Ames for accuracy: Predicting housing prices in Ames, IA

Presenter: Christopher McGowen

What should I expect to pay for a house, and what features increase the cost?

Process:

Data cleaning:

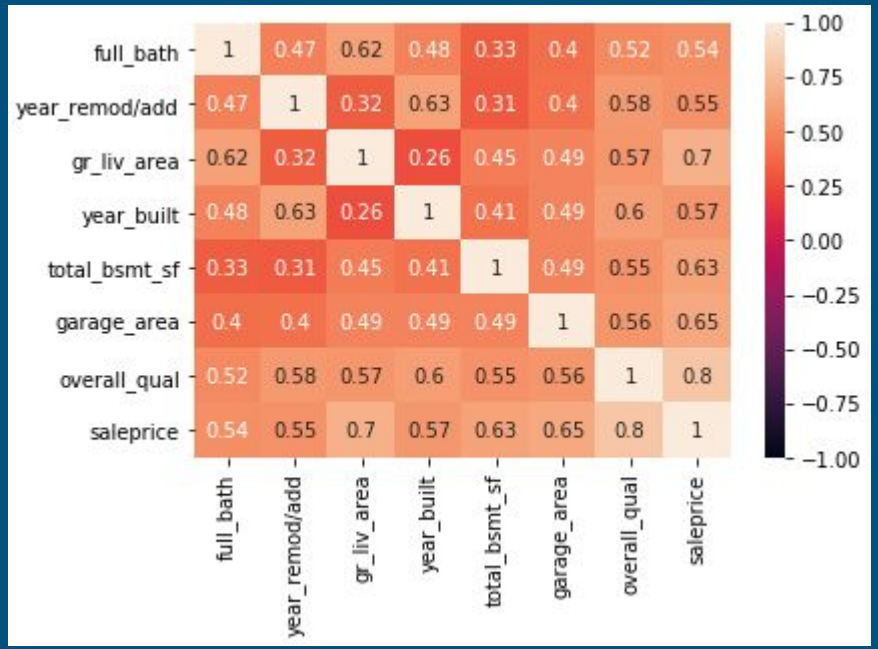
- 9822 missing values!
- Different types of columns
- Sometimes nulls were actually ‘NA’

01	**Ordinal data**	• Columns of quality (Basement condition etc.) • Ranked: Ex, Gd, TA, Fa, Po, NA
02	**Nominal data**	• Columns like zoning etc. • Split into dummy columns of 0, 1

Data analysis:

- Used heatmap and pairplot
- Most columns have no collinearity

Selecting features:



Modeling:

- Using the columns from above = 0.789
- Using all numerical columns = 0.814

Summary:

- Features that significantly increase sale price:
  - Total liveable sqft
  - Overall quality
  - Basement sqft
  - Garage size

Next steps:

- More robust modeling methods (Lasso, Ridge, etc.)
- Feature engineering

Questions?

Presenter: Christopher McGowen