D. Journic Phase 2 Project

Topics

- This presentation will cover:
  - Business Problem
  - Data used
  - Models created and how
  - Further issues
  - Conclusion

Business Problem

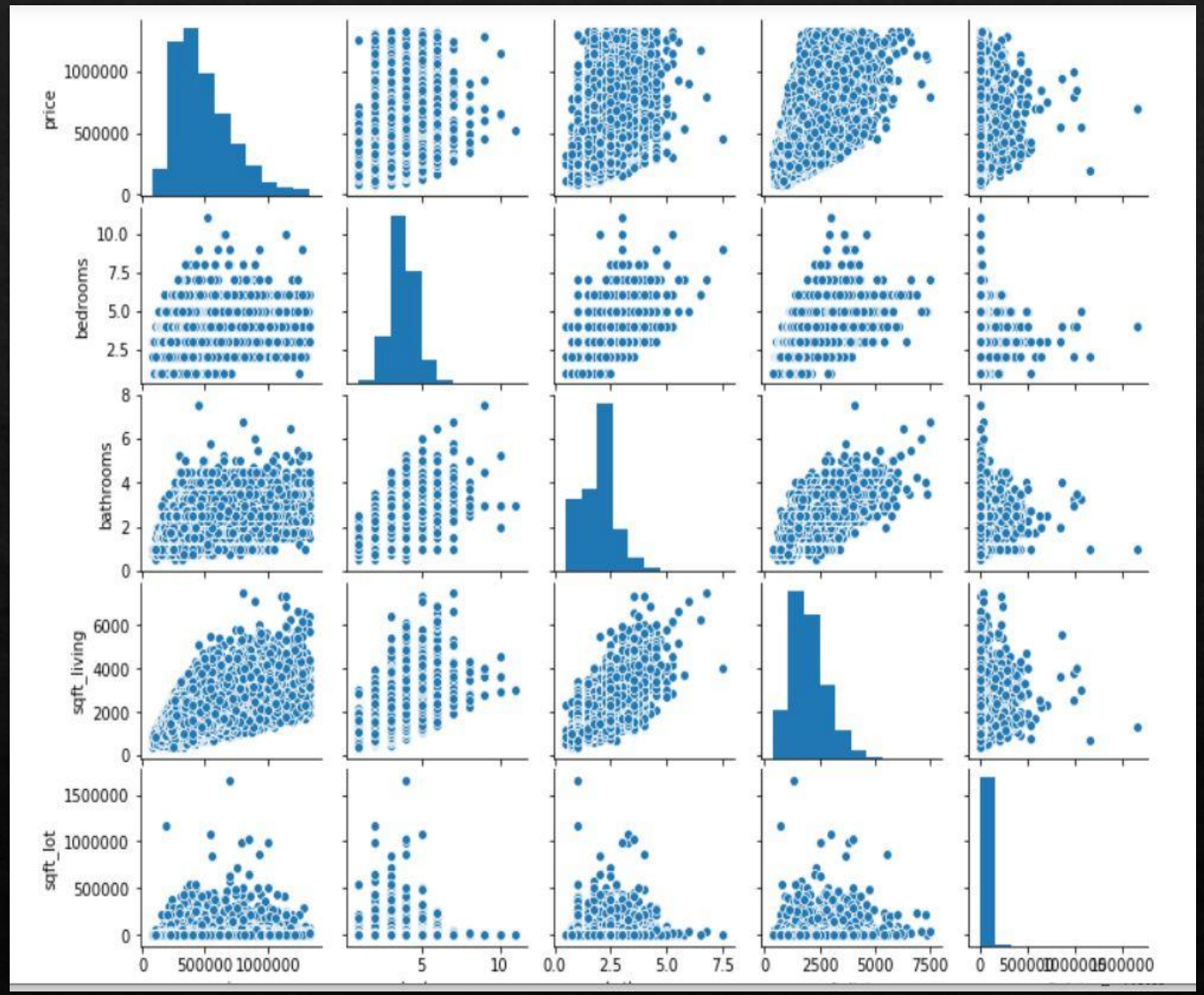
- Assist homebuyer in finding a fair price based on specific criteria:
  - Bedrooms
  - Bathroom
  - Living area
  - Lot size

Data

- Housing data from King County, WA
- Approx 21,600 houses

Variable Correlations

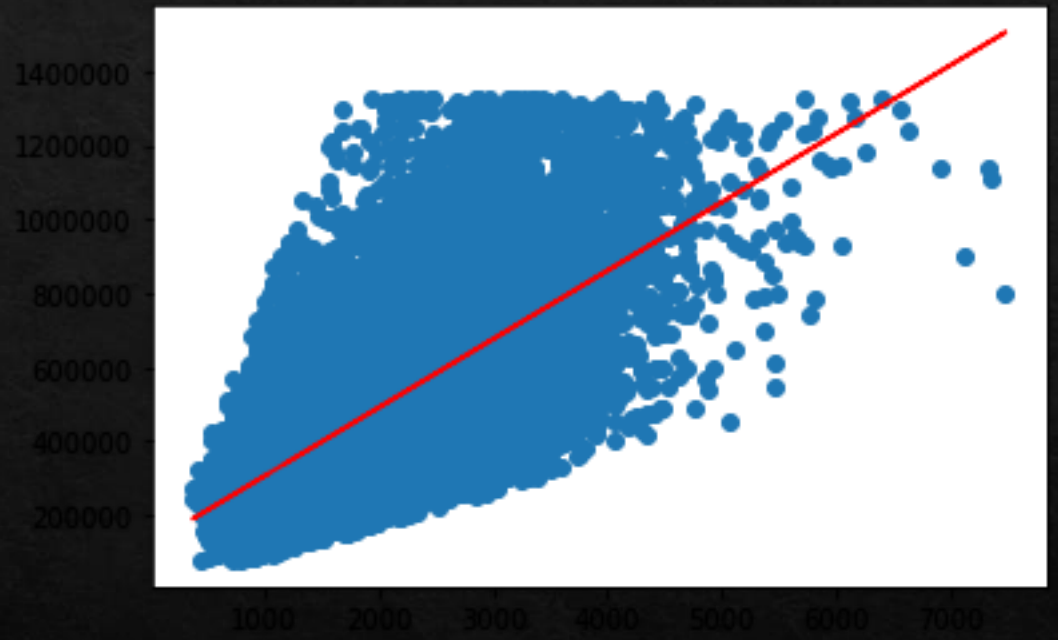
- The highest correlation with price is living area (.64), followed by bathrooms (.47), then bedrooms (.31) and lot size (.089)



Simple linear regression

Living area and price.

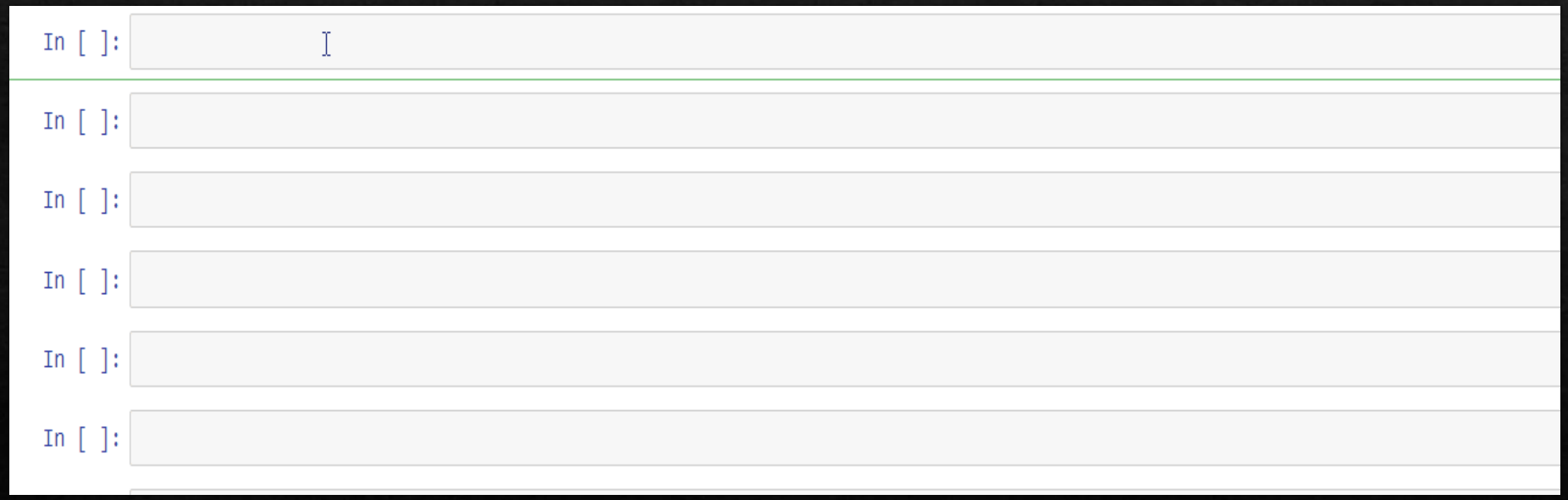
R2= 0.415



- Other models run:
  - Living area, lot, beds and baths Vs. Price (R2= 0.51)
  - LT Living area, LT lot Vs. LT Price (R2= 0.401)
  - LT Living area, LT lot, beds, baths Vs LT Price (R2= 0.402)

- Limits/manipulations used:
  - Capped the house price to $1.33MM
  - Capped bedrooms to 6
  - Capped bathrooms to 5
  - Removed outliers

Estimator function:



Conclusions

- After all the data manipulating and narrowing of scope, the best model is still the first with roughly 50% reliability.
- More work is needed to improve the model:
  - Organizing/limiting the data to specific area code
  - Further limits placed on sqft both in living area and acreage

Thank you.
Questions?

djournic@gmail.com