Analyzing the Impact of Population and Car Crashes on Noise Complaints in NYC

Rameasa Arna

2023-02-28

# Preview

This study uses a linear regression model to examine the relationship between the number of vehicle-related noise complaints in different zip codes and two variables: population and crash count. The dataset used contains information on 181 zip codes, including their population,crash\_count,injured\_count, num\_of\_complaints\_noise\_vehicle, killed\_count and injured\_ped\_count. The model’s findings demonstrate that the number of complaints is statistically influenced by the population. In particular, the model discovered a positive correlation between population and noise complaints of vehicle, whereas a negative correlation between crash count and complaints was revealed. However, the corrected R-squared value of 0.09137 indicating that approximately 9% of the variance in the number of noise complaints related to vehicles can be explained by population and crash counts. This model can still be helpful in identifying zip codes that are more likely to receive vehicle-related complaints based on their population and crash frequency. Such information could be used to inform targeted interventions to address this issue in affected areas.

# Background

This linear regression model is relevant to local governments and policymakers who are interested in understanding the factors that contribute noise complaints related to vehicles in their jurisdictions. By identifying the factors that contribute to higher complaint rates, this model can inform targeted interventions to address this issue. For example, local governments could use the model’s findings to prioritize resources and implement measures to reduce noise and improve road safety in zip codes with a high population and crash count. Additionally, residents and community organizations could use the model’s predictions to advocate for increased attention and resources towards addressing noise and vehicle-related issues in their neighborhoods. Overall, this model provides a valuable tool for stakeholders who are working to improve the quality of life in their communities.

# Data

The analysis uses three datasets to build the model. The first one is NYPD 311 Service Requests from 2010 to Present provided by 311, DoITT and and distributed by [NYC Open Data](https://data.cityofnewyork.us/Social-Services/311-Service-Requests-from-2010-to-Present/erm2-nwe9). This dataset contains information about service requests made to the city’s 311 service for various issues, including noise complaints related to vehicles . The dataset was filtered to include only noise complaints related to vehicles and was aggregated by zip code to count the number of complaints per zip code.

The second dataset used Motor Vehicle Collisions - Crashes provided by the New York City Police Department (NYPD) and and distributed by [NYC Open Data](https://data.cityofnewyork.us/Public-Safety/Motor-Vehicle-Collisions-Crashes/h9gi-nx95). This dataset contains information about motor vehicle collisions that have been reported to the New York City Police Department. The dataset was aggregated by zip code to count the number of collisions, as well as the number of people injured or killed, for each zip code. Both of the dataset has included information from 2019 through 2022.

The third dataset used was New York Zip Codes by Population, which was obtained from (<https://www.newyork-demographics.com/zip_codes_by_population>). This dataset contains information about the population of each zip code in New York City.

All three datasets were merged together based on the zip code field to create a new dataset that contained information about the population , number of noise complaints related to vehicles, and number of motor vehicle collisions, number of people killed, number of pedestrian killed and number of people injured. for each zip code. The merged dataset was then used for further analysis to investigate the relationship between these variables amd to build a model for estimating number of noise complaints based on population and crash\_count , as follow:

num\_of\_complaints\_noise\_vehicle = b0 + b1\*population + b2\*crash\_count

OLS stands for Ordinary Least Squares, and it is a statistical method used for modeling the relationship between a dependent variable and one or more independent variables. In an OLS model, the goal is to find the line of best fit that minimizes the sum of the squared errors between the predicted values and the actual values.The OLS model provides a way to estimate the coefficients (slopes) of the regression line, which describe the relationship between the dependent variable and each independent variable. These coefficients can be interpreted as the expected change in the dependent variable for a one-unit increase in the corresponding independent variable, holding all other independent variables constant. In this study, multiple linear regression used to build the model.Multiple linear regression is used to estimate the relationship between two or more independent variables and one dependent variable.The “b” values are called the regression weights and they measure the association between the predictor variable and the outcome. b1 can be interpreted as the average effect on “num\_of\_complaints\_noise\_vehicle” of a one unit increase in “population”, holding all other predictors fixed and same for the crash\_count.

# Descriptive Statistics

The variables used in the model include population, num\_of\_complaints\_noise\_vehicle, crash\_count. The two predictor variables: population and crash\_count. Population represents the population of the zip code area, while crash\_count represents the number of car accidents reported in the zip code area. The response variable is the number of noise complaints related to vehicles reported in the zip code area. Let’s see the statistics of these variables.

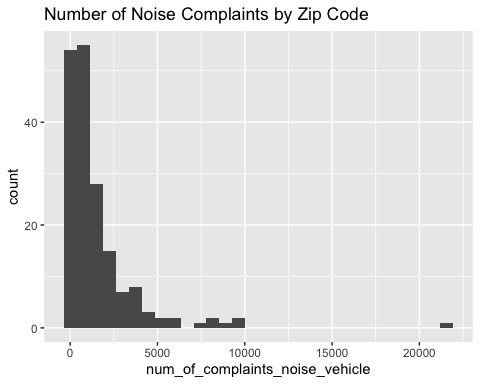
## 1. Response Variable :

The response variable is Number of noise complain related to vehicle. The variable represents the number of noise complaints related to vehicles by zipcode since 2019 till now. In this part, we will describe the variable to understand what the measurements mean, and how to interpret their distribution.

summary(merged\_data$num\_of\_complaints\_noise\_vehicle)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1 290 800 1543 1779 21565

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



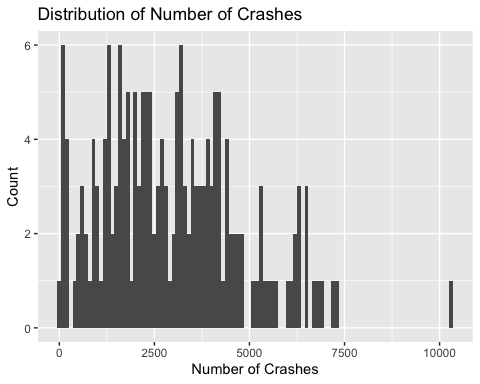
The “num\_of\_complaints\_noise\_vehicle” variable represents the number of noise complaints related to vehicles by zipcode since 2019 till now. The distribution of this variable ranges from 1 to 21565 with a mean of 1543, indicating that the majority of zip codes had a relatively low number of noise complaints related to vehicles. The first quartile (Q1) value of 290 and the third quartile (Q3) value of 1779 indicate that 25% of zip codes had 290 or fewer noise complaints related to vehicles, while 75% of zip codes had 1779 or fewer noise complaints related to vehicles. Moreover, The histogram shows that the majority of the complaints fall within the range of 0 to 5000, with very few complaints above that range.

## 2. predictor variables :

One of the predictor varibles is “crush\_count”. The variable is a measure of the number of crashes that occurred in a particular zip code during the study period.In this section, we will disscuss about what the measurements mean, and how to interpret their distribution.

summary(merged\_data$crash\_count)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 48 1614 2742 2969 4109 10292



The variable “crash\_count” is a measure of the number of crashes that occurred in a particular zip code during the study period. The descriptive statistics reveal that the minimum number of crashes in a zip code is 48, while the maximum is 10292. The mean number of crashes is 2969, which is higher than the median of 2742. This suggests that the distribution is skewed to the right, with some zip codes having a much higher number of crashes than the rest of the dataset. The first quartile (Q1) value of 1614 and the third quartile (Q3) value of 4109 indicate that 25% of zip codes had 1614 or fewer crashes, while 75% of zip codes had 4109 or fewer crashes. Overall, these statistics suggest that the number of crashes varies widely across zip codes, and that there are some areas with a higher incidence of crashes than others.

## 3. predictor variables :

Other predictor variable is Population. It represents the total population in a particular zip code. Now we will look at their distribution.

# Descriptive statistics  
summary(merged\_data$population)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 222 27575 43849 48522 67252 116469

hist(merged\_data$population) Chart, histogram

Description automatically generated

The distribution of this variable ranges from 222 to 116469 with a mean of 48522, indicating that the population sizes of zip codes in the dataset vary widely. The first quartile (Q1) value of 27575 and the third quartile (Q3) value of 67252 indicate that 25% of zip codes had a population size of 27575 or lower, while 75% of zip codes had a population size of 67252 or lower.This information can be useful in identifying any relationships between population size and the number of noise complaints related to vehicles in a particular zip code.

# Fit the multiple linear regression model

model1 <- lm(num\_of\_complaints\_noise\_vehicle ~ population + crash\_count, data = merged\_data)

summary(model1)

##   
## Call:  
## lm(formula = num\_of\_complaints\_noise\_vehicle ~ population + crash\_count,   
## data = merged\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2410.0 -838.3 -483.7 91.2 20957.4   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 257.22100 349.83237 0.735 0.463143   
## population 0.03009 0.00865 3.479 0.000632 \*\*\*  
## crash\_count -0.05877 0.12769 -0.460 0.645926   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2231 on 178 degrees of freedom  
## Multiple R-squared: 0.1015, Adjusted R-squared: 0.09137   
## F-statistic: 10.05 on 2 and 178 DF, p-value: 7.32e-05

# Model

The regression analysis revealed that population and crash count had different effects on the number of noise complaints related to vehicles.

Population had a statistically significant effect on the number of noise complaints related to vehicles, with a coefficient estimate of 0.03009 and a p-value of 0.000632. This means that for every one-unit increase in population, there was a 0.03 increase in noise complaints, holding the crash count constant.

In contrast, crash count did not have a statistically significant effect on the number of noise complaints related to vehicles, with a coefficient estimate of -0.05877 and a p-value of 0.645926. This indicates that changes in crash count were not associated with changes in the number of noise complaints related to vehicles.

The final equation of the regression model is:

num\_of\_complaints\_noise\_vehicle = 257.22100 + 0.03009 \* population - 0.05877 \* crash\_count

Where the intercept is 257.22100 and the coefficients for population and crash count are 0.03009 and -0.05877, respectively. This equation can be used to predict the expected number of noise complaints related to vehicles based on population and crash count values. However, it is important to keep in mind that the model may not capture all the factors that influence the number of noise complaints and should be used with caution.

The adjusted R-squared value of the model was 0.09137, indicating that approximately 9% of the variation in the number of noise complaints could be explained by population and crash count. The multiple R-squared value of the model was 0.1015, indicating that the model explained only a small proportion of the variance in the number of noise complaints.

Overall, the results suggest that population is a more important factor to consider when predicting the number of noise complaints related to vehicles. However, it is important to note that the model has some limitations and may not capture all the factors that influence the number of noise complaints.

# Conclusions

Our regression model provided valuable insights into the factors influencing the number of noise complaints related to vehicles. The results of our analysis showed that population had a statistically significant effect on the number of noise complaints related to vehicles, while crash count did not.

The finding that population is a significant predictor of noise complaints suggests that noise-reducing interventions targeted at densely populated neighborhoods may be effective in mitigating noise pollution. Policymakers and city officials can use this information to prioritize noise mitigation efforts in areas with high population density.

However, our model’s failure to find a significant relationship between crash count and noise complaints indicates that focusing solely on crash-prone areas may not necessarily lead to reductions in noise pollution. Therefore, a more comprehensive approach that considers other factors that contribute to noise pollution, such as traffic volume, road design, and vehicle type, may be necessary.

In conclusion, our findings can inform noise mitigation efforts and help policymakers and city officials make data-driven decisions to improve residents’ quality of life. However, it is important to note that our model has some limitations and should be used with caution. Future studies can build on our findings by considering additional factors that contribute to noise pollution and by collecting more detailed data on noise complaints and their causes.