Commercial Development of Vacant Land Adjacent to Public Transit in Chicago

Introduction

In Chicago, Illinois, the City owns two vacant lots adjacent to the Harlem-O'Hare Blue Line public transit train station and bus depot. The City's proposal to develop these lots into a Starbucks and a car wash was met unfavorably by the community. This project determines the ideal commercial development for this land.

Data

This project utilizes a csv of CTA train station locations provided by the City, and venue information provided by Foursquare. The data must be cleaned to remove duplicate station entries, since a single physical station location has an index for every train line that stops there and in each direction the line travels, which would cause any nearby venues to appear multiple times.

Methodology

This project begins with the following assumptions:
1. The City's primary consideration is establishing businesses that will be successful and increase tax revenue.
2. The most numerous existing businesses near CTA stations are the most likely to be successful near this specific station, provided the business does not already exist.
3. 'Near' is defined as within 250 meters, or a little over one block.

Since this data is almost exclusively categorical, analysis is primarily done through one-hot encoding and k-means clustering. This project examined the aggregate venues throughout the system, as well as the cluster of stations most similar to Harlem, and also the 15 stations along the O'Hare branch of the Blue Line.

Results

The aggregate data set included more than 2,500 rows and 295 unique venue categories.

By far, the most common venue adjacent to CTA stations is Coffee, with 127 venues. The next most popular is a Sandwich shop, with 100 venues. The next top 8 are Pizza (76), Mexican food (69), Gyms (60), Bars (59), Chinese food (57), Donut Shop (50), Hotel (47), and Italian food (46).

When examining the data subsets of similar clusters, and stations along the same section of train line, the most common venues are largely unchanged. One notable exception is the absence of any hotels, except for two near the Clark & Lake superstation where the O'Hare branch of the Blue Line ends and the Loop begins. A potential area for further exploring the data would be identifying redundant venues found in the Loop, since the number and density of train stations could cause overlapping venue results within this geographic zone.

Discussion

The findings of this report recommend the community approve the development of a coffee shop, given the high likelihood of this venue category successfully generating revenue for the city. It further recommends rejecting the proposed car wash, which notably had zero venues adjacent to any station in the entire system. Instead, it proposes one or more of the top 10 venue categories not already adjacent to Harlem (excluding Hotel), which are Sandwich, Pizza, Gym, and Italian.

Conclusion

This report examined existing commercial venues adjacent to CTA train stations, and determined that new development most likely to succeed includes a coffee shop, easy-to-grab food like sandwiches and pizza, and a gym.