# Data

Our primary data source is the [Ames, Iowa housing sales dataset](https://www.kaggle.com/c/house-prices-advanced-regression-techniques/data) from Kaggle. The dataset contains data for 1,460 homes, spans 25 neighborhoods in Ames, and covers the years 2006 through 2010. The attributes include sales price, square footage, number of bedrooms, and 75 other features for each home.

We merged our primary dataset with neighborhood-level data on school quality, crime, walkability, median income, and unemployment rate. We sourced the school quality data from [GreatSchools.org](https://www.greatschools.org/), which assigns a numerical rating of 1 to 10 for each school district. We sourced the crime data from [INSERT SOURCE]. We sourced the walkability data from [WalkScore.com](https://www.walkscore.com/). For each neighborhood, we pulled both the walk score and the bike score, which respectively measure a neighborhood’s walkability and bike-ability on a numerical scale from 0 to 100. Finally, we sourced median income and unemployment rate data from the 2019 5-year American Community Survey.

Combining the neighborhood-level data with our primary dataset yielded a total of 83 features, which we will use to predict the sales price of homes.