**Data Source**

* **Denver Open Data Catalog (Crime)**
  + Run by the city and county of Denver, it is a public access website for various data sets involving transportation, public safety, education, etc.
  + Data is based on the National Incident-Based Reporting System (NIBRS)
  + First published in 2019, currently updated weekly
  + <https://opendata-geospatialdenver.hub.arcgis.com/datasets/geospatialDenver::crime/about>
* **Denver Open Data Catalog (demographic data)**
  + Neighborhood-level data derived from the American Community Survey
  + Data was combined with the Community Planning and Developments Neighborhoods layer.
  + <https://opendata-geospatialdenver.hub.arcgis.com/datasets/geospatialDenver::american-community-survey-nbrhd-2016-2020/about>
* I chose this data set because it has practical applications that could benefit me, and the people around me. It could be more interesting for potential future employers, as I will be looking for jobs in the Denver metro area.

**Data Profile**

* Cleaning and consistency checks
  + **Columns dropped**= ‘geo\_x’ and ‘geo\_y’ since I already had the GPS data in the ‘geo\_lon’ and ‘geo\_lat’ columns.
  + **Missing values**: 15,451 found within the ‘geo\_lon’ and ‘geo\_lat’ columns
  + **Rows dropped**: 15,451 (only 4% of the whole dataset (386,340 rows))
  + **Datasets merged**: Denver Crimes merged with demographic data since both sets shared “neighborhood\_id” and the demographic data would add more detail to the picture.
  + See Jupyter notebook for details on all changes made to the data.
* Basic stats- there are simply too many variables included to do a cursory assessment of all of them.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Mean** | **Median** | **Standard Deviation** | **Max** |
| Victim\_count | 1.02 | 1 | 0.24 | 19 |

A screenshot of a number of neighborhood names

Description automatically generatedA screenshot of a number of cities

Description automatically generated

* Both the top 10 (left) and bottom 10 (right) for the number of crimes in that neighborhood. These are absolute numbers, so they don’t account for differences in population sizes.
* **Limitations and ethics** 
  + Not limited by the age of data, since it is updated weekly
  + Data does present with many missing entries, highlighting the fact that data collection is an imperfect process and not all pertinent info gets gathered
  + Demographic dataset: values provided are estimates, not exact. Unsure of the margin of error.

**Project Questions**

1. Which neighborhoods have the highest crime rate per capita over the period covered within this data set?
2. Are certain demographic profiles associated with higher crime rates (ie, income status, presence of older or younger residents, education levels, etc.)?
3. What types of crime are most common?
4. Which neighborhoods are the safest to live in?
5. Are certain crimes more prevalent depending on the time of year?

**Hypotheses**

1. Neighborhoods with more young people will have higher rates of crime.
2. Lower socioeconomic status neighborhoods will have more crime.
3. Less-educated neighborhoods will have more crime.
4. Crime will peak during months with less severe weather.