



Project Design Report

PREDICTING SALES PRICES OF HOUSES BASED ON
VARIOUS PARAMETERS

Introduction

Housing prices are an important reflection of the economy, and housing price ranges are of great interest for both buyers and sellers. In this project, house prices will be predicted given explanatory variables that cover many aspects of residential houses

Dataset :

| S.No | Scales of Measurement | No. of variables |
|------|-----------------------|------------------|
| 1 | Nominal | 23 |
| 2 | Ordinal | 23 |
| 3 | Discrete | 14 |
| 4 | Continuous | 20 |

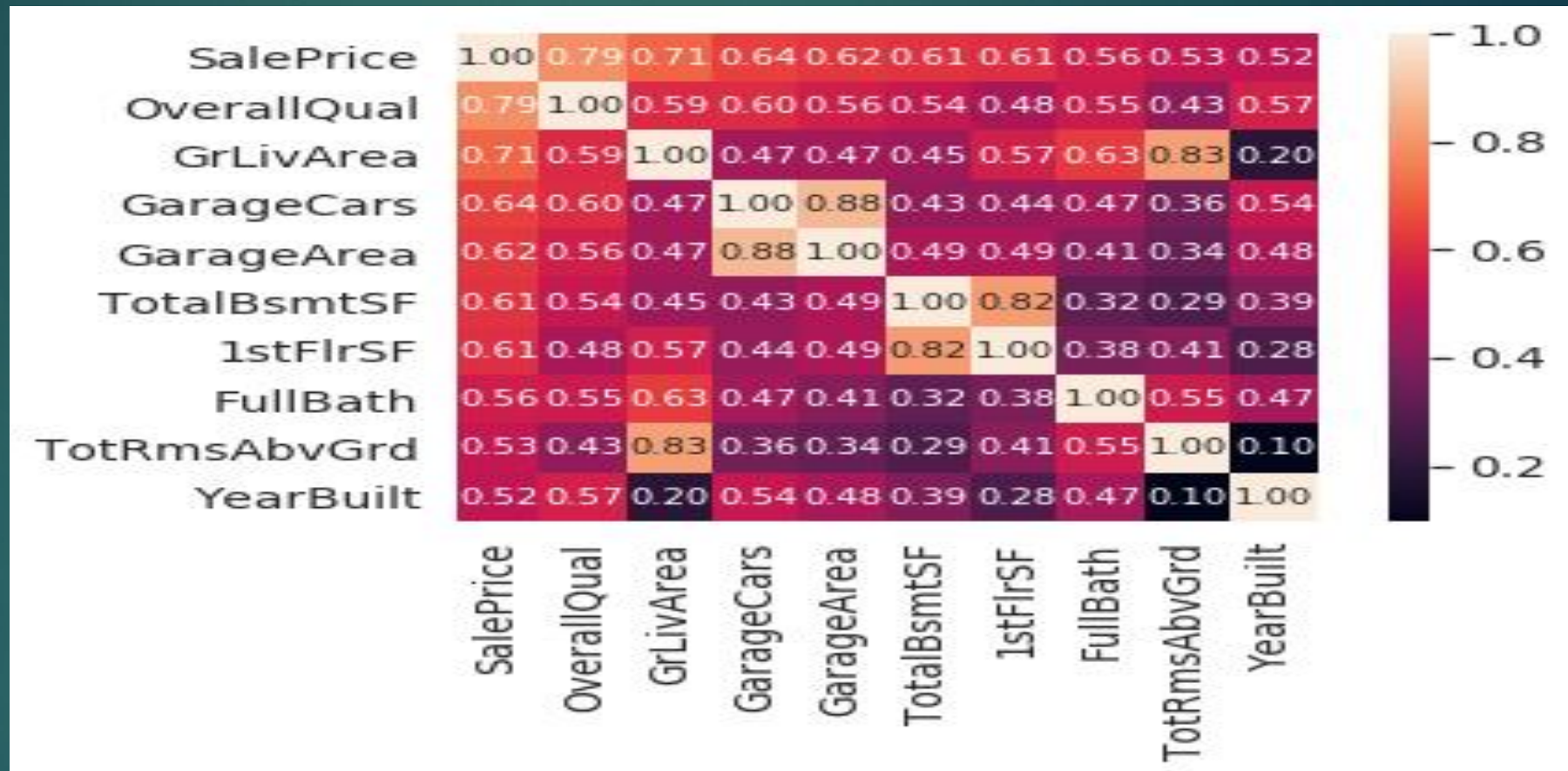
Data and Preprocessing

The dataset is the prices and features of residential houses sold from 2006 to 2010 in Ames, Iowa, obtained from the Ames Assessor's Office. This dataset consists of 79 house features and 1460 houses with sold prices.

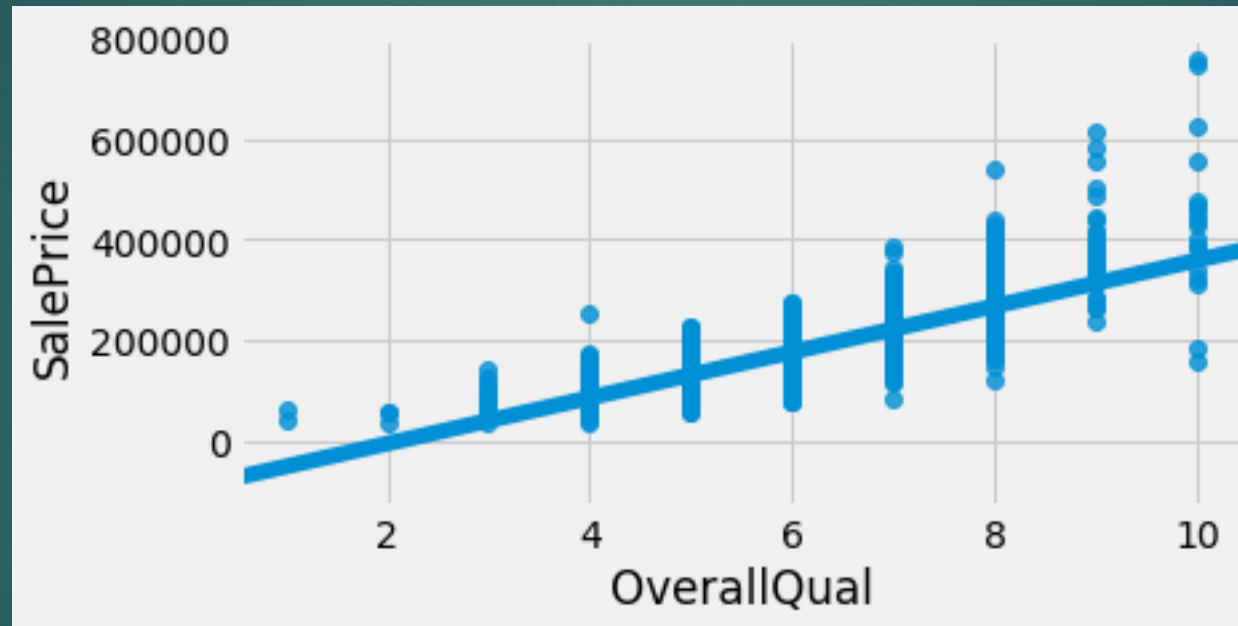
DATA CLEANING: DATA VISUALIZATION

| S.No | Parameters | Measurement Scale | S.No | Parameters | Measurement Scale | S.No | Parameters | Measurement Scale | S.No | Parameters | Measurement Scale |
|------|--------------|-------------------|------|--------------|-------------------|------|--------------|-------------------|------|---------------|-------------------|
| 1 | MSSubClass | Nominal | 21 | RoofStyle | Nominal | 41 | CentralAir | Nominal | 61 | GarageCars | Discrete |
| 2 | MSZoning | Nominal | 22 | RoofMatl | Nominal | 42 | Electrical | Ordinal | 62 | GarageArea | Continuous |
| 3 | LotFrontage | Continuous | 23 | Exterior1st | Nominal | 43 | 1stFlrSF | Continuous | 63 | GarageQual | Ordinal |
| 4 | LotArea | Continuous | 24 | Exterior2nd | Nominal | 44 | 2ndFlrSF | Continuous | 64 | GarageCond | Ordinal |
| 5 | Street | Nominal | 25 | MasVnrType | Nominal | 45 | LowQualFinSF | Continuous | 65 | PavedDrive | Ordinal |
| 6 | Alley | Nominal | 26 | MasVnrArea | Continuous | 46 | GrLivArea | Continuous | 66 | WoodDeckSF | Continuous |
| 7 | LotShape | Ordinal | 27 | ExterQual | Ordinal | 47 | BsmtFullBath | Discrete | 67 | OpenPorchSF | Continuous |
| 8 | LandContour | Nominal | 28 | ExterCond | Ordinal | 48 | BsmtHalfBath | Discrete | 68 | EnclosedPorch | Continuous |
| 9 | Utilities | Ordinal | 29 | Foundation | Nominal | 49 | FullBath | Discrete | 69 | 3SsnPorch | Continuous |
| 10 | LotConfig | Nominal | 30 | BsmtQual | Ordinal | 50 | HalfBath | Discrete | 70 | ScreenPorch | Continuous |
| 11 | LandSlope | Ordinal | 31 | BsmtCond | Ordinal | 51 | BedroomAbvGr | Discrete | 71 | PoolArea | Continuous |
| 12 | Neighborhood | Nominal | 32 | BsmtExposure | Ordinal | 52 | KitchenAbvGr | Discrete | 72 | PoolQC | Ordinal |
| 13 | Condition1 | Nominal | 33 | BsmtFinType1 | Ordinal | 53 | KitchenQual | Ordinal | 73 | Fence | Ordinal |
| 14 | Condition2 | Nominal | 34 | BsmtFinSF1 | Continuous | 54 | TotRmsAbvGrd | Discrete | 74 | MiscFeature | Nominal |
| 15 | BldgType | Nominal | 35 | BsmtFinType2 | Ordinal | 55 | Functional | Ordinal | 75 | MiscVal | Continuous |
| 16 | HouseStyle | Nominal | 36 | BsmtFinSF2 | Continuous | 56 | Fireplaces | Discrete | 76 | MoSold | Discrete |
| 17 | OverallQual | Ordinal | 37 | BsmtUnfSF | Continuous | 57 | FireplaceQu | Ordinal | 77 | YrSold | Discrete |
| 18 | OverallCond | Ordinal | 38 | TotalBsmtSF | Continuous | 58 | GarageType | Nominal | 78 | SaleType | Nominal |
| 19 | YearBuilt | Discrete | 39 | Heating | Nominal | 59 | GarageYrBlt | Discrete | 79 | SaleCondition | Nominal |
| 20 | YearRemodAdd | Discrete | 40 | HeatingQC | Ordinal | 60 | GarageFinish | Ordinal | 80 | SalePrice | Continuous |

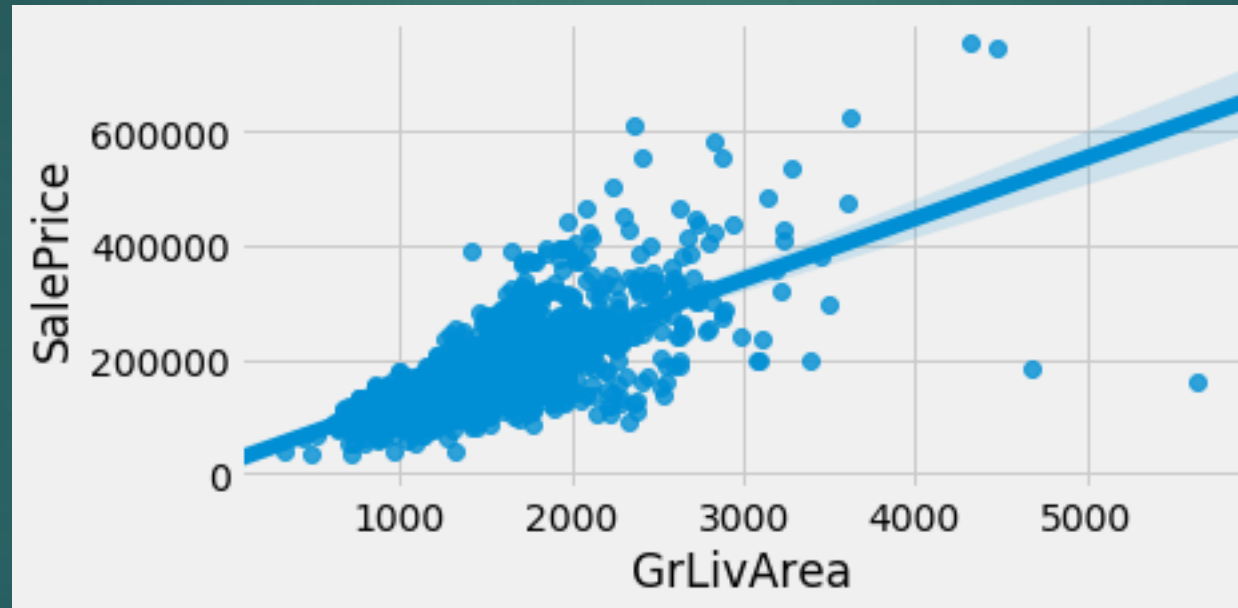
Co-Relation between Metric Data



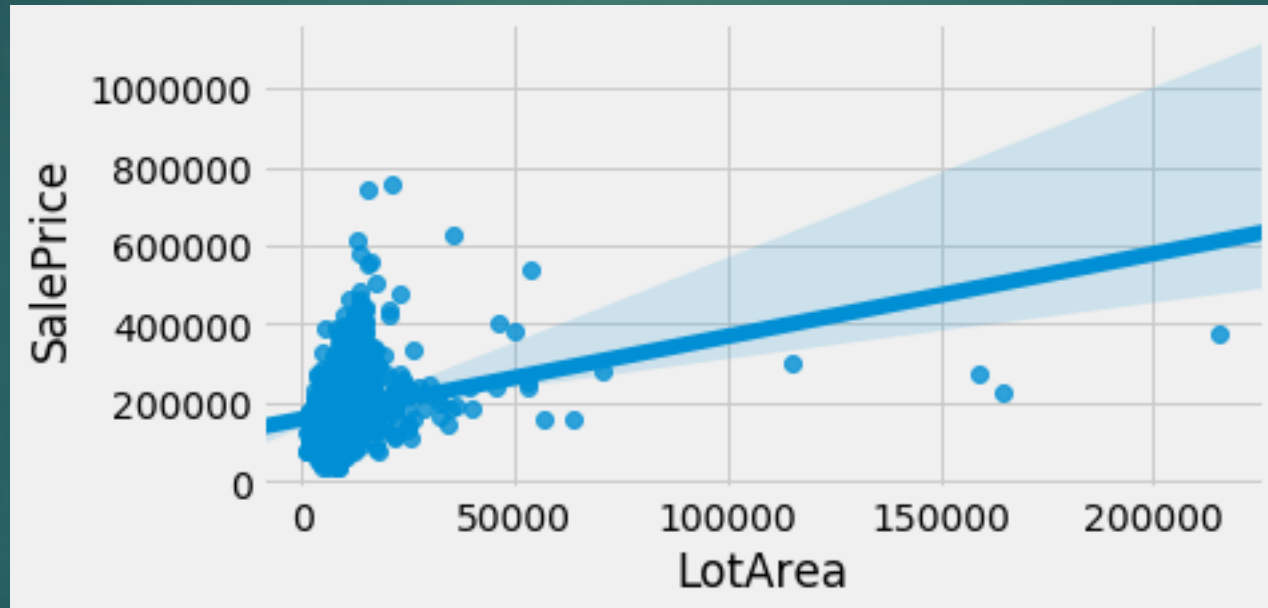
Sale Price against Overall Quality



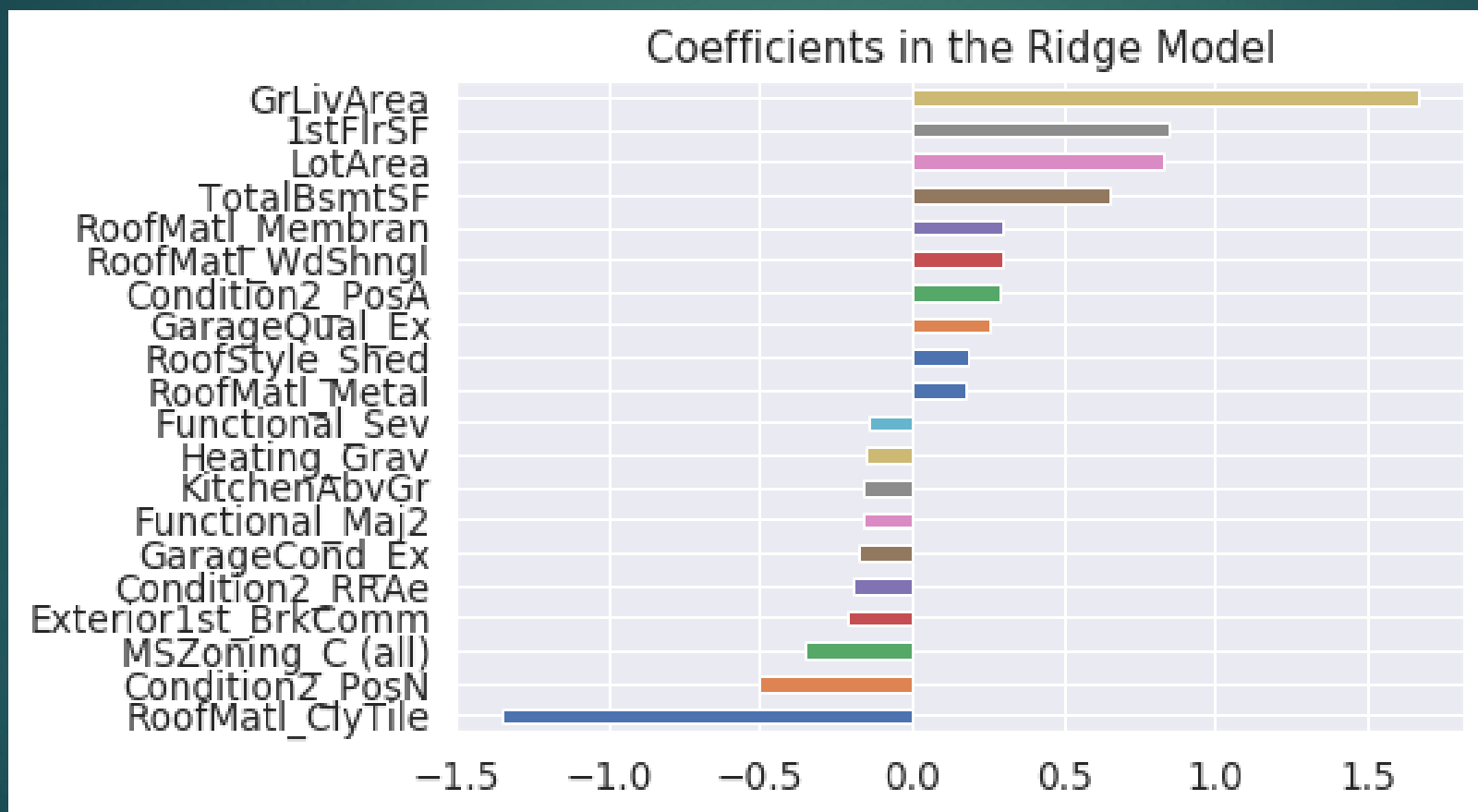
Sale Price against Ground Living Area



Sale Price against Lot Area



Results of Ridge Regression



REGRESSION ANALYSIS:

- ❖ ORDINARY LEAST SQUARES

- ❖ ADVANCED REGRESSION ANALYSIS

FUTURE SCOPE

As continuous house prices, they will be predicted with various regression techniques including Lasso, Ridge, SVM regression, and Random Forest regression; as individual price ranges, they will be predicted with classification methods including Naive Bayes, logistic regression, SVM classification, and Random Forest classification.

Thank you

