Analysis of Traffic Offences Across Toronto Neighborhoods*

The effect of neighbourhoods incomes

Gadiel David Flores

September 23, 2024

In this paper, we will be looking at police annual reports of tickets isssued and analayzing the frequency in tickets issued based on different neighbourhoods. We will focus on which neighbourhoods exhibit the highest number of specific traffic offences (e.g., Aggressive Driving, Speeding). and are there observable geographical patterns or clusters of these offences based on neighbourhood? Additionally, how do the averages relate to the overall frequency of offences across different neighbourhoods?

1 Introduction

Higher police presence could lead to higher traffic violations. It is probable that lower income neighborhoods have a higher police presence and would lead to higher citations overall. Another pattern that may be noticed is more dangerous roads that are prone for human error that leads to more violations given and this could be a potential exploit by officers as they may know this information beforehand. We use R Core Team (2023) and Toronto (2024).

The remainder of this paper is structured as follows. Section 2 will explore our data from Toronto (2024) and show any potential clusters and patterns that may be in the number of offences given, locations, and type of citation and we will use this to further analyze in Section 3. Lastly, we will discuss our findings in **?@sec-result**.

*Code and data are available at: https://open.toronto.ca/dataset/police-annual-statistical-report-tickets-issued/

2 Data

In this section we will be specifically looking at the names of neighbourhoods, and how many citations they got from years [TODO: find the minimum year and the max year]. The data of the frequency of each citation can be seen (Figure 1), and was extract from Toronto (2024).

```
here() starts at C:/Users/david/Documents/Traffic_Offenses_Geography_Toronto
```

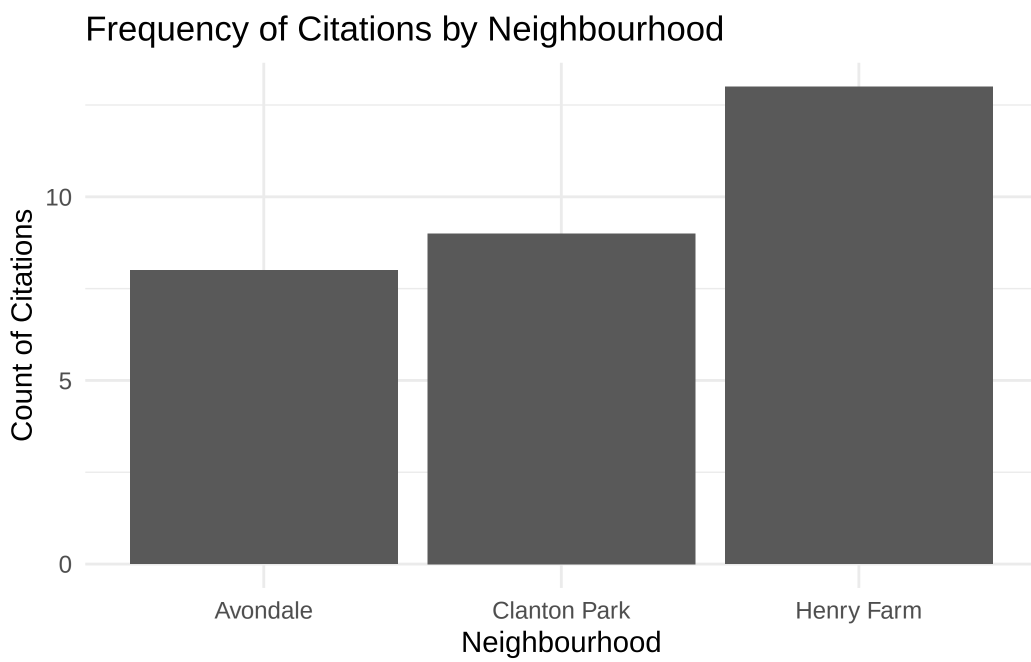


Figure 1: Frequency of Citations by Neighbourhood

As it can be seen in this graph, [TODO: Talk about any patterns or observations from the real dataset]. In (Figure 2) we can see which neighborhoods have a higher income and which ones are lower income.This data came from the latest census of toronto and the only variables taken out from this data was the neighbourhoods and their average income. This data comes from [TODO: add reference to toronto data].

This data shows actual numbers of citations given out in each neighbourhood and the average income of each neighbourhood of each household comebined. Based on these graphs, we can see [TODO: Talk about any patterns or observations from the real dataset]. This raises a question on whether there is a correlation on income and the amount of citations that are given out. In the next seciton we will come with a model that can help us find a correlation.

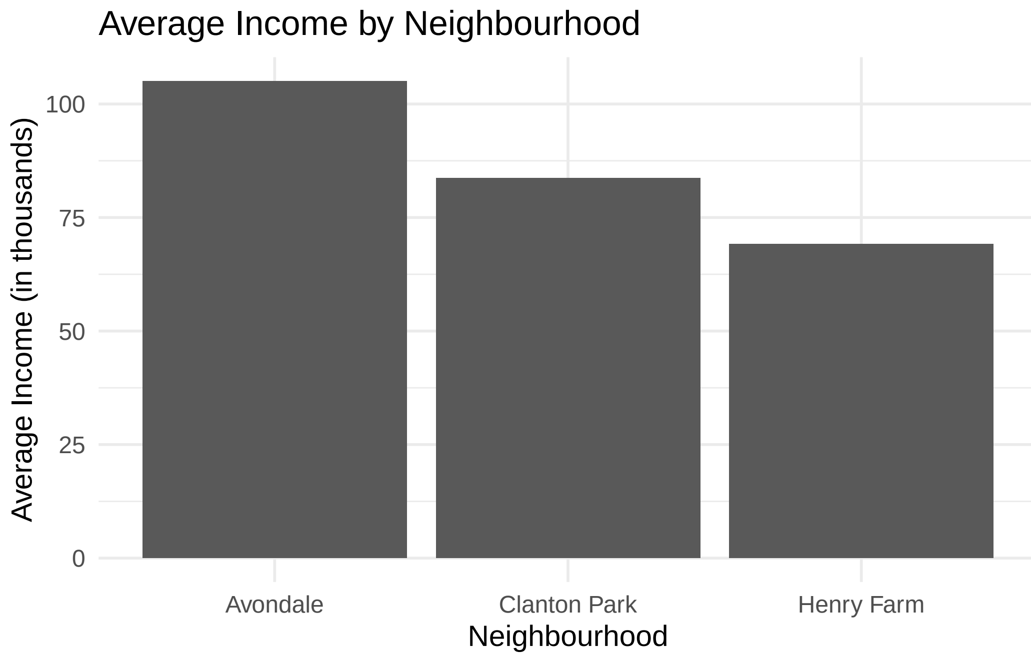


Figure 2: Toronto Neighbourhoods income

3 Model

The goal of our modelling strategy is twofold. Firstly, is there a relationship between the income and how much offences they get. Secondly, how strong is this relationship. We will be using Simple Linear Regression model to see the effect of income and the number of citations given in that neighborhood. Background details and diagnostics are included in Appendix B. ## Model set-up

Define y_i as the number offences done by a certain neighborhood. Then x_i is the average income of a certain neighborhood measured in thousands.

$$E(Y|X)Y_i = \beta_0 + \beta_1(X) \tag{1}$$

3.0.1 Model justification

The reason a linear regression model was specifically chosen was because we want to analyze the effect of an increase of citation, does it mean that the income of the neighborhood goes down or does it go up? This will allow us to see how strong of a relationship there is between citations and average neighbourhood income.

4 Results

The linear regression model was applied to analyze the relationship between neighborhood income and the number of traffic offences. The model's equation is as follows:

[E(Y|X) = _0 + _1 average_income]

where (Y) is the number of tickets issued, and (X) is the average income in thousands. The summary of the linear regression model is shown below:(**ressidual_plots?**).

We expect a positive relationship between the number offences given and the lower the income of the neighborhood. In particular…

Our results are summarized in (**ressidual_plots?**). Residual vs. Predictors, Residuals vs. fitted values and Normal quantile-quantile are all used to verify that our 4 assumptions of linear regression models aren't violation and that the data is works as intended. We can see that Linearity is [TODO: check if linarity is violated]. Uncorrelated errors is [TODO: check if uncorralted isnt violated], and constant error is [TODO: check if constant error isnt varied], lastly nromality of errors [TODO: check if normal]. Give all these assumptions, we can [TODO: check if we can trust results of linear regression]

5 Discussion

5.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

5.2 Second discussion point

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Additional data details

B Model details

B.1 Linear Regression

In **?@fig-ppcheckandposteriorvsprior-1** we implement a p
Simple Linear Regression Model. This shows...

Examining how the model fits, and is affected by, the data

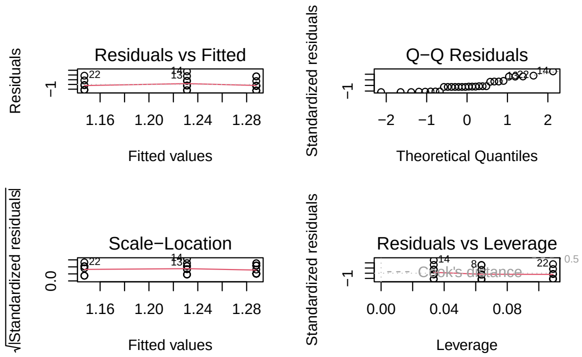
```
Call:
lm(formula = tickets_issued ~ average_income, data = merged_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.2879 -0.9323 -0.2595  0.7547  2.7689 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     1.559434   1.185444   1.315    0.199
average_income -0.003925   0.014053  -0.279    0.782

Residual standard error: 1.122 on 28 degrees of freedom
Multiple R-squared:  0.002778,  Adjusted R-squared:  -0.03284 
F-statistic: 0.078 on 1 and 28 DF,  p-value: 0.7821
```

B.2 Diagnostics

?@fig-ResidualDiagnostics-1 is a Residual vs. Fitted Values Plot. It shows... This suggests...

?@fig-ResidualDiagnostics-2 is a Normal Q-Q Plot. It shows... This suggests...

?@fig-ResidualDiagnostics-3 is a . Residuals vs. Predictors Plot. It shows... This suggests...



(a) Residuals vs. Fitted Values

Figure 3: Diagnostic plots to assess residuals.

References

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Toronto, Open Data. 2024. "Police Annual Statistical Report - Tickets Issued." https://open.toronto.ca/dataset/police-annual-statistical-report-tickets-issued/.