

# FDS Fall 2017 – Kaggle Project

Kaggle Account: vamsikrishna90

URL: <https://www.kaggle.com/vamsikrishna90>

Vamsi Krishna Varma Gunturi / 1794653

Dastagiri Dudekula / 1826239

Pavan Kumar Alikana / 1826777

## 1st attempt:

- Linear regression of SalePrice vs GrLivArea

Model Selection:

- Linear Model

**Kaggle Score: 0.28918**

## 2nd attempt:

Data Tidying:

- Combining train and test dataset by removing fields like Id and Saleprice for data cleaning

- Replacing all the NA values with the mean for integer features and 0 for all the string features after converting all the string values in to integers using automated integer vectors based on number of levels for a particular feature

- Splitting the fullDataset into train and test again after data pre-processing

Model Selection:

- Linear Model with all features

**Kaggle Score: 0.16012**

## 3rd attempt:

Data Tidying: Same as above (2<sup>nd</sup> attempt)

Model Selection:

- XGBoost

**Kaggle Score: 0.22064**

## 4th attempt:

Data Tidying: Same as above (2<sup>nd</sup> attempt)

Model Selection:

- Random Forest

**Kaggle Score: 0.15444**

## 5th attempt:

Data Tidying: Same as above (2<sup>nd</sup> attempt)

Model Selection:

- Stacking both Multilinear regression and Random Forest model results and finding mean from them

**Kaggle Score: 0.13561**