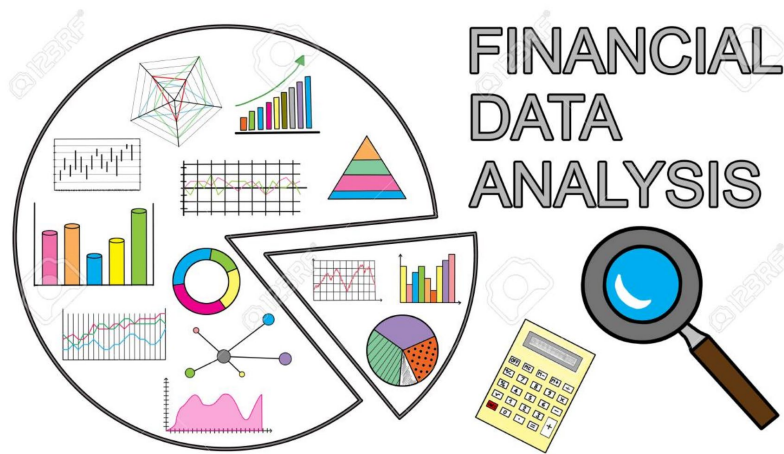


Final Project - ACC 9993

NYC Budget Data Analysis



Team Data Sharks

Aqsa Naeem

Bong Oh Park

Dechen Hu

Meng Li

Michael Friedlich

Shobhit Ratan

Sory Fofana

May 08, 2019

Table of Contents

1. Executive Summary	2
2. Background Overview	3
3. Role and Motivation	4
4. Identify the Questions	4
5. Master the Data	5
6. Perform the Test Plan	6
7. Address and Refine Results	6
8. Communicate Insights and Recommendations	7
9. Extensions	8
Appendix 1 : Master the Data description	9
Appendix 2 - Data Visualization	11

1. Executive Summary

New York City has by far the largest population of any US cities, and thus its agencies by far receive the most funding. As such, for a governmental entity, the report measures its performance by analyzing how well it spends its tax dollars to improve the daily lives of its citizens. The main role and motivation of the report were to address the issue of whether the agencies are using the budgets assigned to them effectively. In line with this, the main issues identified in this report included the problem of budgeting and resource allocation, in which analyzing budget variance was useful. In an attempt to understand the big picture of this issue, the report sought to answer two questions: Which agency has the largest budget and which expense category is most frequently used? To analyze the trend over time and detect noteworthy deviations, the report analyzed the last five years of the budget of New York while focusing on these questions: “Has the proportion of budget allocation by agency remained constant and has the overall NYC budget increased, decreased or remained constant?” Finally, in order to better provide more context and understand the data, the report sought to view how the findings align with the overall understanding of the needs of NYC. The report mastered the data, which is a dataset that contains financial information of 151 agencies in NYC from 2015-2019. This was downloaded from checkbooknyc.com. From the dataset, the report deployed descriptive statistics for modified and adopted budget variables to better understand the data and enabled the identification of the range of data like mean, maximum and minimum. Also, the report created a new budget variance for the 2019 Fiscal Year to determine which agencies are likely to go under or over the budget.

Two criteria were used to select a suitable analysis tool to answer the report questions. First is the criterion of the performance for processing large data in a reliable and robust way. The second criterion entailed the visualization capability to convert data that is unstructured into logical result with graphical dashboards. Accordingly, the report performed the analysis using Tableau to investigate budget variances for every agency and expense categories. The major limitations with the report analyses include some of the agencies with high remaining budgeted amounts may not be using their budgets poorly. Also, the analysis was predicted not to be able to completely reflect the details of the budgeting process. The key insights of this report include an extension which entailed two things. One, besides analyzing how the agencies can spend their budgets effectively and correctly, the report suggested an attempt to analyze how the NYC Tax Revenue may affect the amount of NYC budget for each year. This could be performed by combining the NYC budget and the dataset of NYC Tax Revenue, and analyze the question using prediction and regression model to prove the assumption. Second, the report also suggested that it is valuable to analyze how the major changes such Amazon’s headquarter cancellation affect the Tax Revenue of NYC as well as the budgetary deficit if it does not take place.

2. Background Overview

The key facts about the city of New York are that it has by far the largest population of any city in the United States. It has more than double the population of the second largest city, Los Angeles, and three times the population of Chicago, the third largest city. Because it is such a large city, it needs to be well-funded. In regard to funding, education, including the City University of New York and other city-related agencies, receives by far the most funding. The NYPD and FDNY also get a lot of funding from the city government. These two areas of the city receive such a large portion of funds because of the necessity to grow and develop future citizens so that these well-developed citizens can then create a better society. Further, one of the city's highest priorities is maintaining law and order and keeping its citizens safe. Finally, because there are a vast number of agencies needed to run the city, New York City has to hire a large number of people to run these agencies. Therefore, the city has very high payroll expenses in order to pay all of these employees. Of all types of expenses, New York City's payroll expenses are more than any other type of expense.

The key factors that affect the New York City government are somewhat different from those of a business because it is a governmental entity. A governmental entity like the New York City government is mainly focused on measuring performance by analyzing whether it is spending its citizens tax dollars in a way that improves their daily lives. While this may be similar to a customer satisfaction index in a business, the government is serving the people, where a business is not. Overall, a governmental entity should be analyzed by its operational performance and how it is helping its citizens. Further, the city's government should analyze its spending by whether it is adequately meeting the needs of its taxpayers and whether the spending is effective in helping the city. Some ways to measure these factors are to look at spending on education, health care, law and order, parks and recreation, social services, sanitation, etc. These expenses are all clearly aimed at helping the citizens. The city could also examine administrative and other than personal expenses that are not wasteful but may not be as useful as those described above. The city government could also look at how pleased citizens are with services provided by the government to its citizens. For example, when the Metropolitan Transportation Authority (MTA) recently implemented a fare hike, many riders may think a fare increase is unreasonable when they do not believe that the service of subways and other transportation has been good enough. Another way that the performance of the New York City government can be measured is through examining its ability to attract businesses to open corporate headquarters in the city. This could bring a lot of tax revenue to the city with more jobs and more people moving to the city, and thus could allow the city to spend more on its citizens. However, the city recently lost Amazon's second headquarters after protesters drove the company away, so not only was that a loss in potential tax revenue but also a potential factor that may keep other companies away from opening their headquarters in the city.

3. Role and Motivation

The problem we are being asked to address is whether the agencies are using their assigned budgets effectively given the limited amount of funding available to all agencies. We are analyzing the amount of money these agencies are receiving and how much of this money is used for the purpose it was designated to be spend on. For example, if the City University of New York is given \$1 million for the purpose of training employees and increasing safety, we would want to see whether that money is actually used for its designed purpose. To ensure that all agencies are spending their money in a way that helps the people of New York City, we want to identify those agencies that consistently have high uncommitted budget amounts or have the lowest committed amounts.

This is an important issue to address because the citizens of New York City pay taxes to the city's government. Thus, the city needs to use the taxpayers' money in a way that makes New York City a better place, thus helping the taxpayers. Further, when certain agencies are consistently spending less than their assigned budgets, the city government could then adapt its resource allocation based on the agencies that are spending more or less than their budgets. This will enable the city to become more efficient as an organization. For example, if the city government identifies over time that the City University of New York is regularly spending over budget, but the Department of Buildings is typically under budget, then the city may want to increase the City University of New York's budget while decreasing the Department of Building's budget.

4. Identify the Questions

The problem we have identified is one of budgeting and allocation of resources. Given the limited amount of funding, agencies have to compete for resources and carefully monitor the use of their funds. This makes the analysis of budget variances quite important. But before we can identify variances, we must first identify the current state of New York City's financial budget. This includes the current fund allocation by agency and, more specifically, the designated purpose (indicated by the budget code). We seek to answer the following questions in an attempt to understand the big picture:

Q1. What are the Adopted, Committed and Modified Budget during the last five years and how the trends can be used to forecast estimates for 2020?

Q2. Which are the Agencies with maximum Committed budget? Is there a trend?

Once we have determined the relative expenses, we will observe if there is a trend over time. Much like auditing, this is done in order to detect significant deviations. If there is a

blip, it may not be as troublesome as a small, but consistent variance. We will be analyzing the last five years of New York City's budget and the questions we will be focusing on are as follows:

Q 3. Which Expense Categories use maximum Committed budget. What is the trend for last 5 years? Can we use the trend to forecast numbers for 2019? Has the overall NYC budget increased, decreased or remained constant?

Q 4. Which Agencies have maximum increase in Committed budget compared to their Adopted Budget? Is there a trend which suggests Diagnostic Analysis?

Lastly, in order to better understand the data and provide more context, we want to see how these findings align with our overall understanding of NYC's needs. We will also be looking at major events that have taken place in the past five years, including the Metropolitan Transportation Authority fare hikes as well as the loss of Amazon deal.

5. Master the Data

The purpose of this data analysis project is to identify poorly managed agencies and make recommendations to improve them.

This dataset contains financial information about different agencies in the city of New York from 2015-2019. The dataset was downloaded from checkbooknyc.com. It has 767,936 rows and 15 attributes such as the name of the city agencies, departments within agencies, adopted budget, modified budget, expense category, fiscal year, budget code, committed, budget name, encumbered budget, pre-encumbered, accrued expense, cash expense, post adjustment, remaining. Please refer to the table (data description) for the definition of different attributes.

Tasks performed in mastering the Data:

- We compared the number of records of the data we extracted from the website with the number of records we downloaded to ensure we didn't miss any records while downloading the file.
- The dataset was stored in a comma values format, we changed it to an excel workbook format.
- We used pivot table and tableau to aggregate data by agencies to facilitate manipulating the data.
 - Row: agency
 - Values: all attributes with numerical values
 - Column: Year
- Sorted the data from largest to smallest to easily identify agencies with 0s across all attributes.

- Filtered out agencies with 0s across all attributes because they didn't receive any funding from the city. Out of the 151 agencies 14 agencies' records were removed because they had only 0s as values which decreases the number of records from 767, 936 to 746,771.
- Used Box and Whisker Plot using 3 different variables (Modified, Committed, Remaining) to detect potential outliers and the range of the dataset.
 - 2016 remaining contains an unusual high number (will need to investigate further, we do not think it is an outlier, so we did not remove it).
- Used Descriptive Statistics using the same 3 variables mentioned above.
 - The mean and sum for Modified and committed budget has been increasing (excluded 2019 from the analysis because the period has not end yet).
 - The mean and sum for the remaining has been decreasing for the past 2 years (excluded 2019 from the analysis because the period has not end yet).

Note : Refer Appendix for description on data items

6. Perform the Test Plan

There are two criteria in selecting a suitable analysis tool to get answers to our questions. The first criterion is the performance to process the big data in robust and reliable way, because we need to be able to process the big amount of budget data for each agency within the NYC government. We also need to be able to handle each agency data in conjunction with expense category, so our tool must support creating relational data model within it. As a tool that meets those requirements, we have two choices which are MS Excel and Tableau. Both of those tools can process the big amount of data and create data models within itself with database relations.

The second criterion is the visualization capability to convert the unstructured statistical information into logical result with graphical dashboards. We concluded that Tableau outperforms the MS Excel in the visualization capability because Tableau allows the user to create analysis dashboards easily with high performance. Also, Tableau is capable of representing data in different dimensions such as color and size, which works as advantage to our project as we can apply color and size dimensions to our analysis report on agency budget and category expense.

Accordingly, we will perform the analysis using Tableau to answer the questions in Identify the Questions section.

7. Address and Refine Results

The main limitations of our analyses are that some of the agencies with high remaining budgeted amounts may not be using their budgets poorly. While we are assuming that a high remaining budget means that the agency is not using its allocated budget efficiently, the New York City government may have given that agency a high budget for a multiple year project. Therefore, a high remaining budget would not be an issue, as that was meant to be spent in a future year. We are concerned here with bias in our analysis, as all high remaining budgets do not indicate a problem with the agency. Additionally, our analysis will not be able to completely reflect the intricacies of the budgeting process. Budgets are future predictions and they may not account for unforeseen circumstances. For example, an agency may not end up using funds it received for a planned project after the project is canceled. Therefore, we hope to look for patterns of behavior over the years, instead of focusing on one year that may simply be an outlier or anomaly.

8. Communicate Insights and Recommendations

Various Data Visualizations were used to analyze the budget data and gain insights about the questions identified in Section 4 above. The data visualizations along with questions are placed at Appendix 2.

The recommendations and insights based on the data analysis and data visualizations are as follows :

Insights

- Adopted, Committed and Modified have consistently increased over the past 5 years. Based on the trend these budgets have been forecasted for year 2020.
- Dept of Education (DOE) has the highest committed budget from 2015-2018. This trend is likely to continue in 2019 also. For 2015-2018, same names feature in list of top 5 agencies with maximum committed budget and trend is likely to continue for 2019.
- Expenses for Dept of Education and Pension Contributions consistently show an increasing trend
- Same names feature in the list of top 10 expense categories with maximum Committed budget for period 2015-2019. Following expense categories show a consistently increasing trend for committed budget:
 - Full time Pedagogical personnel
 - Full year positions
 - Interest on Bonds General
 - Full time uniformed personnel
 - Teach Retsys

- *Debt services* shows a constant uptrend and consistently tops the list of agencies with highest increase in Committed in comparison to Adopted budget. It is an outlier and shows a steep rise since 2017.

Recommendations

- Increase in Committed over Adopted budget is highest for 'Debt Services' and shows a steep increase since 2017. This could be because of unbalanced Budget necessitating increased borrowing. This may require diagnostic analysis as continued budget deficit may continue to propel the cycle of increased borrowing leading to fiscal imbalance.
- Most agencies showing consistently increasing trend for Adopted, Committed and Modified budget. This needs further investigation and diagnostic analysis of agencies and expense categories to reduce discretionary spending. Top 5 agencies and expense categories with highest Committed budget can be a good starting point for further analysis.

9. Extensions

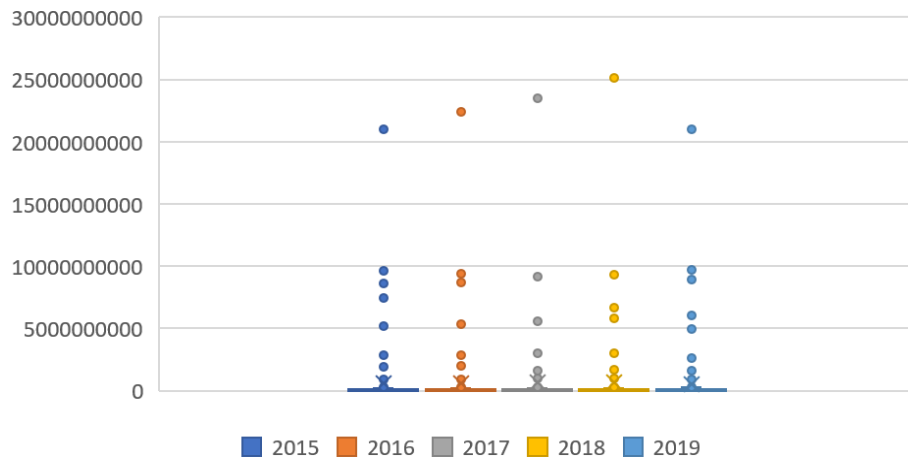
Besides the question we are trying to in this case that whether the agencies can utilize their budget correctly and effectively, we can also try to analyze how the NYC Tax Revenue may influence the amount of NYC budget for each year. The assumption is the growth trend of NYC budget may be the same as the growth trend of NYC tax revenue. In this case, we can combine NYC Tax Revenue dataset with NYC budget and analyze our question with regression and prediction model to prove our assumption.

Also, it is useful to analyze how the big changes happened in NYC such as the cancellation of Amazon's second headquarters would decrease the NYC Tax Revenue and how these would decrease the budgetary deficit if it does not happen. To analyze this question, we can add up the data of the tax revenue that Amazon contributes to Seattle to NYC to determine whether our politicians made a correct decision or not.

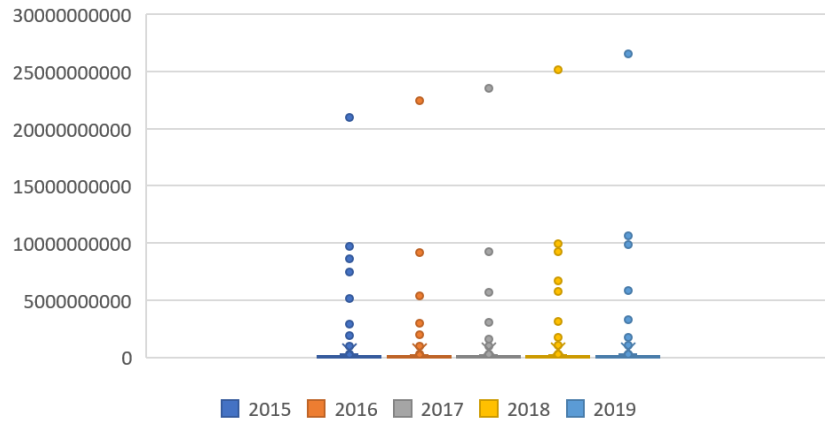
Appendix 1 : Master the Data description

Agency	Names of different agencies in the city
Department	Names of different departments within agencies
Adopted	The financial plan adopted by the City Council which forms the basis for appropriations.
Committed	Is the actual consumption or the total expenses (Committed Budget = Pre Encumbered + Encumbered + Accrued Expense + Cash Expense + Post Adjustments
Modified	Increases or decreases of adopted budget over the fiscal year in response to changing conditions.
Year	Fiscal year
Expense Category	Different types of expense
Budget Code	Budget Code
Budget name	Name of the budget
Pre-encumbered	A pre-encumbrance is a request to reserve funds for planned expenditures. The funds have been requested but have not yet been approved for the expenditures.
Encumbered	Encumbrance The commitment of all or part of an appropriation. Encumbrances represent valid obligations related to unfilled purchase orders or unfulfilled contracts.
Accrued Expense	An expense incurred but not yet paid
Cash Expense	Paid expense
Post adjustment	Post adjustment
Budget Variance	Modified budget – actual budget (in this case accrued expense)

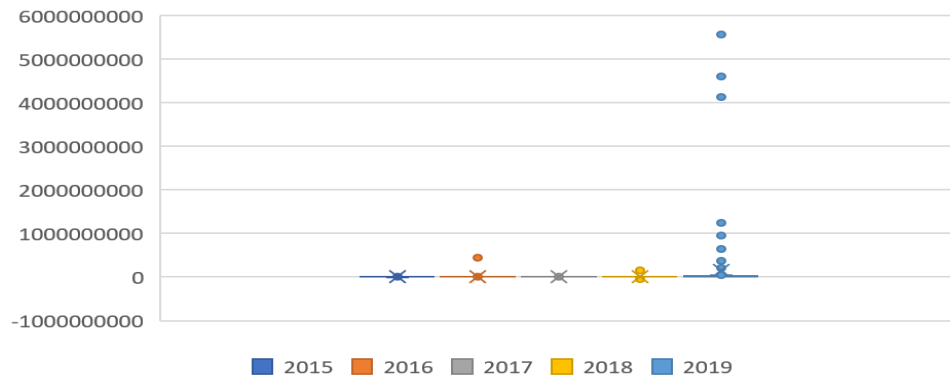
Committed Budget



Modified Budget



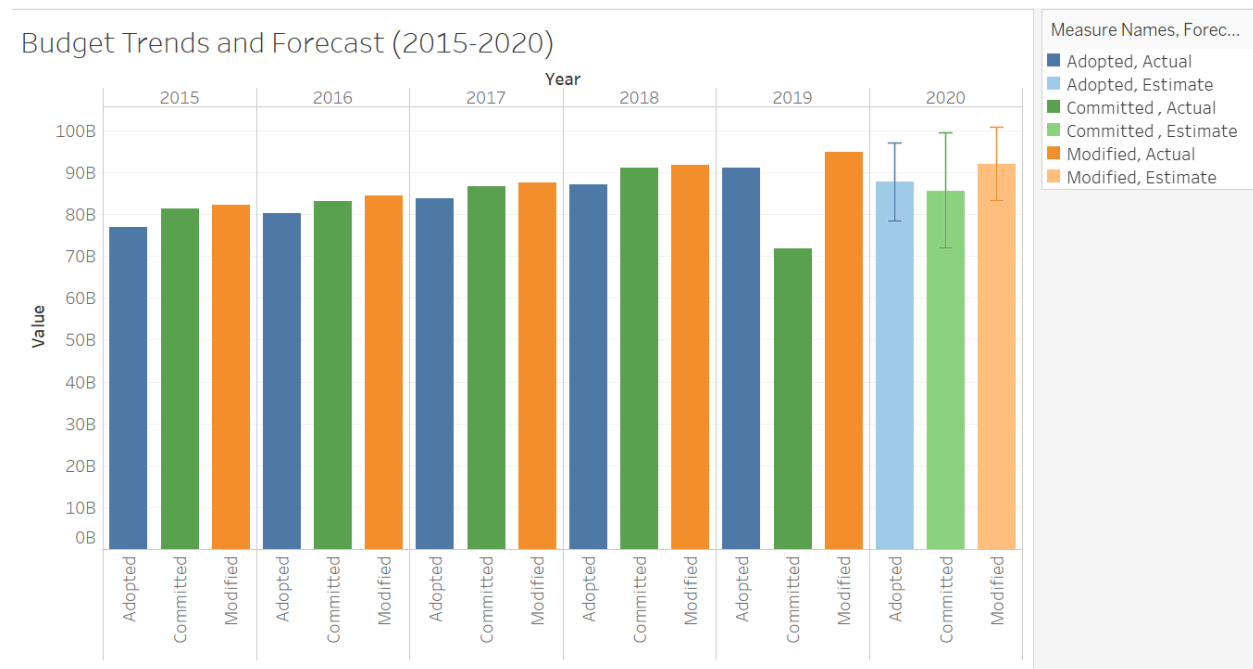
Remaining Budget



Modified budget	2015	2016	2017	2018	2019
Mean	\$ 604,240,902.94	\$ 620,463,557.71	\$ 638,730,661.59	\$ 670,238,620.55	\$ 693,317,625.82
Minimum	\$ 216,895.00	\$ 233,911.00	\$ 233,911.00	\$ 233,911.00	\$ 288,364.00
Maximum	\$ 21,003,585,311.00	\$ 22,421,945,726.00	\$ 23,511,738,815.00	\$ 25,144,670,132.00	\$ 26,539,187,393.00
Sum	\$ 82,176,762,800.00	\$ 84,383,043,848.00	\$ 87,506,100,638.00	\$ 91,822,691,016.00	\$ 94,984,514,738.00
Committed budget					
Mean	\$ 598,825,354.29	\$ 611,585,502.79	\$ 632,370,933.27	\$ 665,195,049.36	\$ 538,975,843.08
Minimum	\$ 149,626.31	\$ 171,928.32	\$ 155,207.58	\$ 165,901.88	\$ 151,085.78
Maximum	\$ 20,999,365,395.32	\$ 22,421,559,952.92	\$ 23,508,038,000.69	\$ 25,144,008,626.78	\$ 20,974,481,920.10
Sum	\$ 81,440,248,184.01	\$ 83,175,628,379.37	\$ 86,634,817,857.53	\$ 91,131,721,762.15	\$ 73,839,690,501.34
Remaining Budget					
Mean	\$ 5,415,548.65	\$ 8,878,054.92	\$ 6,359,728.32	\$ 5,043,571.20	\$ 154,341,782.75
Minimum	\$ (969,324.12)	\$ (79,701.99)	\$ 9.39	\$ (44,157,928.38)	\$ 39,625.87
Maximum	\$ 117,553,997.85	\$ 451,290,978.27	\$ 150,452,945.73	\$ 145,453,077.50	\$ 5,564,705,472.90
Sum	\$ 736,514,615.99	\$ 1,207,415,468.63	\$ 871,282,780.47	\$ 690,969,253.85	\$ 21,144,824,236.66

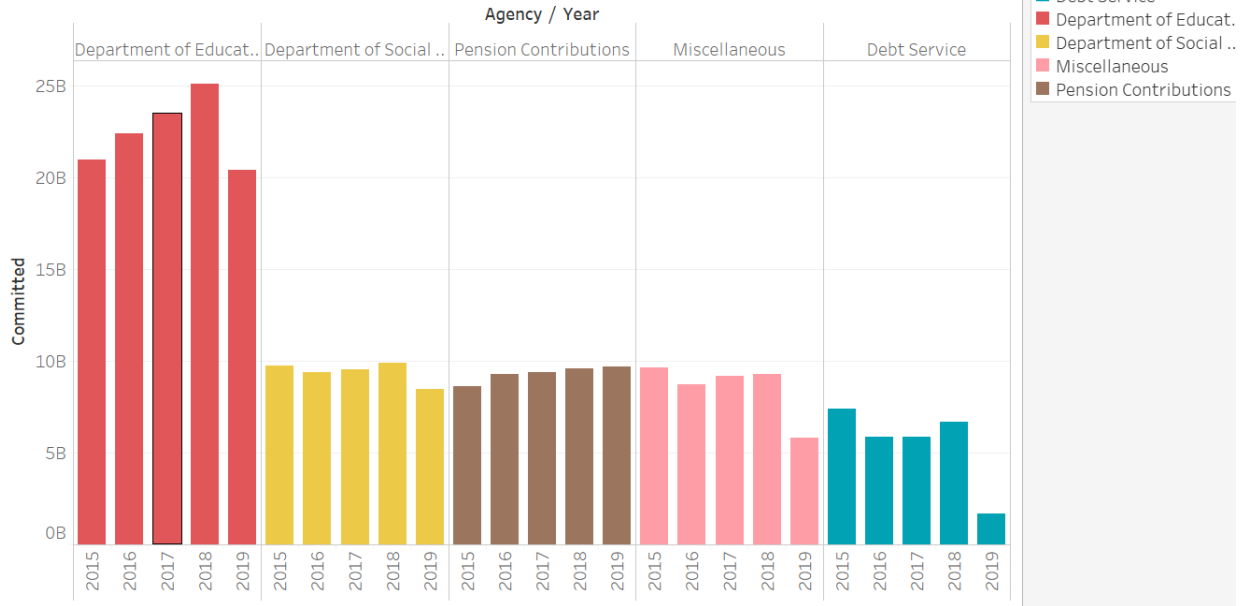
Appendix 2 - Data Visualization

Q1 : What are the Adopted, Committed and Modified Budget during the last five years and how the trends can be used to forecast estimates for 2020



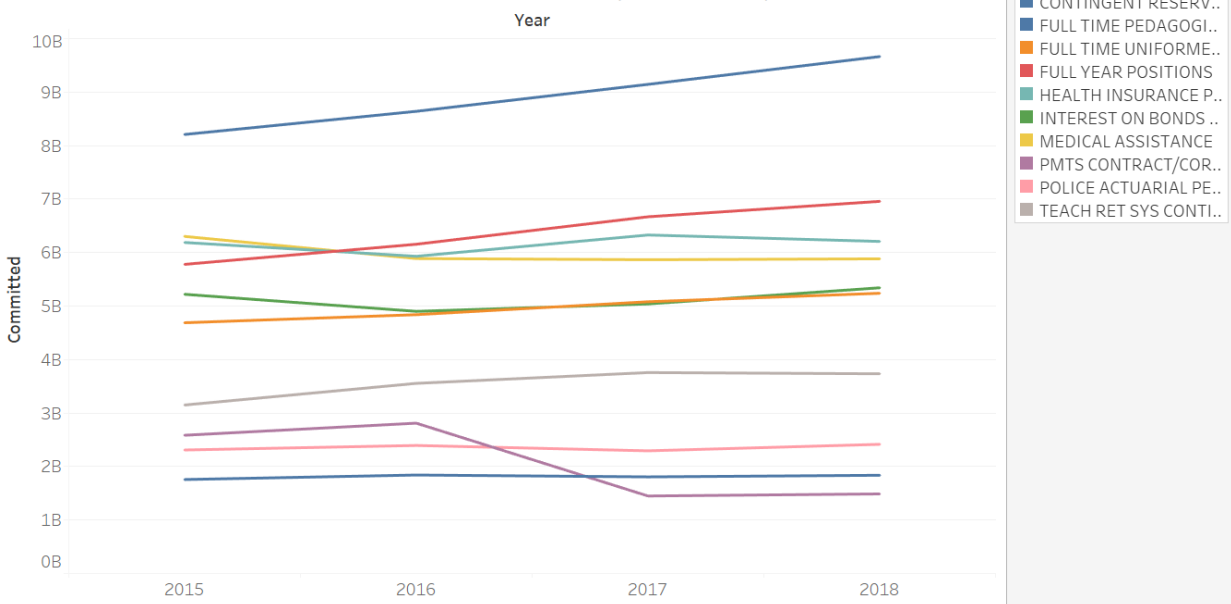
Q2. Which are the Agencies with maximum Committed budget? Is there a trend?

Top 5 Agencies - Committed (2015-2019)

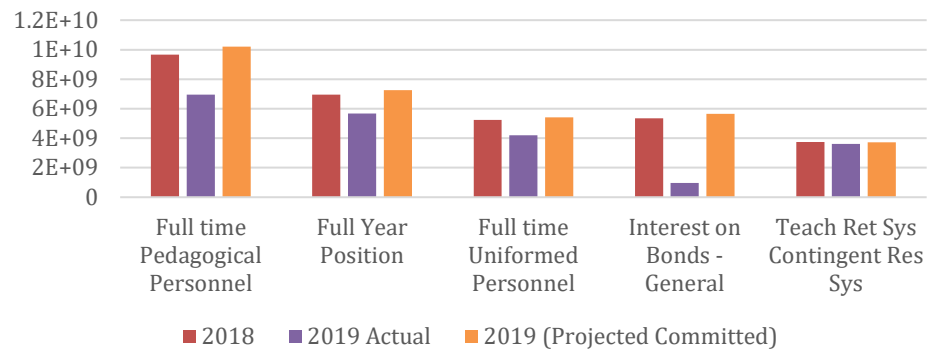


Q3 : Which Expense Categories use maximum Committed budget. What is the trend for last 5 years? Can we do a predictive analysis for 2019?

Trend - Top 10 Expense Categories - Committed (2015-2018)



Trend and Forecast - Top 5 Expense Categories - Committed (2018 - 2019)



Trend Analysis used to Predict Committed Budget of Top 5 Expense Categories the year 2019

Q 4: Which Agencies have maximum increase in Committed budget compared to their Adopted Budget? Is there a trend which suggests Diagnostic Analysis?

Top 5 Agencies - Committed exceeds Adopted (2015-2018)

