Boston Housing Prices

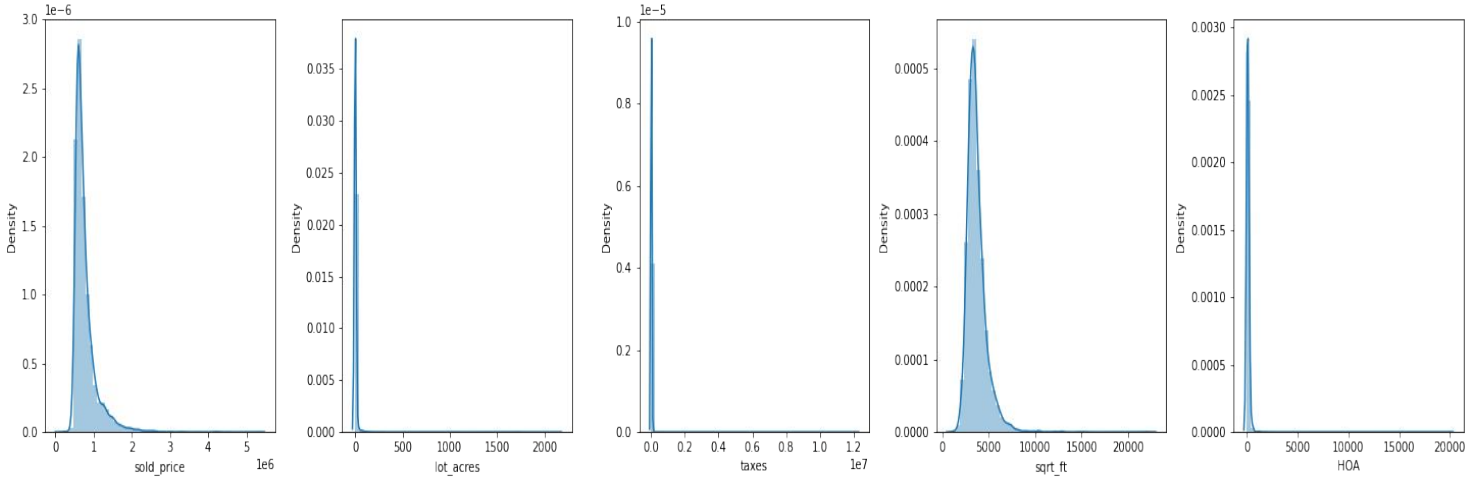
Ezra J. Cook

Imputation

- ➔ Imputed many columns with zero where it made sense.
- ➔ When not imputing with zero used a combination of mean and groupby
  - sqrt_ft
  - taxes
  - lot_acres
- ➔ Some categorical features had over one hundred values, others were encoded as int or float where numerical
  - kitchen_features > 100
  - floor_covering > 100
  - bedrooms as float
  - bathrooms as float
  - garage as float
  - year_built as int > 100

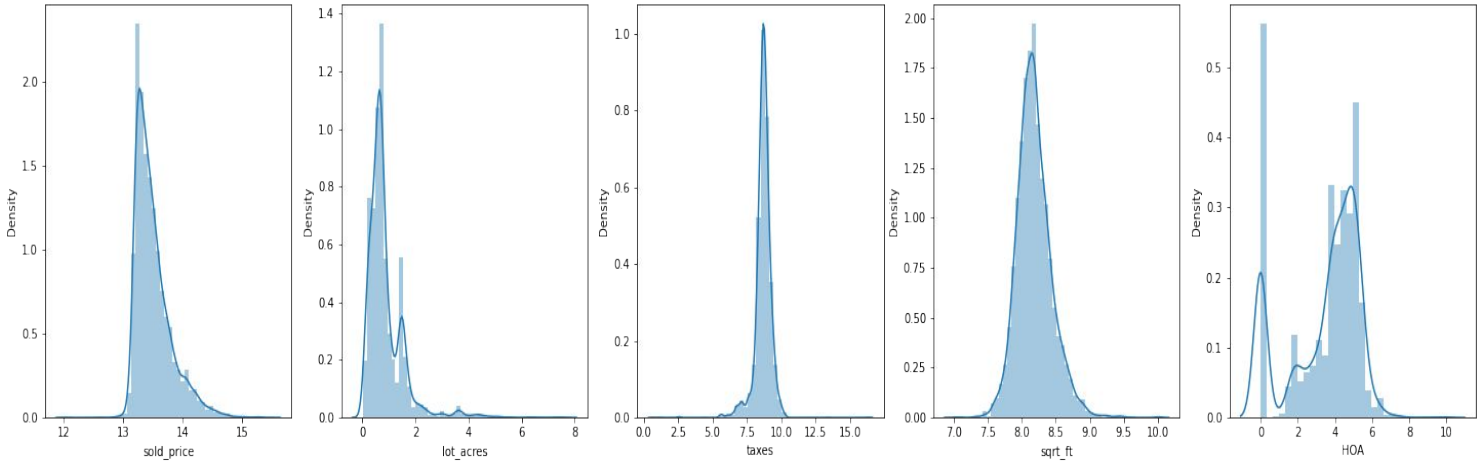
Visualization - Numerical

Numerical Features - Density Plot



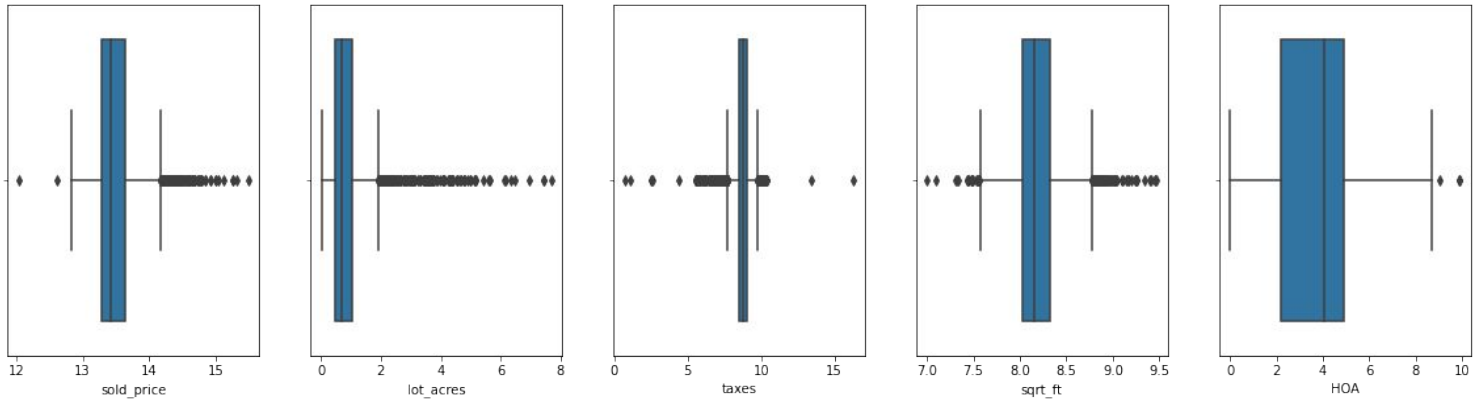
Visualization - Numerical

Numerical Features - Log Scale

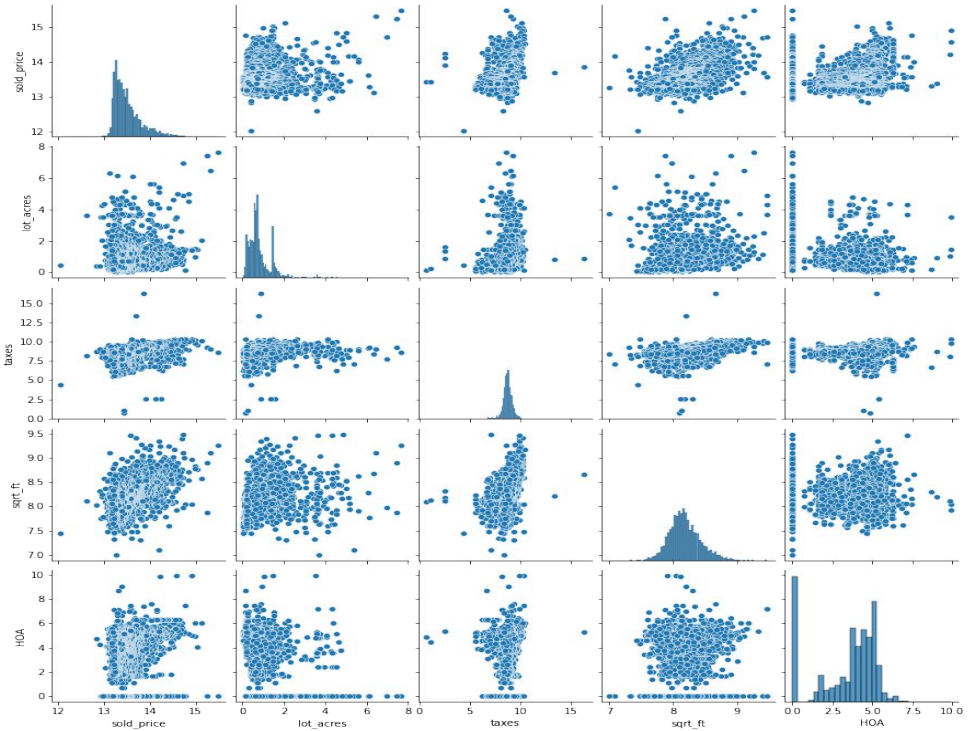


Visualization - Numerical

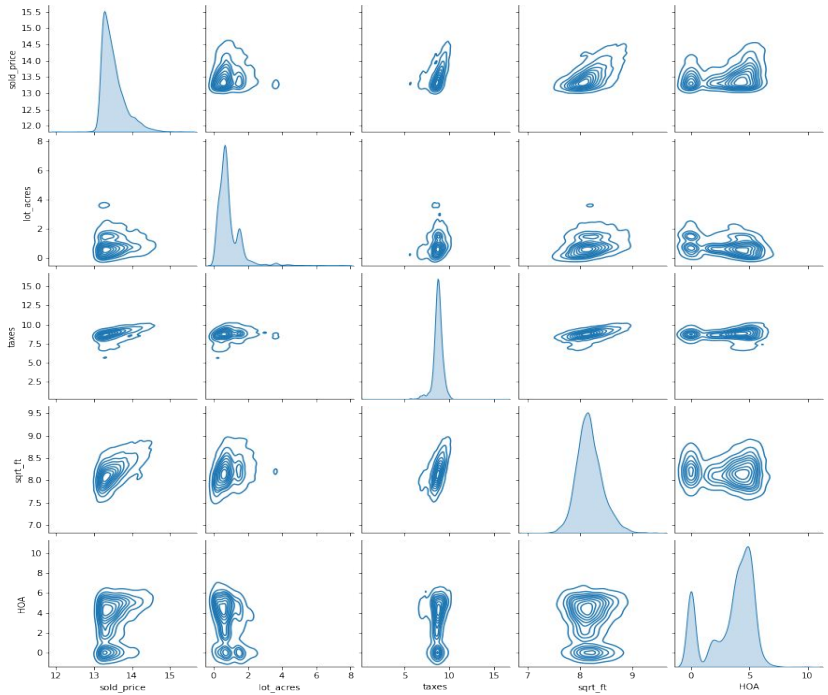
Numerical Features - Box Plots



Visualization - Scatter Matrix

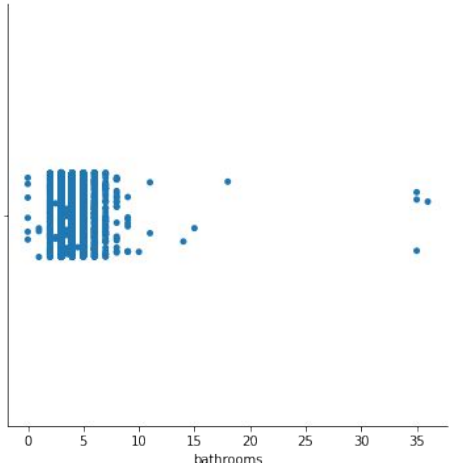


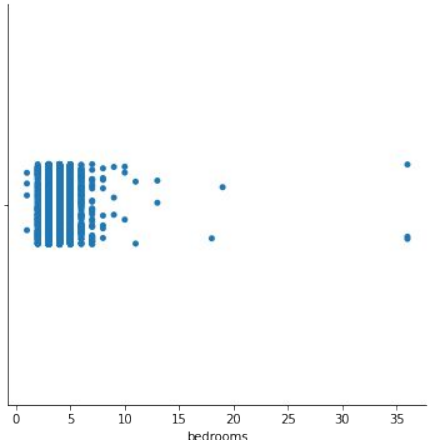
Visualization - KDE

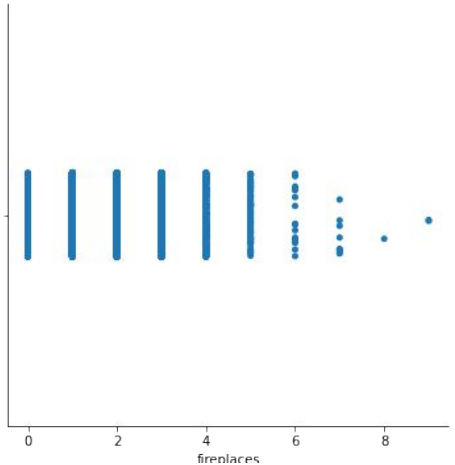


Visualization - Cat Plots

Outliers Were Removed

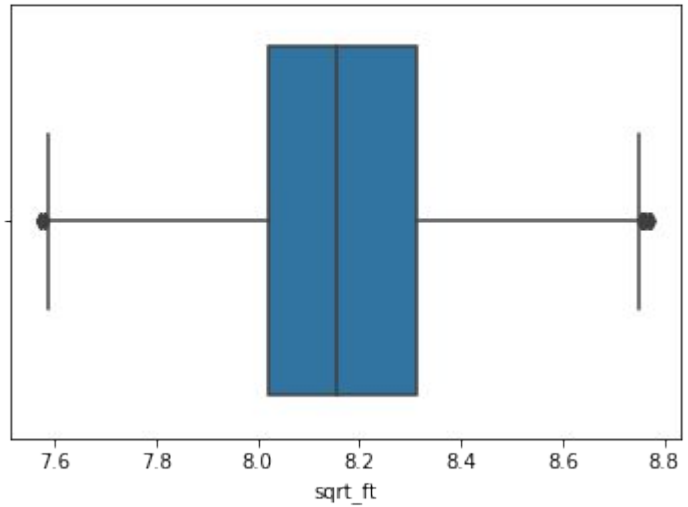






Visualization - sqr_ft No Outliers

Removed 1.5 * IQR



Summary

- ➔ Removed null values
- ➔ Visualized data
- ➔ Log scaled numerical values with right skew
- ➔ Removed outliers
- ➔ 97% of data retained