

City Finances Results

[Table of Contents](#)








Original Proposal versus Actual Work Completed:	2
About the Data:	2
Description:	2
Database Columns Summary:	2
Data to Use in this Study:.....	3
Exploring and Visualization:	3
Metric 1: Cities with a Deficit vs Cities with a Surplus	3
Metric 2: Annual Remaining Balance:	4
Metric 3: View of Fiscal Responsibility by cities in each state	5
Machine Learning	6
Models:	6
Summary of Key Findings:	6
Recommendations:	6

City Finances Results

Original Proposal versus Actual Work Completed:

The objective of this section is to compare the objectives of the proposal to the work actually performed, and explain the deltas between the two.

Analyze 150 cities across the United States using data from 2002 to 2016 to determine...

Original Scope	Accomplished	Reason not accomplished
Percent of cities that are managing expenses within revenues (sound budget practices)		
If annual remaining balance (deficit or surplus), after adjusting for inflation, is increasing, stable, or decreasing for each city		
Which states have the highest percent of FSCBally responsible and not responsible cities		
Using (revenue – spend) apply clustering methods to look for ways to group cities that over-spend versus cities that maintain budget or generate a surplus.		Given the disparity in year over year deficit/surplus percent change, clustering did not work for this study
Build a model that can be used to predict the probability that a city will remain solvent or go bankrupt		Insufficient data: only 5 of the 150 cities in this study declared (or came close to declaring) bankruptcy.
Create a Tableau viz that can be used to assess the health of a given city and whether the city is going to go bankrupt		Insufficient data: same as above
Modified Scope:		
In lieu of clustering, Regression Classification, Random Forest, and XGBoost algorithms were used to assess categories of (1) went bankrupt, (2) deficit, (3) surplus		

About the Data:

Description:

- The “[FSCBally Standardized Cities Database \(FSCB\)](#)” is the primary source for this study
- A listing of cities that require bailouts or went bankrupt (source is tbd)

About the FSCB:

This database makes it possible to compare local government finances for 150 of the largest U.S. cities across more than 120 categories of revenues, expenditures, debt, and assets.

Database Columns Summary:

The City Finances spreadsheet contains 662 columns and 6,080 rows. The analysis of columns (features) resulted in the following observations...

- Columns A thru E are “core data” columns containing year, state/city, FSCB city id, city population, and CPI. These core columns will be included in the data analysis.

City Finances Results

- Columns F thru TE (520 columns) are supporting data used to arrive at the “FSCB data columns”, which spans columns TF thru YL (137 columns). Since the FSCB data is the basis for this study, the supporting data columns will not be included in the data analysis.
- The FSCB data columns can be divided into Revenue (TF thru WT) and Expenditures (WU thru YK)

Data to Use in this Study:

This data analysis will have access to 142 columns...

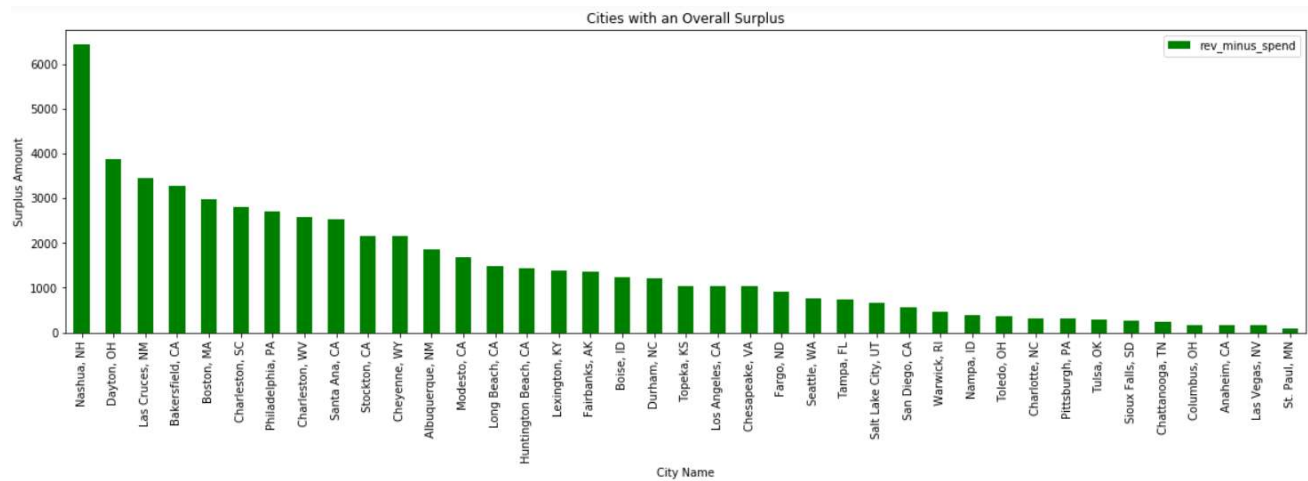
- Core data that spans columns A thru E (5 columns)
- FSCB Revenue data that spans columns TG thru WW (93 columns)
- FSCB Expense data that spans columns WX thru YM (44 columns)

*** Summary based on the FSCB excel spreadsheet “Variable List” sheet and the [“Explanation of FSCBally Standardized Cities”](#).*

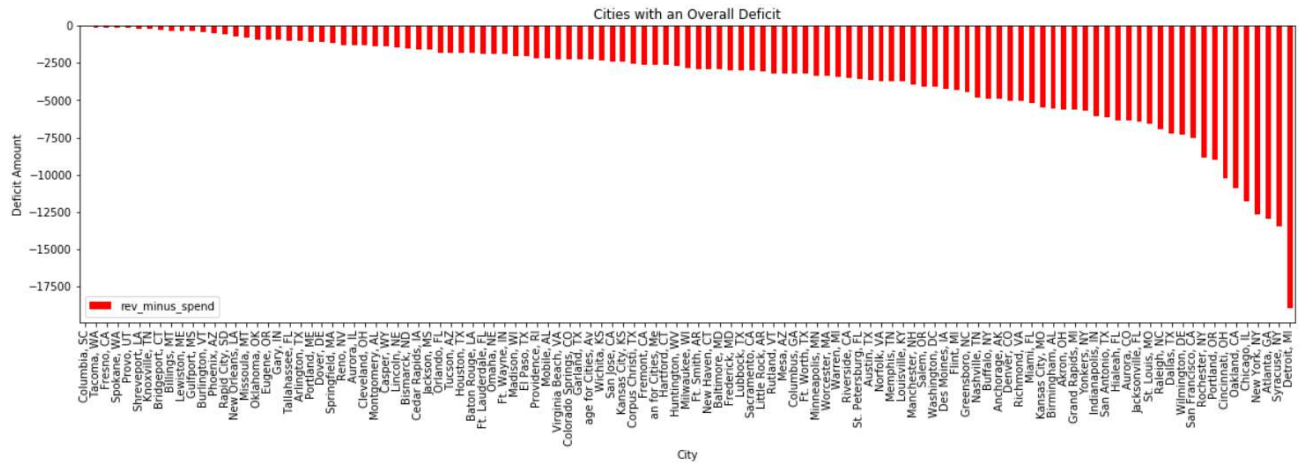
Exploring and Visualization:

Metric 1: Cities with a Deficit vs Cities with a Surplus

26% (39) of the 150 cities ran an overall surplus between 2002 – 2016, while 74% (111) ran a surplus.



City Finances Results

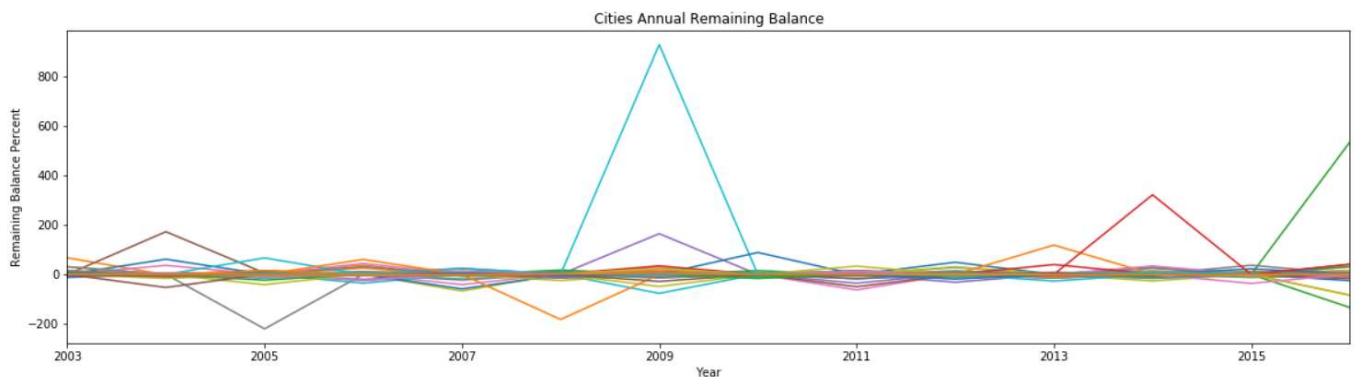


Detroit run the highest deficit total at \$17,500,000 while Columbia, SC runs the lowest deficit of \$12,000.

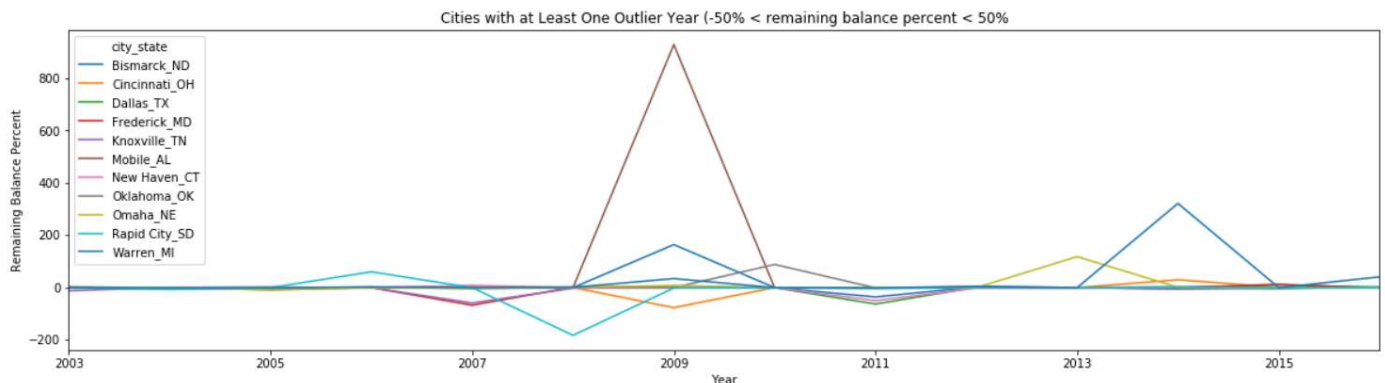
Nashua, NH run the highest surplus of \$6,400,000 while St. Paul, MN runs the lowest surplus of \$89,000

Metric 2: Annual Remaining Balance:

Viewing the year by year percent change in each cities remaining balance, yields a view that shows a great deal of disparity and some very large outliers...

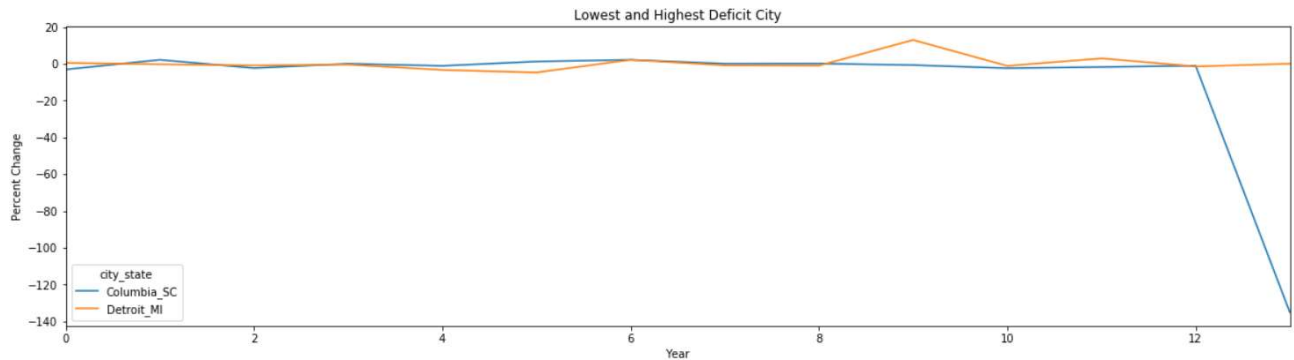


Defining outliers as at least one year with a spending increase or decrease of at least 50%, highlights 11 cities...

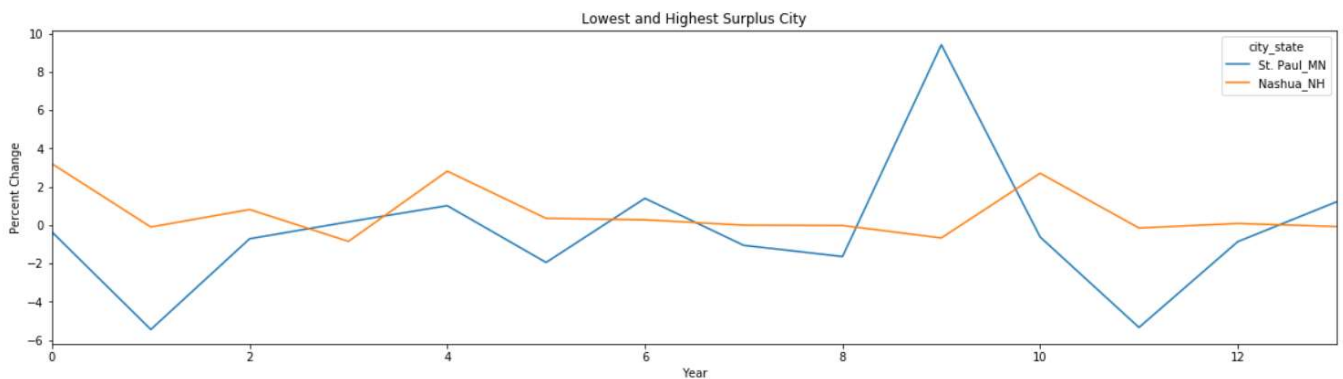


City Finances Results

Metric 1 above identified Columbia, SC as the lowest deficit and Detroit, MI as the highest. A view of their year over year percent change is shown below...

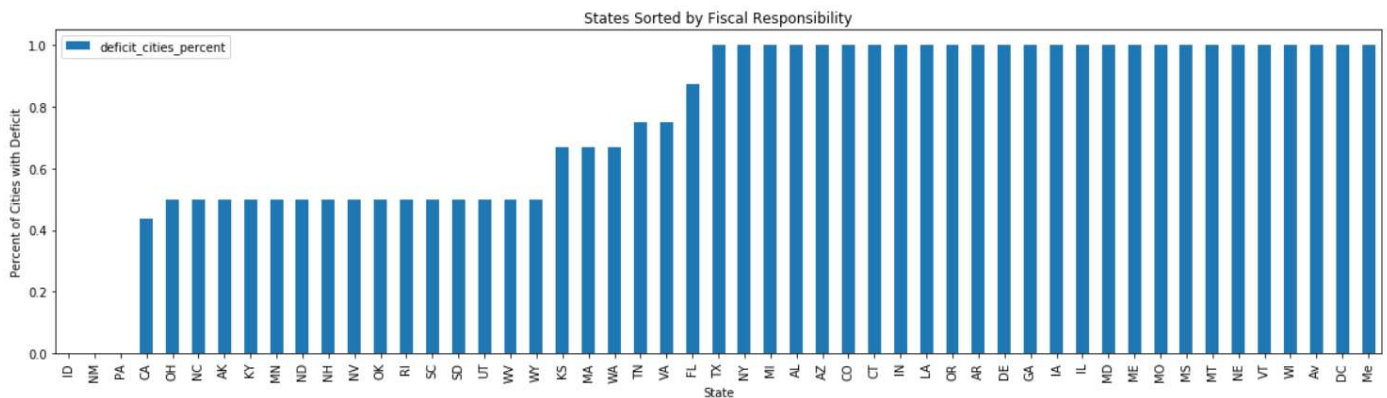


Metric 1 above identified St. Paul, MN as the lowest surplus and Nashua, NH as the highest. A view of their year over year percent change is shown below...



Metric 3: View of Fiscal Responsibility by cities in each state

The view below only represents 150 cities (average of 3 per state). Given this low number of cities per state and that each state on average has dozens of cities, this view may not be an accurate rating of states with most to least Fiscally responsible cities.



City Finances Results

Machine Learning

Models:

Clustering did not work well for this study (see above section, “Original Proposal versus Actual Work Completed”), so supervised learning algorithms of Logistic Regression, Random Forest, and XG Boost were selected instead to produce models.

Logistic Regression was selected simply because regression is a good (and easy) starting point for any machine learning activity. Random Forest and XGBoost were selected because given the high variance across cities and within each city’s percent change in spending per year, these ensemble methods seem best suited to handle the randomness of the data.

This table summarizes the results showing XG Boost with the best results.

Algorithm	Accuracy Score	Precision Score	Recall
Logistic Regression	54%	42%	47%
Random Forest	68%	58%	61%
XG Boost	68%	62%	63%

Summary of Key Findings:

1. Initial analysis of the data for this study showed a data set that overall required very little cleaning/scrubbing of data to get it ready for further analysis. This is a credit to the maintainers of the FSCB database.
2. This data is very challenging to draw conclusions from...
 - a. On the FSCB database website they state that due to the variety of financial structures that cities (even within a given state) employ, it is challenging to create apples-to-apples comparisons of the data.
 - b. In addition, the disparity in revenues/expenses between cities and the often large percent changes in year over year revenue/spending for each city makes it challenging for algorithms to build models of high accuracy. For example, achieving accuracy scores of only 68% for what is primarily a binary (deficits and surpluses) classification problem is not very good.
3. 74% (111) of the 150 cities are deficit spenders. This supports what is often reported in the media – that being that government spending and the accumulation of debt is a problem in the United States.

Recommendations:

1. The first step for this study should have been to assess how many of the 150 cities actually have gone bankrupt or come close to going bankrupt. Discovering there are only 5 cities after the effort to achieve data familiarity, and complete exploration, & visualization was a draw back to the original goals of this study and changed the scope significantly (i.e. went from build model to discover cities trending toward bankruptcy to build model to predict deficit/surplus trends).
2. Percent change per year in spending is obviously not the only way to apply data to predicting whether a city will be a deficit or surplus city. Other way not investigated that could be tried are to take categories of spending (i.e. amounts applied to pensions, public parks & recreation, police & fire protection, etc...) and determine if a particular category or combination of categories could be a better means to predict whether a city will continue to run a deficit or surplus.