Problem Statement

Bob the Baker is a Chicago native that wants to start a business. He's passionate about baking and has developed quite a following from his award winning carrot cake. Usually Bob sells his cakes and other goods at the local farmer's market, but with business picking up he wants to expand to his own shop where he can build an industrial kitchen and hire a staff of passionate bakers. He's contracted a local data science consultancy to help him pick the best location to start his business. Bob has been selling his baked goods for years and knows his consumer base is generally families. Typically, Bob has the highest sales on the weekend and wants to be in a location with plenty of foot traffic during this time. Given his success at farmer's markets, Bob knows that he wants his new shop to be near a market and would prefer to open in the spring when the markets start up for the season.

Approach

I'm a big proponent of starting simple and when addressing a problem it's best to begin with the basics. In this case, exploratory data analysis (EDA) is going to be a lot more helpful to our client, Bob, than trying to impress him with a complex neural network.

In the warm-up, we already saw that Clark/Lake is likely one of the busiest stations in Chicago. It's tempting to tell Bob to set up shop there, but in doing so we ignore a lot of other factors: Who is his target demographic? When is foot traffic near the bakery most important? Fortunately, Bob has a clear vision that gives us an indication of what we need to look at in the data.

To begin addressing the problem, I found an outside dataset that includes the location of farmer's markets in Chicago. This way we can identify train stations near farmer's markets. In order to do this, we need the location of our train stations. I brought in a dataset that includes the latitude and longitude of each station and then used an API to connect each coordinate to a zip code. The same is done for the coordinates of the farmer's markets. The train station and farmer's market datasets are then merged by zip code. We have now identified train stations near farmers markets.

Next, I filtered down the data by looking at weekend ridership in the spring for the past 3 years. These are times Bob indicated as when he needs the most foot traffic outside of his shop. Once

the data was filtered, I identified the top five stations with the highest ridership. These are our highest value stations given the information that we have. I also built a map for Bob to reference the location of the stations. Based on the analysis, Bob would likely want to consider opening his Bakery near the Fullerton, Roosevelt or Addison-North Main stations.

Top Five Stations with Highest Average Weekend Ridership in Spring