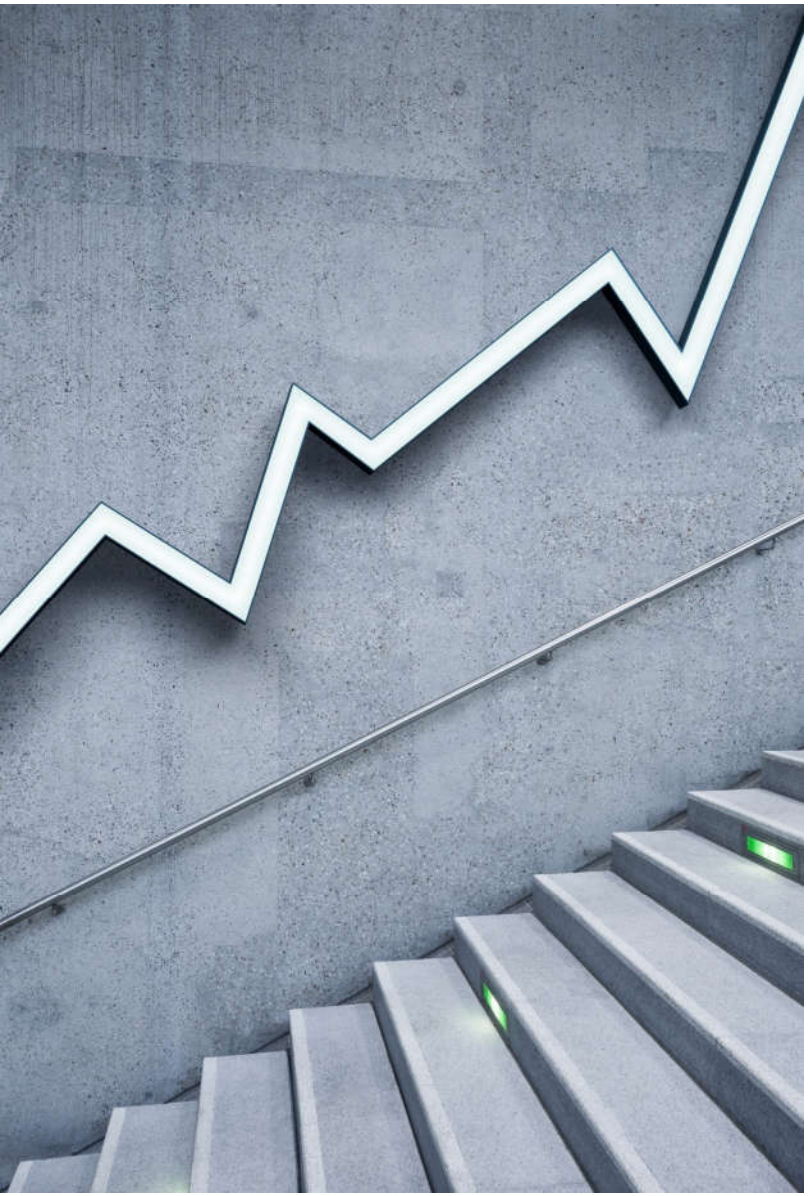

Model driven investing

IN KING COUNTY, WASHINGTON

BRIAN TRACY

DECEMBER 10, 2021

Outline

- Business Problem
- Data & Methods
- Modeling Process
- Regression Results
- Recommendations
- Next Steps
- Questions

Business Problem

- A new bitcoin millionaire in Seattle has decided they want to diversify into real estate
- They want our company to build a model to predict home sales in King County, Washington
- The goal is to use our model to predict the price of a home currently for sale in order to influence the decision to buy.

Data & Methods

- Data provided on home sale information from homes sold in King County between May 2014 and May 2015
- 21,597 entries in the data
- Features include:
  - Sale price
  - Home composition, condition and quality
  - Location
- Using python and jupyter notebook for work environment

Modeling Process

Baseline Model:

- After conversion of categorical features
- Includes all features available
- Provides a reference point for future iteration of model to compare
- Attempts to explain the relationship between all the features and the price of the home.

Successive iterations:

- The first step is pruning features based on linear regression assumptions and multicollinearity
- After pruning, attempted to include removed features by engineering new features
- Evaluate after each addition, decide to keep or remove more features
- Continue until unable to improve model further
- Final iteration transforms price to attempt even better fit

Regression Results

- The final model can explain about 72% of the variation when scored against testing data
- 32 predictors were used in the final model
- Square footage of the home seems to be the primary driving predictor of price
- Other predictors which have a positive effect on price include:
  - Waterfront property
  - Property with a view
  - Property recently renovated
  - Property in excellent condition
- Scoring against test data supports these results

Recommendations

- Evaluating against test data shows a high error
- Model should not be the only decision maker for purchasing a home.
- Likely additional features that affect home price we either don't have the data for or haven't engineered yet.

Next steps

- Other features that may affect home price could include:
  - Relative location to amenities (like grocery stores, entertainment, restaurants, etc)
  - Schools
  - Crime
  - Public services
- Gather more and newer home sale data
  - The data we have is about 6 years old
  - Much has changed in the real estate market in that time
  - May help identify trends

Questions?

GITHUB: HTTPS://GITHUB.COM/BRTRACY/DSC-PHASE-2-PROJECT-V2-3