Applied Data Science Capstone Project

Cassandra Martinez

## The Battle of the Neighborhoods (Week 1)

### Introduction

What is the trend between diabetes rates and food deserts in Northern New Mexico? With the transition from volume-based payment toward value-based payment methods in American healthcare, there is also a shift of focus in patient care toward outcomes. Patient health outcomes have already been a critical element in funding for Federally Qualified Health Centers (FQHCs), and Patient-Centered Medical Homes (PCMHs). Funding is typically awarded using the Uniform Data System (UDS) measures reported quarterly to the federal government. One such measure focuses on abnormal A1C levels, blood glucose, known as Diabetes (types 1, 2, and Gestational). Though there are some genetic risk factors associated with diabetes, there are also many prevention and treatment measures one can consider. These measures include regular physical exercise, maintaining a healthy body weight (BMI), and avoiding tobacco use. The prevention and treatment measure I hope to explore more extensively through data exploration and location data is diet.

Diabetes is a costly disease to patients, communities, and medical providers on all levels, from administrative to insurance and down to providers and other clinical staff. As of 2017, 9.4% of Americans are now living with a diabetes diagnosis, more than 100 million people and in New Mexico the rate is above the national rate at 10.7% or 223,951 people. In 2017, the overall food insecurity rate (percentage of the population that is food insecure) was 16%, and 23% of residents participate in the Supplemental Nutrition Assistance Program (SNAP). The question is of just how impactful the access to healthy foods is on overall diabetes rates. New Mexico is a predominantly rural state with many residents living in smaller communities within counties with a large land area. Given that many of these communities face food deserts, areas that have limited access to affordable, nutritious food, the strain placed on many FQHCs and PCMHs (typically the only source of healthcare for these communities) to manage patients with abnormal A1C is inexorable. The only question to answer is what the quantifiable measure is.

### Data

I extracted the data necessary for this exploration from public health data available from the Center for Disease Control, US Census Bureau, and The State of Obesity for diabetes statistics. Much of this data is public health-related and open for public viewing. The New Mexico Health Indicator-Based Information System provided much of the county used to explore the rate of diabetes versus proximity to grocery stores. Initially, there is a mapping of diabetes rates by county using a choropleth method. Additional location data was extracted using Foursquare and venues explored using "Grocery Store" labeled with blue markers, and "Supermarket" labeled with red markers. For the sake of maintaining the definition of "Food Desert," I excluded "Farmers Market" and "Health food Store" as these are typically more costly than the average Grocery or Supermarket.

I needed to extract diabetes rates by county in New Mexico, as well as longitudinal and latitudinal data to build the initial choropleth. After, I needed to determine the appropriate clustering method to map food source points and their associations with the rates of diabetes. Finally, I attempted to predict the diabetes rate by zip code based on the existent data.