Michael Rocchio

06/24/2023

Assignment #1

**Data Quality**

To begin our assessment of this dataset for modeling potential, let us first describe the data set.

Dataset shape:

2930 rows

87 columns

Rough Schema:

A screenshot of a computer

Description automatically generated with low confidence

As evidenced by the third column above there are several fields that have null records. A further visual can be shown here with the null values being shown in white:

A picture containing screenshot, text

Description automatically generated

As most models cannot handle null field inputs we need to handle each of the following columns:

Column: LotFrontage Null Count: 490

Column: Alley Null Count: 2732

Column: MasVnrType Null Count: 23

Column: MasVnrArea Null Count: 23

Column: BsmtQual Null Count: 80

Column: BsmtCond Null Count: 80

Column: BsmtExposure Null Count: 83

Column: BsmtFinType1 Null Count: 80

Column: BsmtFinSF1 Null Count: 1

Column: BsmtFinType2 Null Count: 81

Column: BsmtFinSF2 Null Count: 1

Column: BsmtUnfSF Null Count: 1

Column: TotalBsmtSF Null Count: 1

Column: Electrical Null Count: 1

Column: BsmtFullBath Null Count: 2

Column: BsmtHalfBath Null Count: 2

Column: FireplaceQu Null Count: 1422

Column: GarageType Null Count: 157

Column: GarageYrBlt Null Count: 159

Column: GarageFinish Null Count: 159

Column: GarageCars Null Count: 1

Column: GarageArea Null Count: 1

Column: GarageQual Null Count: 159

Column: GarageCond Null Count: 159

Column: PoolQC Null Count: 2917

Column: Fence Null Count: 2358

Column: MiscFeature Null Count: 2824

**Column LotFrontage:**

Definition: Linear feet of street connected to property.

This is effectively the size of the driveway.