



Predicting Tax Assessed Property Value

Presented by
Anisha Joshi
Gabriella Sussman



Problem Setting

- Property taxes are an important part of a local government's revenue.
- It helps pay for things like public schools, libraries, local government employees' salaries, parks and recreation, sanitation, sewer, police and fire protection, roads, and other local needs
- It is crucial that local governments be able to increase the accuracy and fairness of assessing properties, while also increasing the efficiency in assessing properties' value.
- By utilizing Machine Learning Models, local government employees can enhance their judgement by comparing the properties' assessment with the assessed value provided by the Models, allowing them to spend more time and resources on other value-added activities.
- The overall result is a more objective, timely, accurate, and equitable property valuation, providing more confidence to both the local government and the taxpayers.



Dataset



3 California
counties

From
Kaggle



2.9M
records
with 58
feature
columns

LA, Orange &
Ventura

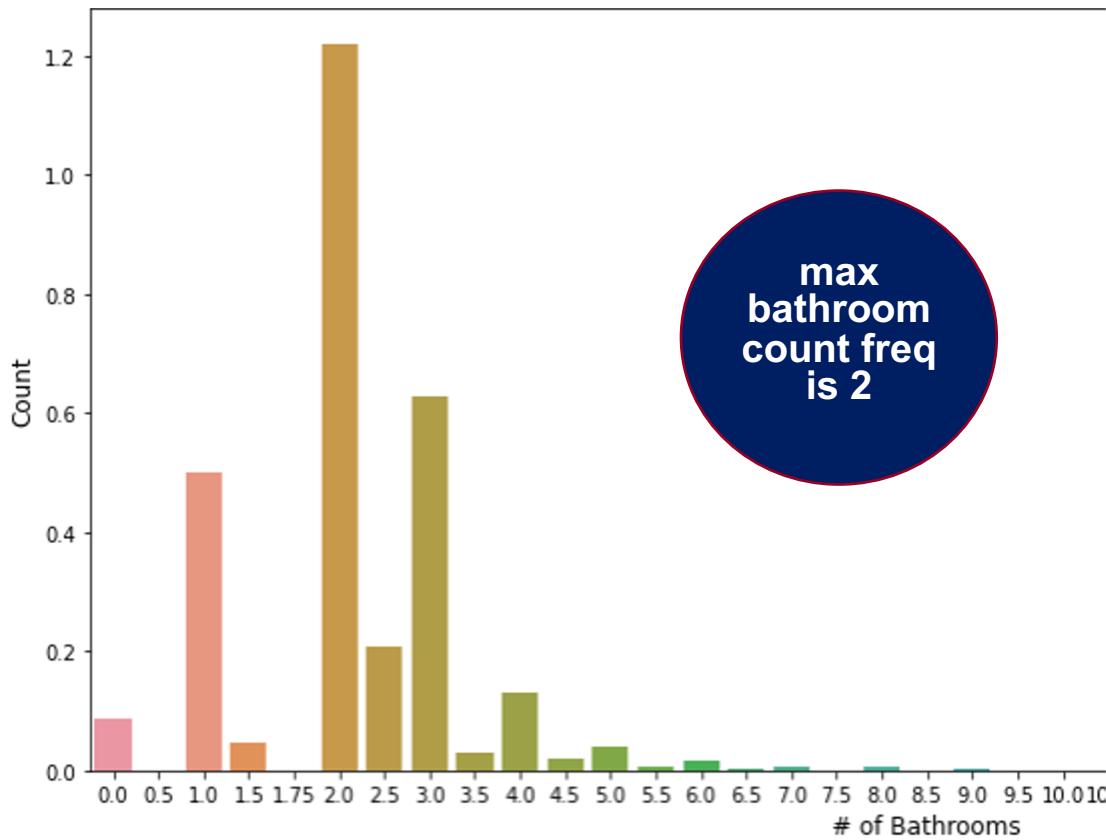
Dependent Variable

“taxvaluedollarcnt”
(The total tax
assessed value of
the parcel)

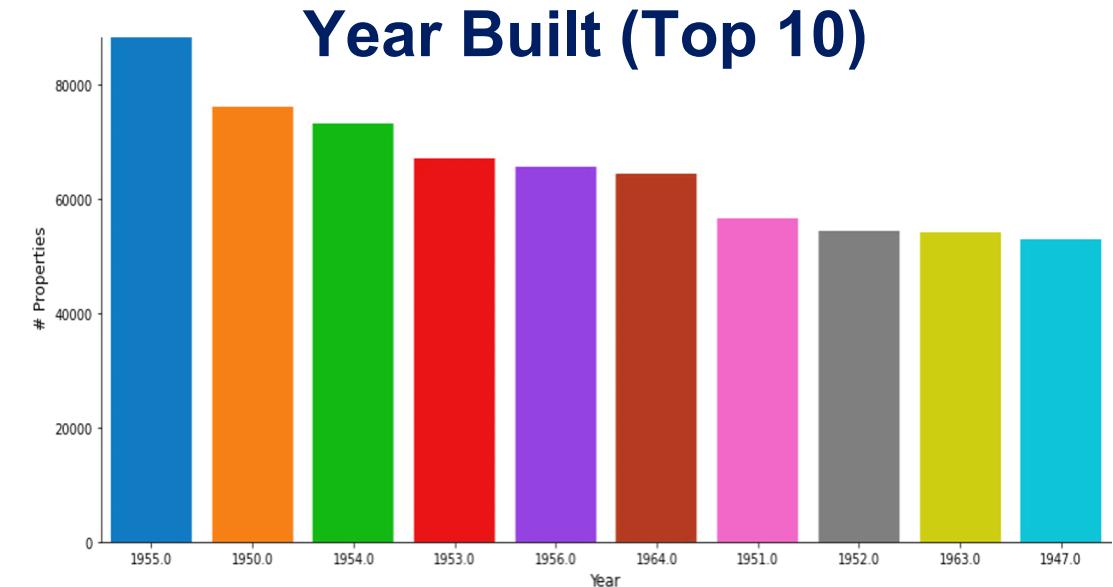
Independent Variables

```
-- airconditioningtypeid: integer (nullable = true)
-- architecturalstyletypeid: integer (nullable = true)
-- basementsqft: integer (nullable = true)
-- bathroomcnt: double (nullable = true)
-- bedroomcnt: double (nullable = true)
-- buildingclass typeid: integer (nullable = true)
-- buildingqualitytypeid: integer (nullable = true)
-- calculatedbathnbr: double (nullable = true)
-- decktypeid: integer (nullable = true)
-- finishedfloor1squarefeet: integer (nullable = true)
-- calculatedfinishedsquarefeet: double (nullable = true)
-- finishedsquarefeet12: integer (nullable = true)
-- finishedsquarefeet13: integer (nullable = true)
-- finishedsquarefeet15: integer (nullable = true)
-- finishedsquarefeet50: integer (nullable = true)
-- finishedsquarefeet6: integer (nullable = true)
-- fips: integer (nullable = true)
-- fireplacecnt: integer (nullable = true)
-- fullbathcnt: integer (nullable = true)
-- garagecarcnt: integer (nullable = true)
-- garagetotalsqft: integer (nullable = true)
-- hashottuborspa: boolean (nullable = true)
-- heatingorsystemtypeid: integer (nullable = true)
-- latitude: integer (nullable = true)
```

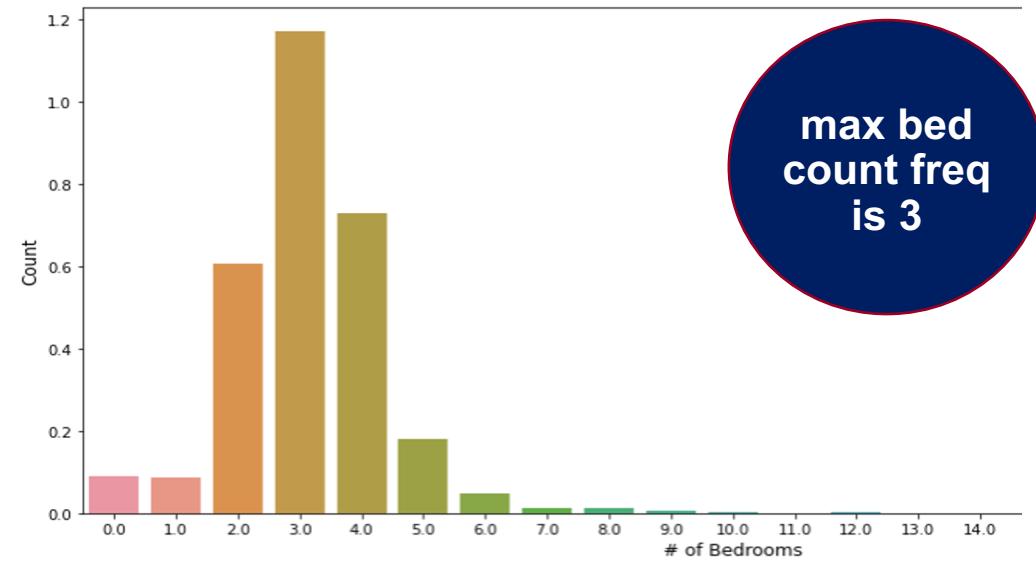
Data Exploration



max
bathroom
count freq
is 2

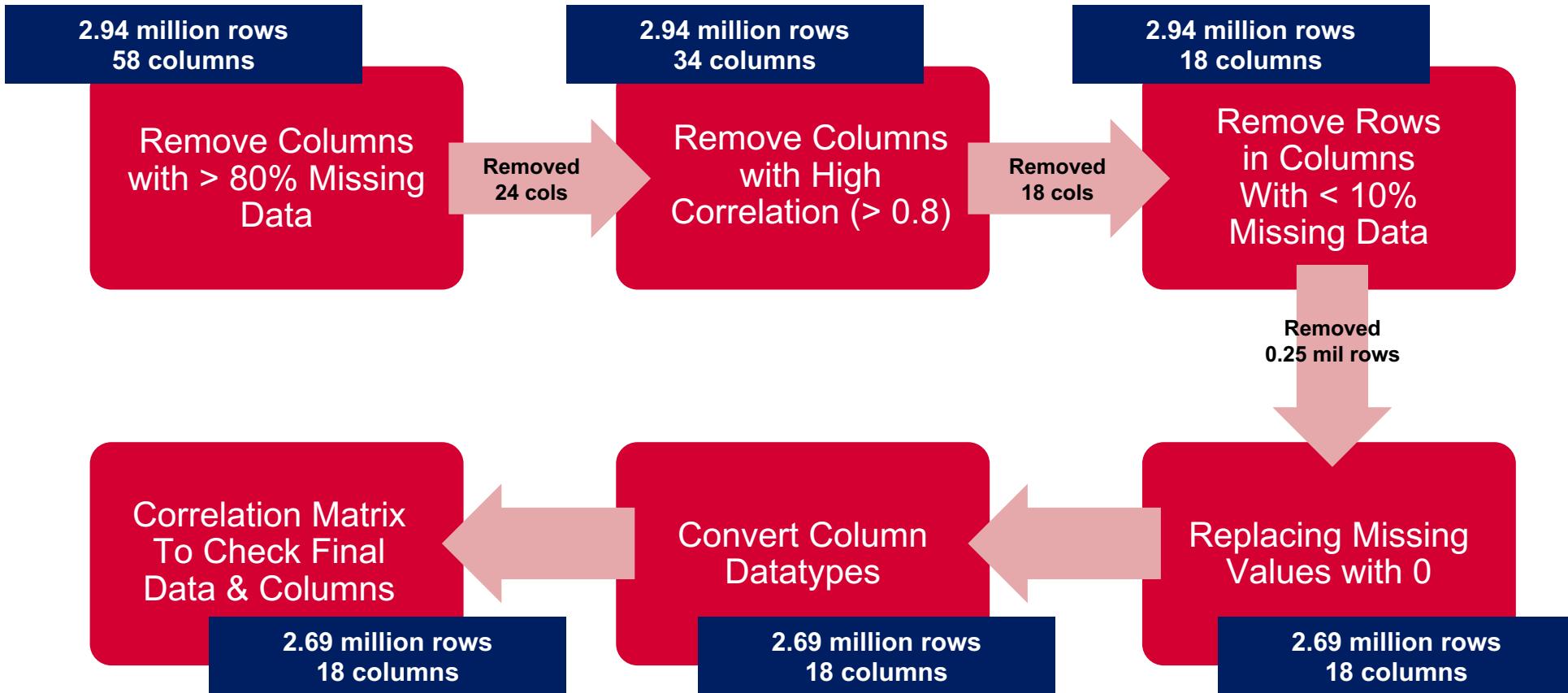


Year Built (Top 10)



max bed
count freq
is 3

Data Cleaning



Data PreProcessing



Models



Decision Tree

DecisionTreeRegressor



Random Forest

RandomForestRegressor



Linear Regression

LinearRegression



Gradient Boosted Tree

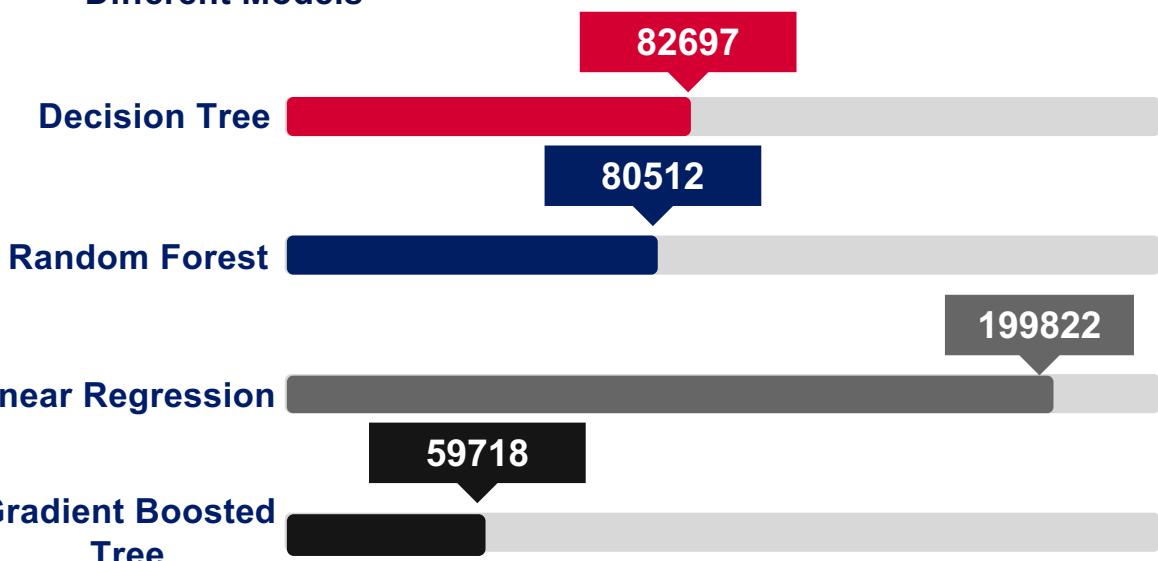
GBTRegressor

Model Evaluation Results

RMSE

Root mean squared error (RMSE) is the square root of the mean of the square of all of the error, and a smaller value indicates a smaller error between prediction and actual value. The use of RMSE is very common, and it is considered an excellent general-purpose error metric for numerical predictions.

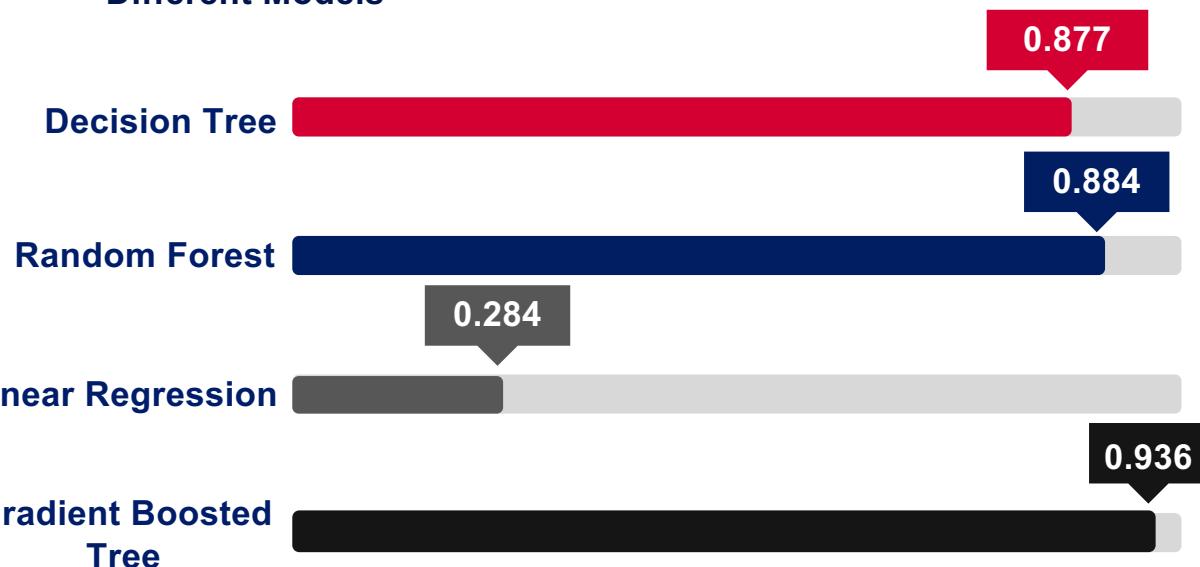
Different Models



R2 (R-Square)

R Square value is between 0 to 1 and a bigger value indicates a better fit between prediction and actual value. It is a good measure to determine how well the model fits the dependent variables.

Different Models



Best Model: Gradient Boosted Tree (GBTRRegressor)

Top 4 Important Features: buildingqualitytypeid, unitcnt, calculatedfinishedsquarefeet, propertylandusetypeid



Thank You