Refining City of Houston Budget Predictions

Joe Down

Flatiron School Final Project
September 12th, 2019

Project Overview

- Houston Budget Data 2010-2016
  - Format - Excel spreadsheets
  - Source - data.houstontx.gov
- Goal: Create a refined prediction of actual spending based on the given budget.
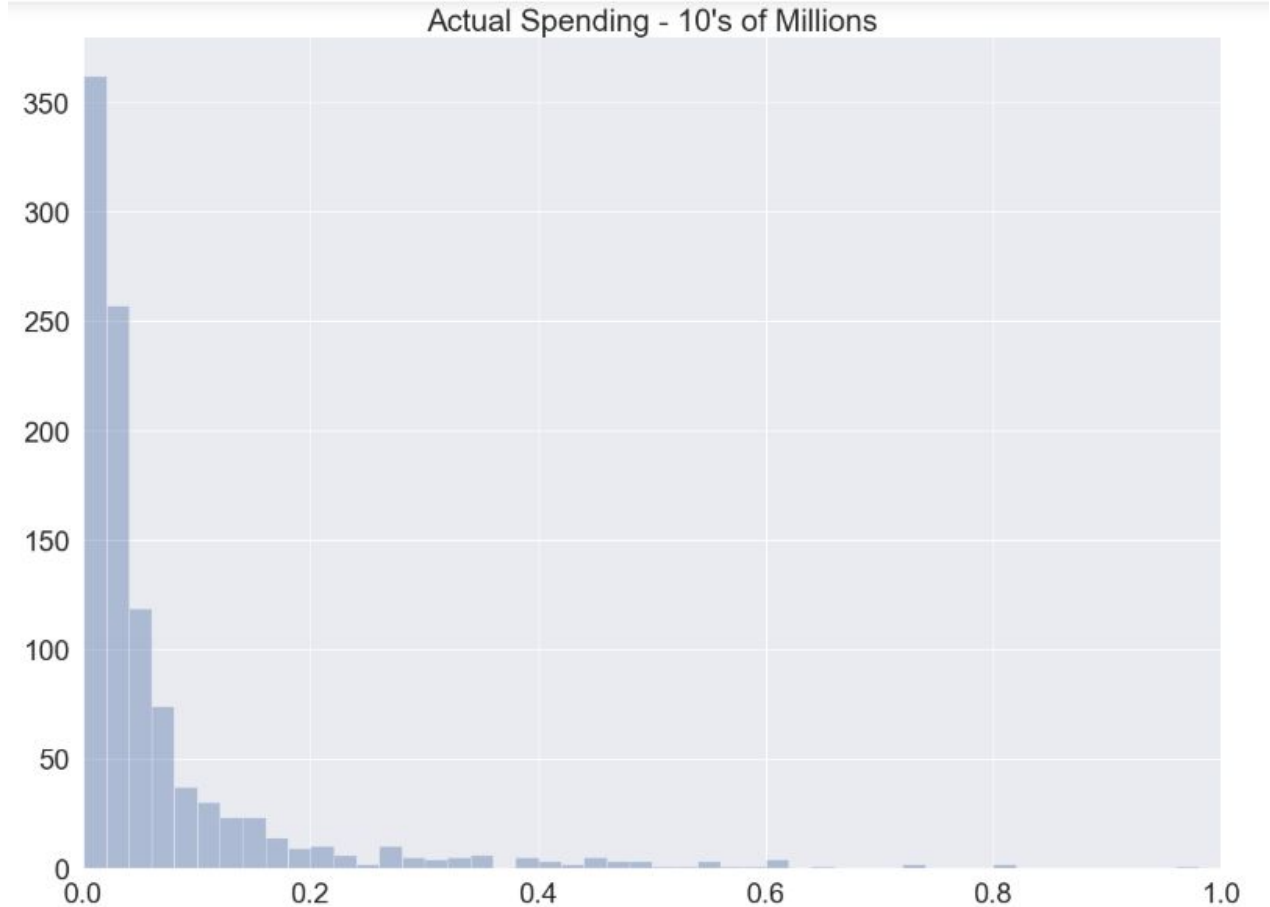
Example Data

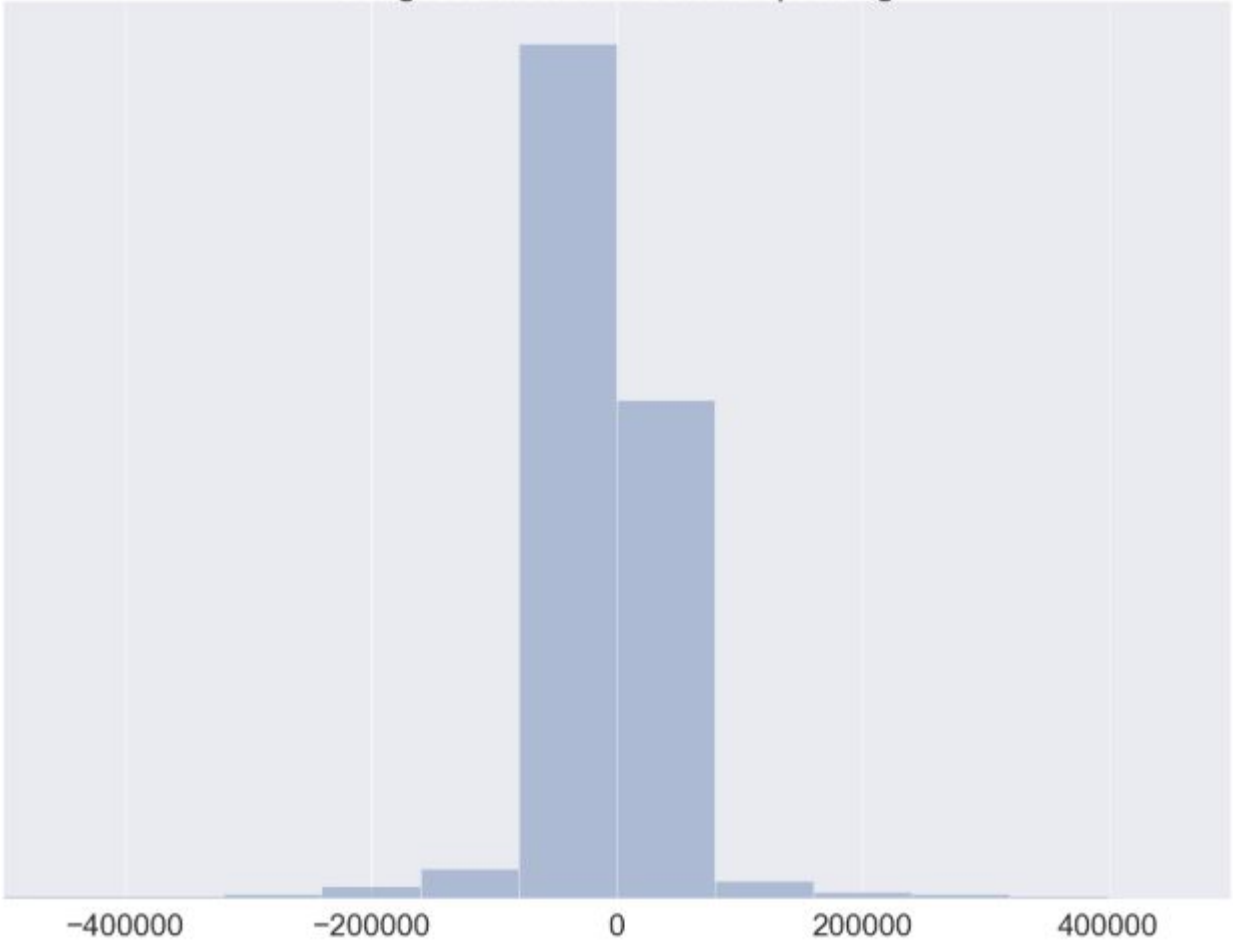| GL Account	GL Description	GL Category	GL Category Description	Revenue or Expenditure	Fiscal Year	Fund Name	Fund Type Descr	Business Area Name	Fund Center Name	Commitment Set Description	Original Budget	Current Budget	Actuals
501070	501070 - Pension - Civilian	500	500 - Personnel Services	Expenditures	2010	General Fund	General Funds	Health & Human Services	HHS-Office Prog Supp	50 - Total Expenditures	113579.0	113579.0	121344.82
511085	511085 - Drugs & Medical Chemicals	510	510 - Supplies	Expenditures	2010	General Fund	General Funds	Health & Human Services	HHS-Department Wide	50 - Total Expenditures	25000.0	25000.0	0.00
520765	520765 - Membership & Professional Fees	520	520 - Other Services and Charges	Expenditures	2010	General Fund	General Funds	Houston Police Department	HPD-Tactical Ops.	50 - Total Expenditures	140.0	140.0	813.00
500010	500010 - Salary Base Pay - Civilian	500	500 - Personnel Services	Expenditures	2010	General Fund	General Funds	Parks & Recreation	PRD-Park Fac. Permit	50 - Total Expenditures	159241.0	161056.0	164955.57
500010	500010 - Salary Base Pay - Civilian	500	500 - Personnel Services	Expenditures	2010	General Fund	General Funds	Solid Waste Management	SWM-SW Yard Waste	50 - Total Expenditures	512310.0	519021.0	454980.25

Data Properties

- 7 Budgets around 28,000 rows each
- Altogether 196,731 Rows
  - Only 22 rows with null values - dropped
- Kept only identical shared line items across all seven budgets - 80,493 rows
- Average annual Houston budget is around $5 billion annually, $3.3 billion of which is represented by the 80,493 line items shared identically
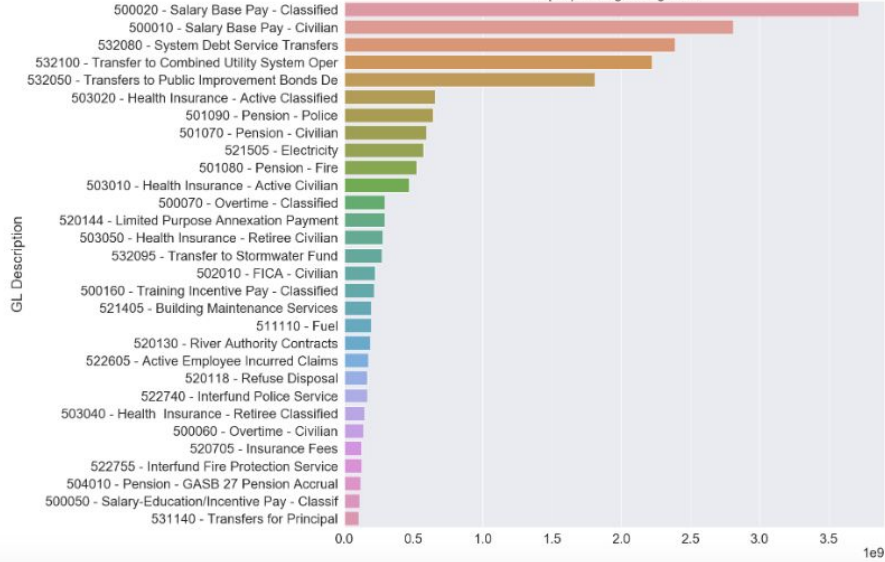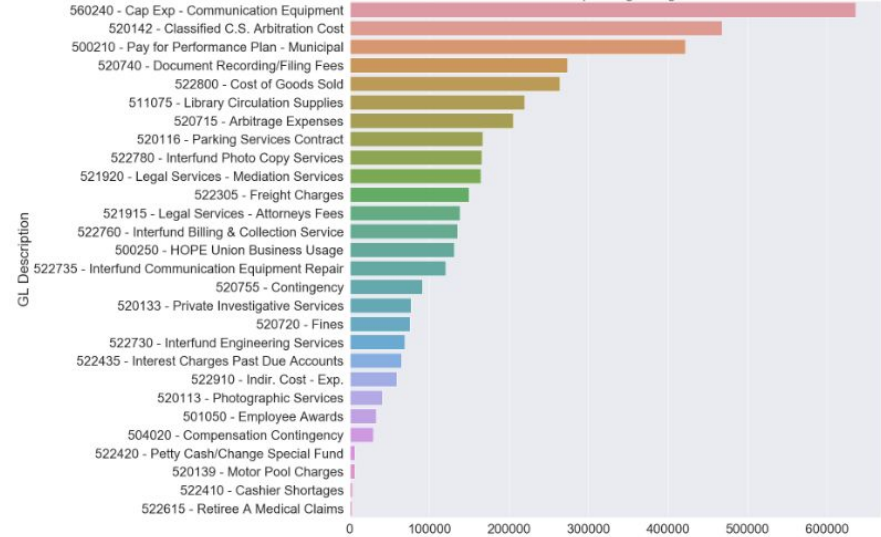
Actual Spending Distribution

Actual Spending - 10's of Millions

Budget Residuals on Actual Spending

Top Spending Categories

GL Description	
500020 - Salary Base Pay - Classified	
500010 - Salary Base Pay - Civilian	
532080 - System Debt Service Transfers	
532100 - Transfer to Combined Utility System Oper	
532050 - Transfers to Public Improvement Bonds De	
503020 - Health Insurance - Active Classified	
501090 - Pension - Police	
501070 - Pension - Civilian	
521505 - Electricity	
501080 - Pension - Fire	
503010 - Health Insurance - Active Civilian	
500070 - Overtime - Classified	
520144 - Limited Purpose Annexation Payment	
503050 - Health Insurance - Retiree Civilian	
532095 - Transfer to Stormwater Fund	
502010 - FICA - Civilian	
500160 - Training Incentive Pay - Classified	
521405 - Building Maintenance Services	
511110 - Fuel	
520130 - River Authority Contracts	
522605 - Active Employee Incurred Claims	
520118 - Refuse Disposal	
522740 - Interfund Police Service	
503040 - Health Insurance - Retiree Classified	
500060 - Overtime - Civilian	
520705 - Insurance Fees	
522755 - Interfund Fire Protection Service	
504010 - Pension - GASB 27 Pension Accrual	
500050 - Salary-Education/Incentive Pay - Classif	
531140 - Transfers for Principal	

Bottom Spending Categories

GL Description	
560240 - Cap Exp - Communication Equipment	
520142 - Classified C.S. Arbitration Cost	
500210 - Pay for Performance Plan - Municipal	
520740 - Document Recording/Filing Fees	
522800 - Cost of Goods Sold	
511075 - Library Circulation Supplies	
520715 - Arbitrage Expenses	
520116 - Parking Services Contract	
522780 - Interfund Photo Copy Services	
521920 - Legal Services - Mediation Services	
522305 - Freight Charges	
521915 - Legal Services - Attorneys Fees	
522760 - Interfund Billing & Collection Service	
500250 - HOPE Union Business Usage	
522735 - Interfund Communication Equipment Repair	
520755 - Contingency	
520133 - Private Investigative Services	
520720 - Fines	
522730 - Interfund Engineering Services	
522435 - Interest Charges Past Due Accounts	
522910 - Indir. Cost - Exp.	
520113 - Photographic Services	
501050 - Employee Awards	
504020 - Compensation Contingency	
522420 - Petty Cash/Change Special Fund	
520139 - Motor Pool Charges	
522410 - Cashier Shortages	
522615 - Retiree A Medical Claims	

Added Features

- City Council Yearly Roster - One Hot Encoded
- Astros, Rockets, and Texans yearly attendance and playoff rounds
- Yearly Rainfall, Yearly Crime Index, Avg Temps

Derived Features

- 1Y, 2Y, 3Y past Actuals
- 1Y, 2Y, 3Y past Actuals mean
- 1Y, 2Y, 3Y past difference between budget and actual spending
- 1Y, 2Y, 3Y past difference between budget and actual spending means
- One hot encoding for Fund Name, Fund Type Description, Business Area Name, Commitment Set Description

Modelling Objective:

Using all data from 2010-2014 and only the original budget from 2015 and 2016 make a binary (yes/no) prediction as to whether actual spending in 2015 and 2016 will be over or under the budget based on past years.

Model Used: XGBoost Classifier

Modeling Techniques:

- Weight of Evidence encoding - Information Value
- Principal Component Analysis

Model Used: XGBoost Classifier
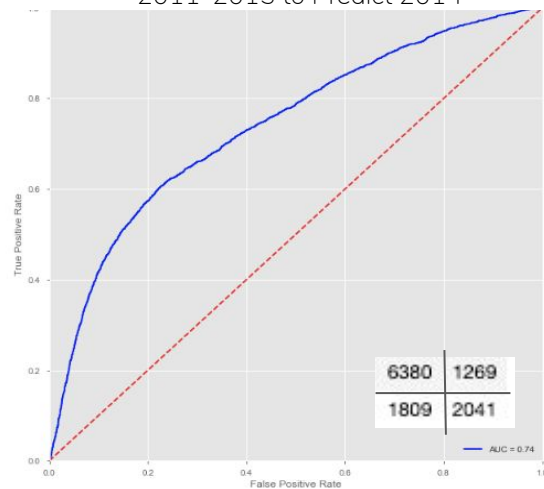
Area Under the Curve Graphs

2010 to Predict 2011

5452	2058
3268	721

AUC = 0.42
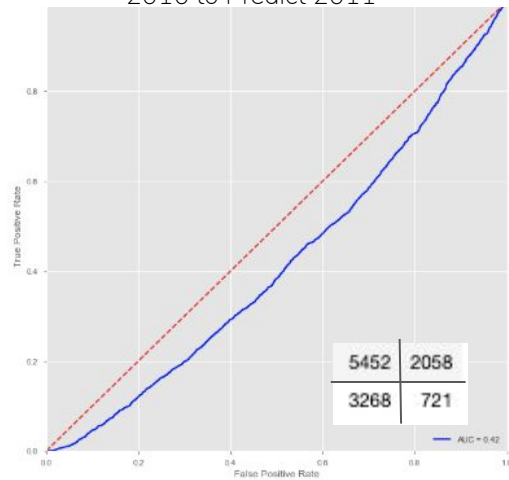
2010-2011 to Predict 2012

True Negatives	False Positives
False Negative	True Positives

5091	1859
3347	1202

AUC = 0.47
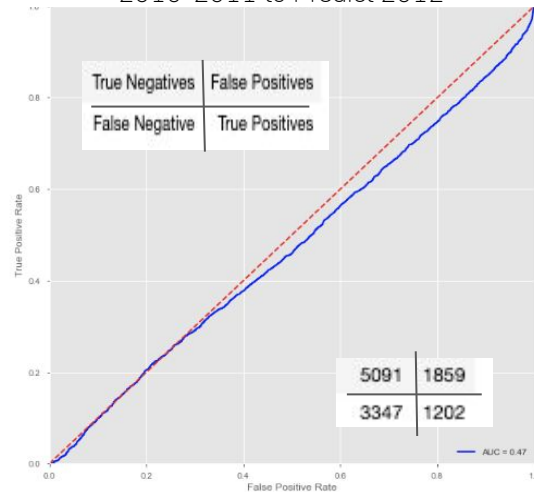
2010-2012 to Predict 2013

6382	1120
2650	1347

AUC = 0.62
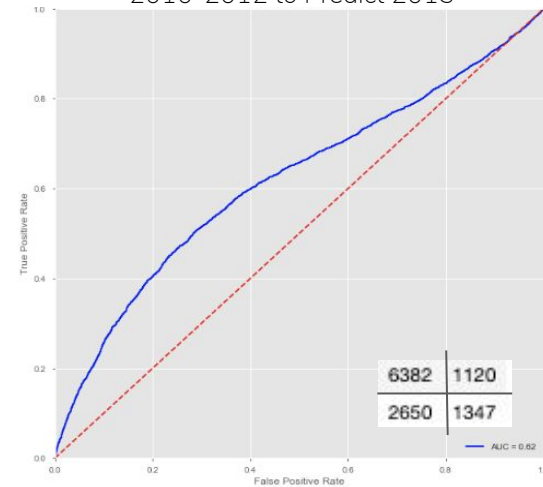
2011-2013 to Predict 2014

6380	1269
1809	2041

AUC = 0.74

2012-2014 to Predict 2015

6117	1346
1854	2182

AUC = 0.76

2013-2015 to Predict 2016

| 6684	899
2304	1612

AUC = 0.77
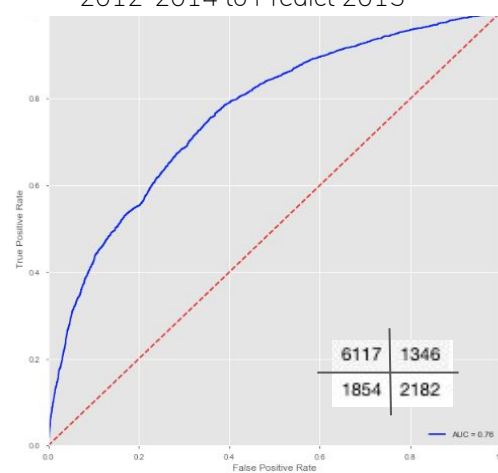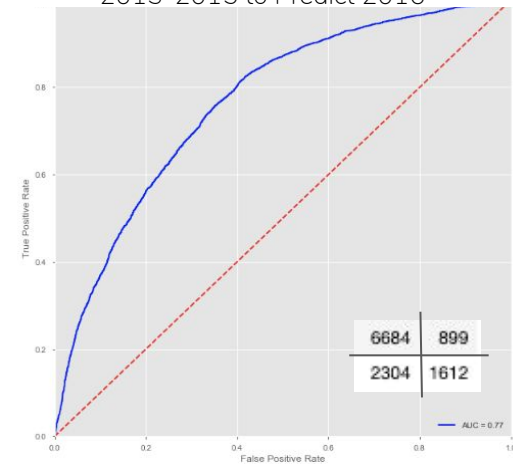
Additional Modelling Objective:

Using all data from 2010-2014 and only the original budget from 2015 and 2016 make a numeric prediction of the actual 2015 and 2016 spending that is more accurate than the budget
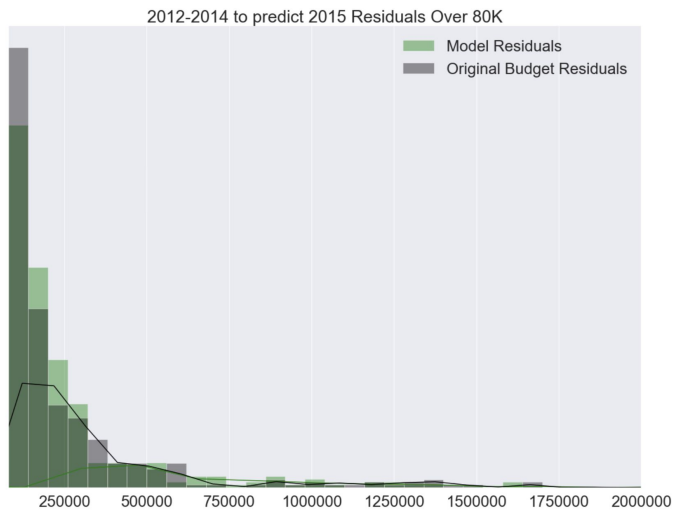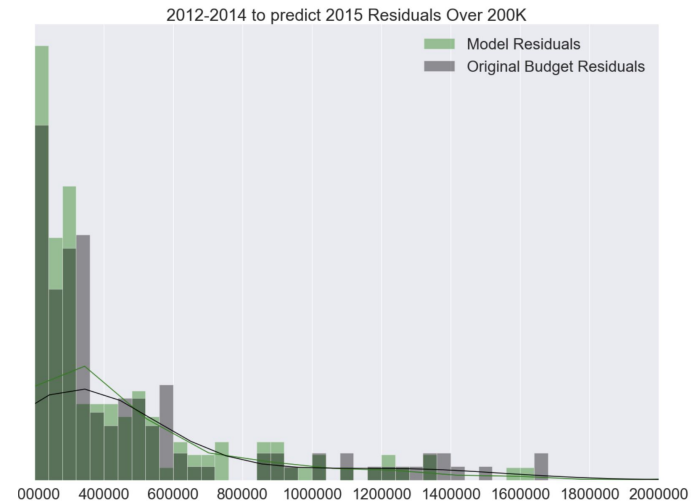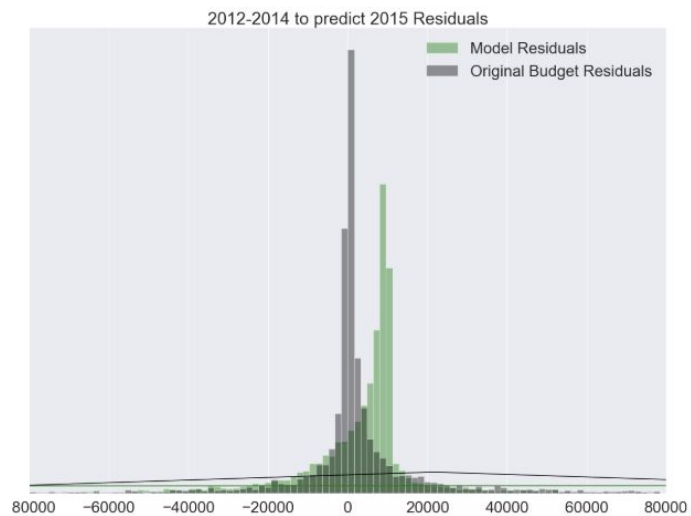
Model Used: XGBoost Regressor
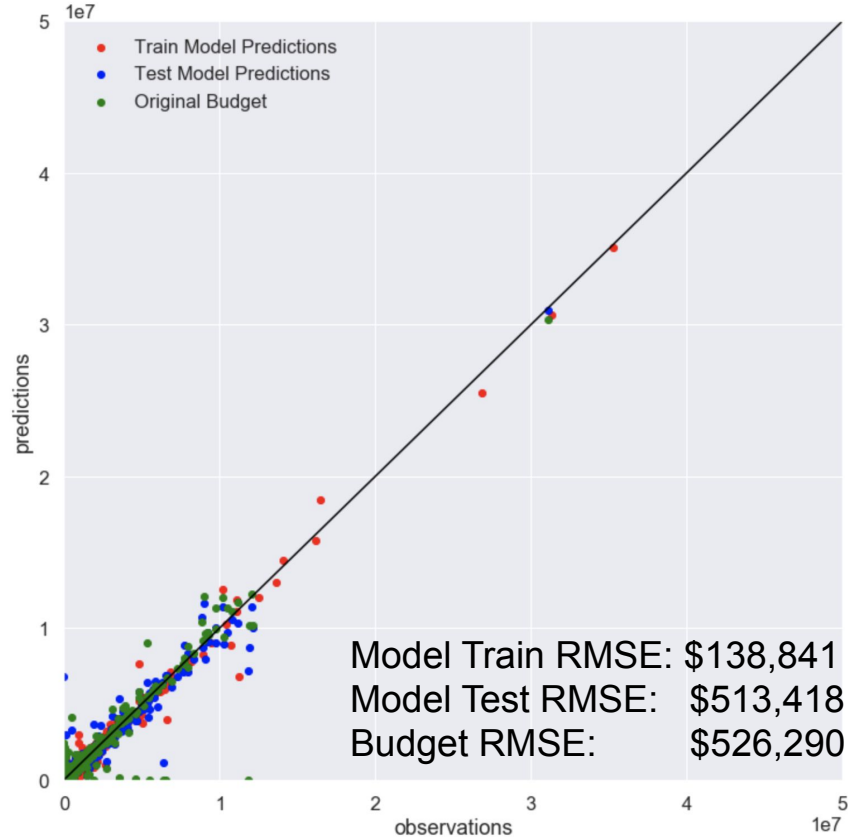
2012-2014 to Predict 2015 Model

Model Train RMSE: $140,000
Model Test RMSE: $446,811
Budget RMSE: $511,971
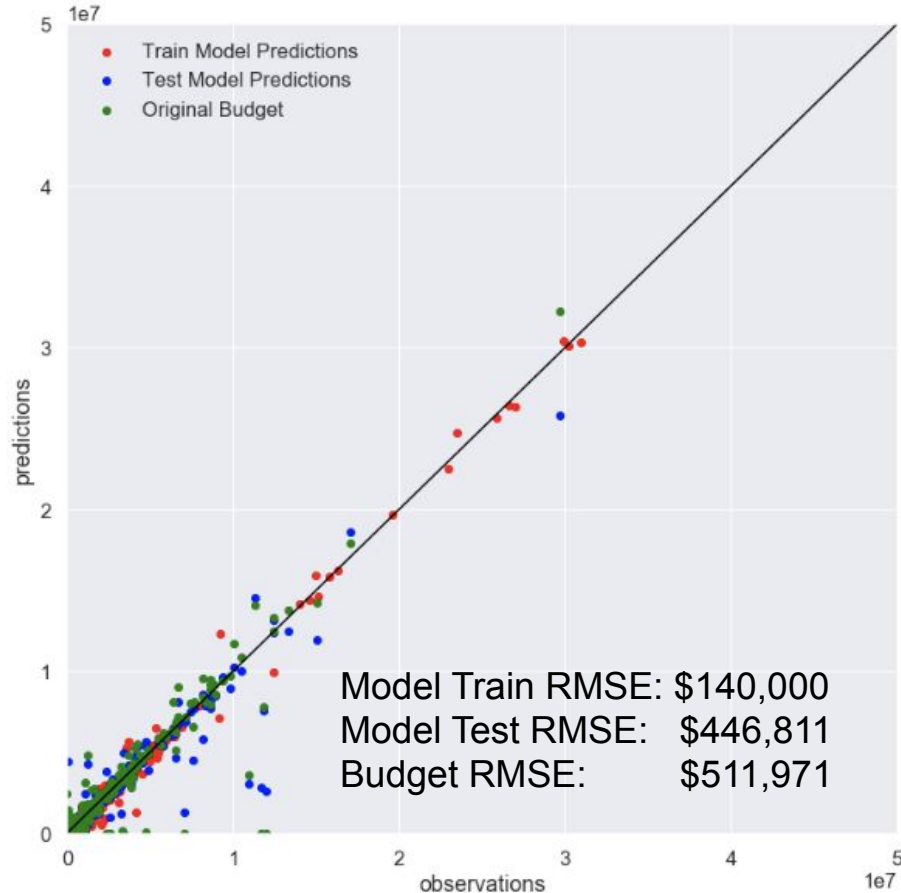
Feature Importance

Original Budget	240
1Y_Past_Actuals	99
1Y_Past_OA_Difference	53
Line Item	34
2Y_Past_OA_Difference	31
2Y_Past_Actuals	30
2Y_1Y_OA_Difference_Mean	29
2Y_1Y_Actuals_Mean	22
City Council District A_B Stardig	21
3Y_Past_Actuals	19
2Y_Past_Original_Budget	16
3Y_2Y_1Y_OA_Difference_Mean	13
Fund Id	11
3Y_Past_Original_Budget	11
3Y_2Y_1Y_Actuals_Mean	10
1Y_Past_Original_Budget	7
GL Account	6
2Y_1Y_Original_Budget_Mean	6
1YP_OA_Diff_Over_80K	6
Business Area	4
3Y_Past_OA_Difference	3
Over_Budget_Prediction	2
Fund Center Id	2
Fiscal Year	2
Fund Name_PWE-Combined Utility System Oper Fund	2
Fund Name_HAS-Revenue Fund	1
3Y_2Y_1Y_Original_Budget_Mean	1
Fund Name_Building Inspection Fund	1

2012-2014 to predict 2015 Residuals

2012-2014 to predict 2015 Residuals Over 200K

2012-2014 to predict 2015 Residuals Over 80K

2012-2014 to predict 2015 Residuals Over 2M

2013-2015 to Predict 2016 Model

Model Train RMSE: $138,841
Model Test RMSE: $513,418
Budget RMSE: $526,290

Feature Importance

Original Budget	237
1Y_Past_Actuals	109
1Y_Past_OA_Difference	54
3Y_Past_Actuals	52
Line Item	45
3Y_Past_OA_Difference	40
2Y_1Y_OA_Difference_Mean	30
2Y_Past_Actuals	29
2Y_Past_OA_Difference	27
2Y_1Y_Actuals_Mean	16
3Y_Past_Original_Budget	16
GL Account	15
Fund Id	14
Business Area	12
1Y_Past_Original_Budget	11
City Council District A_B Stardig	10
Controller_C Brown	9
Over_Budget_Prediction	9
Fund Name_PWE-Combined Utility System Oper Fund	6
2Y_1Y_Original_Budget_Mean	4
2Y_Past_Original_Budget	4
Business Area Name_Houston Fire Department (HFD)	2
1YP_OA_Diff_Over_80K	1
Fund Center Id	1
Business Area Name_General Debt Services	1