Crime and Demographics in the Mile High City

Correlations in Rates of Violent Crime and Comparative Demographics by Neighborhood in Denver, CO

Naureen Bharwani
Computer Science
University of Colorado Boulder
Boulder, CO USA
naureen.bharwani@colorado.edu

Sean Mulligan
Computer Science
University of Colorado Boulder
Boulder, CO USA
sean.mulligan@colorado.edu

Cody Thornton
Computer Science
University of Colorado Boulder
Boulder, CO USA
cody.thornton@colorado.edu

ABSTRACT

We sought to study correlations between violent crime and the demographic characteristics of Denver, CO. Our initial literature survey suggested that, in general, places with higher levels of violent crime have poorer health outcomes, higher levels of poverty, lower rates of higher educational attainment, and more racial diversity.

Based on this research, we sought to discover if certain neighborhoods in Denver experienced higher levels of violent crime than others. If such neighborhoods existed, we also planned to seeks answers to the following questions about these neighborhoods:

- Do these areas have poorer health outcomes?
- Do these areas have higher rates of poverty?
- Do these areas have lower levels of education?
- Do these areas have larger minority populations?

Using k-Means clustering we confirmed that there are neighborhoods and regions of Denver with disproportionately high densities of violent crimes. Comparing these locations to the state of Colorado, we found that these locations on average had the following characteristics:

- 2-3 years lower life expectancy
- More than 7% higher rates of obesity
- More than 7% higher rates of poverty
- 15-16% lower rates of higher education
- Nearly 42% more racial diversity

These findings suggest affirmative answers to all the questions we posed above, and corroborate our research on correlations between violent crime and demographics.

INTRODUCTION

Identifying the connections between violent crimes and demographics is important because it can provide insight on the true costs of violent crime, and how to potentially prevent it in the first place.

Like any complex social phenomenon, violent crimes are not the result of a single cause. Violent crimes are the aggregate result of well studied social issues including, for example, poverty and education. By examining the violent crime centroids, at both a neighborhood and regional level, we hope to see how higher levels of poverty, or conversely lower levels of education, affect violent crime rates.

Moreover, violent crimes have costs beyond the crimes themselves, including impacts to health. By studying the violent crime centroids, again at a neighborhood and regional level, we hope to see how violent crime may impact rates of obesity and the average life expectancy.

Interwoven between these layers are additional societal constructs and forces that serve as both catalysts and inhibitors to themselves and all other variables, such as income, race, religion, etc. By identifying trends in the primary crime and demographic data, resources can be better targeted to communities in need.

LITERATURE SURVEY

The Denver Department of Public Safety [1]. The Denver Department of Public Safety (DOS) launched an initiative called The Denver Opportunity Index. The initiative aimed at identifying areas within the city, where residents' opportunities were less likely than other areas within the city. The Department of Public Safety's object is to focus on these neighborhoods with less opportunity to increase their overall quality of life. Within the initiative the DOS deems quality of life based on financial security, behavioral health (Do they have access to health insurance? Does violent crime impact my life?), and people left behind (barriers to employment, housing, education, and access to transportation).

The US Department of Housing and Urban Development [2]. In 2016, the U.S. Department of Housing and Urban Development published a review of studies related to crime in their publication, *Evidence Matters*. This review notes that violent crime in the United States has dropped by half, a truly "massive" amount, between 1995 and 2014. Decreases in the crime rate have been most dramatic in disadvantaged communities. However, significant disparities still exist in the rate of violent crime between different neighborhoods in American cities. Predominantly African-American communities average five times more violent crimes than predominantly white communities, and predominantly Latino neighborhoods average about two and a half times as many violent crimes as predominantly white neighborhoods. Similarly, low-income people are much more likely than others to experience crime, including violent crime. Furthermore, within neighborhoods, research has indicated that violent crime occurs in a small number of “hot spots" - small geographic areas like street intersections or street segments (two block faces on both sides of a street between two intersections). Policing these hot spots appears to be an effective way to reduce crime. In general, exposure to violence puts youth at significant risk for psychological, social, academic, and physical challenges and also makes them more likely to commit violent crimes themselves. Studies also suggest that violent crime decreases property values.

The National Institutes of Health [3]. A study of the children of 119 families over three years with assessments occurring one year apart. This study focused on three types of violent crime: marital aggression, parent-to-child aggression, and community violence. The authors of the study ranked cumulative exposure to violence into three bins: Low, Moderate, and High. The greater the cumulative exposure of participants to violence, the greater the incidence of somatic complaints, depressive symptoms, anxiety, aggressive and delinquent behavior, and academic failure. The authors conclude that exposure to violence significantly disrupts adolescents' development.

9News (KUSA-TV) [6]. As reported by 9News, a study by the Denver Police Department identified five specific areas of Denver that they say account for 26% of murders and nearly 50% of shootings. These areas are:

- South Federal Boulevard and West Alameda Avenue
- Colfax Avenue and Broadway
- East Colfax Avenue and Yosemite Street
- East 47th Avenue and North Peoria Street
- Martin Luther King Jr. Boulevard and North Holly Street

9News (KUSA-TV) [7]. An earlier article which reiterates the Denver Police Department study discussed above. This article also discusses a study by the Colorado Bureau of Investigation that found that violent crimes in Denver increased 6.52% percent from 2019 to 2020 whereas crimes against society (such as drug offenses, gambling, pornography, prostitution and animal cruelty) decreased 26.1% over that same period.

DATASETS

For our datasets we used the **City of Denver's Open Data Catalog [4]**. The specific datasets we used were:

• **Crime in the City of Denver**. Criminal offenses in the City and County of Denver, CO for the current year to date as well as for the last five calendar years. Data attributes include a brief description of the criminal offense, the category of the offense, as well as the date and neighborhood where the offense occurred.

https://www.denvergov.org/opendata/dataset/city-and-county-of-denver-crime

• **Hate Crimes in the City of Denver**. Hate crime offenses in the City of Denver, CO for ranging from Jan 2010 to Jan 2021. Dataset explores criminal offenses that target an individual(s) based on the offender's preconception of which group(s) the victim belongs to. Data attributes include a date, time month of year and neighborhood, as well as an offense description and bias type involved in the offense.

https://www.denvergov.org/opendata/dataset/hate-crimes

• **Traffic Accidents in the City of Denver**. Traffic Accidents in the City of Denver, CO for the previous five calendar years plus the current year to date. Relevant attributes include the specific incident address and neighbourhood associated with the accident.

https://www.denvergov.org/opendata/dataset/city-and-county-of-denver-traffic-accidents

• **Police Pedestrian Stops and Vehicle Stops in the City of Denver**. Police pedestrian stops and vehicle stops in the City of Denver, CO for the last four calendar years and the current year to date. Data attributes include an address and neighborhood name, as well as a problem attribute for reason of stop. Data has been cleaned by excluding data without a valid address.

https://www.denvergov.org/opendata/dataset/city-and-county-of-denver-police-pedestrian-stops-and-vehicle-stops

• **American Community Survey in the City of Denver**. Neighborhood level data in the City and County of Denver, CO for a 5 year average, years include 2013 - 2017. Data attributes include a neighborhood name, varying attributes on race, age and education levels as well as varying attributes on poverty levels regionally.

https://www.denvergov.org/opendata/dataset/city-and-county-of-denver-american-community-survey-nbrhd-2013-2017

• **Life Expectancy 2010 - 2015 in the City of Denver**. Average life expectancy in 78 distinct sections of Denver based on U.S. census data from 2010-2015. Relevant attributes include the Denver area name, life expectancy in years, and Federal Information Processing Standards (FIPS) codes which precisely identify the census tract referred to by the area name.

https://www.denvergov.org/opendata/dataset/city-and-county-of-denver-life-expectancy-2010-2015

• **Adult Obesity 2014-2016 in the City of Denver**. Estimated numbers of obese individuals older than two by neighborhood in Denver, CO. These estimates were arrived at by looking at individuals who sought care at health care institutions participating in the Colorado Department of Public Health and Environment's BMI Monitoring System between 2014 and 2016. As part of their routine care, individuals had their height and weight measured and their BMI calculated from these measurements. Obese individuals in the study are defined as having a BMI of 30 kilograms per meter squared (kg/m2) or greater. Findings are generalized to neighborhood-wide estimates by dividing the total number of individuals in the BMI Monitoring System in a given neighborhood with the total estimated population of that neighborhood as estimated by census data. Relevant attributes include neighborhood, total population in BMI registry, percent obese, number of obese adults, and confidence interval.

https://www.denvergov.org/opendata/dataset/city-and-county-of-denver-adult-obesity-2014-2016

TOOLS

• **Python.** Python is the base programming language we will use to perform most of our analysis techniques. Many of the other libraries and platforms we will utilize are built on or around python.

• **NumPy.** NumPy is a library used for the python programming language. We will use NumPy as support to explore our large, multidimensional datasets by performing various mathematical operations such as creating a subset array of our data for manipulation, reshaping and combining to explore our results.

• **Pandas.** Pandas is a library used for the python programming language. We will use Pandas to perform data manipulation and analysis on our large, multidimensional datasets. By utilizing Pandas, it will allow us to easily handle missing values within our dataset, and flexibly slice, merge or reshape our data to discover interesting results.

• **matplotlib.** matplotlib is a plotting library for the python programming language, which will employ NumPy for its mathematical extension. We will use matplotlib to explore our large, multidimensional datasets by allowing us to create easily digestible visuals from our dataset. Through these visualizations we will be able to interpret and answer interesting questions that arise throughout our exploration.

• **GitHub.** GitHub is a web-based interface that utilizes Git, which allows for version control and open source code. We will use GitHub to make real-time separate edits, updates and uploads of our source code to collaboratively work on our project as a team.

• **Trello.** Trello is a web-based tool that enables collaboration between teammates to organize a project into storyboards. Within each board a team member will be able to tell what tasks are in process of or have been completed, who is working on or finished which tasks, and milestones that still need to be met.

• **Tableau.** Tableau is an interactive data visualization software. We will use Tableau to explore our large, multidimensional datasets by allowing us to create easily digestible visuals from our dataset. Through these visualizations we will be able to interpret and answer interesting questions that arise throughout our exploration.

• **Scikit-learn.** Scikit-learn is a machine learning library for Python. It contains many built-in implementations of popular machine learning algorithms like DBSCAN and k-Means clustering.

TECHNIQUES APPLIED

1) Data Collection.

Data was collected by the City of Denver and the dataset was made available by the City of Denver's Open Data Catalog. Once the data was cleaned we were able to utilize it to begin the analysis of our project.

2) Data Preprocessing.

The preprocessing phase was broken down into 3 sections: Data Discovery, Data Cleaning, and Data Integration. In our preprocessing of the dataset, we addressed and improved the factors of *data quality.*

a) Data Discovery. In our Initial exploration of the datasets we specifically reviewed the 7 aspects of data quality: accuracy, completeness, consistency, timeliness, believability, and interpretability. Additionally, we Identified common(crime) and comparative(health) datasets as well as identified shared attribute values between the common and comparative datasets. Crucially, we decided to join datasets based on the neighborhood level attribute, which was present within all of our datasets.

b) Data Cleaning. Standard 2 part iterative cleaning process of discrepancy detection and data transformation.

We went through each of our datasets individually and narrowed down the attributes we thought to be the most relevant to the interesting questions we wanted to explore. Within our life expectancy health dataset, we removed several attributes including:

- COUNTY
- LIFE_EXPEC
- NBHD_ID(numerical not alphabetical).

We also combined FIPS and STFID into one attribute since they both corresponded to the same value. In addition we renamed the LE_2010201 to LIFE_EXPEC giving an easier interruption of what the attribute represented.

While cleaning our datasets, we identified and handled missing values by removing them from our dataset. For example, within our main Crime dataset, we removed the Central Park neighborhood due to its absence from the other datasets, in which this geographic area is considered a part of a larger neighborhood called Stapleton. We also found all rows with null values and removed them, and deleted one nonsense data point in the Crime dataset.

In our initial exploration of the Crime dataset, we found four key geographical attributes:

- GEO_X
- GEO_Y
- GEO_LON
- GEO_LAT

We removed the GEO_X and GEO_Y attributes to narrow down our scope, and utilized the GEO_LON and GEO_LAT later to create several visualizations based on the offense category ID within our crime dataset.

In addition to narrowing the scope of attributes for each dataset, there was normalization required on the neighborhood attributes for our main Crime dataset and comparative health datasets. We utilized Pandas and Numpy libraries for the normalization. We normalized our datasets by:

- deleting the Central Park neighborhood
- changing cbd to CBD capitalized
- changing dia to DIA capitalized
- changed title case i.e. auraria to Auraria
- replacing '-' with ' ' i.e athmar-park to Athmar Park

c) Data Integration.

Originally we sought to combine our crime datasets, which included: Crime in the City of Denver, Hate Crimes in the City of Denver, Traffic Accidents in the City of Denver, and Police Pedestrian Stops and Vehicle Stops in the City of Denver. After inspecting our Crime in the City of Denver dataset, we found our Hate Crime, Traffic Accidents and Police Pedestrian Stops and Vehicle Stops were subsets of the Crime in the City of Denver dataset. Due to this overlap, we were able to focus on the Crime in the City of Denver dataset and use subsets individually when relevant.

Given the datasets were all from the same source and relevant to the same domain, they had many common attribute values that inherently helped with the integration process. However, due to the dissimilar attribute values, the comparative(health) datasets could not be easily integrated with one another or with the common(crime) datasets.

Through our normalization on the neighborhood level, we integrated data originating from several sources including: the main Crime dataset, American Community Survey in the City of Denver, Life Expectancy 2010 - 2015 in the City of Denver, and Adult Obesity 2014-2016 in the City of Denver. This allowed us to have a repository of information consolidated under a unified schema, which formed our data warehouse [5].

2) <u>Initial Analysis</u>. Once the data had been preprocessed, we began by doing a more thorough standard examination of the data to gain a better understanding of interesting questions and relevant attribute values. This phase allowed us to ensure the quality of the neighborhood level analysis.

a) General Dataset Exploration. Using Tableau, we were able to answer some general questions related to our demographics within the city of Denver.

- What is the breakdown of gender within the City of Denver? What are the varying ranges of age, income and poverty within the City of Denver?

See Appendix ***Figure 1a*** and ***1b -*** To obtain primary demographic information at the neighborhood level of analysis, we examined the Neighborhood Values Survey. With the neighborhood on the x axis, the remaining attribute values were grouped into relevant categories, to include sex, race, age, education, income, etc. For example, see figures 1a and 1b in the appendix, which include the general summary and sex tabs. The summary tab again encodes neighborhoods to hue and x axis, with total population, median age, median family income, median home value, and percentage of poverty. Note that several attribute values were created for this visualization using existing attribute values. For example, the percentage male/female was created utilizing the counts of total population, and male/female population. Several additional derived attributes values will be created for the finalized neighborhood level analyses.

b) Common(crime) Datasets. Using Tableau, we were able to answer general questions about the crime dataset such as:

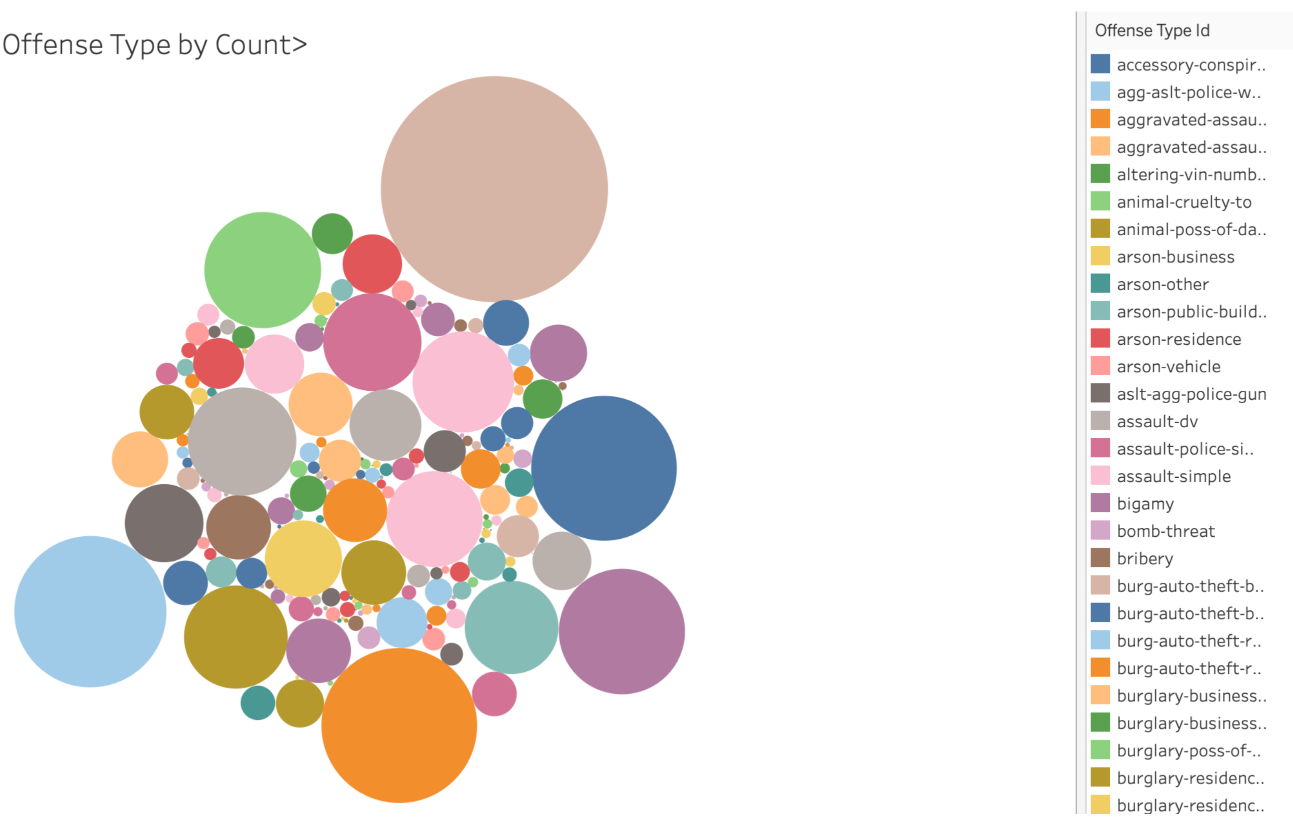
- What is the frequency of these different crimes?

See figure “Offense Type by Count" below - The frequency of the Offense Type ID is encoded by size. This gave us a rough idea of which crimes would be the most common such as traffic-accidents, theft-motor-vehicles, traffic-accident-hit-and-run, and theft-items-from-vehicles.

- Which crimes are most prevalent across neighborhoods?

See Appendix ***Figure 2 -*** To obtain primary Offense Type information at the Offense Category ID level of analysis, we examined the Crime dataset. We split on the Offense Category first to provide a general breakdown for the user to analyze the data on. The Offense Category ID is encoded by hue. After this we were able to split on the Offense Type next. This summarizes the Offense Category based on Offense Type. For example, from the visual we can easily see that traffic-accident is the most common crime. We are also able to see within the drug-alcohol Offense Category ID the second most common Offense Type is liquor-possession with a count of 4,900.

<Offense Type by Count>



- Which of these crimes constitute violent crimes?

We decided to classify violent crimes as a subset of our main Crime dataset based on the offense category ID level. We determined which crimes classifed as violent crimes by referencing one of our literature surveys, "*DPD names 5 areas that accounted for 49% of city's shooting victims in 2020*" [7], which identified four main types of violent crime: murder, aggravated assault, robbery and sexual assault. This conclusion was found within our exploration on the neighborhood analysis level.

c) Comparative(health) Datasets. Using Tableau, we answered general questions about the health dataset such as:

- What is the age of life expectancy in Denver?

See Appendix ***Figure 3 -*** This figure shows a visual representation for the life expectancy at the neighborhood level of analysis, where we examined the life expectancy 2010-2015 in the City of Denver dataset. This visualization was a general graph we made early on, which encodes by size what the life expectancy was in Denver based on neighborhood. Within our dataset we have an attribute LE_STATE that explicitly states the percentage of Denver's life expectancy to be 80.5%.

- What percentage of the population in Denver is obese?

Exploring the Adult Obesity 2014-2016 dataset, we were able to derive the percentage of the population in Denver that is obese. We calculated it using the TOTALPOP_INREGISTRY and COUNT_ADULTS_OBESE attributes. Totaling up the TOTALPOP_INREGISTRY and COUNT_ADULTS_OBESE for all neighborhoods, and then finding the percentage by taking the Total COUNT_ADULTS_OBESE / Total TOTALPOP_INREGISTRY. We found that an estimate of 29.98% of the population in Denver is obese. Note this is only taking into account a fraction of the population that elected to participate in this study.

During this process we took note of any information that provided insight into building the neighborhood level analysis.

3) <u>Neighborhood Analysis</u>.

After the completion of our Initial Analysis, we began the neighborhood level analysis of interesting attributes and correlations discovered through the previous phase exploration. This Neighborhood Analysis has become the base of comparison between the common(crime) and comparative(health) datasets. In this phase, we created a summarizing dataset along with our data warehouse that allowed us to consolidate our findings from Phase 2. Using this summarizing dataset, we were able to make arguments about correlations between the violent crime dataset and health datasets.

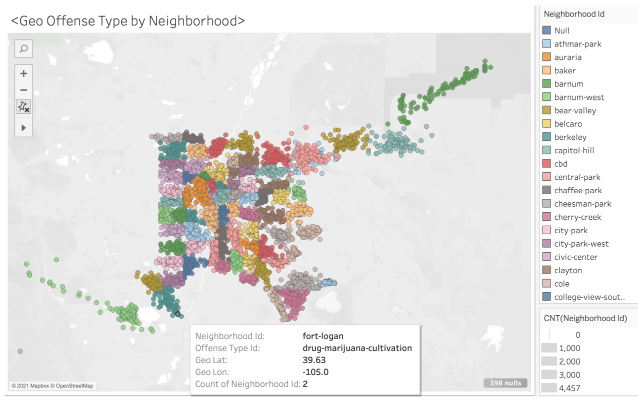
a) Common(crime) Datasets. Using Python and Tableau, we answered specific questions about crime at the neighborhood level such as:

- What is the frequency of the different crimes in different neighborhoods? Do particular neighborhoods suffer disproportionately from certain kinds of crime?

See Appendix ***Figure 4 -*** To obtain primary Offense Type information at the neighborhood level of analysis, we examined the Crime dataset. With the neighborhood encoded by color and each having their own individual tree we can explore the most common Offense Type within that neighborhood. Utilizing the total count of an Offense Type broken down on the neighborhood level allows the user to explore not just most common crimes across neighborhoods, but also shows which neighborhoods have the most crime associated with them.

- How are the neighborhoods within the city of Denver oriented? What neighborhoods border others?

To obtain primary geographical information at the neighborhood level of analysis, we examined the Crime dataset. We created the generic geo map shown below, where each neighborhood is encoded by color and the individual circles represent a specific type of Offense within that neighborhood. The geo maps goal is to give a user an idea of how the Denver neighborhoods are positioned, i.e. what neighborhoods are next to each other, do they border another county, and gives an overview of Offense Types within neighborhoods.



b) Comparative(health) Datasets. Using Python and Tableau, we answered specific questions about health at the neighborhood level such as:

- Which neighborhoods have the most obese population? Which neighborhoods have the least obese population?

See Appendix ***Figure 5 -*** To obtain primary health information at the neighborhood level of analysis, we examined the adult obesity 2014-2016 in the City of Denver dataset. With the neighborhood plotted on the y axis, they were plotted based on the percentage of obesity within the neighborhood. The size of the dot was encoded based on the sum of the count of how many people were obese within each neighborhood, this helps make the user aware of a potential skew in the dataset.

4) <u>Summary Analysis</u>.

a) Data Integration

We separated our data warehouse into three sections: count, percentage and neighborhood. From our integration we were able to update our count of crimes based on the offense category ID for our entire crime dataset and then derived a new attribute called Total Crimes that represented the total count of all crimes together based on the neighborhood level. We then decided to classify a new subset of crime on the offense category ID level called violent crimes based on our literature survey, *"DPD names 5 areas that accounted for 49% of city's shooting victims in 2020"* [7], which identified four main types of violent crime: murder, aggravated assault, robbery and sexual assault.

From this classification we derived another attribute called Total Violent Crimes that broke down the count of violent crime on the neighborhood level. Our second section, percentage, summarized the percentage of life expectancy and percentage of obesity by neighborhood. Finally, our third section represented neighborhoods ranked by violent crime, life expectancy and obesity. Note some neighborhoods such as Kennedy, DIA, Country Club, Civic Center and Auraria had estimates not available due to the small number of deaths.

Exploring our violent crime dataset, we were able to visualize a question from our initial analysis:

- Which of these crimes constitute violent crimes?

See Appendix ***Figure 6 -*** Using Tableau, we created a geo map, where we broke down our dataset by hue, where each color represented each of our violent crimes. The size of the data point shows the count of that violent crime within a particular neighborhood. While we were creating this visual we discovered violent crimes that were categorized as sexual assault crimes did not report any GEO_LON or GEO_LAT. Due to the insufficient geographical information we were not able to include sexual assault crimes within the geo map or in our k-Means analysis of violent crimes.

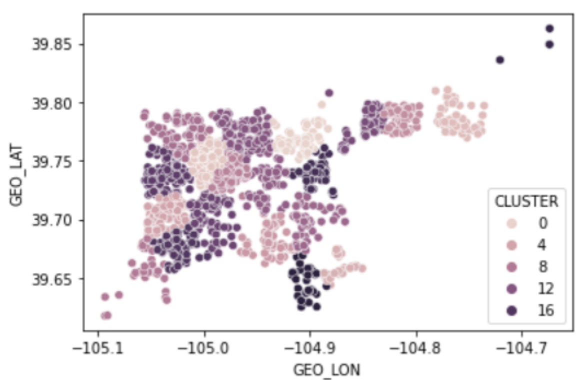
5) <u>Preliminary Results</u>.

a) k-Means Clustering. Utilizing the data preparation we performed within our summary analysis section, using Pandas and NumPy to isolate violent crimes, and clean the dataset by removing values such as:

- Removing data objects from 'violent_crime' that have no location data

- Removing data objects with location data north of Broomfield, CO (Latitude > 40)
- Removing data objects with location data south of Colorado Springs, CO (Latitude < 39)

Applying Scikit-learn to our cleaned violent crime dataset, we performed a k-Means clustering technique using each violent crime's latitude as the x value and longitude as the y value. After experimenting with various values for k, we decided k = 5 clusters was most helpful for our study because it mirrored a similar study done by the Denver Police Department that identified five hot spots for violent crime in Denver [7].

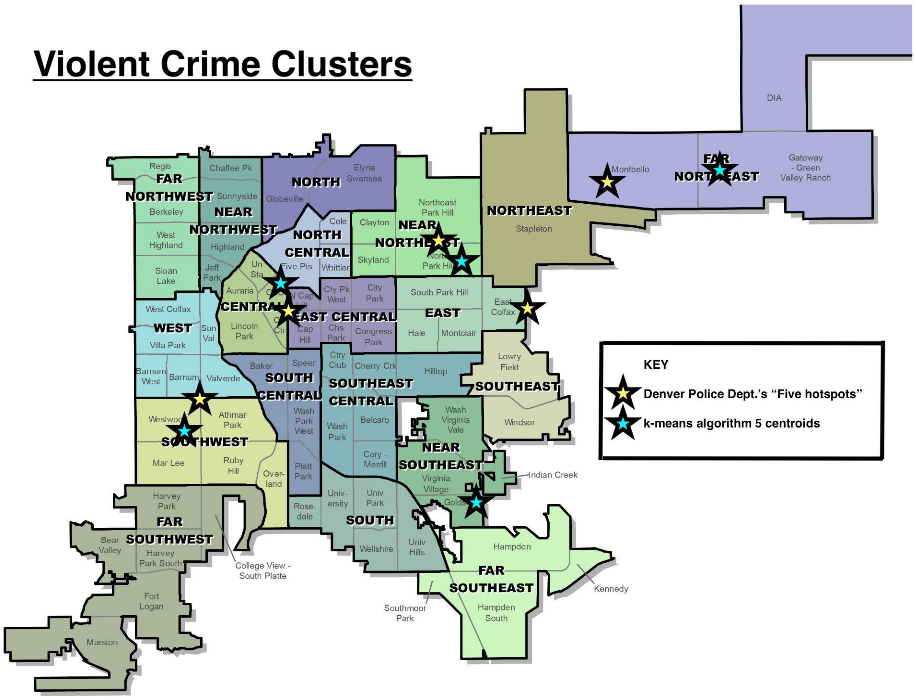


See figure "Violent Crime Cluster" on the next page - Utilizing the five centroids' geo-coordinates output from performing k-Means, we generated five centroids in the City of Denver located in the following neighborhoods: Gateway - Green Valley Ranch, CBD, North Park Hill, Goldsmith and Westwood. We plotted these centroids and the Denver Police Department's top five crime hot spots on a map of Denver's neighborhoods to compare our findings with those of the Denver Police Department. We discovered that our cluster centers closely matched the Denver Police Department's findings in the Southwest, Central and Near Northeast regions of the city.

Combining our data warehouse and the discoveries made from performing k-Means, we were able to do some basic statistical analysis to explore comparisons across centroids based on the neighborhood and region level. We created a Centroid Comparisons dataset that showed across the five hotspots derived from our k-Means clustering technique, what these neighborhoods' health outcomes looked like. We will further explore our findings for the Centroid Comparison dataset more in our key results section.

b) Predictive Tool. As many of the standard techniques we applied have a predictive nature, we attempted to create a python based predictive tool for user functionality and the potential processing of dynamic data.

• **Bayesian Classification.** To begin the bayesian analysis we read in the crime dataset in its entirety with Pandas. Relevant attribute values, such as date/time, were split, properly formatted and combined back into the dataset as additional attribute values. For a binary naive bayes proof of concept, the boolean value of is_crime was chosen as it was already present in the dataset and as such an efficient way to check the functionality of the model. The attributes of neighborhood_id, hour the crime was committed, and offence_category_id were given integer values for each unique element utilizing Scikit-learn, and then reinserted back into the dataset as additional attributes. Note that by including offense_category_id, the model should be able to achieve 100% accuracy for predicting is_crime, as there is a direct relationship between the attributes, again making it suitable for testing purposes. The model was then verified by submitting several neighborhoods, hours, and offense categories and correctly predicting the resulting is_crime bool as true or false. Next a categorical bayes classifier was created to attempt to predict the category of the crime, based on the neighborhood and the hour. To do so, the dataset was randomly split by item so that 70% of the dataset could be used for training, and 30% for testing. Utilizing just the variables of neighborhood and hour, the model can currently predict the category of crime with ~26% accuracy. Note that it was originally intended to add more comparative attribute values to the crime dataset, so that the accuracy level could be further increased, but time constraints prevented us from doing so.



KEY RESULTS

Violent crime in Denver clusters in particular areas. The Southwest, Central, and Near Northeast regions of Denver are particularly affected.

On average, neighborhoods and regions around which violent crimes cluster have diminished health outcomes, higher levels of poverty, lower levels of higher education attainment and greater racial diversity. On average, the neighborhoods containing our centroids had:

- ***Figure 7a -*** A life expectancy 2.8 years lower than the state average
- ***Figure 7b -*** Adult obesity rates 7.1% higher than the state average
- ***Figure 7c -*** Poverty rates 7% higher than the state average
- ***Figure 7d -*** Higher educational attainment as measured by the percentage of the population with a bachelor's degree or higher 15.2% less than the state average
- ***Figure 7e -*** 41.4% more racial diversity than the state average

See Appendix for Figures.

Although these general outcomes square with existing research into the correlations between violent crime and demographics, at the level of the individual neighborhood, certain neighborhoods defy easy classification. For instance, although our k-Means algorithm identified the Central Business District as containing the centroid with the largest number of violent crimes, this neighborhood is actually 7% less obese than the state average and has 10% more of its population holding a bachelor's degree than the state average.

We believe that this indicates that although general correlations can be drawn between violent crime and demographic characteristics across many geographic areas, researchers must be careful when studying particular neighborhoods. It is likely that any given locale has characteristics that are unique to it, and although its demographic characteristics may share many similarities with other areas with similar rates of violent crime, unexpected demographic characteristics are still possible.

APPLICATIONS

Our results can help law enforcement identify areas in Denver that require greater attention. The neighborhoods and regions containing the centroids produced by our k-Means algorithm are good candidates for greater policing to reduce violent crime in Denver. Our literature survey suggested that policing these areas could be effective at crime prevention [2].

Correlation does not equal causation, and the relationship between violent crime and the demographic characteristics we investigated is unclear. However, our results provide rich fodder for future research aimed at identifying causality. Among the chief questions our results might inspire are:

- Does the presence of violent crime in a community lead to higher levels of pessimism or nihilism and thus lower motivation for healthy living?
- Does lack of employment opportunities yield higher rates of poverty and a higher likelihood to be involved in criminal activities that result in violence?
- Does the presence of violent crime in a community lead to greater mental stress and thus lower bandwidth for education?
- Are there opportunities that are denied to racially diverse communities that lead to greater desperation and rates of violence?

ACKNOWLEDGMENTS

This project was completed as part of the Fall 2021 course CSBP 4502 Data Mining at the University of Colorado at Boulder. The course instructor was Kristy Peterson and the textbook used was ***Data Mining: Concepts and Techniques*** **[5]**. The terminology and methods of analysis used in this project come from this textbook.

REFERENCES

[1] City and County of Denver, Department of Public Safety. 2021. Denver Opportunity Index. https://geospatialdenver.maps.arcgis.com/apps/MapSeries/index.html?appid=ff9da000f1a344beb341ce839a720021

[2] Chase Sackett. 2016. Neighborhoods and Violent Crime. In *Evidence Matters* (Summer 2016). Office of Policy Development and Research, Washington, D.C., USA, 13 pages. https://www.huduser.gov/portal/periodicals/em/summer16/highlight2.html

[3] Gayla Margolin, PhD, Katrina A. Vickerman, MA, Pamella H. Oliver, PhD, and Elana B. Gordis, PhD. 2010. Violence Exposure in Multiple Interpersonal Domains: Cumulative and Differential Effects. In J Adolesc Health. PMC 47, 2 (Aug, 2010), 198–205. DOI: 10.1016/j.jadohealth.2010.01.020

[4] City and County of Denver. 2015. City of Denver Open Data Catalog. http://data.denvergov.org Licensed under the Creative Commons Attribution 3.0 license (CC BY 3.0) http://creativecommons.org/licenses/by/3.0

[5] Jiawei Han, Micheline Kamber, and Jian Pei. 2011. *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann, Burlington, MA.

[6] Darius Johnson. 2021. “Here's where Denver Police says are the city's crime hot spots”. https://www.9news.com/article/news/crime/denver-crime-hotspots/73-258c20a4-b6d3-495e-abb9-2a2661590192

[7] Allison Sylte. 2021. “DPD names 5 areas that accounted for 49% of city’s shooting victims in 2020”. https://www.9news.com/article/news/crime/denver-crime-hotspots/73-1e73ef1c-1a96-433c-88ad-e7b7b2b10ea3

Appendix

Figure 1a -

Figure 1b -

Figure 2 -

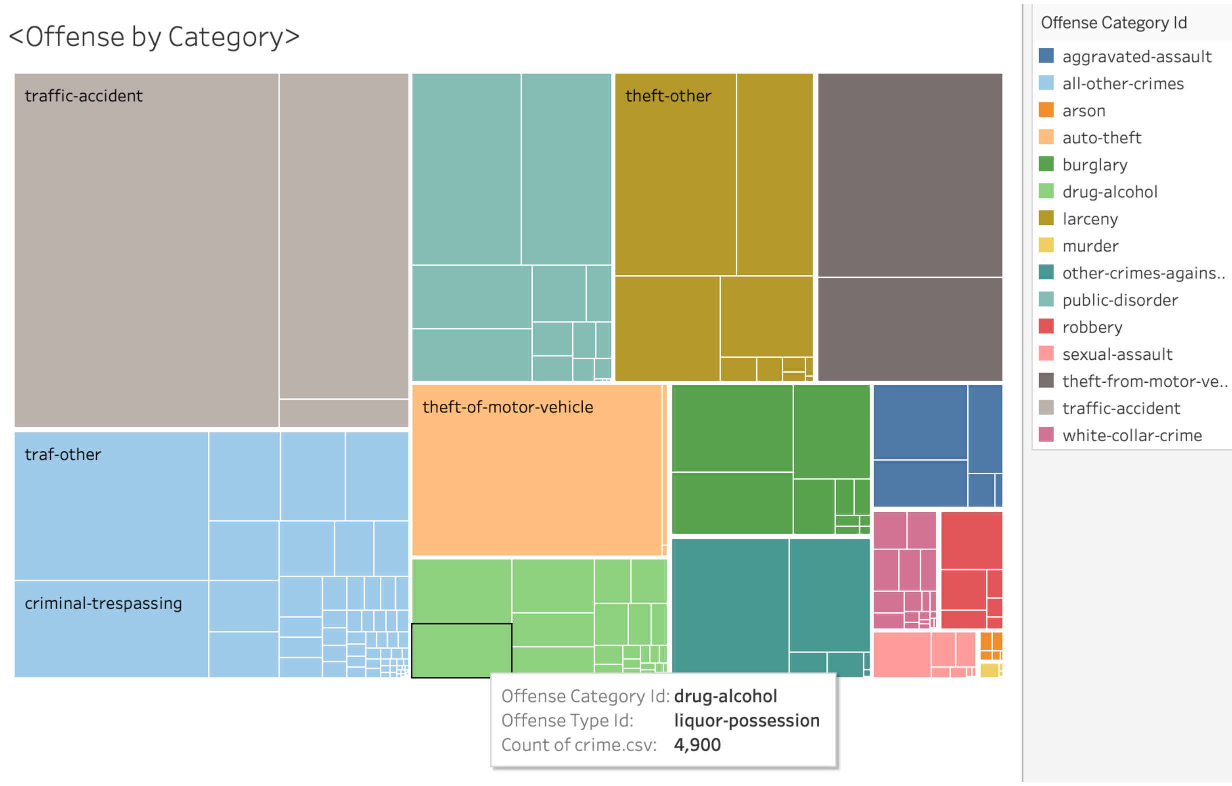


Figure 3 -

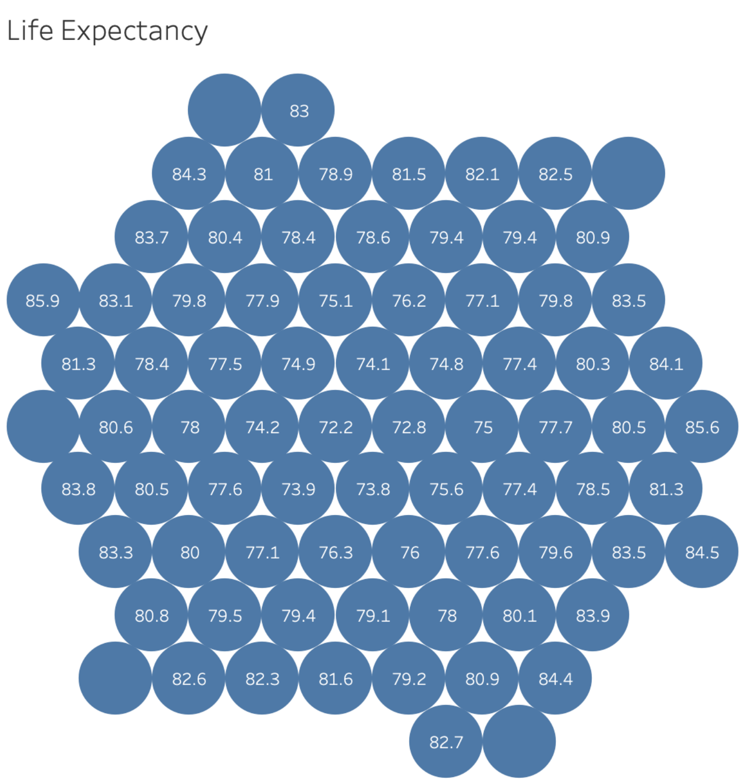


Figure 4 -

<Offense Type by Neighborhood>

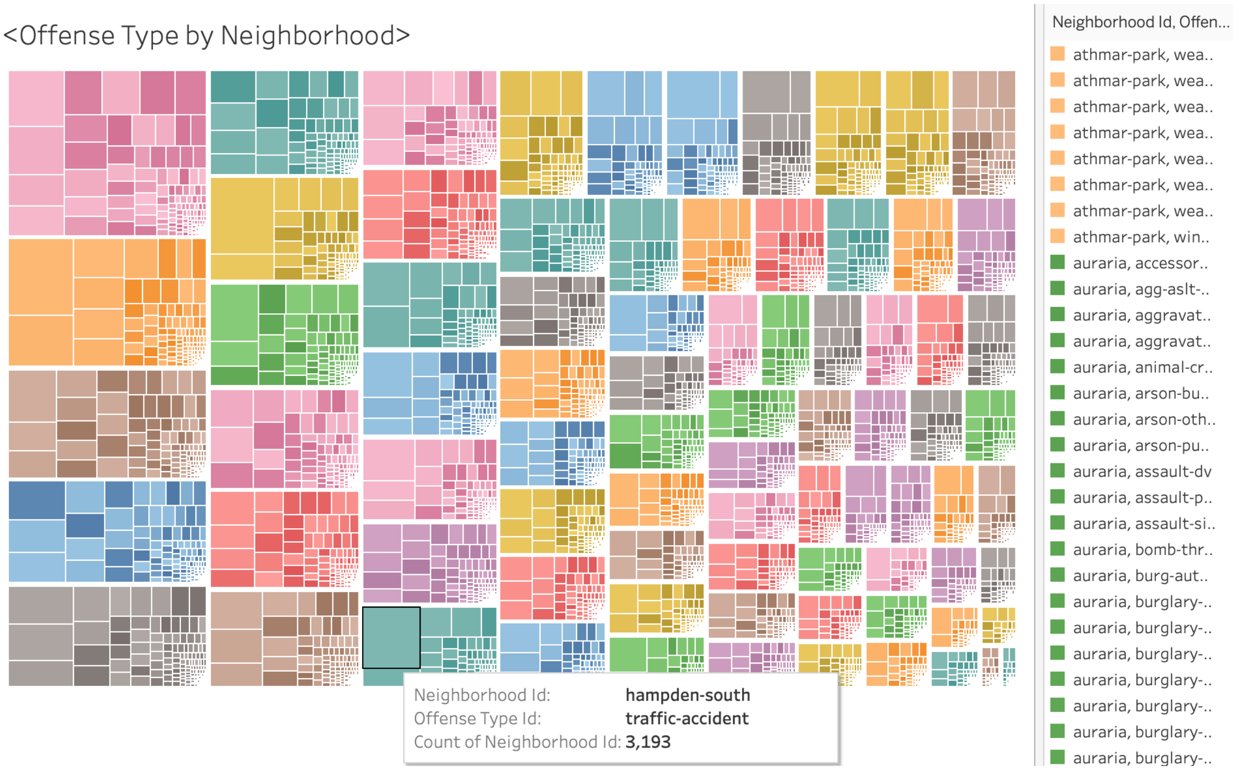


Figure 5 -

<Box Plot Percent Obese>

Figure 6 -

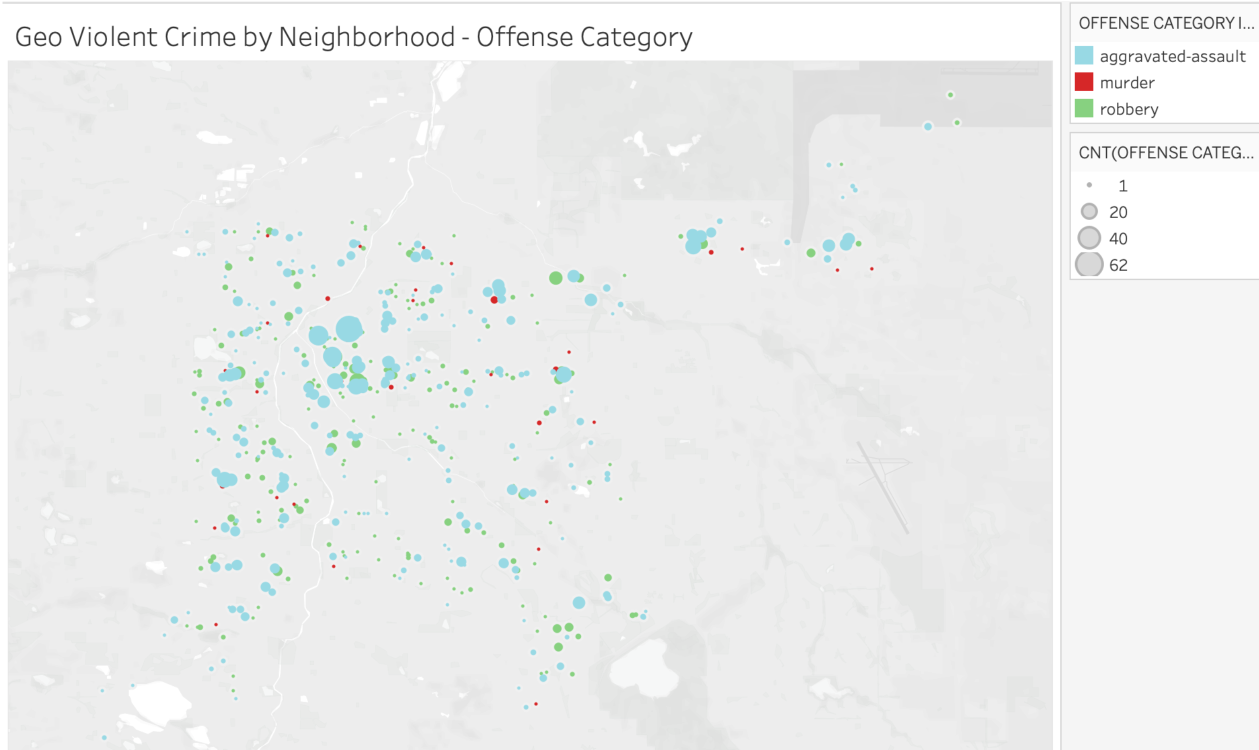

Geo Violent Crime by Neighborhood - Offense Category
OFFENSE CATEGORY I...
aggravated-assault
murder
robbery
CNT(OFFENSE CATEG...
1
20
40
62

Figure 7a -

Life Expectancy
Neighborhoods Containing Centroids

Centroid	Adjacent Neighborhoods	Life Expectancy of Neighborhood	Life Expectancy +/- State Average*
0	Gateway - Green Valley Ranch	80.4	-0.1
1	CBD	76	-4.5
2	North Park Hill	80.5	0
3	Goldsmith	73.8	-6.7
4	Westwood	77.9	-2.6

Cumulative Average	**-2.8**

Life Expectancy
Regions Containing Centroids

Centroid	Region	Life Expectancy of Region	Life Expectancy +/- State Average*
0	Far Northeast	80.5	0
1	Central	76.4	-4.1
2	Near Northeast	77.2	-3.3
3	Near Southeast	78.5	-2
4	Southwest	78.45	-2.05

Cumulative Average	**-2.3**

***State Average:** 80.5 years
Source: Kaiser Family Foundation *Life Expectancy at Birth*
https://www.kff.org/state-category/health-status/life-expectancy/

Figure 7b -

Adult Obesity Rates

Neighborhoods Containing Centroids

Centroid	Adjacent Neighborhoods	Percent Obesity of Neighborhood	Percent Obesity +/- State Average*
0	Gateway - Green Valley Ranch	38.12	13.92
1	CBD	16.7	-7.5
2	North Park Hill	29.09	4.89
3	Goldsmith	32.22	8.02
4	Westwood	40.44	16.24

Cumulative Average	**7.1**

Adult Obesity Rates

Regions Containing Centroids

Centroid	Region	Percent Obesity of Region	Percent Obesity +/- State Average*
0	Far Northeast	40.2	16.0
1	Central	19.6	-4.6
2	Near Northeast	34.2	10.0
3	Near Southeast	27.9	3.7
4	Southwest	37.1	12.9

Cumulative Average	**7.6**

***State Average:** 24.2%

Source: The Robert Wood Johnson Foundation, *Adult Obesity Rate by State*

https://stateofchildhoodobesity.org/adult-obesity/

Figure 7c -

Poverty Rates
Neighborhoods Containing Centroids

Centroid	Adjacent Neighborhoods	Poverty Rate of Neighborhood	Poverty Rate +/- State Average*
0	Gateway - Green Valley Ranch	7	-2.3
1	CBD	17.5	8.2
2	North Park Hill	7.3	-2
3	Goldsmith	18	8.7
4	Westwood	31.55	22.25
		Cumulative Average	**7.0**

Poverty Rates
Regions Containing Centroids

Centroid	Region	Poverty Rate of Region	Poverty Rate +/- State Average*
0	Far Northeast	12.2	2.9
1	Central	19.2	9.9
2	Near Northeast	17.9	8.6
3	Near Southeast	13.6	4.3
4	Southwest	20.3	11.0
		Cumulative Average	**7.3**

***State Average:** 9.3%
Source: The United States Census Bureau, *Quick Facts*
https://www.census.gov/quickfacts/CO

Figure 7d -

Higher Education Rates
Neighborhoods Containing Centroids

Centroid	Adjacent Neighborhoods	% Pop with Bachelors in Neighborhood	% Pop with Bachelors +/- State Average*
0	Gateway - Green Valley Ranch	13.8	-28.9
1	CBD	52.7	10
2	North Park Hill	41	-1.7
3	Goldsmith	25.3	-17.4
4	Westwood	4.7	-38
		Cumulative Average	**-15.2**

Higher Education Rates
Regions Containing Centroids

Centroid	Region	% Pop with Bachelors in Region	% Pop with Bachelors +/- State Average*
0	Far Northeast	11.9	-30.8
1	Central	41.2	-1.5
2	Near Northeast	30.0	-12.8
3	Near Southeast	34.4	-8.3
4	Southwest	12.9	-29.8
		Cumulative Average	**-16.6**

***State Average:** 42.7%
Source: The Metro Denver Economic Development Corporation, *Educational Attainment*
https://www.metrodenver.org/regional-data/demographics/educational-attainment

Figure 7e -

Percent of Population White
Neighborhoods Containing Centroids

Centroid	Adjacent Neighborhoods	% Pop White in Neighborhood	% Pop White +/- State Average*
0	Gateway - Green Valley Ranch	23.1	-63.8
1	CBD	75.6	-11.3
2	North Park Hill	54.9	-32
3	Goldsmith	61.1	-25.8
4	Westwood	12.8	-74.1

Cumulative Average	**-41.4**

Percent of Population White
Regions Containing Centroids

Centroid	Region	% Pop White in Region	% Pop White +/- State Average*
0	Far Northeast	23.9	-63.0
1	Central	70.2	-16.7
2	Near Northeast	40.0	-46.9
3	Near Southeast	63.5	-23.4
4	Southwest	27.3	-59.6

Cumulative Average	**-41.9**

***State Average:** 86.9%
Source: The United States Census Bureau, *Quick Facts*
https://www.census.gov/quickfacts/CO

Figure 7f -

Rankings Across Metrics

Neighborhoods Containing Centroids

Centroid	Adjacent Neighborhoods	Life Expectancy Lowest to Highest	% Obesity Highest to Lowest	Poverty Rate Highest to Lowest	% Pop w Bachelors Lowest to Highest	% Pop White Lowest to Highest
0	Gateway - Green Valley Ranch	4	2	5	2	2
1	CBD	2	5	3	5	5
2	North Park Hill	5	4	4	4	3
3	Goldsmith	1	3	2	3	4
4	Westwood	3	1	1	1	1

Rankings Across Metrics

Regions Containing Centroids

Centroid	Region	Life Expectancy Lowest to Highest	% Obesity Highest to Lowest	Poverty Rate Highest to Lowest	% Pop w Bachelors Lowest to Highest	% Pop White Lowest to Highest
0	Far Northeast	5	1	5	1	1
1	Central	1	5	2	5	5
2	Near Northeast	2	3	3	3	3
3	Near Southeast	4	4	4	4	4
4	Southwest	3	2	1	2	2