Targeting Family Counseling Services Across Kansas City

1 Introduction

1.1 Background

As the United States confronts an ongoing coronavirus, more people are adapting to new social habitation standards. With families in closer proximity during isolation, quarantine, job lay-off, there has been a rise in domestic violence (Maclen, 2020). Kansas City is no exception to the rise in domestic violence seen across the world (Kraske, 2020). As the pandemic persists more families are falling into cycles of domestic violence, often with no other choice than to lockdown with their abusers.

1.2 Problem

This rise in domestic violence requires the attention of family counseling services, battered females' shelters, and other organizations that can intervene, prevent, and deescalate tensions between family members. The goal of this data analysis project is to identify where the abuse takes place and bring attention to the zip codes that most need the services these organizations provide. To examine the need for services, each zip code across Kansas City will undergo a domestic violence increase analysis, k-means clustering to bin the neighborhoods, and identification of the clusters with the highest need.

2 Developing Data and Cleaning

2.1 Data Sources

Kansas City maintains an open data portal. This open data portal contains a repository of crime data (Mesa, 2020). An csv is available for download. The crime data table contains a column that indicates if the offense was a domestic violence incident. It also contains information on the area where the crime took place. Isolating only the domestic violence incidents with complete zip code information there are 10,141 occurrences in 2020, for Kansas City – shown in Figure 1.

	Reported_Date	Zip Code	DVFlag
0	2/5/2020	66206	Y
1	2/5/2020	66206	Y
2	5/18/2020	66110	Y
3	5/18/2020	66110	Y
4	5/18/2020	66106	Y

Figure 1 The head() of the long table of domestic violence occurrences in Kansas City for 2020. DVFlag 'Y' indicates that, "yes" the crime was domestic violence.

This table, however, lacks latitude and longitude for the domestic violence occurrences. Location data for all of Kansas City's zip codes is available in CSV format (*US Zip Code Latitude and Longitude*, n.d.). The joined tables appear in Figure 2.

	Unnamed: 0	Reported_Date	Zip Code	DVFlag	Latitude	Longitude
0	2089	1/12/2020	64153	Y	39.281602	-94.71439
1	3309	2/24/2020	64153	Y	39.281602	-94.71439
2	3498	3/3/2020	64153	Y	39.281602	-94.71439
3	3930	1/31/2020	64153	Y	39.281602	-94.71439
4	4527	1/24/2020	64153	Y	39.281602	-94.71439

Figure 2 DVFlag here indicates a domestic violence incident in Kansas City, MO. This table represents the first 5 rows in the 2020 Kansas City domestic violence crimes list, made in the data preparation phase.

2.2 Data Cleaning

The original extent of the 2020 domestic violence crimes spreadsheet for Kansas City was over 10,100 rows long. This also contained incorrect zip codes, NaN values, among others. I was able to prepare the spreadsheet in Python by dropping NaN values. When it came to connecting with a geocoding API, none of my calls were returned. I reduced the scope from individual addresses to zip codes. There were about 50 unique zip codes that I could connect a latitude and longitude to, however, no calls were returned from the Google API or ArcGIS. I opted instead to manually pull another csv file from OpenDataSoft.com (*US Zip Code Latitude and Longitude*, n.d.). The result was Figure 2. This contains 10,098 rows of domestic violence crimes with errant zip codes removed.

In the future it would be wise of Coursera and the IBM Professional instructors to develop a consistent means of prescribing APIs, geocode, data, and merging and join lessons that are more appropriate for this capstone project overall. It would greatly help with the data gathering and cleaning process, as I came here to learn to code, not spend hours running into paywalls trying to learn what I should have already been taught before being expected to perform.

2.3 Data Summary

This data is prepared for analysis in part 2 of this capstone project. In the next phase I intend to determine the areas with higher levels of domestic violence taking place. This information could be related to other externalities, such as the topology or composition of the neighborhood and the services found within it.