

# Kaggle Project - Team Fat Tails

- [Kaggle Project - Team Fat Tails](#)
  - [Introduction](#)
  - [Data Synopsis](#)
  - [Analysis Question 1](#)
    - [Restatement of Problem](#)
    - [Build the Model](#)
      - [Interrogate the Data](#)
      - [Fit the Model](#)
      - [Check Assumptions](#)
      - [Residual Diagnostics](#)
      - [Outlier Analysis](#)
      - [Model Assumptions](#)
        - [Homogeneity of Variances](#)
  - [Analysis Question 2](#)
  - [Appendix A](#) \* [SAS Program](#) \* [main.sas](#) \* [dataimport.sas](#) \* [procmeans.sas](#)
  - [Appendix B](#)
  - [Appendix C](#)

## Introduction

Ask a home buyer to describe their dream house, and they probably won't begin with the height of the basement ceiling or the proximity to an east-west railroad. However, it is essential to review the data because it proves that there are many other influences in price negotiations than the number of bedrooms or a white-picket fence.

## Data Synopsis

The Ames House dataset was compiled by Dean De Cock and contains 79 explanatory variables describing almost every aspect of residual home in Ames Iowa from 2006 to 2010. The data set contains 2930 observations involved in assessing home values.

```
@import "../figs/procmeans.html"
```

## Analysis Question 1

### Restatement of Problem

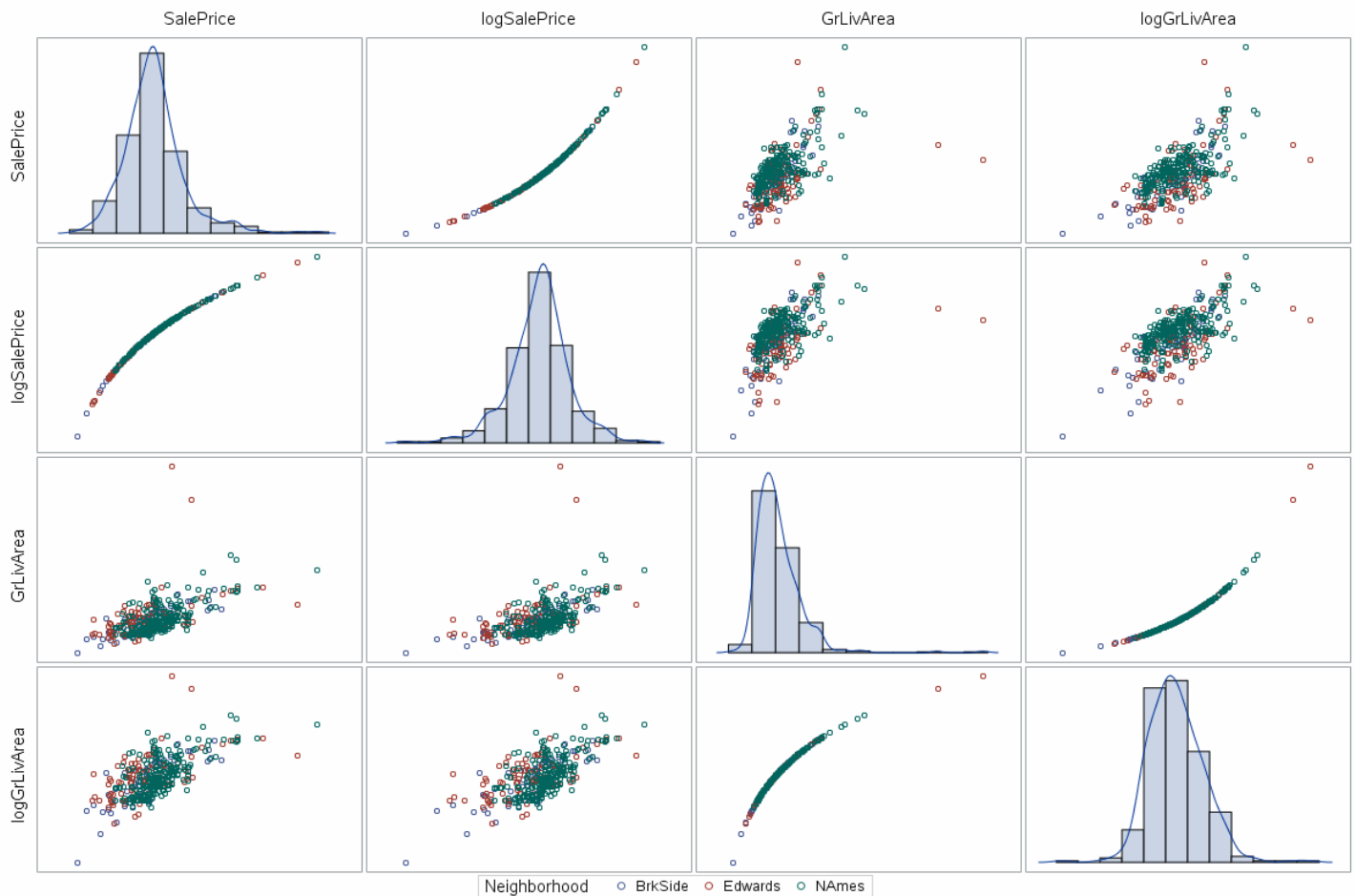
To build and fit a model, an analysis must be performed to identify features of the dataset that are statistically significant in their relation to, and prediction of, the sales price.

### Build the Model

#### Interrogate the Data

To build and fit a model, an analysis must be performed to identify features of the dataset that are statistically significant in their relation to, and prediction of, the sales price.

- Plot the data.
- Develop a tentative model(s).
  - Using the question(s) of interest (QOI).
  - Accounting for confounders.
  - Accounting for relationships ( $X^2$ ,  $X^3$ , etc).
- Fit the model(s).
- Evaluate residual plots.
  - Constant SD.
  - Normality and zero mean.
  - Identify any influential observations.

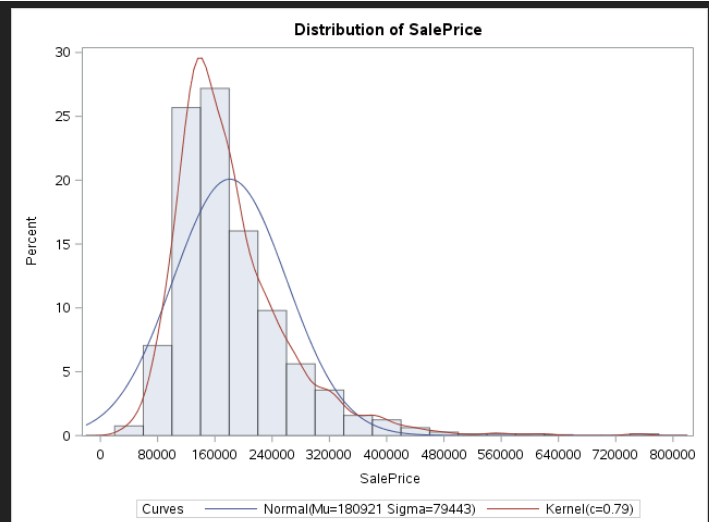
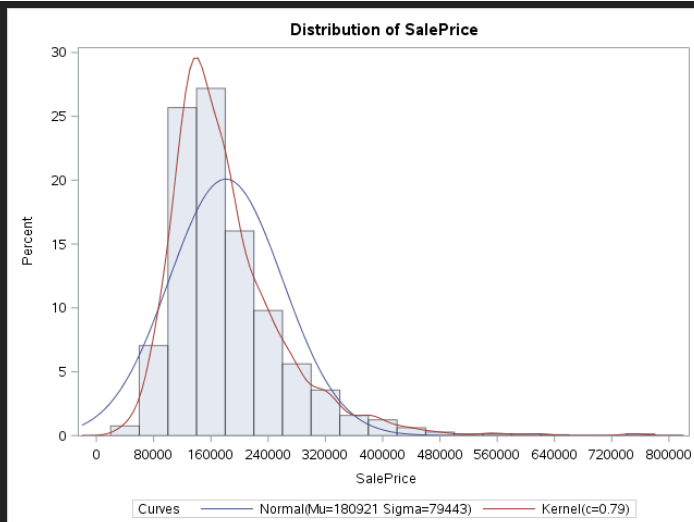


Fit the Model

Check Assumptions

Solarized dark

Solarized Ocean

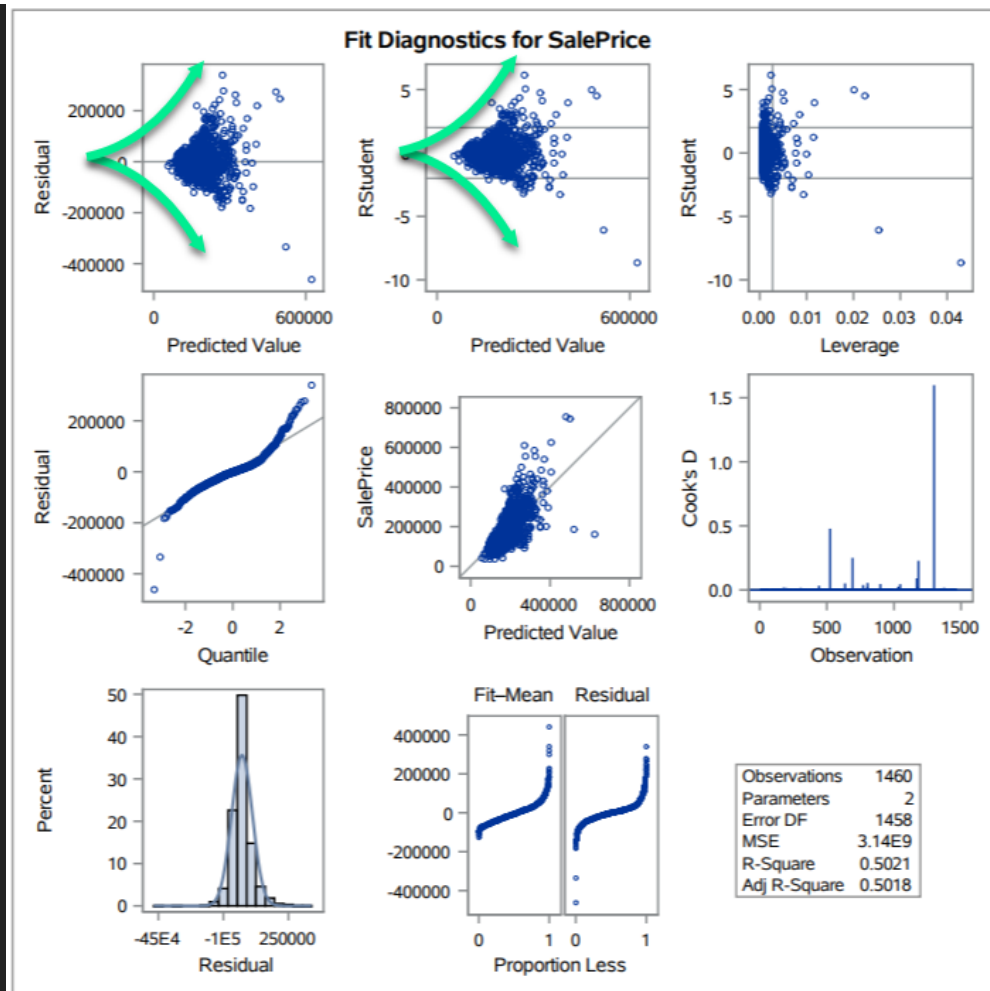


Residual Diagnostics

Outlier Analysis

Model Assumptions

Homogeneity of Variances



Code (will move to appendix later):

Neighborhoods: \$x\_1 = \text{BrkSide}\$ \$x\_2 = \text{NAmes}\$ \$x\_3 = \text{Edwards}\$

$$\hat{\mu}(\text{SalesPrice}) = \beta_0 + \beta_1 \cdot \text{BrkSide} + \beta_2 \cdot \text{Edwards} + \beta_3 \cdot \text{NAmes} + \beta_4(\text{LivingArea}_{\log}) \cdot \text{BrkSide} + \beta_5 \cdot x_1(\text{LivingArea}_{\log}) \cdot x_2(\text{Edwards})$$

$$\hat{\mu}(\text{SalesPrice}_{\text{Ames}}) = \beta_0 + \beta_1 \cdot x_1(\text{BrkSide}) + \beta_2 \cdot \text{Edwards} + \beta_3 \cdot \text{NAmes} + \beta_4(\text{LivingArea}_{\log}) \cdot x_1(\text{BrkSide}) + \beta_5 \cdot x_1(\text{LivingArea}_{\log}) \cdot x_2(\text{Edwards})$$

## Analysis Question 2

## Appendix A

### SAS Program

1. [main.sas](#)
2. [Example2](#)
3. [Third Example](#)

#### main.sas

```
@import "main.sas"
```

#### dataimport.sas

```
@import "dataimport.sas"
```

#### procmeans.sas

```
@import "procmeans.sas"
```

## Appendix B

```
@import "../data/data_description.md"
```

# Appendix C

[Downloading from the Kaggle API](#)

[Using Code Blocks in Markdown](#)

[Using SAS in Markdown Code Blocks](#)