Investigation of the Police Budgets Within the Toronto Community*

Jing Li

April 27, 2022

Contents

---

1 Abstract

This paper explores the police budgets within the Toronto community and how the budgets within the police departments are distributed. In this paper, graphs including tables and histograms will be used to visualize the data, while multiple linear regression will be used to investigate the influence of different uses of police budgets as well as give predictions for expected budgets. These findings will provide insights for the police departments on the planning of the allocation of their budgets.

Keywords:Police, Budget, Visualizations, Histograms, Multiple Linear Regression
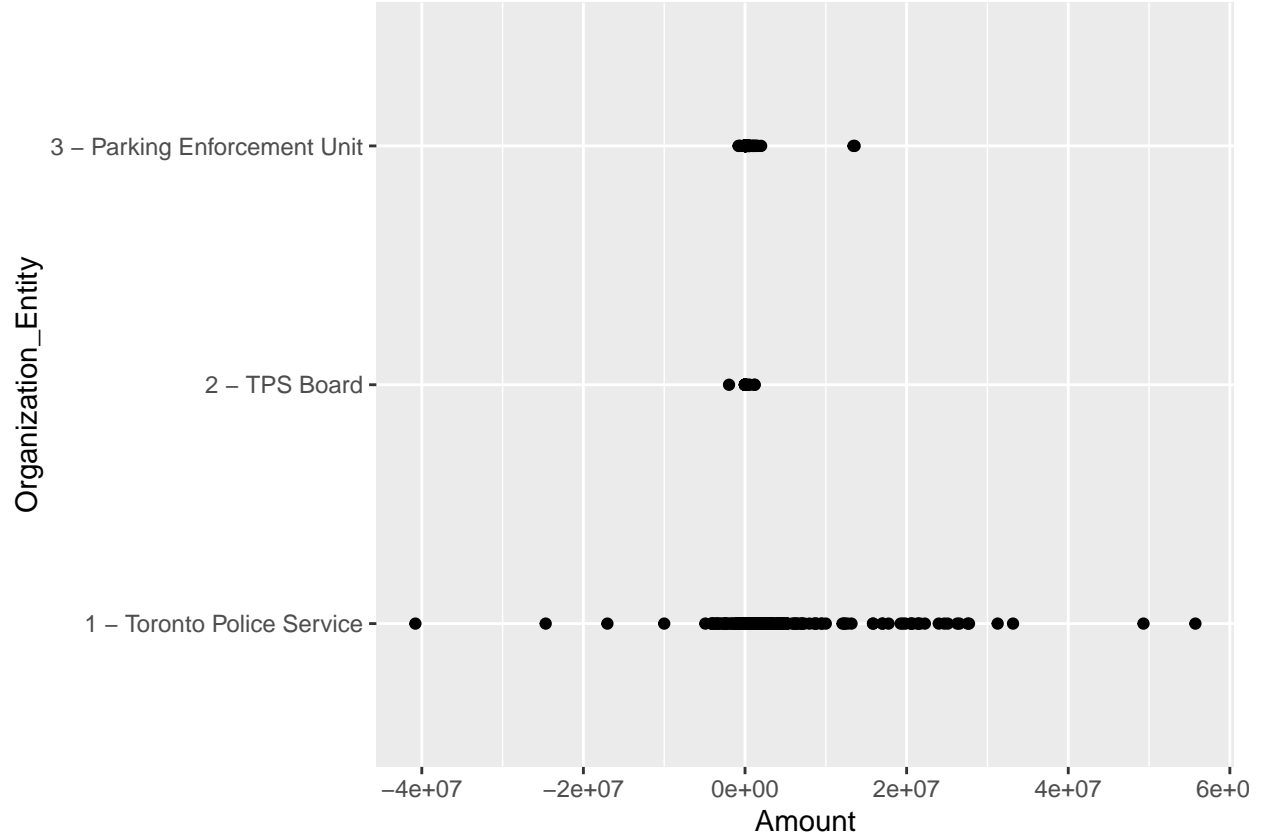
2 Introduction

Every year, police departments will be given budgets for the allocation of police uses around the Toronto community. It is up to workers within the police department for deciding on the budget purposes, however, careful planning would be vital in order for the budget to maximize its purpose. It is important for all the budget to be correctly allocated in order for police stations to run properly for the purpose of protecting Toronto citizens. In this paper, we would be investigating the influencing factors of the allocation of budgets as well as build a multiple linear regression model to predict the estimated budget for each department. The data used in this paper would be from Open Data Toronto and this dataset includes a summarized breakdown of the Toronto Police Service approved budget and actual expenditures by command. The data has twelve variables and 2755 observations. In this paper, I hypothesize that the main influencing factors of police budgets would be allocated at services including equipments and materials, as well as salaries. This is a interesting topic to study because it would help contribute to the help of the community as well as provide insights for the planning of budget allocations.

3 Data

This data was obtained from the Open Data Toronto portal. It is a data on Toronto Police Service approved budget and actual expenditures. The data was collected in the 2021 and was published by Toronto Police Services.The data includes 2,755 observations of 12 variables.

3.1 Variables in the Data

The data collected by the Toronto Police Service includes many variables, these variables include Fiscal_Year, Budget_Type, Organization_Entity, Command_Name, Pillar_Name, District_Name, Unit_Name, Feature_Category, Cost_Element , Cost_Element_Long_Name, Amount. To explore the data, an exploratory data analysis was done on the data.

The above plot displays the distribution of budget expenditures. As we can see, most of the budget is used for salaries and services, with benefits and materials and supplies following behind. This seems resonable as we know that salaries and services are usually the top expenses in any industry.

The above plot is a scatter plot of the relationship between the budget amount and organization entity. As we can see from the plot, most of the budget is given for the Toronto Police Service.

3.2 Methods

In this study we will be using multiple linear regression. Multiple linear regression because we have a continuous response variable and more than one predictors.

The multiple linear regression model is expressed as

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 ... + \hat{\beta}_n x_n$$

Where:
y = the predicted value of the dependent variable
B0 = the y-intercept (value of y when all other parameters are set to 0)
B1X1 = the regression coefficient (B_1) of the first independent variable (X1) (the effect that increasing the
= do the same for however many independent variables you are testing
BnXn = the regression coefficient of the last independent variable
e = model error (a.k.a. how much variation there is in our estimate of y)

4 Results

We performed multiple linear regression on four variables, which are whether it is by command, the organization entity, and whether it is revenue of salary, and the feature salary, and below we have our equation.

Table 1: Coefficients of Regression

	Model 1
(Intercept)	461 880.660
	(198 199.561)
is_command	118 590.829
	(112 987.133)
organization_num	−4663.404
	(123 446.891)
rev_or_sal	2 882 793.173
	(235 822.807)
Feature_CategoryEquipment	−401 584.043
	(385 700.313)
Feature_CategoryMaterials & Supplies	−437 204.835
	(198 495.342)
Feature_CategoryPremium Pay	−309 342.804
	(206 948.526)
Feature_CategoryRevenues	−3 163 292.946
	(389 221.838)
Feature_CategorySalaries	−357 846.198
	(209 512.692)
Feature_CategoryServices	−388 355.574
	(154 753.092)
Num.Obs.	2751
R2	0.108
R2 Adj.	0.105
AIC	89 428.8
BIC	89 494.0
Log.Lik.	−44 703.417
F	36.939
RMSE	2 765 580.51

$$\hat{Y_i} = \hat{\beta}_0 + \hat{\beta}_1 IsCommand + \hat{\beta}_2 OrganizationNum + \hat{\beta}_3 RevOrSal + \hat{\beta}_4 FeatureCategory$$

In this model, we can see that the overall model is significant since it has a p-value of less than 0.05, this means that we can use these results to intepret our model. Furthermore, we see that the variables "rev_or_sal", "Feature_CategoryMaterials & Supplies", "Feature_CategoryRevenues", and "Feature_CategoryServices" are significant, hence we say that the overall police budget can be explained by revenue and salary, that is, we would to see an increase in the budget by 2882793 if the amount is specified as salaries. Furthermore, if the budget is in the category of materials & supplies, sevenues, and services, we would see a decrease in the overall budget.

5 Discussion

5.1 Conclusion

In conclusion, in this paper, we studied the main influencing factors of police budgets and we found that the variables "rev_or_sal", "Feature_CategoryMaterials & Supplies", "Feature_CategoryRevenues", and "Feature_CategoryServices" have significant impact on the budget allocation of the police budgets.

5.2 weaknesses

In this paper, there are limitations, one limitation is that we did not consider model assumptions, thus, if model assumptions does not satisfy, we may not have a very accurate model. Furthermore, we are limited to only the data of one year, which would be difficut for us to make an overall conclusion on the allocation of budget for the police stations.

5.3 References

Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2022). dplyr: A Grammar of Data Manipulation. https://dplyr.tidyverse.org, https://github.com/tidyverse/dplyr.

R Core Team. 2020. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Sharla Gelfand (2022). opendatatoronto: Access the City of Toronto Open Data Portal. https://sharlagelfand.github.io/opendatatoronto/, https://github.com/sharlagelfand/opendatatoronto/.

Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686