

Restaurant Genre and Its Relation To Public Health

Logan Rivera

20 July 2019

Introduction

Human nutrition has long been known to play a critical role in the maintenance of long-term health. A diet with a complete and balanced spectrum of nutrients can aid in the prevention of illness. Our diet is created from a number of factors both controllable and uncontrollable, and the not least of which is the general types of restaurants that are located within a reasonable distance from one's home. All types of restaurants from a Mediterranean restaurant to Pizzeria serve a unique style of food to their costumers that may focus on certain types and nutrients. By determining correlations between the types of restaurants that are located in areas with high rates of illness and in low rates of illness Physicians will be able to recommend specific types of restaurants that patients with a record of frequent illness should eat and which types of restaurants they should avoid. The goal of this study is to aid doctors and individuals alike in decreasing rates of illness through healthier eating habits.

Data

Two main sets of data will be the used for the purpose of this study. The first is a set of location based health data from five hundred different cities put together by the Centers for Disease Control and Prevention. This dataset consists of local level data from five hundred cities on a variety of different health related topics from current smoking habits among adults to rates of cancer in adults. The section of data that will be of interest for this study will be the percentage of people over the age of eighteen that have had poor physical health for more than fourteen days during 2016. The second main set of data that will be used for this study is the Foursquare database api. The Foursquare database contains geographical data on the types of venues that are located near particular areas. The Foursquare database api will be combined with the location based health data to find the types of restaurants that are located in areas with poor levels of health and in areas of high levels of health respectively. The study will also a have geojson file that contains location data on the different districts of San Antonio. The geojson file will aid in the visualization of data on a map of San Antonio.

Methodology

Introductory Data Preparation and Exploration -

The health data for 500 hundred different cities was imported using the pandas python library. The data that was not related to illness among adults for greater than fourteen days was filtered out and rows without a value for the percentage of illness for a given location were removed. All but five cities (San Antonio, Austin, Salt Lake City, San Francisco, and Atlanta) were filtered out of the data set to simplify initial data exploration. A violin chart was generated for each of the five cities in order to view the distribution of the variable of interest (figure 1).

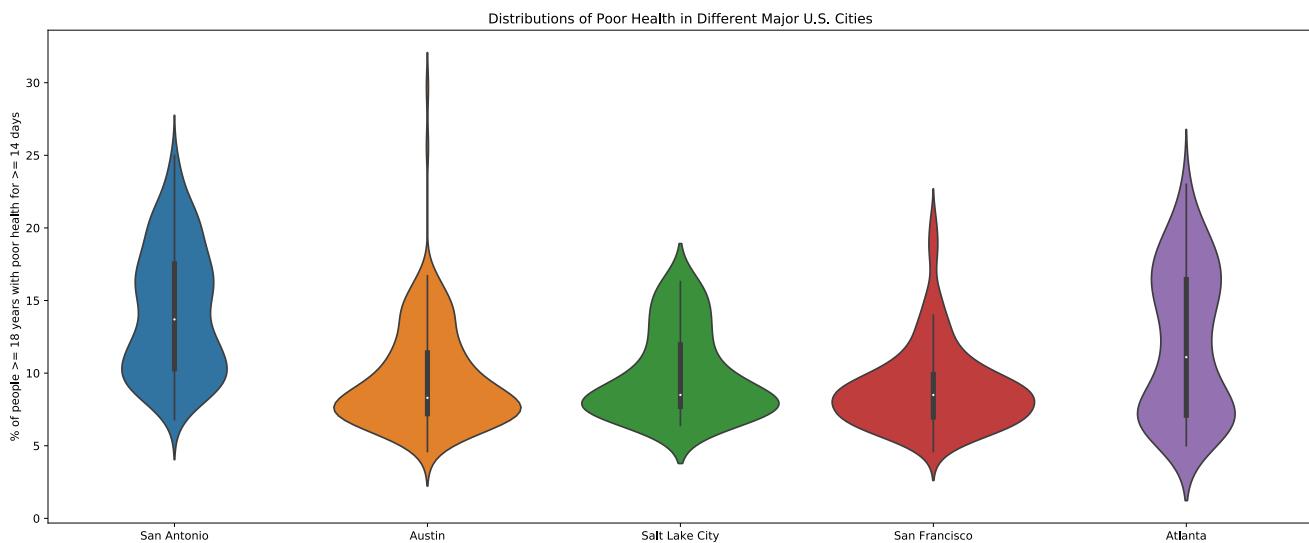


Figure 1 - The distribution of five major cities is shown above. Three of the cities above (Austin, Salt Lake City, and San Francisco) have skewed distributions toward the lower percentages.

The distribution of San Antonio was chosen for this study because of the large number of data points that existed for it and because of its broad distribution of data. San Antonio's distribution was then looked at in closer detail since it would be the city of focus for the remainder of the project. A box plot of the San Antonio health data was created (figure 2).

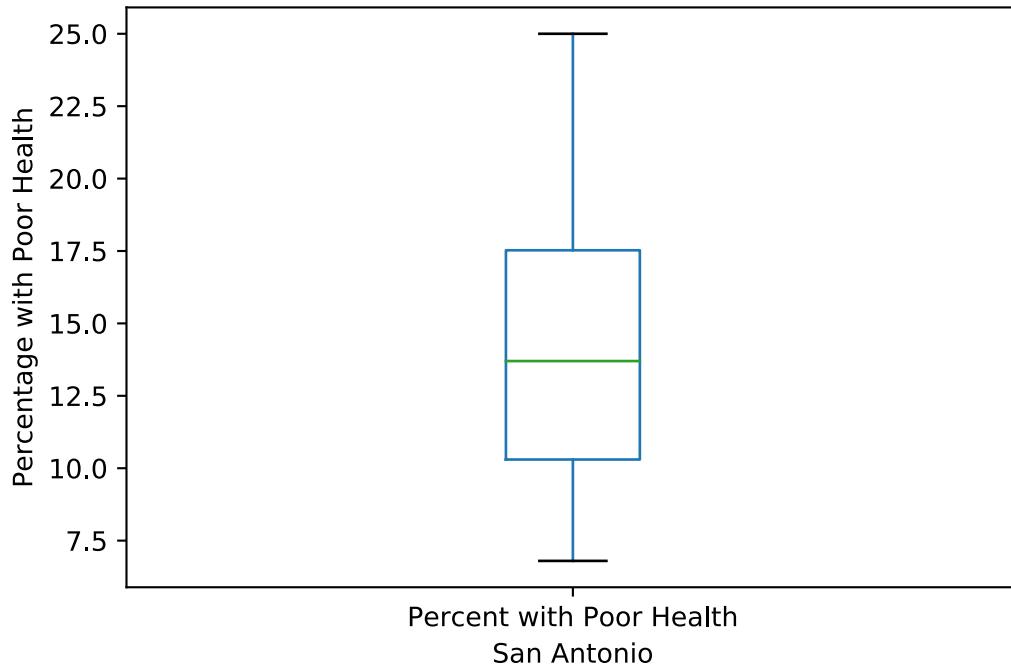


Figure 2 - A box plot showing the distribution of San Antonio health data. The difference between the third quartile and the maximum appears to be fairly large, so it is likely that some parts of San Antonio have a relatively higher rate of poor health than other areas. For the purpose of this study the genres of restaurants that exist in those areas would be of particular interest.

The San Antonio health data must be broken into different geographical subgroups so that the average percentages of poor physical health can be seen as a variable of location. San Antonio can be broken into ten different districts with areas of relatively equal size. Each datapoint in the dataset will be placed into one of the ten San Antonio political districts based on the longitude and latitude where the sample was taken. Each political area will be labeled by the representative that oversees their respective political district. All geographically labeled datapoint were grouped by the political district and averaged. A choropleth map of the data was created using the folium python package in order to visualize the geographical distribution of data based on political district (figure 3).

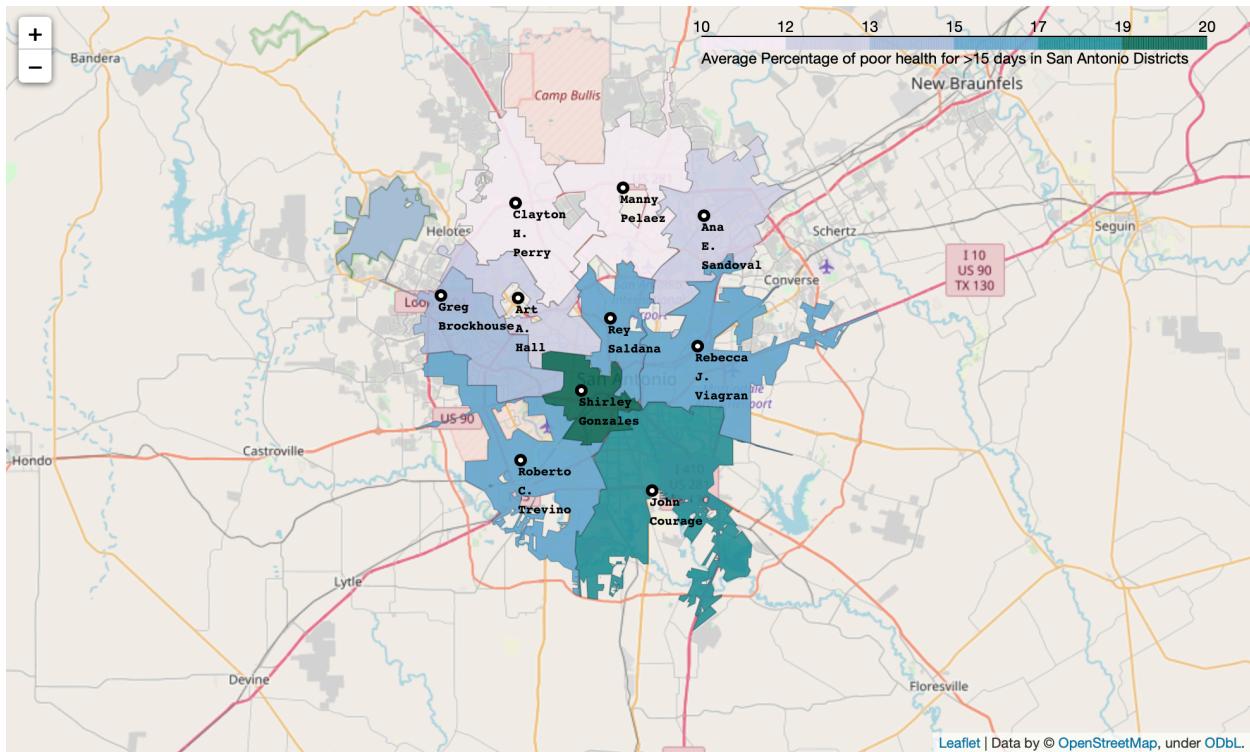
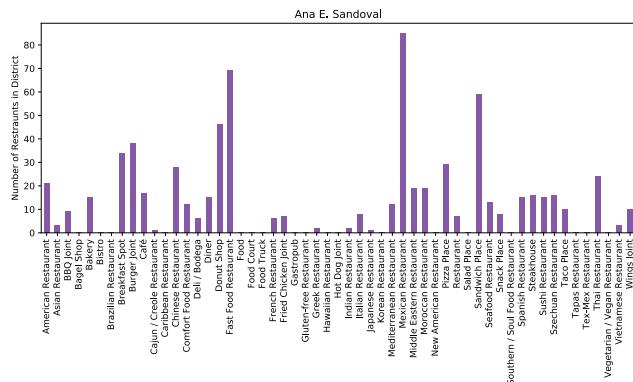


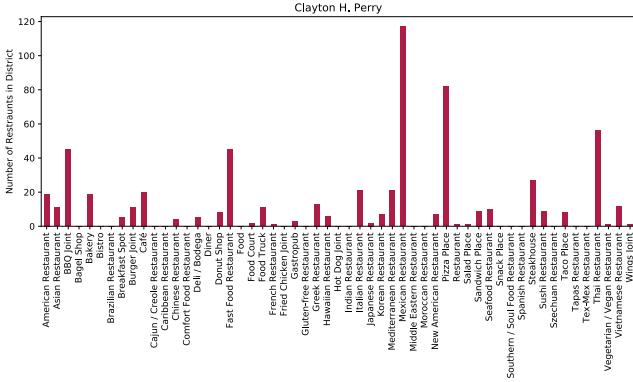
Figure 3 - A choropleth map of San Antonio that shows the average percent of poor physical health grouped by political district. It is evident from the map that the district represented by Shirley Gonzales and John Courage have a higher rate of poor physical health as compared to other districts.

Data Analysis with Foursquare Database api -

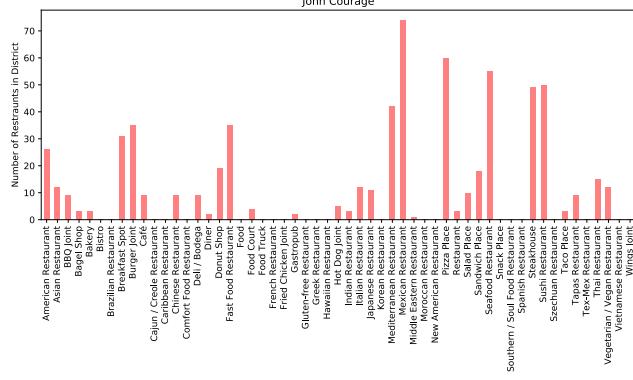
The Foursquare api was used to collect data about the type of venues that existed within each district. A call was made to the api to search for a maximum of fifty venues within a ten mile radius around each of locations in the dataset. The number of each genre of restaurant was encoded as a feature variable for each of the data points and saved as a local csv file so that future api calls would not be necessary in the future. Each genre of restaurant was organized by the San Antonio political district that it appeared within so that the most common genres of restaurants for each area could be determined. A bar graph (figure 4) was constructed that shows the counts of each type of restaurant organized by the district that it is located.



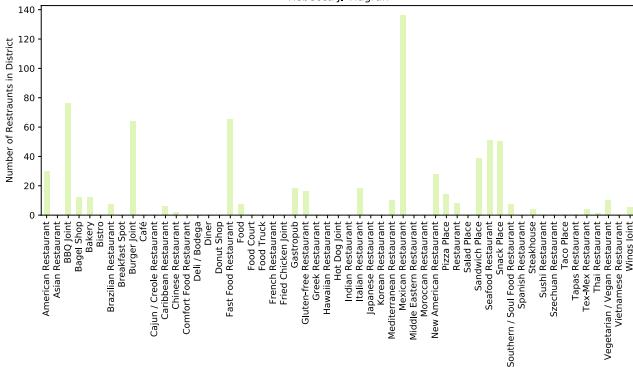
Ana E. Sandoval



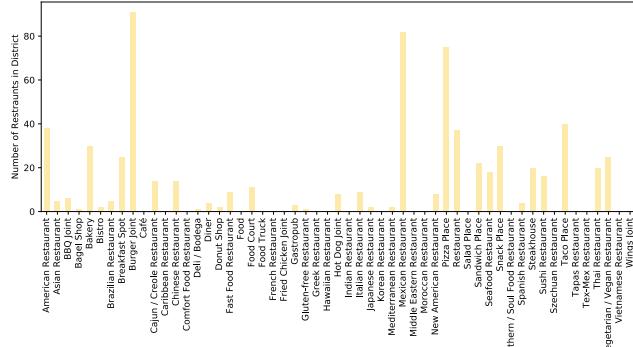
Clayton H. Perry



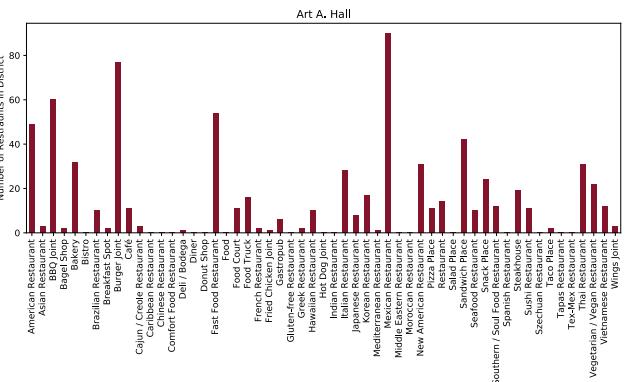
John Courage



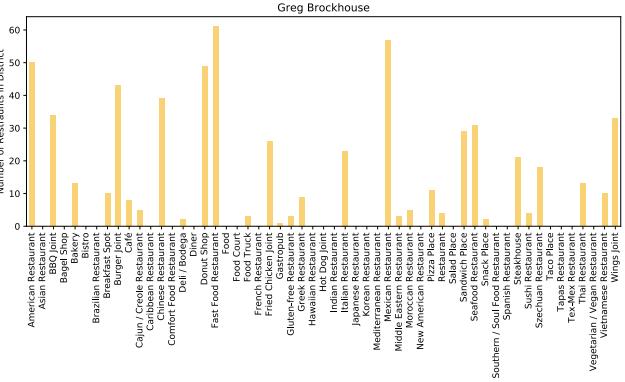
www.ijer.org



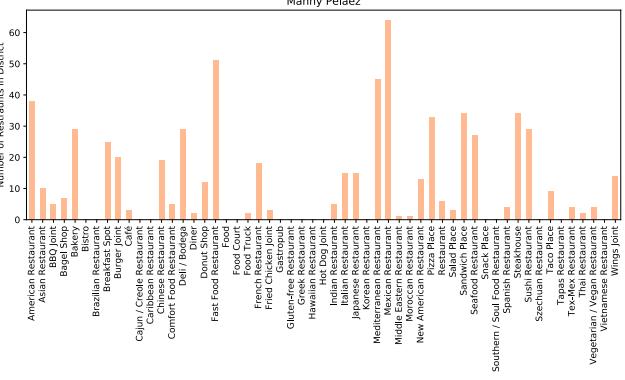
Page 1



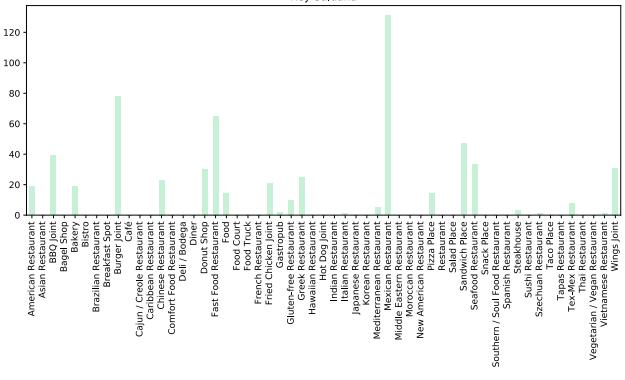
Art A. Hall



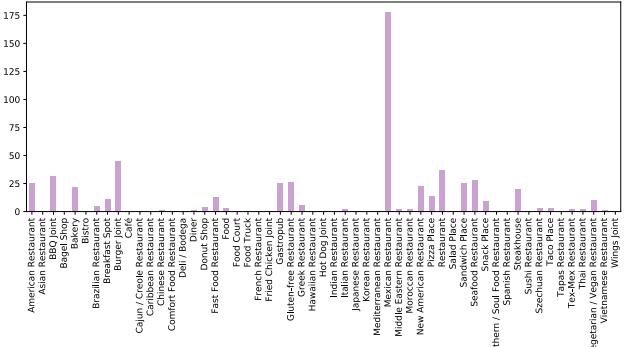
Greg Brockhouse



Manny Peña



• 100 •



10 of 10

Figure 4 - The bar graphs on the previous page details the number of each type of genre that existed within each San Antonio district. The most prevalent type of restaurant for each district was ‘Mexican Restaurant’ which is due to the strong prevalence of Mexican culture in San Antonio.

The types of restaurants that were most prevalent in each area were then compared with the rates of physical sickness within each district to search for correlations. The rates of physical sickness within each district were visualized in a bar graph (figure 5).

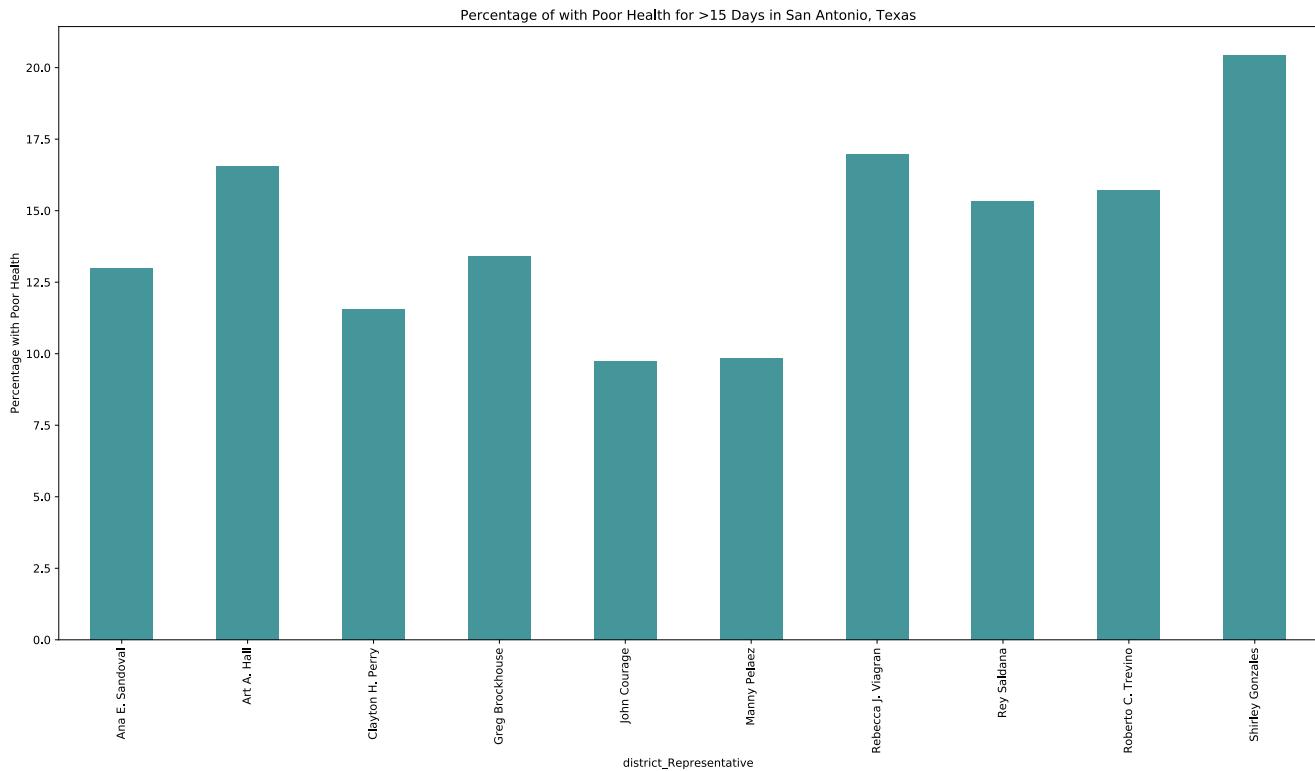


Figure 5- The bar graph above displays the different percentages of poor health grouped by district. This graph can be used in conjunction with figure 4 to determine correlations between restaurant types and rates of poor health.

Results

The top five restaurants for each of the districts are shown below along with the average percent with poor health per district (figure 6).

Top Five Restaurants for each district

	Most Common Restraunt	2nd Most Common Restraunt	3rd Most Common Restraunt	4th Most Common Restraunt	5th Most Common Restraunt	Average Percent Poor Health
Shirley Gonzales	Mexican Restaurant	Burger Joint	Restaurant	BBQ Joint	Seafood Restaurant	20.41
Rebecca J. Viagran	Mexican Restaurant	BBQ Joint	Fast Food Restaurant	Burger Joint	Seafood Restaurant	16.98
Art A. Hall	Mexican Restaurant	Burger Joint	BBQ Joint	Fast Food Restaurant	American Restaurant	16.54
Roberto C. Trevino	Burger Joint	Mexican Restaurant	Pizza Place	Taco Place	American Restaurant	15.72
Rey Saldana	Mexican Restaurant	Burger Joint	Fast Food Restaurant	Sandwich Place	BBQ Joint	15.34
Greg Brockhouse	Fast Food Restaurant	Mexican Restaurant	American Restaurant	Donut Shop	Burger Joint	13.42
Ana E. Sandoval	Mexican Restaurant	Fast Food Restaurant	Sandwich Place	Donut Shop	Burger Joint	12.99
Clayton H. Perry	Mexican Restaurant	Pizza Place	Thai Restaurant	Fast Food Restaurant	BBQ Joint	11.55
Manny Pelaez	Mexican Restaurant	Fast Food Restaurant	Mediterranean Restaurant	American Restaurant	Steakhouse	9.84
John Courage	Mexican Restaurant	Pizza Place	Seafood Restaurant	Sushi Restaurant	Steakhouse	9.74

Figure 6 - This table shows the top five restaurants for each district along with the percent with poor health in each district.

For nine of the ten districts ‘Mexican Restaurant’ was the most frequent type of restaurant. Among the top five restaurants for the districts with the five greatest averages for poor health, ‘Burger Joint’, ‘Fast Food Restaurant’, ‘BBQ Joint’ and ‘Fast Food Restaurant’ appeared

numerous time. Many of the same restaurants appeared in the bottom five districts except for ‘Thai Restaurant’, ‘Steakhouse’, ‘Sushi Restaurant’, and ‘Sandwich Restaurant.’

Discussion

The districts with the highest rates of poor health in general seemed to have restaurants that served mostly high-caloric food (burger joints that serve burgers, Fast Food Restaurants that more than likely also serve fried foods). Districts with lower rates of poor health also contained many of the same restaurants as the districts with high rates, but they also contained restaurants in their top five that focus on foods that contain vegetables or protein that is grilled rather than fried (Mediterranean Restaurant, Sushi Restaurant, Steakhouse, Thai Restaurant). Foods that contain vegetables and protein that is grilled on average tend to have an average lower caloric value than fried foods and foods that contain large amounts of carbohydrates. Excess amounts of calories in ones diet can lead to a number of health conditions such as obesity, high cholesterol, and diabetes. It can thus be assumed that living in areas that are located around restaurants that focus on genres of food that contain food with low caloric averages can raise one’s chance of living a healthy life style. The next step for analysis would be to find the average caloric values or most common macronutrient for each genre of restaurant so that the differences between the diets of those living in different districts can be further quantified.

Conclusion

The purpose of this study was to determine the types of restaurants that correlate with high rates of good health. The different rates of poor health and the most common types of restaurants in each of the San Antonio political districts served as a model to achieve the goal of this study. It was found that districts that contained genres that focused on foods with relatively low caloric values in their top five most prevalent types of genres tended to have lower rates of poor health than those districts that contained genres that focused foods with high caloric values.