**Days On Market ML Model:**

**Description of Preliminary Data Preprocessing:**

To preprocess data, it was important to look at all data to find what needs to be cleaned and used. Looking at all tables, it was found that the prop\_charac, pub\_rec, and sales\_data tables. I used the .info() method to find the data types and non-null counts for all columns in each table. Using the majority of the columns without any null values, I was able to then determine which features to include. After choosing features, I cleaned the specific features. Any objects were changed into categories, and then each category point was changed into a number, since all data points need to be numeric.

**Description of Preliminary feature engineering and preliminary feature selection, including their decision-making process:**

The features were determined using both market knowledge and non-null values in columns. After running the machine learning model, I was able to use feature\_importances\_ from the BalancedRandomForestClassifier module to see which features were mostly impacting the model results. This provided me with more information to drop unnecessary rows as well as any rows with repeated information.

**Description of how data was split into training and testing sets:**

The data was split into training and testing sets using train\_test\_split(X, y). This splits the data into two sections: 80% fit to the machine learning model and 20% used to test the machine learning model.

**Explanation of model choice, including limitations and benefits:**

The best model at this point in the analysis is the Balanced Random Forest Classifier. A classifier model is the type of machine learning model we are using because we are trying to split the data into one of four groups, depending on the range of predicted days on market. This model was able to obtain an accuracy score of 33.8% after dropping some columns and only including 13 columns.