

A top-down view of a white ceramic cup filled with dark coffee. The coffee surface is covered in a thick layer of brown foam with many small, dark bubbles. A silver spoon is partially visible on the left side of the cup. The background is a soft, out-of-focus grey.

D. Journic Phase 2 Project

Topics

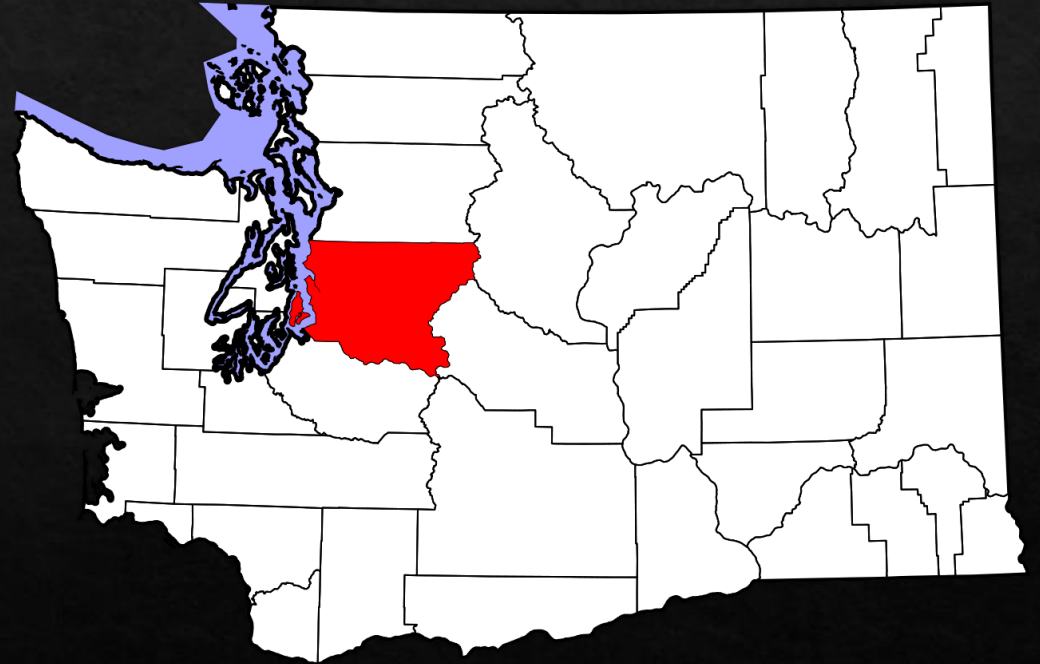
- ◆ This presentation will cover:
 - ◆ Business Problem
 - ◆ Data used
 - ◆ Models created and how
 - ◆ Further issues
 - ◆ Conclusion

Business Problem

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

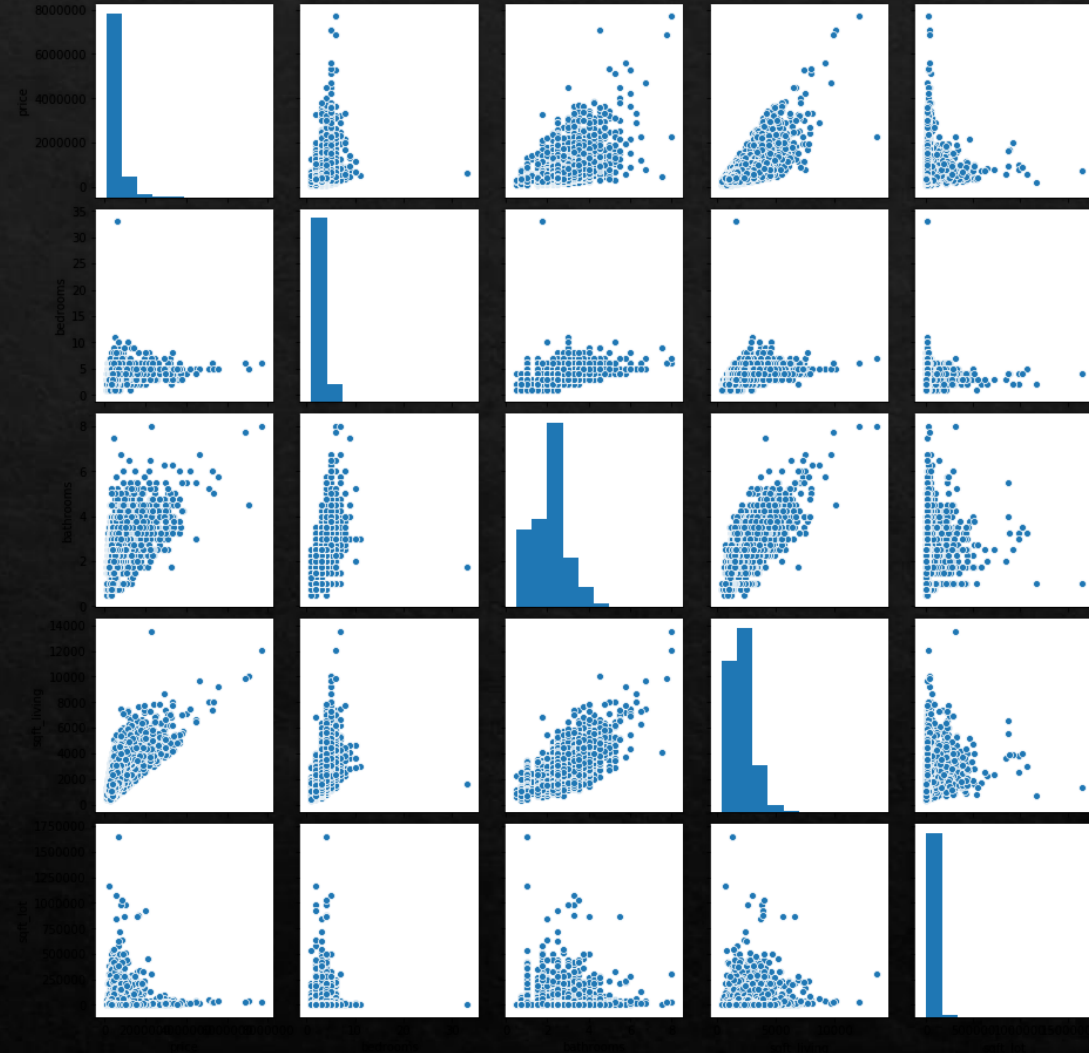
Data

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



Variable Correlations

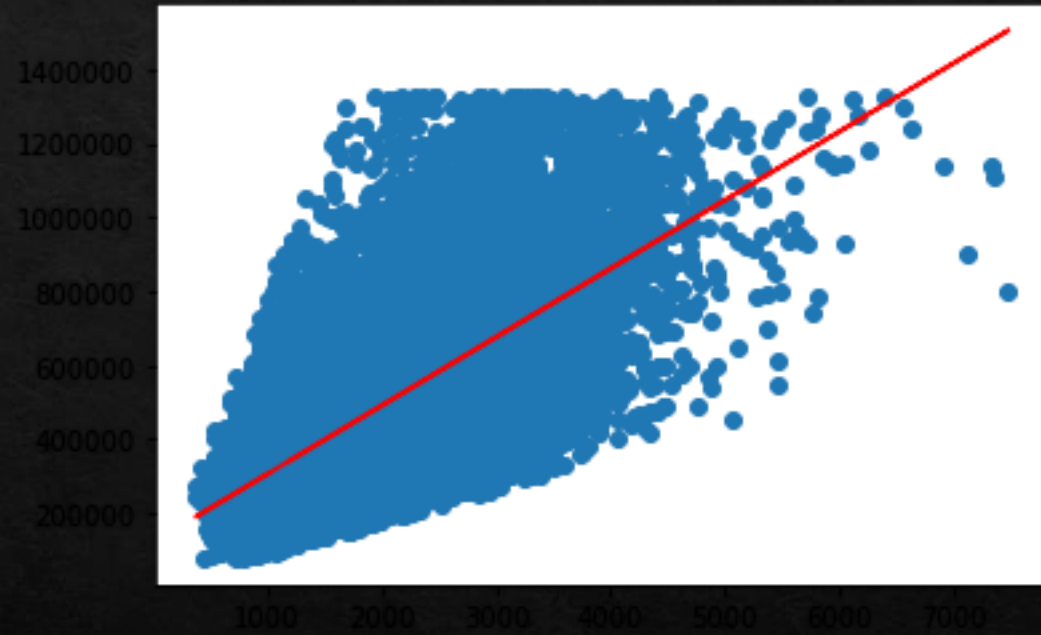
- The highest correlation with price is living area (.64), followed by bathrooms (.47), then bedrooms (.31) and lot size (.089)



Simple linear regression

Living area and price.

$R^2 = 0.415$



◇ Other models run:

◇ Living area, lot, beds and baths Vs. Price ($R^2 = 0.51$)

◇ LT Living area, LT lot Vs. LT Price ($R^2 = 0.401$)

◇ LT Living area, LT lot, beds, baths Vs LT Price ($R^2 = 0.402$)

◇ And many, many more.....

Estimator function:

In []:

In []:

In []:

In []:

In []:

In []:

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.

Thank you.
Questions?

djournalic@gmail.com