is a residential neighborhood. Clustering (summing) coffeeshops and all businesses by 1km2 grid square revealed that businesses are much more evenly distributed than coffeeshops, suggesting that some business-friendly areas must exist which have no coffeeshops. Mapping these distributions (Fig. 2) showed that while some parts of Denver's southwest indeed have few businesses at all, other parts do have numerous businesses but simply few coffeeshops, suggesting that this would be a great place to put a coffeeshop, e.g. the Athmar Park neighborhood. Conversely, while the northeast has few coffeeshops, it has very few businesses in total, so it is not an optimum location for a coffeeshop. K-Means clustering (Fig. 4) indicates that there are additional smaller areas due north and due east of downtown with few coffeeshops. While the area due north can be seen, by examining the Folium map, to be a busy railway/highway intersection where no one can stop for coffee, the area due east might also be worth considering for a new coffeeshop, since there are two parks nearby that might draw customers.

Conclusion section

The objective of this report was to find suitable areas within the city of Denver for siting a new coffeeshop. Plotting individual coffeeshops, and coffeeshop density, revealed that the southwest is devoid of coffeeshops. By additionally examining the total number of businesses, a sub-section of Denver's southwest, the Athmar Park neighborhood and a few blocks immediately north of it, was identified as the ideal place to site a new coffeeshop, given the scarcity of coffeeshops and reasonable density of other businesses in the area. The northeast is devoid of coffeeshops and businesses and a whole and would therefore not be suitable. Downtown Denver already has too many coffeeshops, whereas an area due east of downtown, with few coffeeshops but two parks, would be a secondary viable location for a new coffeeshop.