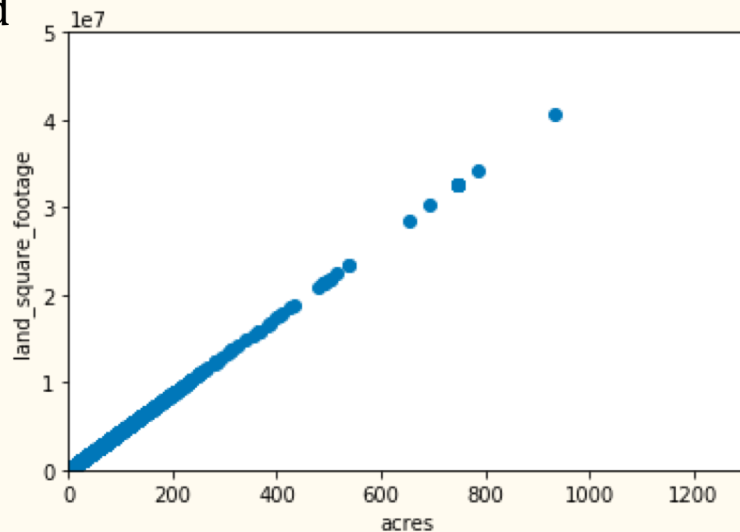# Predictions of house price
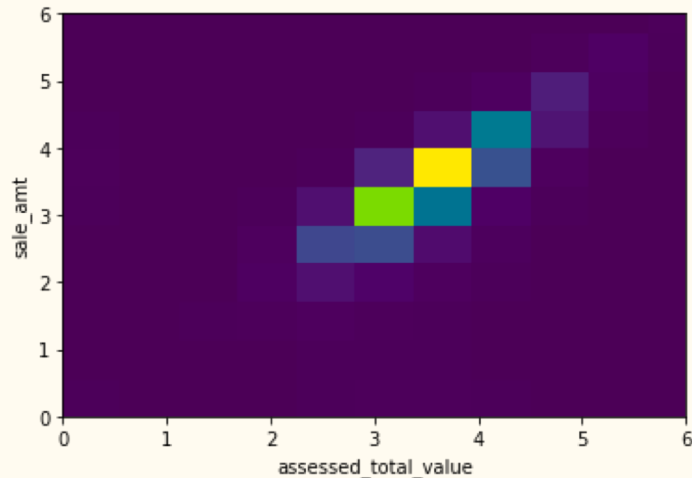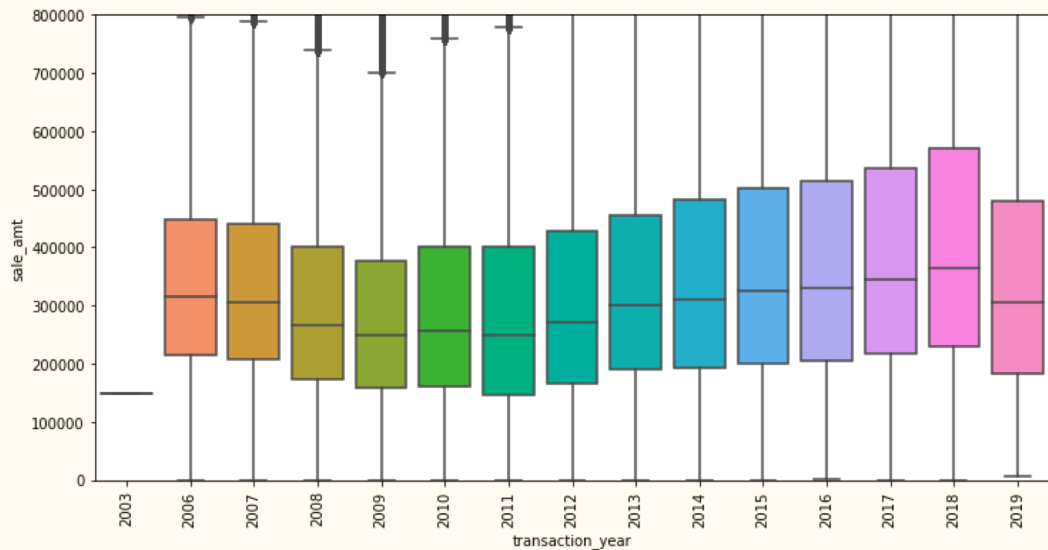
—

By Yaqi Zhang

Data Preprocessing

- Columns with only NAN are deleted.

- Columns with duplicate information are deleted

- Pick 1 column with correlated features (shown

with figure on the right)

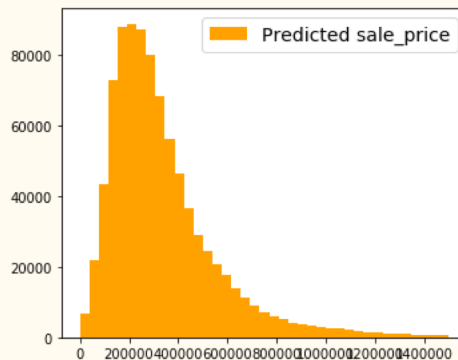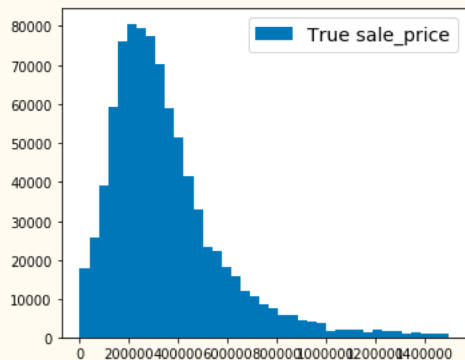- Log transformation of numeric features

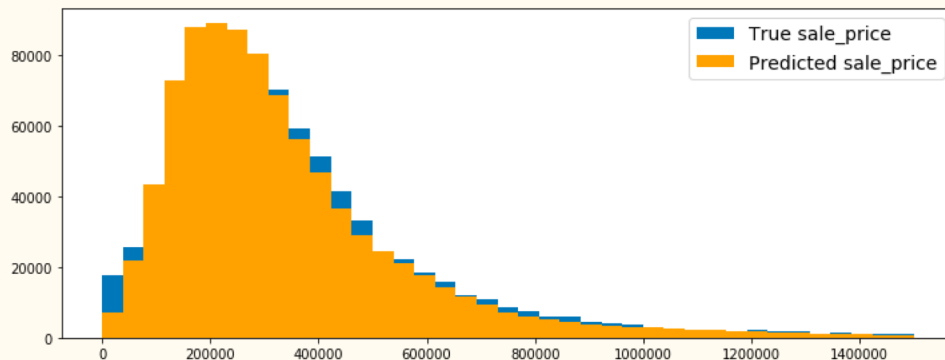# Relations between sale_amt and other features

Results

LightGBM model, dataMA

Histograms of true sale_price and predicted sale_price using LightGBM model

Percentage Results (within 10%)

LightGBM	dataMA	dataRI	dataPA
training	50%	41%	
testing	53%	50%	

dataRI	LightGBM	linear_model	K-NN
training	41%	18%	19%
testing	50%	22%	16%