Housing Data Analysis

Using Kings County Data

Christie Sarver

1/26/21

Business Problem

Business Opportunity

Middle class buyers are currently facing a shortage of available and affordable inventory due to the increase in demand during 2020, driven by Covid

Project Goal

Understand which features a home should include at the median price range in order to effectively design new homes for this consumer

Data Understanding

❖ The data used includes house sale prices and conditions from houses sold through 2014-2015 in the Seattle area

❖ These houses were built as early as 1900 and as late as 2015

❖ The data includes several variables that describe the house, its location, and its condition
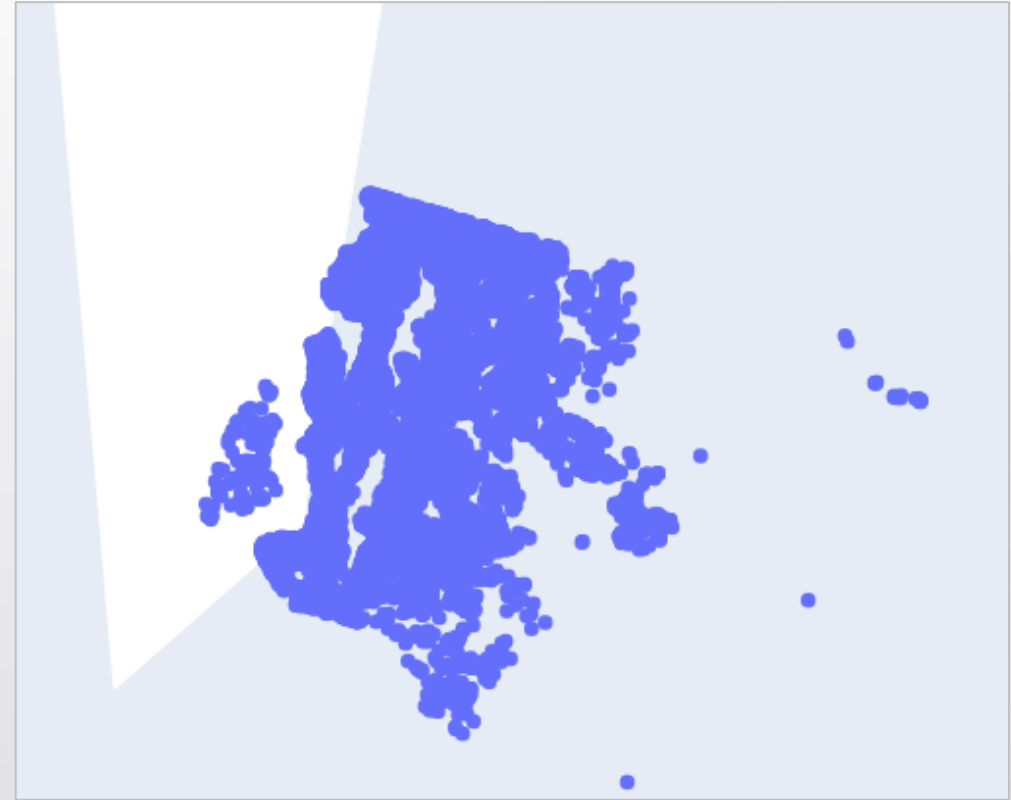
Methodology

❖ This project follows the cross-industry process for data mining or CRISP-DM

❖ Business understanding was applied to the data in order to create a relevant model*

❖ Data was cleaned and edited until model contained most accurate and relevant information

Model analyzes houses within a price range of between $89K and $1.15MM
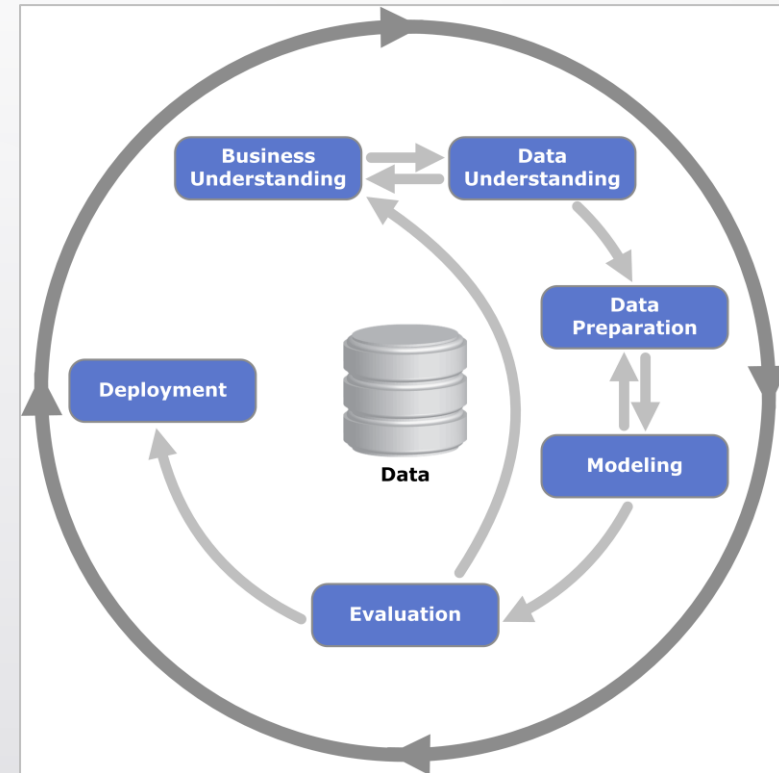
Results

❖ Features that drive value in homes for our target buyer are:

  ❖ Size of the home interior

  ❖ Number of bathrooms (more so than bedrooms)

  ❖ High construction quality and materials

❖ The value is significantly decreased where:

  ❖ Construction quality is average or low

  ❖ House is further away from the city center
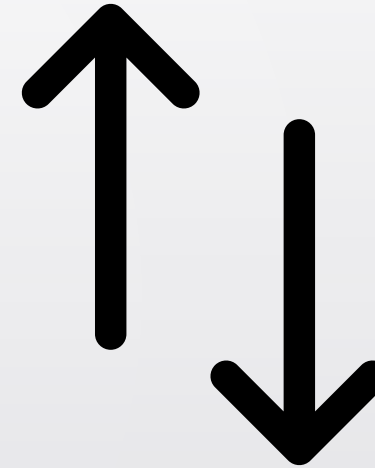
Recommendations

- ❖ Focus on maximizing the living area of the house over the yard size, or adding a basement
  - ❖ Buyers will likely sacrifice lot size to be closer to city
- ❖ Use high quality construction methods
  - ❖ This may increase the budget but will be worth the investment
- ❖ Build multi-floor homes and include ample bathrooms to reflect what buyers are looking for

Suggestions for Future Work

❖ Gather more recent data with a wider date range

❖ To make the model more generalizable, include data from multiple markets

❖ Pull in additional variables on neighboring lots to make recommendations on structure of development

THANK YOU!