Statistical Inferences into the Property Development Industry

Nancy Ho

Outline

- Business Problem
- Data and Methods
- Results
- Conclusions
- Next Steps

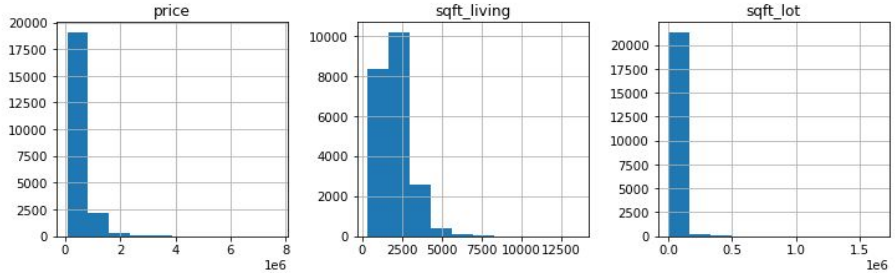
Business Problem

- Property development -- rewarding yet risky business
  - Encompasses many processes in building and selling property
  - Can be hard to make decisions when starting out as a new developer
- Gain foundational understanding of best choices to make by asking questions:
  - What kinds of factors contribute to a house's base value?
  - When building a house, what factors should we look out for before beginning the construction of property?
  - How does the condition of a house at the point of sale influence the value of a house?
  - Are there specific features that add significant value to houses (e.g. rooms, floors)?
- Goal: to be able to make informed decisions starting out as new developer

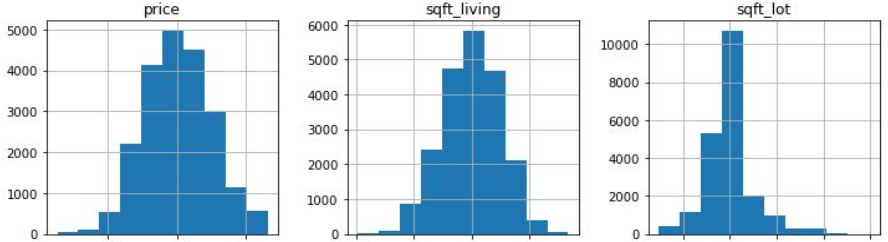
Data and Methods

- Dataset containing information about house sales in Kings County, California
- Use of feature scaling and transformation tools to normalize distribution of data and improve reliability of data

Before transformation:

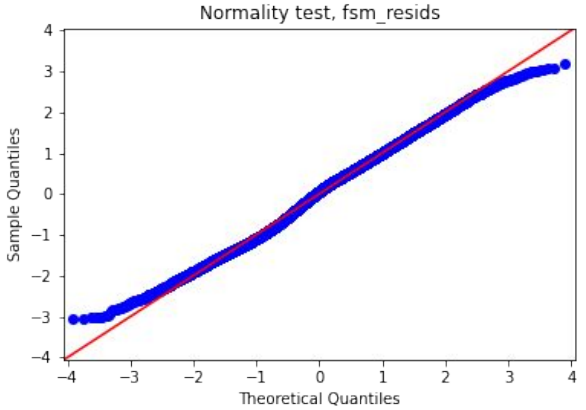


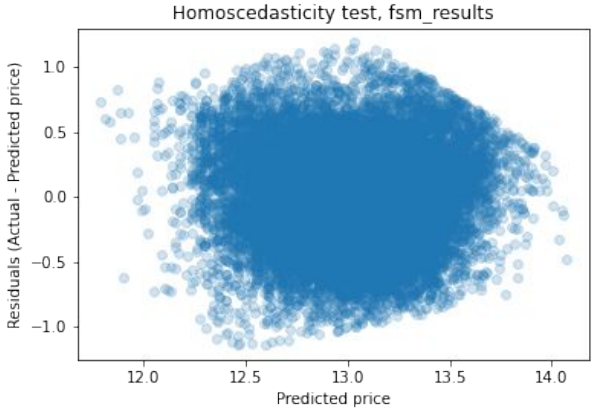
After transformation:



Regression Analysis

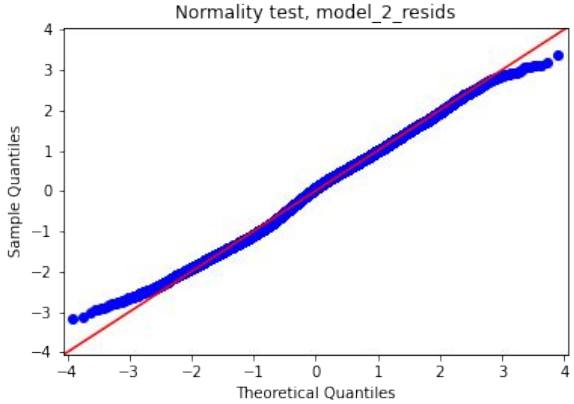
- Target variable = price
- Features:
  - sqft_living
- Regression statistic = 0.412
- Assumptions of linear regression:
  - ✓ Linearity
  - X Normality
  - X Homoscedasticity
  - ✓ Independence

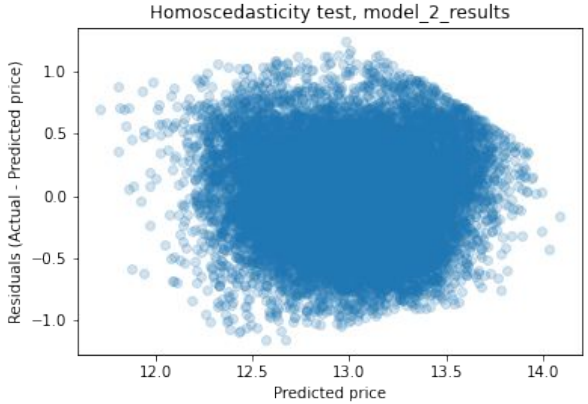




Regression Analysis

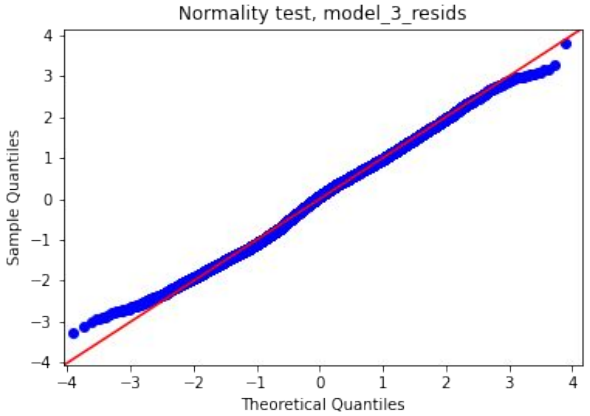
- Target variable = price
- Features:
  - sqft_living
  - sqft_lot
- Regression statistic = 0.422
- Assumptions of linear regression:
  - ✓ Linearity
  - X Normality
  - X Homoscedasticity
  - ✓ Independence

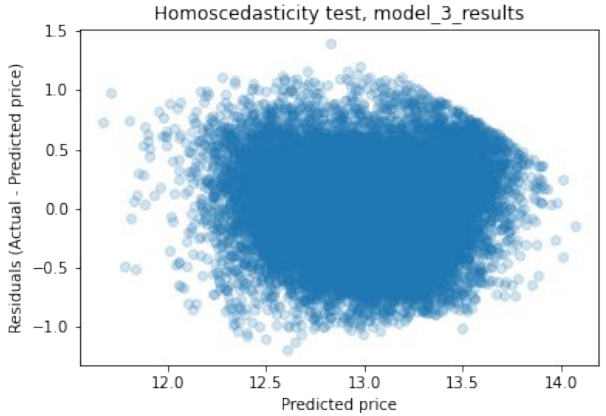




Regression Analysis

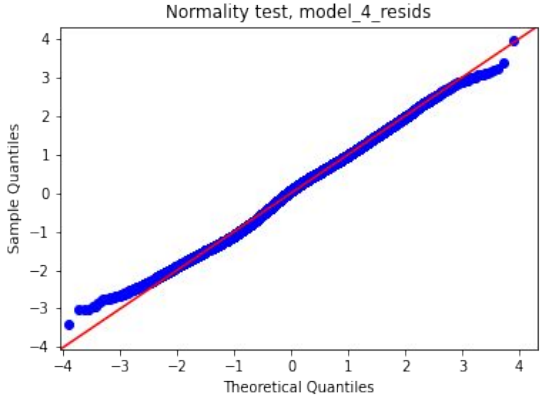
- Target variable = price
- Features:
  - sqft_living
  - sqft_lot
  - condition
- Regression statistic = 0.429
- Assumptions of linear regression:
  - ✓ Linearity
  - X Normality
  - X Homoscedasticity
  - ✓ Independence

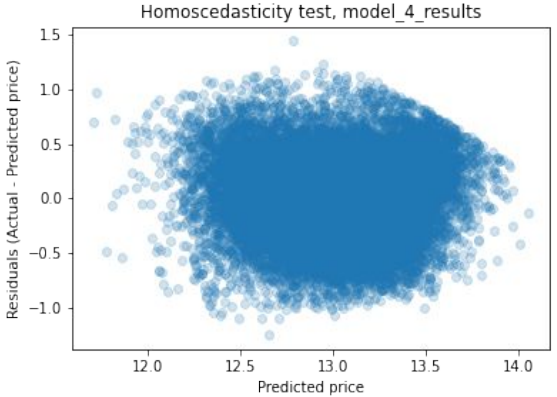




Regression Analysis

- Target variable = price
- Features:
  - sqft_living
  - sqft_lot
  - condition
  - floors
- Regression statistic = 0.432
- Assumptions of linear regression:
  - ✓ Linearity
  - X Normality
  - X Homoscedasticity
  - X Independence





Results and Conclusion

- Third model is our best model -- gives us best regression statistic while meeting most assumptions of linear regression
  - A house's base value with no features would be about 12.670
  - For 1 unit increase in square footage of house, price increases by 0.313
  - For 1 unit increase in square footage of lot, price decreases by 0.041
  - For 1 unit increase in house's condition, price increases by 0.061
- Important factor: proportion between square footage of lot and house
- Condition and number of floors do have significant impact, but smaller
  - Rooms were not observed because of low correlation with price and multicollinearity with other variables

Next Steps

- Finding other methods to scale and transform data to gain better regression statistics and parameters
- Additional analysis on housing data in other counties
- Analyzing data about how housing prices change over time in response to economic factors (i.e. inflation)
    - May help us be proactive in decisions as prospective property developers

Thank you!

Email: nancyho83@yahoo.com
GitHub: @nancyho83
LinkedIn: linkedin.com/in/nho3/