

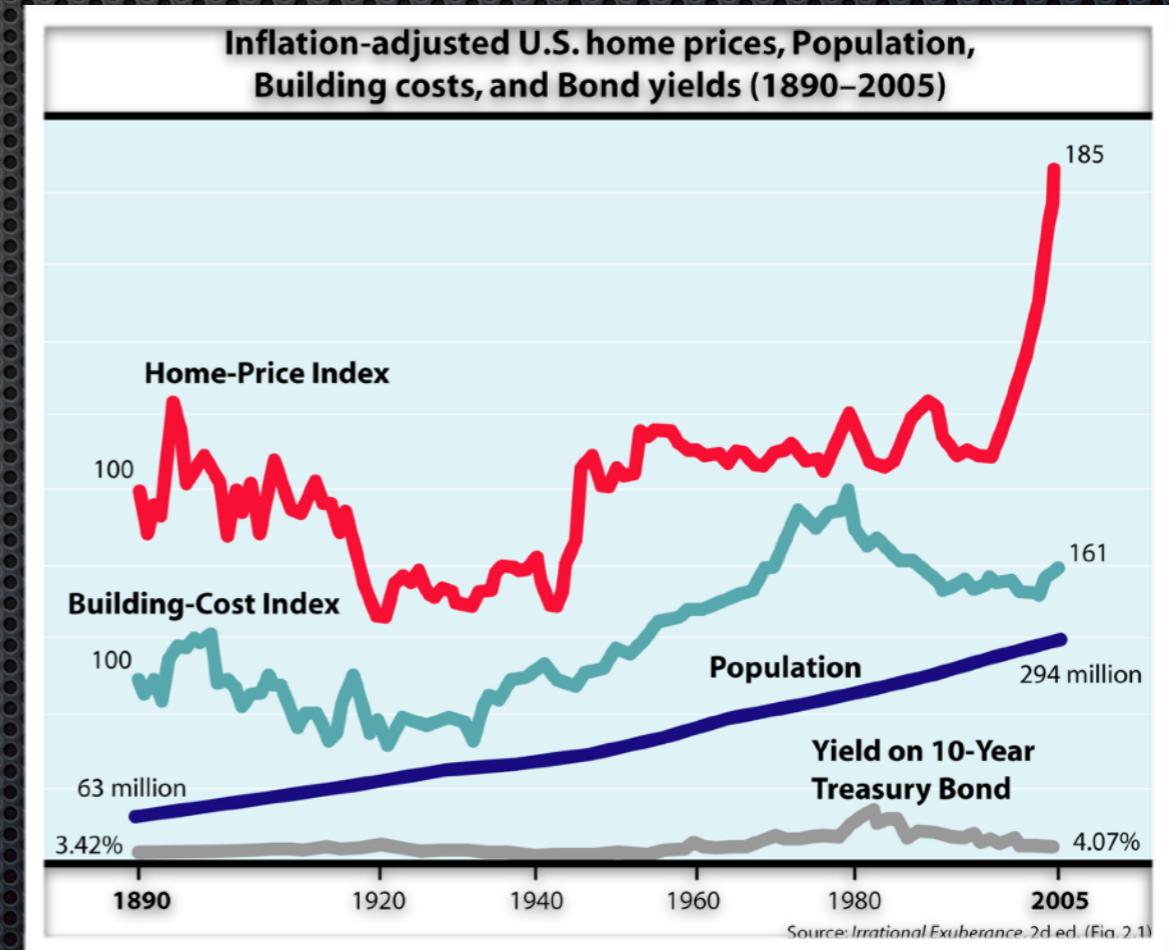
Predicting Sales Prices for the Kaggle Ames, Iowa Housing Dataset

Team Proxima Centauri - Wenchang, Kenny,
James, Danny

Brief Overview of Data Set

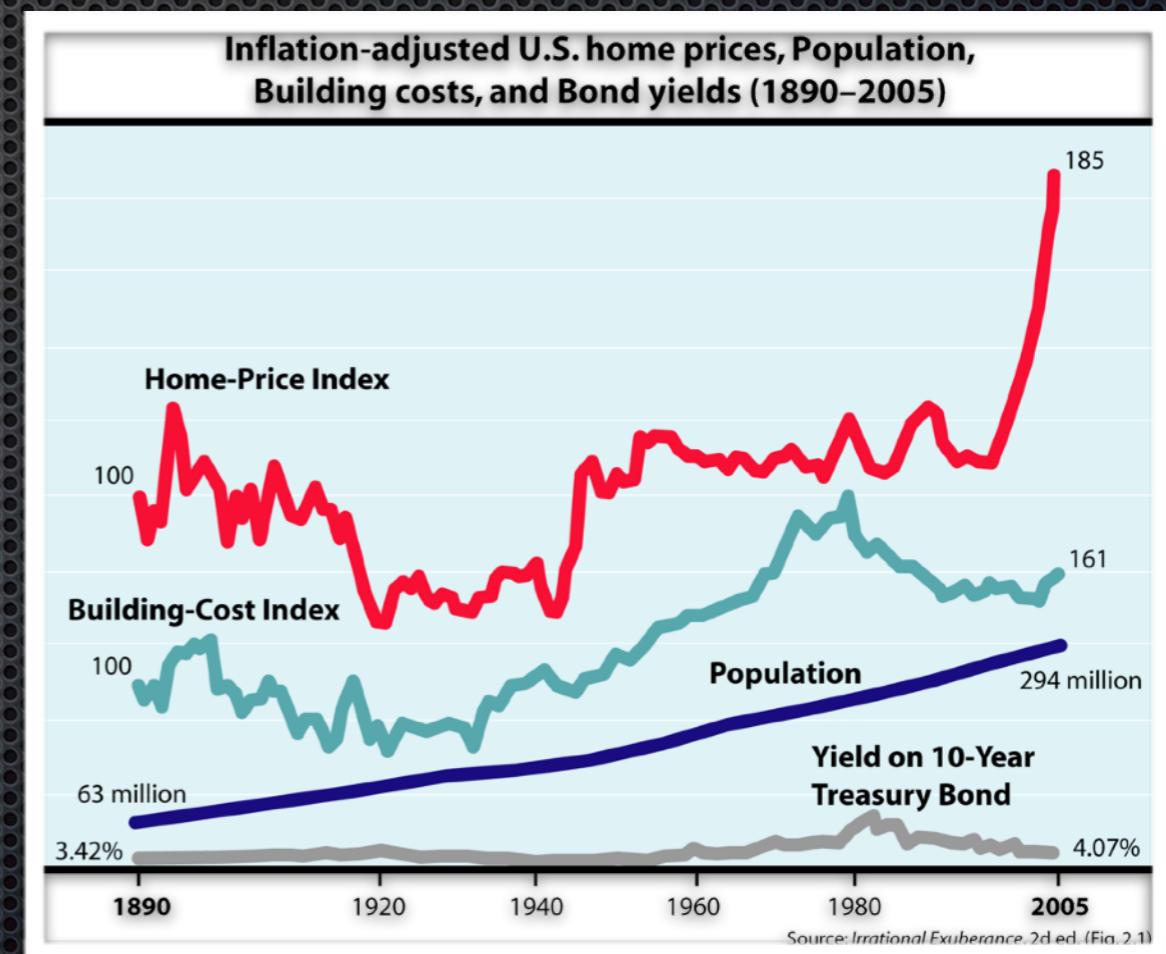
- 80 variables tied to price of home sales in Ames, Iowa between 2006-2010
- Initially obtained from Ames City Assessor's Office with 113 variables that described 3970 property sales -> cleaned and collated by Dean De Cock {Journal of Statistics Ed., **19**, (3) 2011}
- Reduced to 80 predictors:
 - 20 continuous variables {total dwelling sq. footage etc}
 - 14 discrete variables {number of bathrooms, kitchens}
 - 46 categorical variables - 23 nominal/23 ordinal {street pavement, neighborhood}

The US Housing Bubble



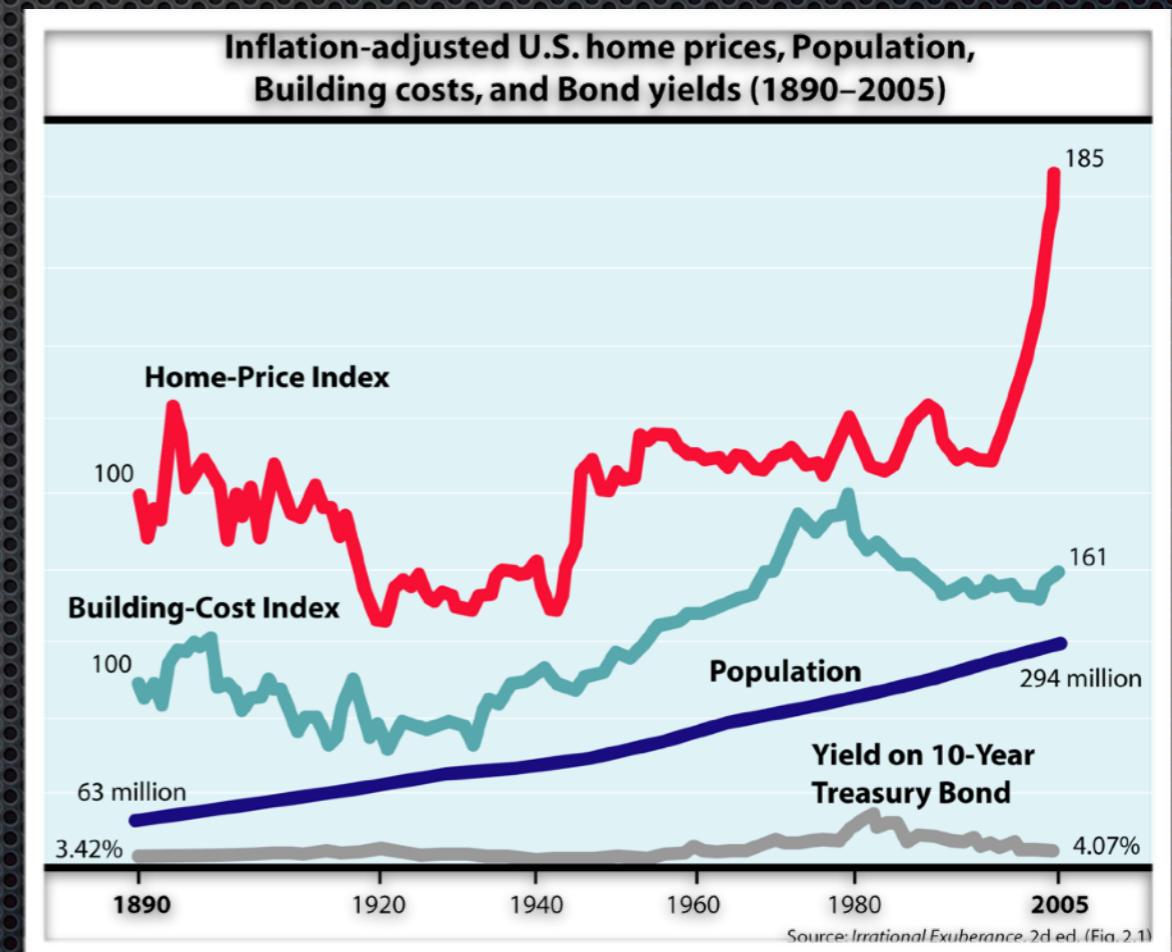
The US Housing Bubble

- What was going on between 2006-2010 in our country?
- Will there be a noticeable effect?



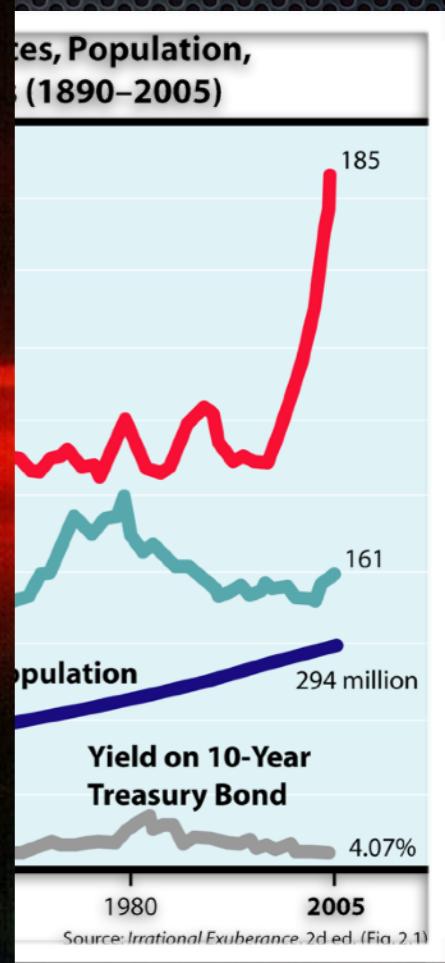
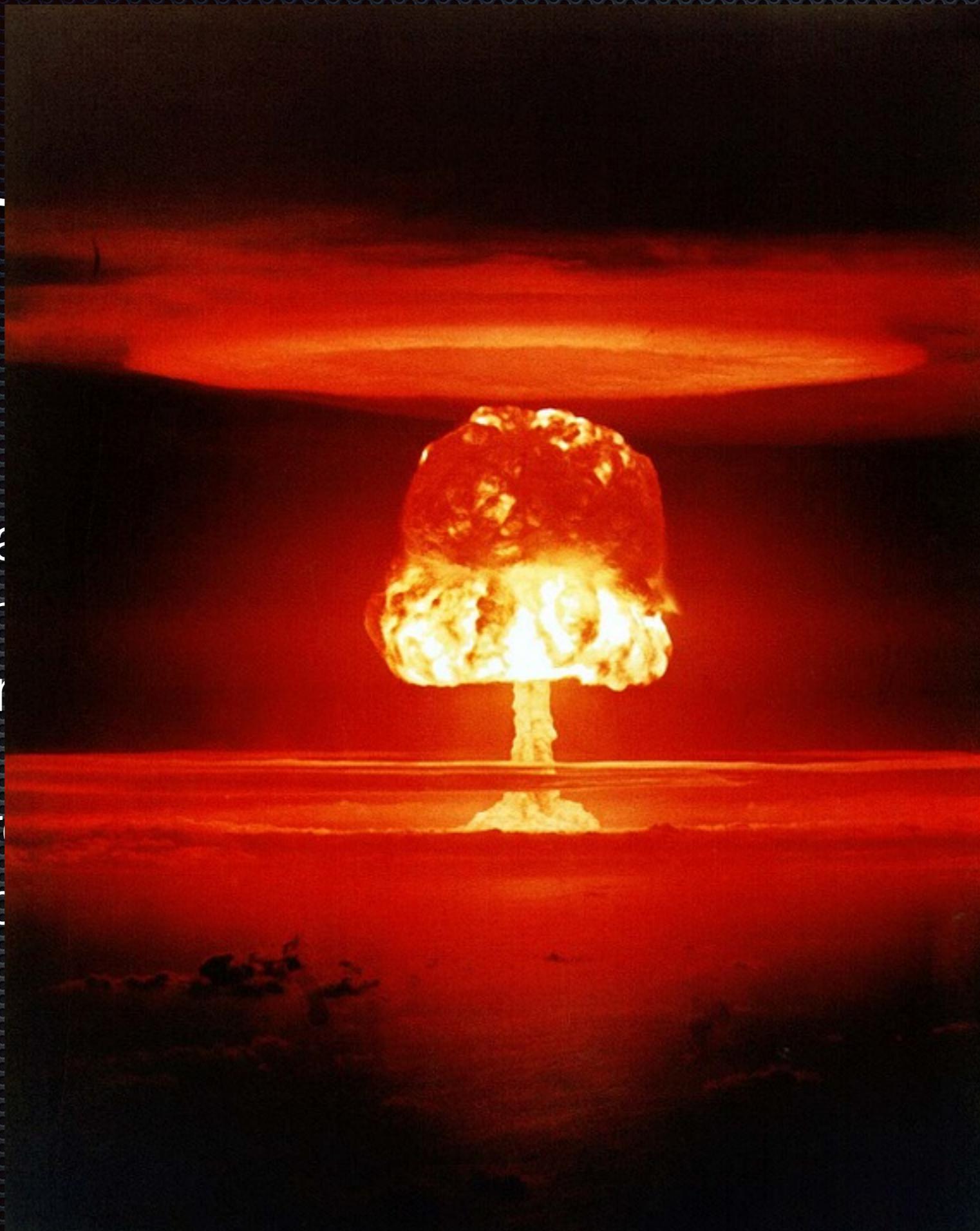
The US Housing Bubble

- What was going on between 2006-2010 in our country?
- Will there be a noticeable effect?



The U.S.

- What was between our country
- Will there be noticeable





LiveSlides web content

To view

Download the add-in.

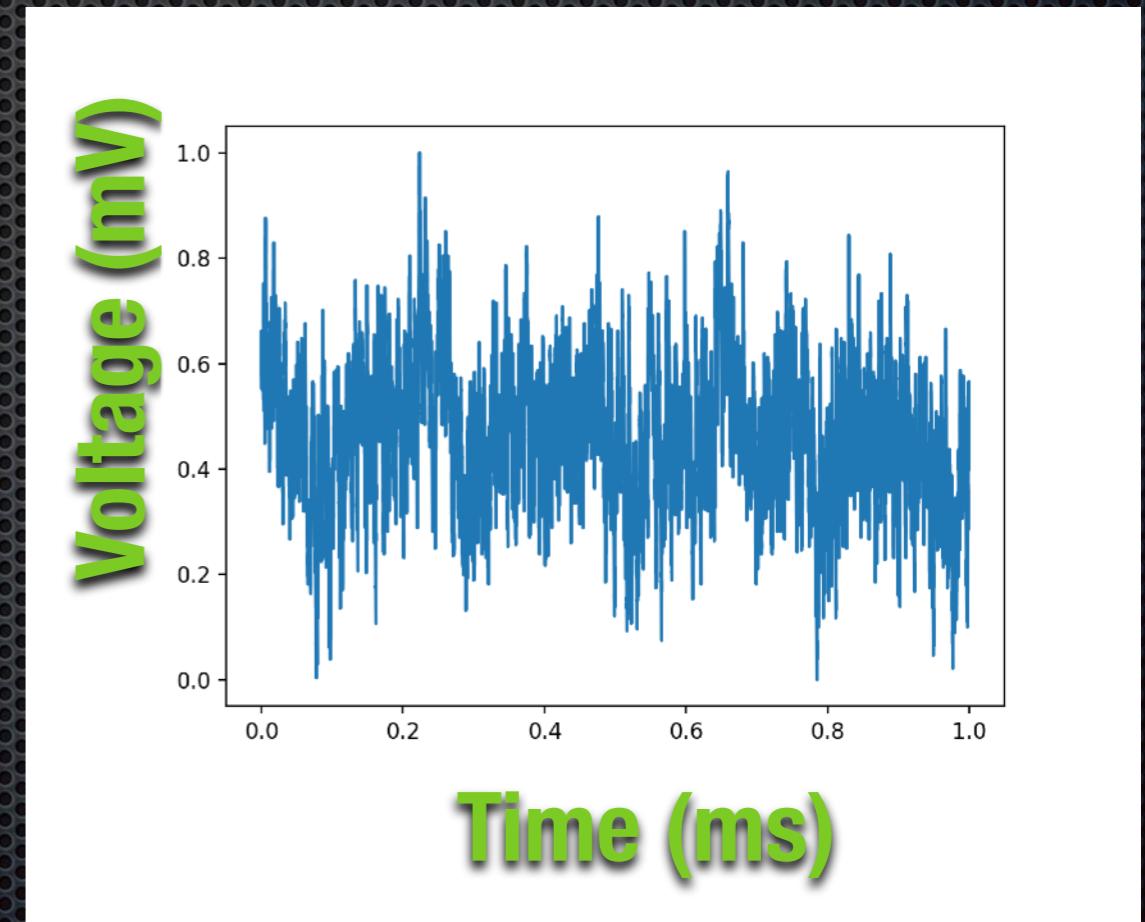
liveslides.com/download

Start the presentation.

Hurst exponent

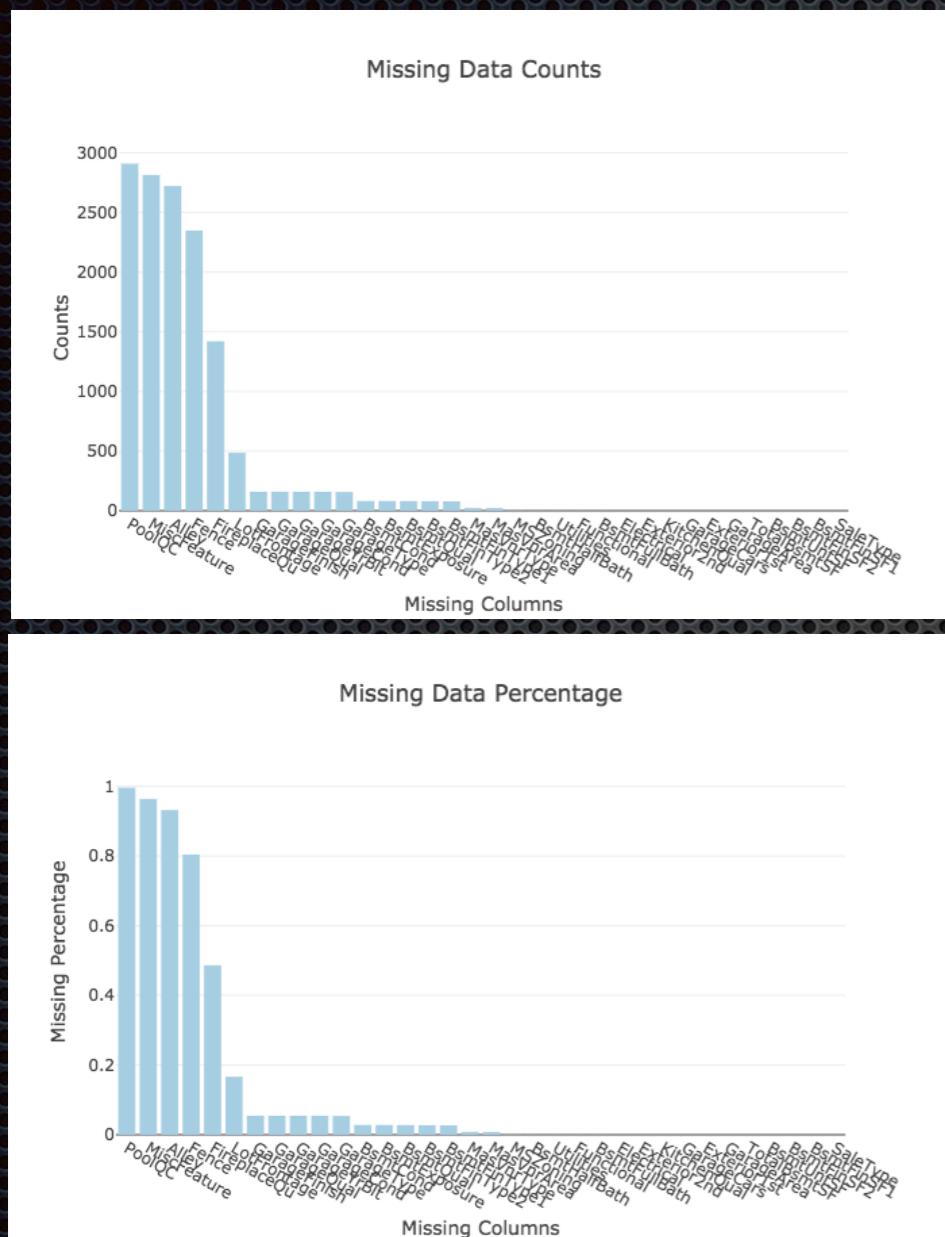
- $H = 2 - D_f = 0.536 \sim$ Brownian Motion
- $0 < H < 0.5$ - mean-reverting, anti-persistent
- $0.5 < H < 1$ - persistent
 \rightarrow increases tend to follow increases, same for decreases...

$$D_f = -\lim_{\epsilon \rightarrow 0} \frac{\ln(N(\epsilon))}{\ln(\epsilon)}$$

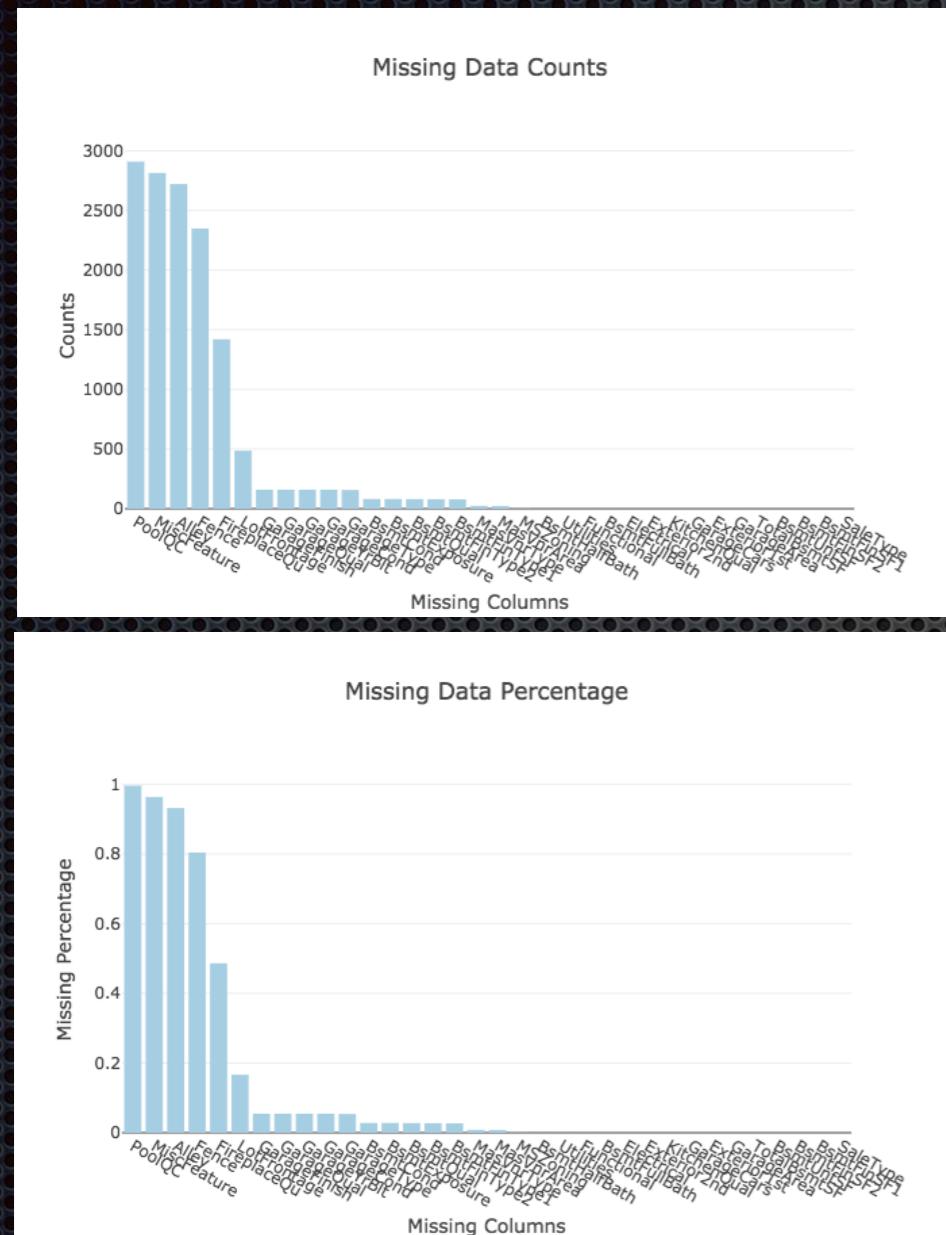


Good discussion [here](#) re: S&P 500

Data Cleaning, Feature Eng.

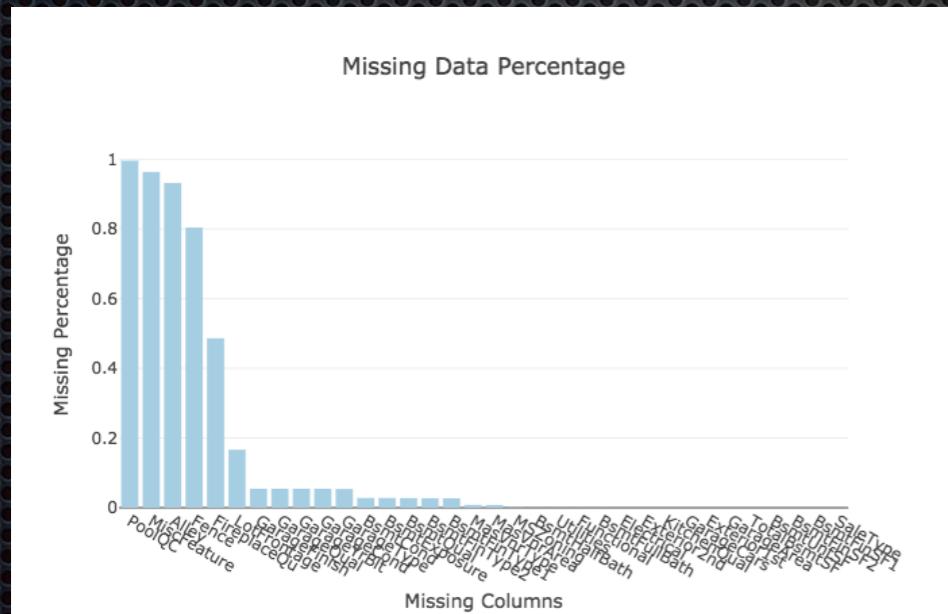


Data Cleaning, Feature Eng.



- Missingess
 - Imputation
 - Added features
 - Normalization

Data Cleaning, Feature Eng.



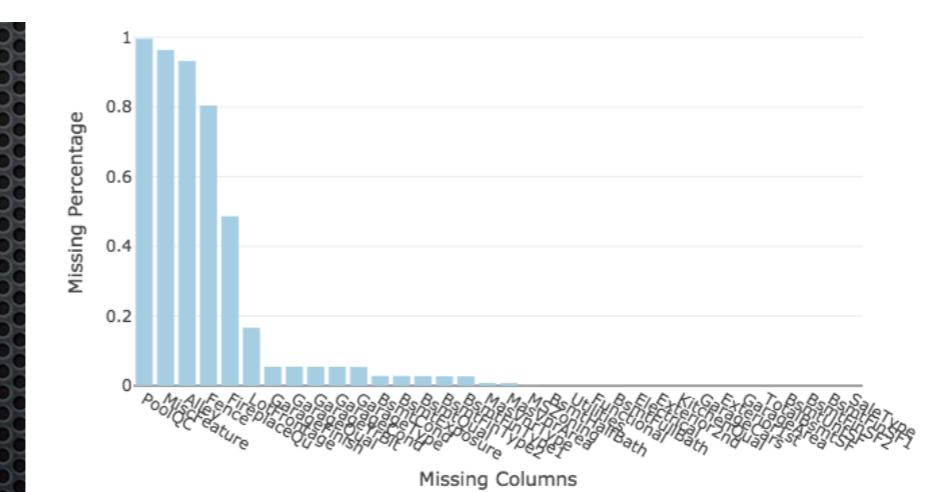
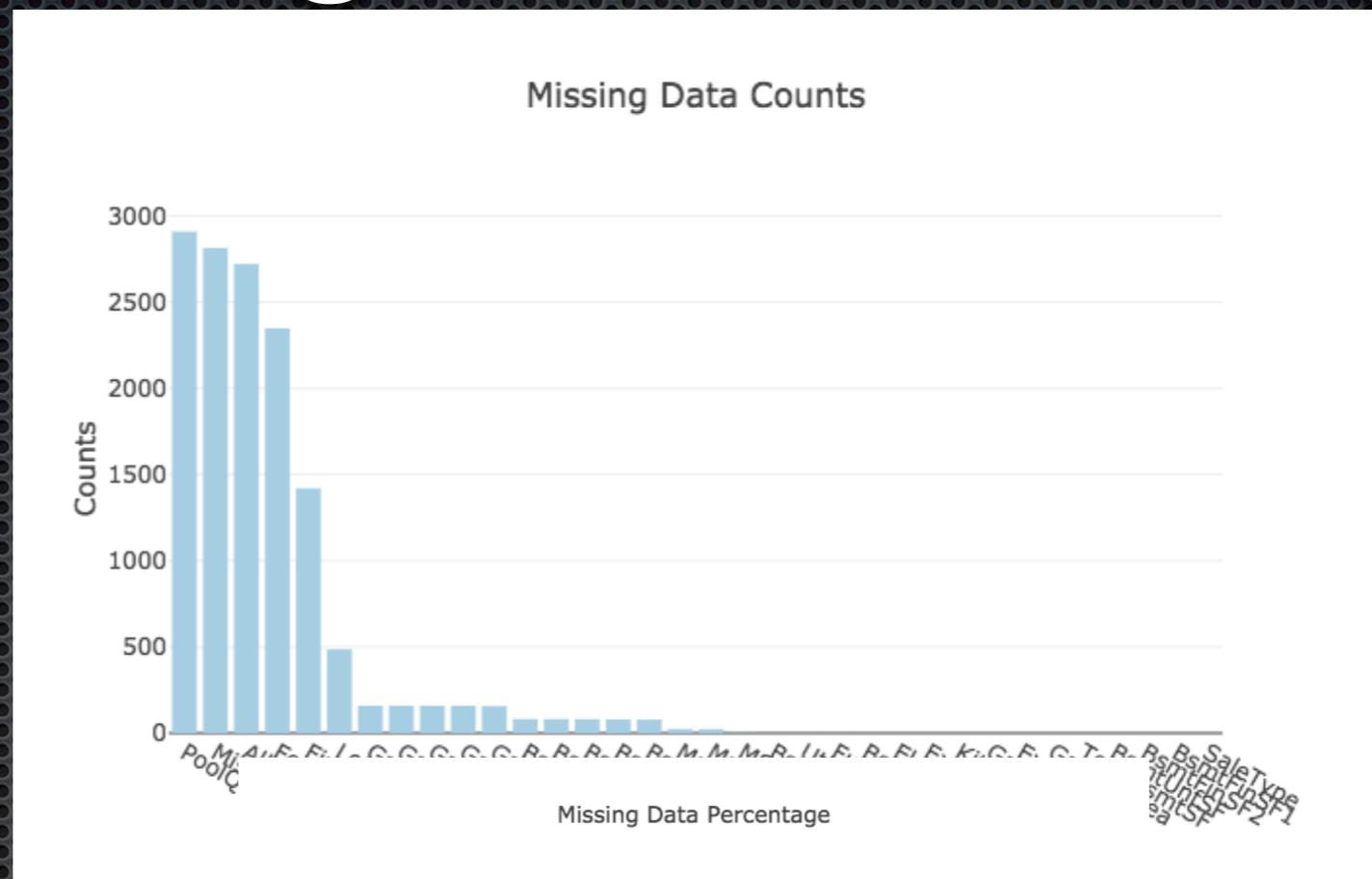
- Normalization

Data Cleaning, Feature Eng.

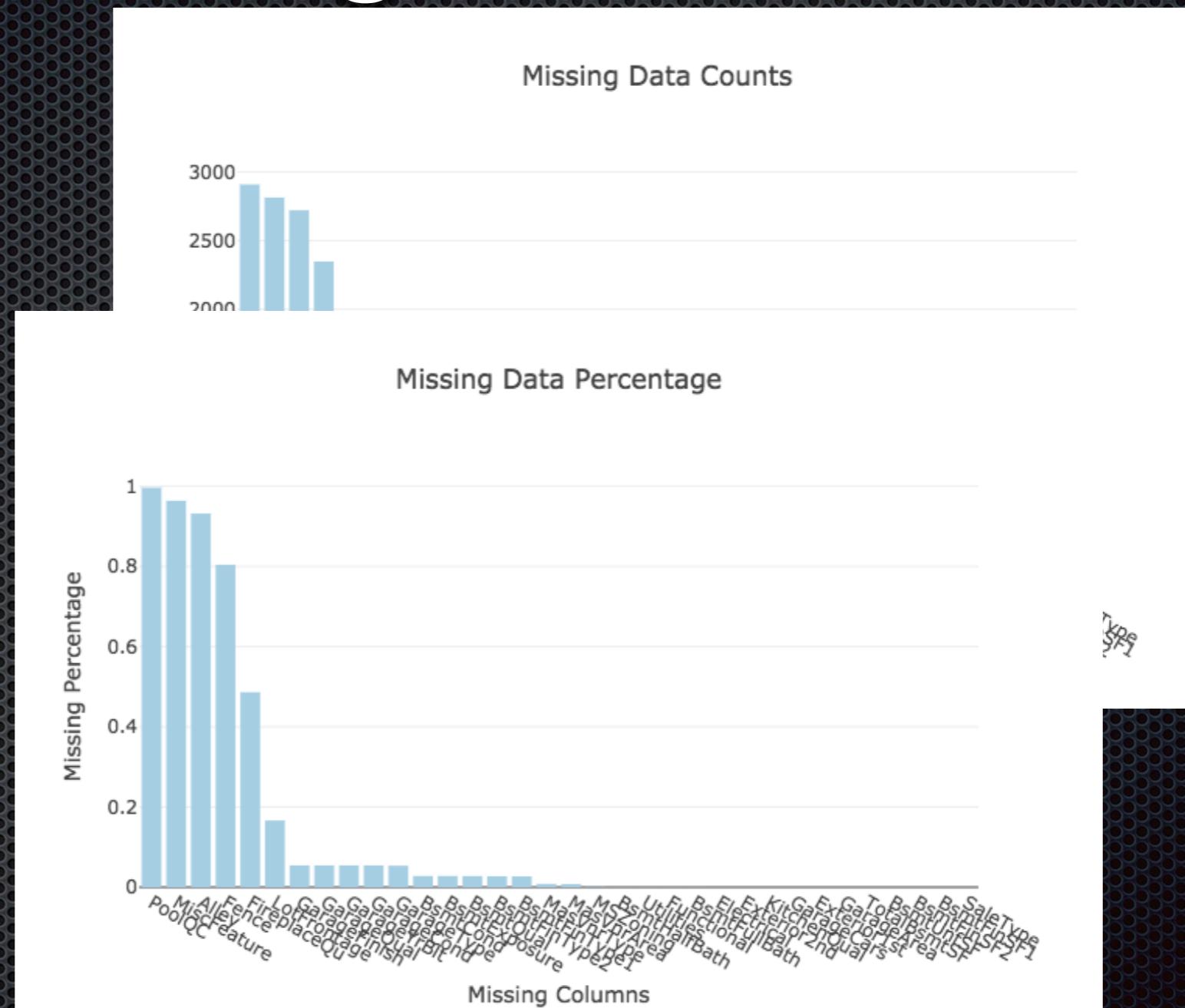


- Normalization

Data Cleaning, Feature Eng.



Data Cleaning, Feature Eng.

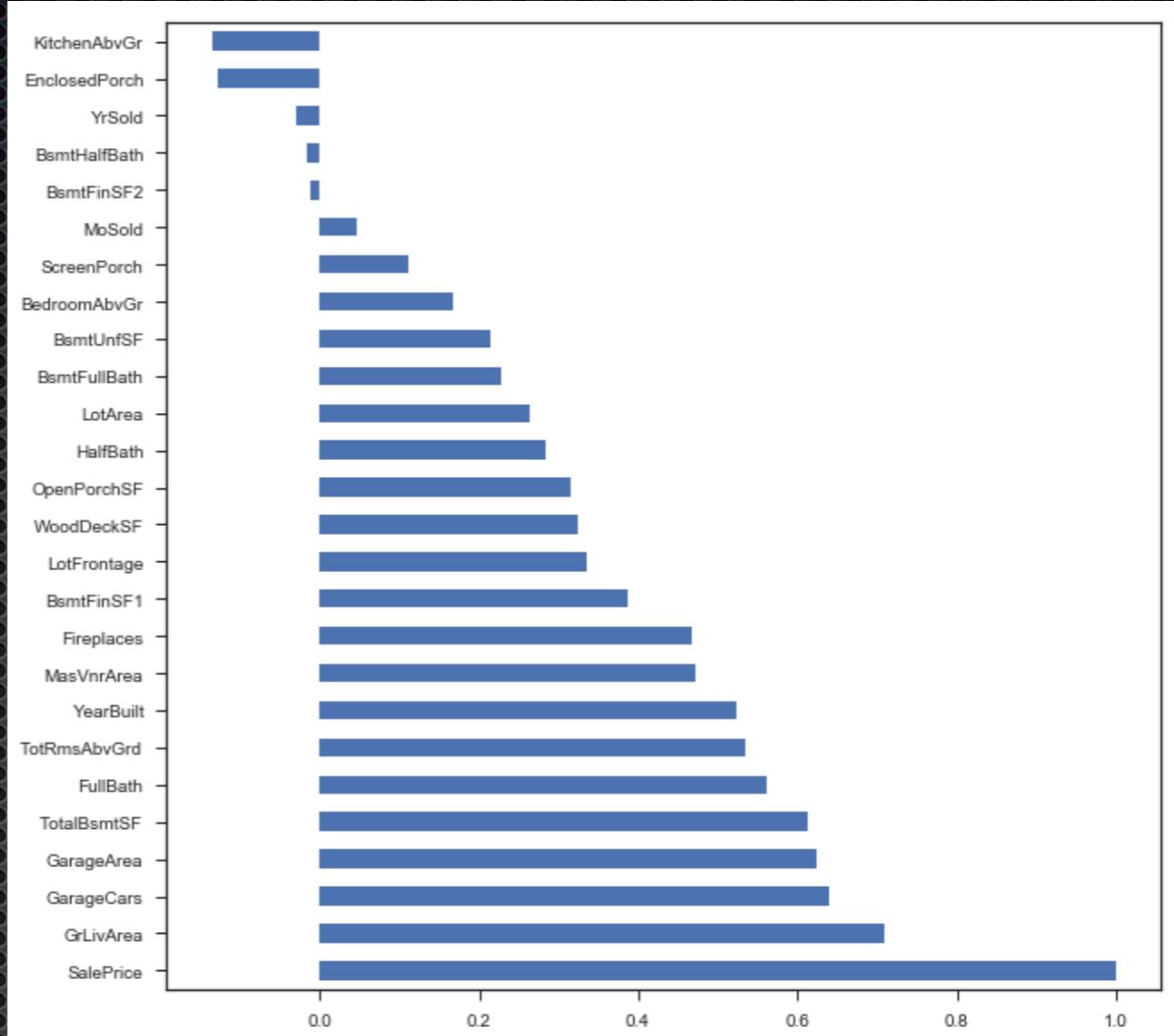
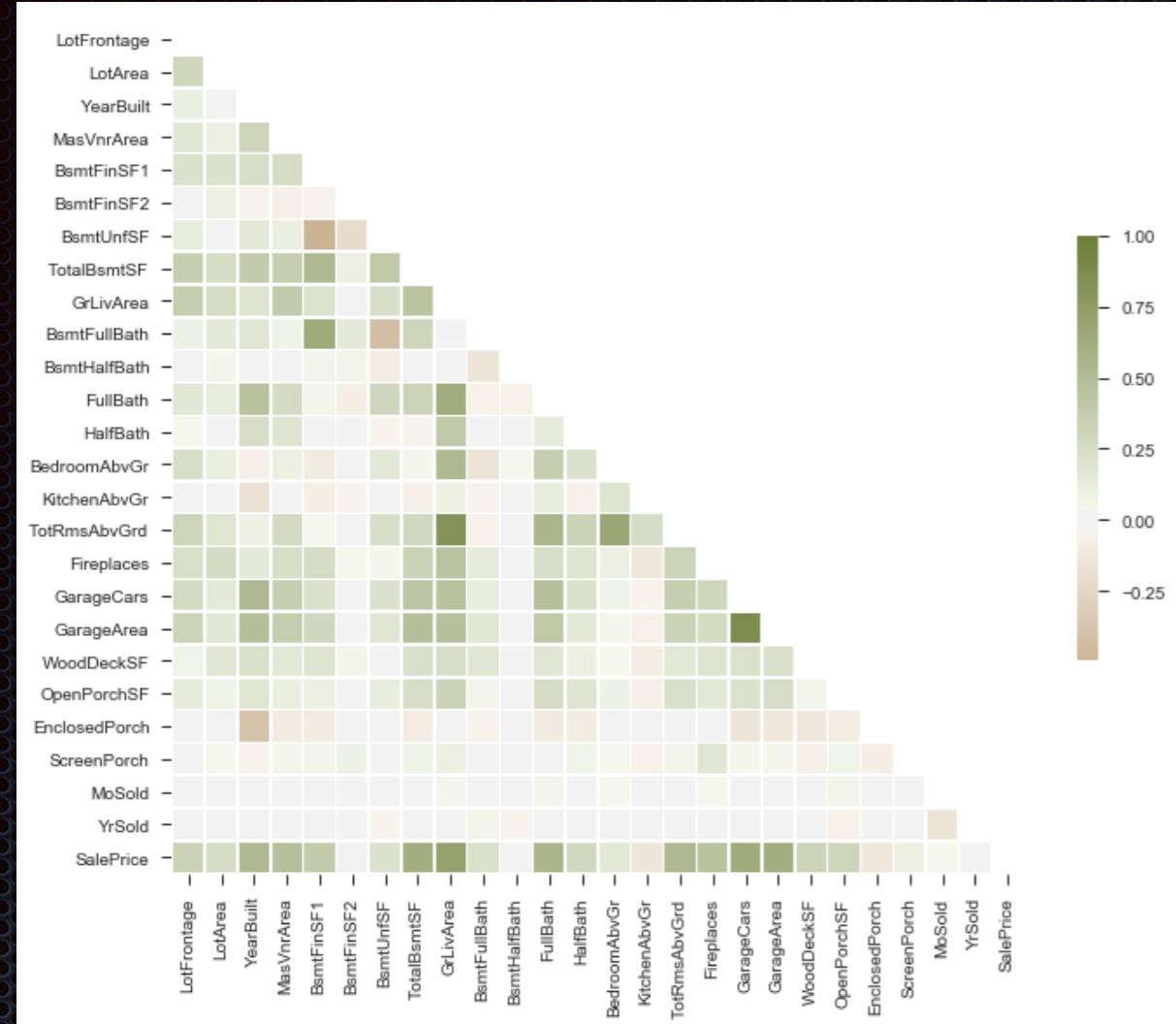


Data Cleaning, Feature Eng.



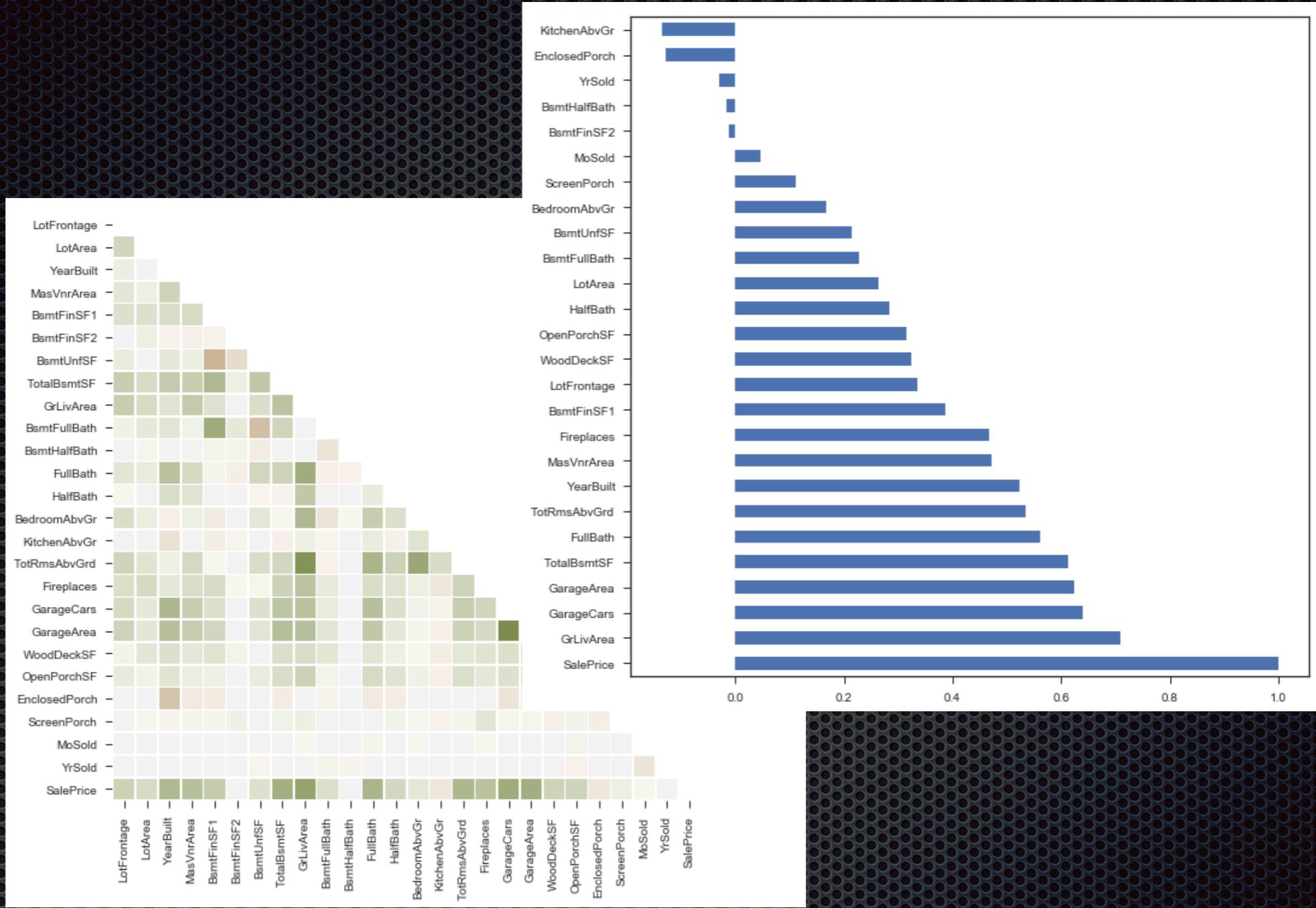


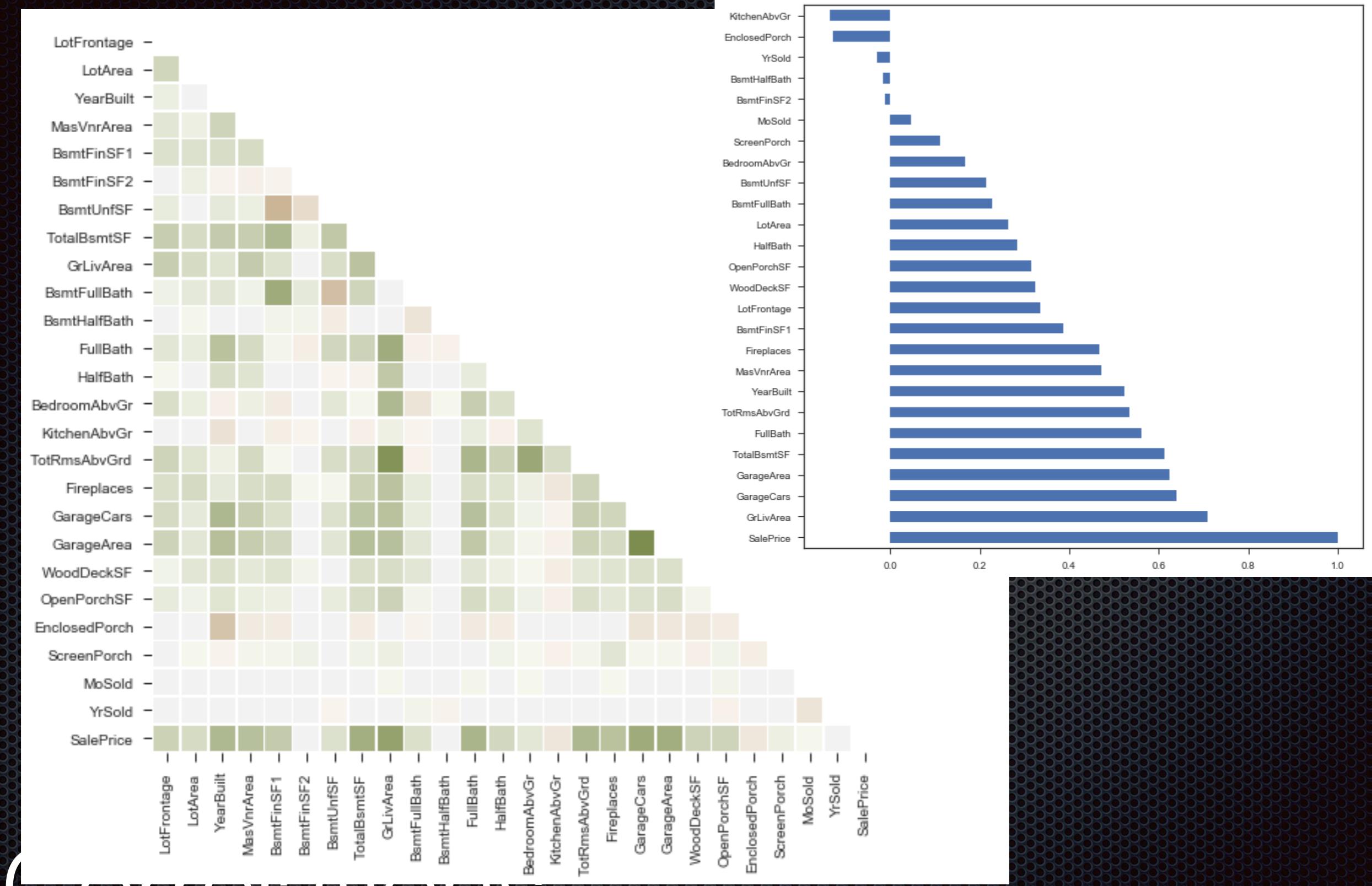
Correlated Missingness?

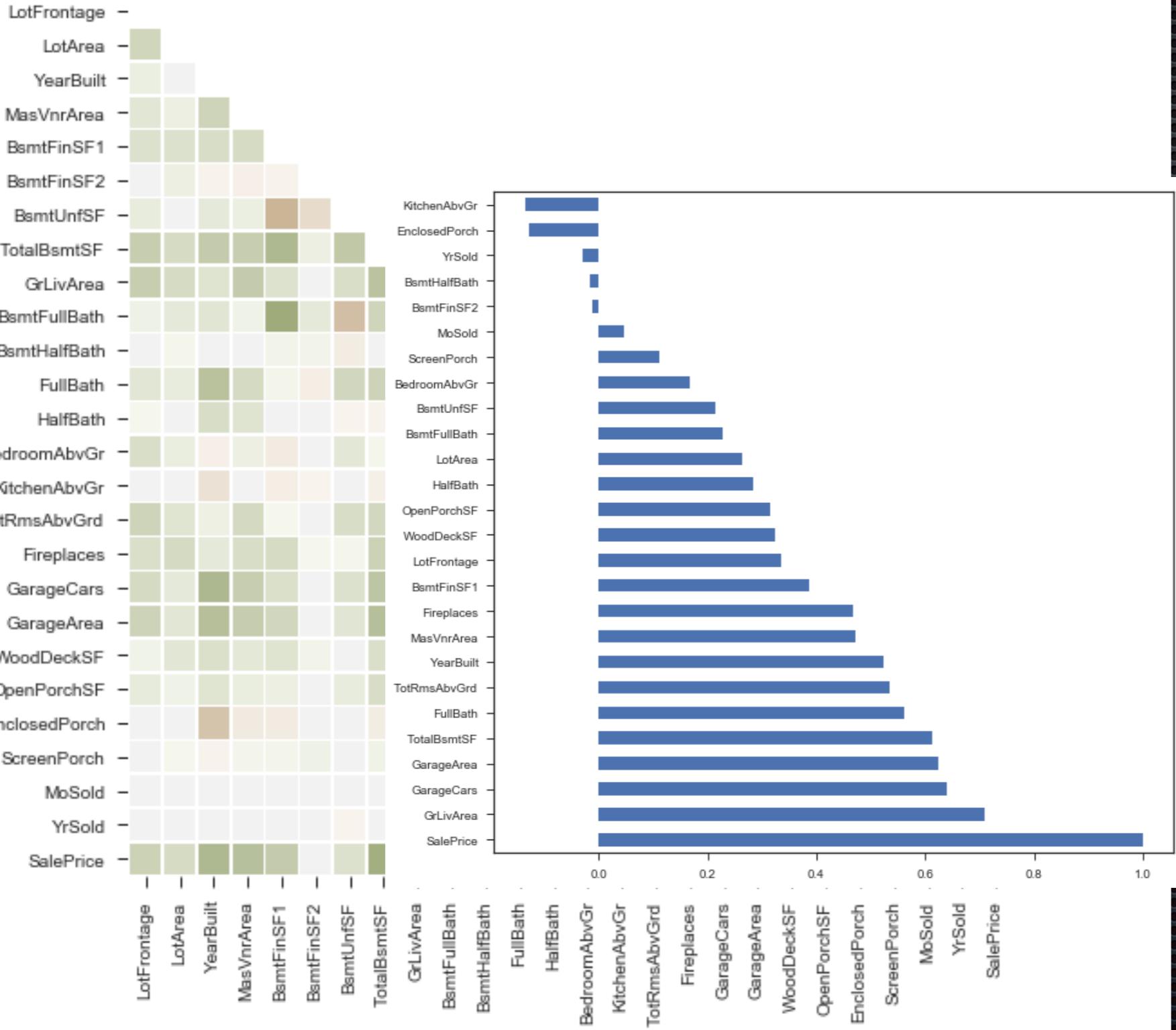


Correlations

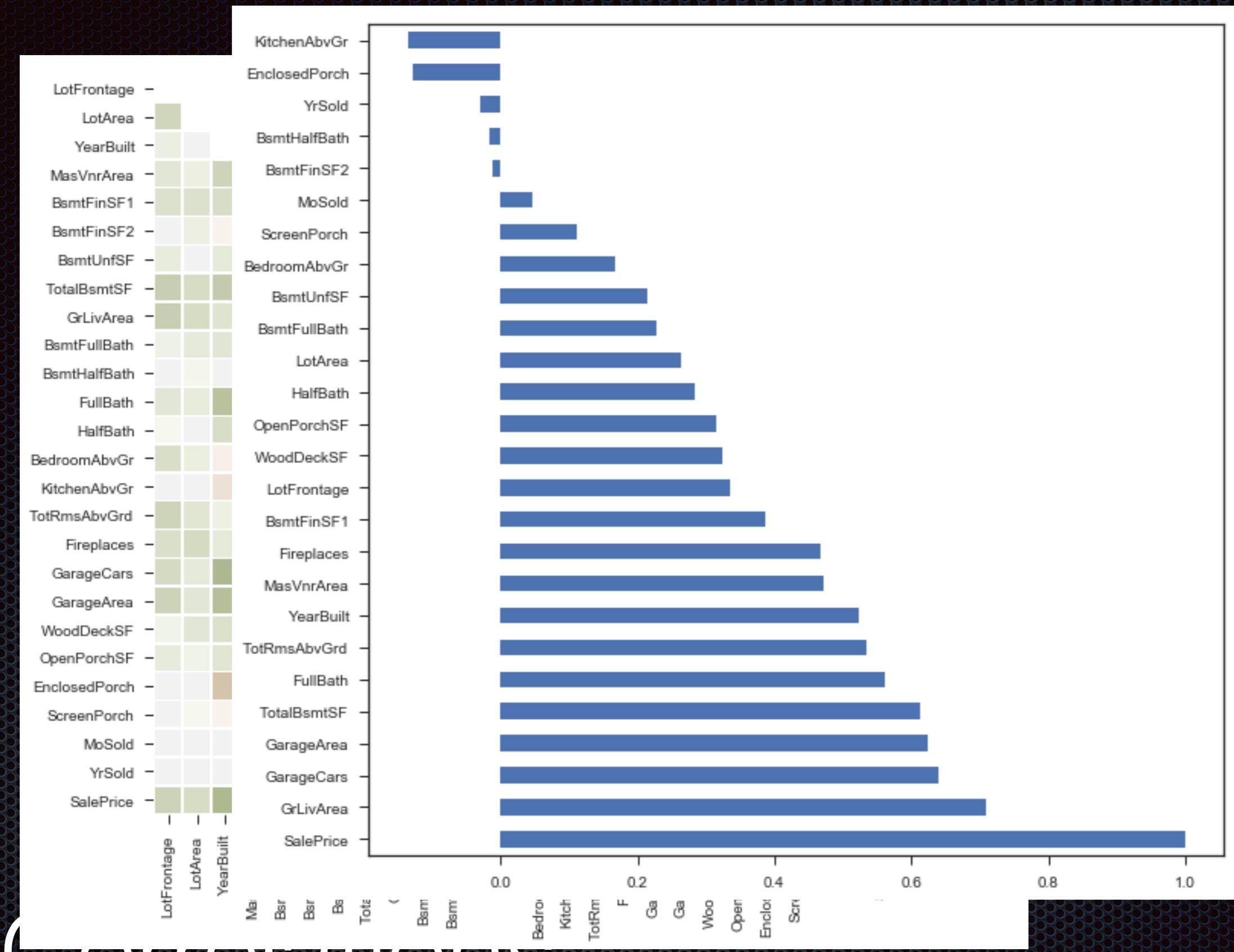
Correlations



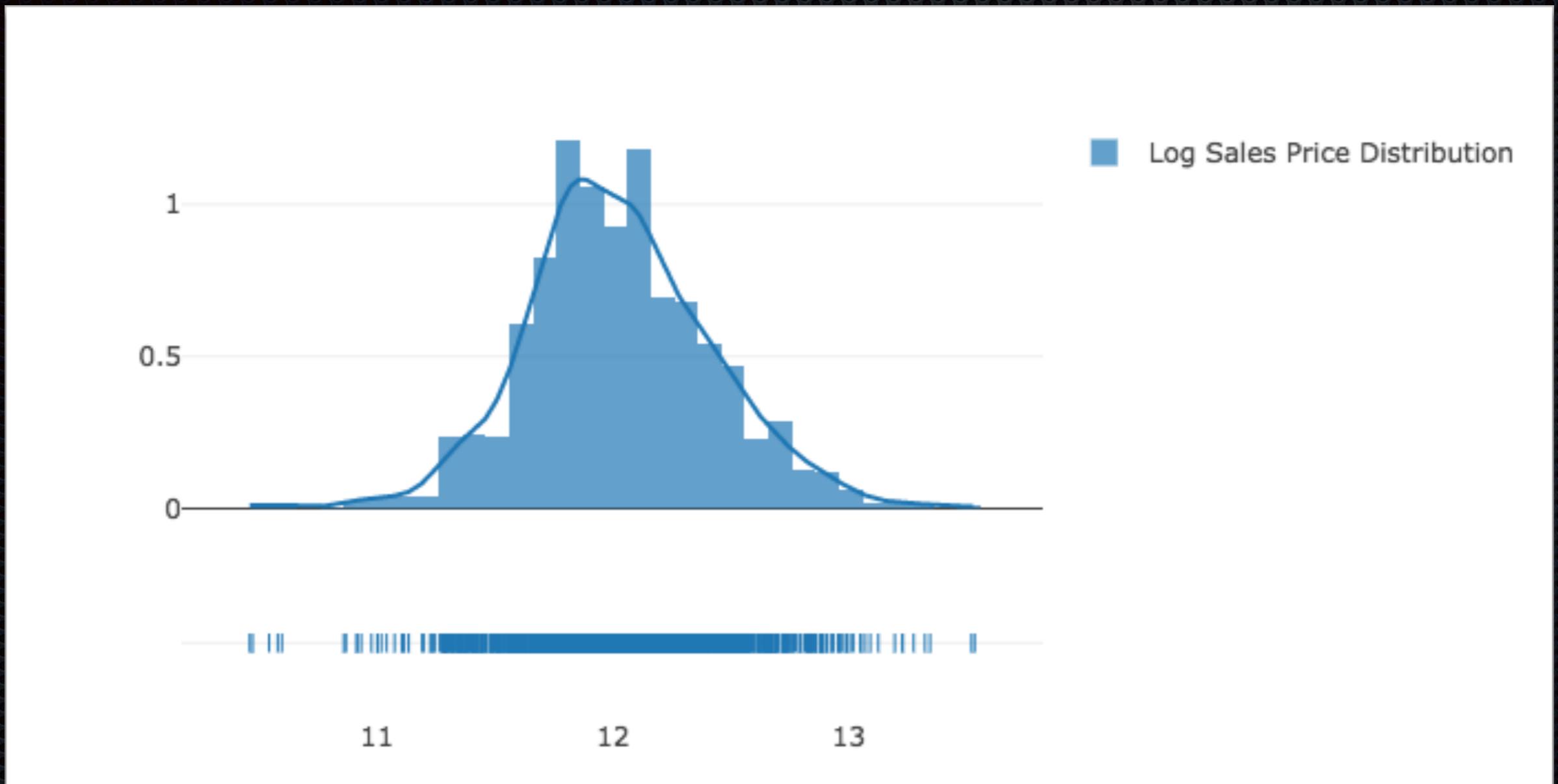




Correlations



Correlations



Normalizing Sales Price Distribution

Models Chosen

- Regularization
- Ridge
- Lasso
- PCA -
- Random Forest
- XGBoost

