UNDERSTANDING FREQUENCY OF MOTOR ACCIDENTS

JONATHAN NGAI

UC BERKELEY

EXPLORER TRANSPORT DATA SCIENCE PROJECT

USING TIME SERIES AND POISSON PROCESS TO ANALYZE MOTOR ACCIDENT TRENDS AND FREQUENCY

INTRODUCTION

The subject of this project is better understanding the frequency of motor accidents across boroughs in New York City. I wanted to extend existing analysis regarding progression of motor accident counts in New York City to do a more granular analysis of those trends by borough. I also wanted to use a model to predict the frequency of motor accidents by borough.

Thus my research questions are:

- Which borough has the most motor accidents vs the least?
- Are trends in motor accident counts across boroughs similar?
- What's the expected number of motor accidents by borough?

OBJECTIVE

- Understand trends in frequency of motor accidents by borough across the yearrs
- Predict the number of motor in NYC in 2025 by borough

METHODOLOGY

I used a time series decomposition to first gain insight on differences in trends regarding motor accident counts by borough. Then I constructed a poisson distribution and accompanying normal approximation to model number of crashes per second.

To verify whether or not to use the poisson process[1], I calculated the gap in seconds between each observation for entries from 2024 only (the most comparable year) then plot the gap to verify that the density distribution looked exponential.[2]

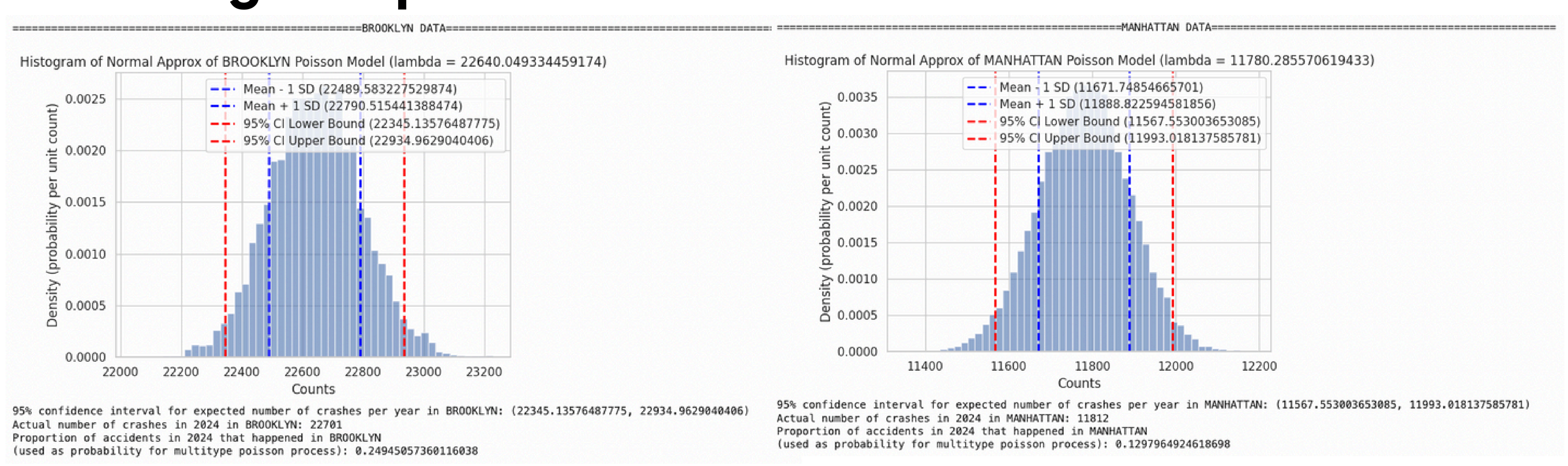
The poisson lamda parameter for the poisson distribution modeling crashes in New York City as a whole was constructed by calculating 1 / (mean gap sizes in seconds).

The lambda parameters for each borough were constructed via a multi-type poisson process with borough being encoded as a multionmial variable and the 'chance' of each borough 'appearing' with each observation encoded as the proportion of observations in 2024 that belonged to said borough.[3]

The figure for 'expected' number of motor accidents taken from the expectation of all those poisson distributions.

ANALYSIS

Fig. 1 Expected Number of Motor Accidents




```
MSE incorporating 'Unspecified': 1553.9692782801637
MSE not incorporating 'Unspecified': 1929.4980318620303
```

Fig. 1 to the the bottom left has the normal approximations and the 95% confidence intervals for the number of motor accidents expected in Brooklyn and Manhattan. For comparison, the actual number of motor accidents in 2024 is also listed along with the Means Squared Error (MSE). The MSE was 1554 when including 'Unspecified' boroughs and 1929 when excluding 'Unspecified' boroughs. The figure also indicates the exact probabilities used to 'simulate' picking from different boroughs.

Fig. 2 Trends in Motor Accident Counts

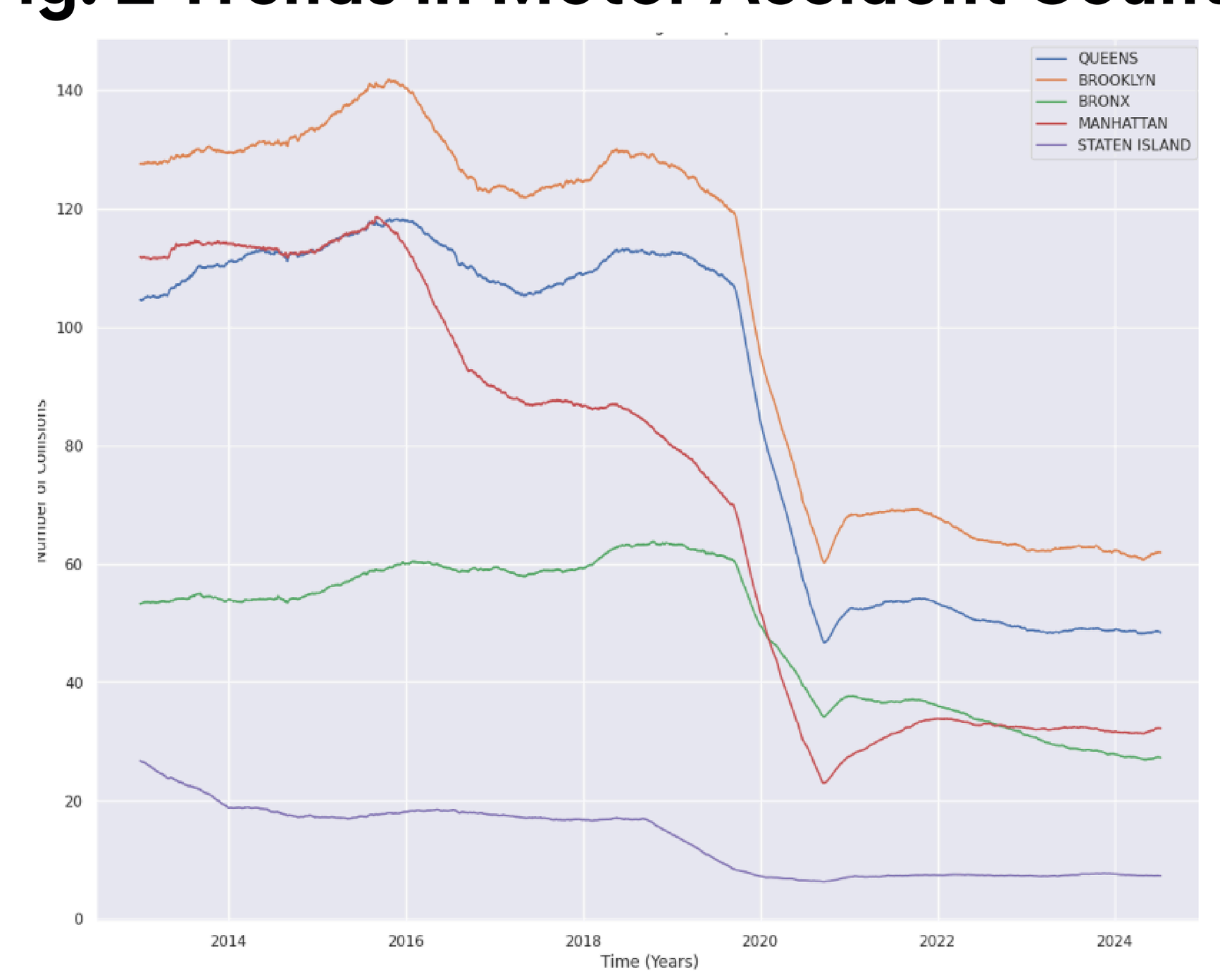


Fig. 2 shows the general trend in number of motor accidents across time. All boroughs saw a stark drop in motor accident coinciding with COVID-19 lockdowns. Some notable trends:

- Staten Island by far had the lowest count of motor accidents
- Manhatten moved from having some of the highest accident counts to some of the lowest dipping lower than the Bronx at a few points
- A sudden spike in motor accidents began in Queens, Brooklyn and Manhattan in 2015 that hit its peak in 2016 and it's trough near mid-2017. This trend was entirely absent in the other two boroughs
- All boroughs experienced a sharp v-shaped trough a bit before 2021 except for Staten Island which maintained flat since a quarter in 2020
- Trends largely stabilize past 2022, making those years more valid to base predictions for 2025 on with the Bronx being the exception seemingly embraking on a continuing downward trend

Fig. 3 Expected Number of Motor Accidents

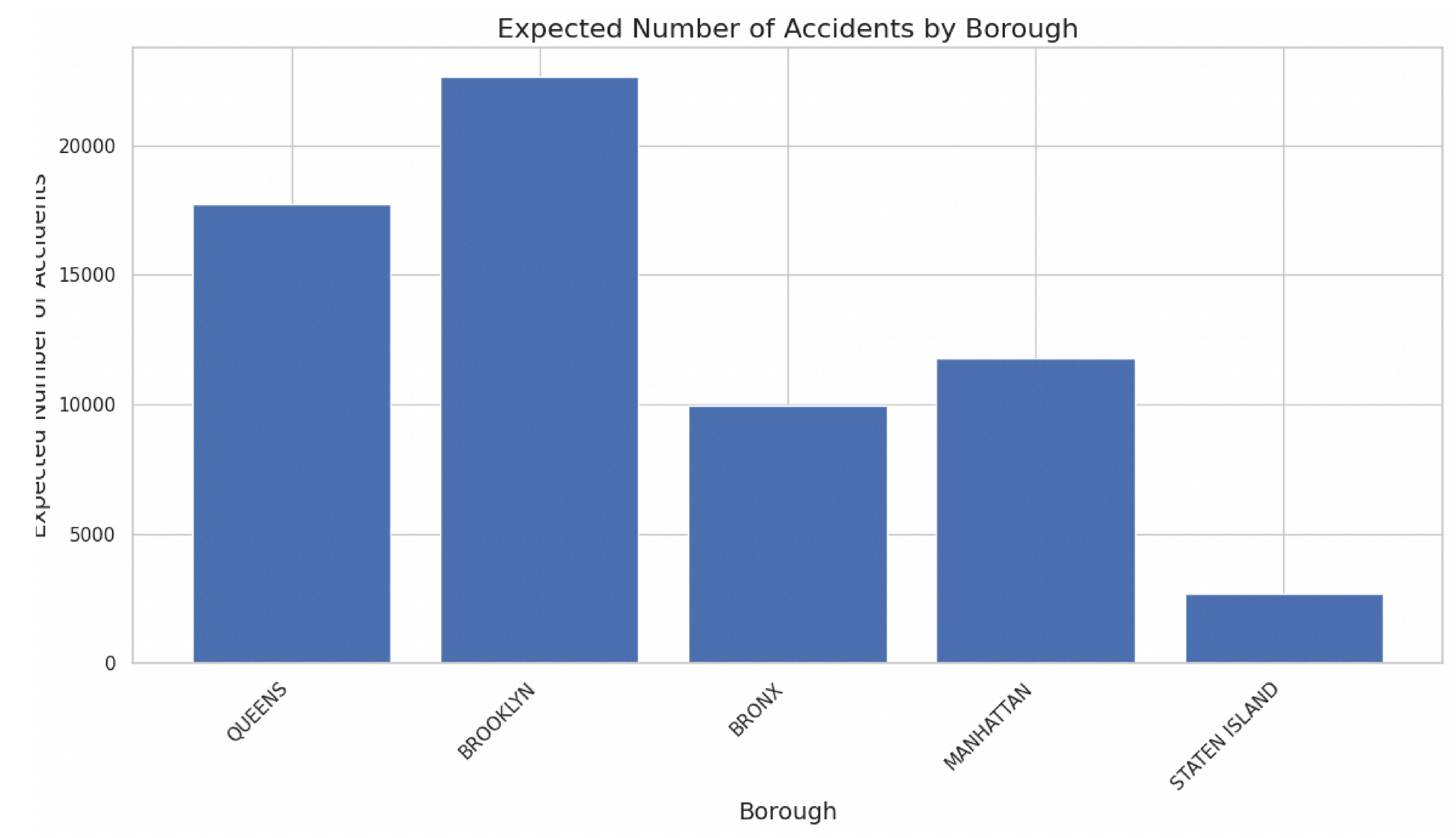


Fig. 3 to the center bottom compares the exact expected number of motor accidents to occur by borough. Three rough 'groups' emerge based on this chart. Queens and Brooklyn having the highest motor accident counts at around a bit over 18k and 23k respectively. The Bronx and Manhattan are next with roughly 10k and slightly under 12k expected motor accidents. Finally Staten Island trails last with just under 2.7k motor accidents expected which is less than a fourth of what's expected in Manhattan and an eighth of what's expected in Brooklyn. The total number of expected motor accidents in NYC is just under 91k. Relative to that figure, Brooklyn is expected to make up about 25% of all motor accidents in NYC, Manhattan a roughly 13% and Staten Island around 3%.

KEY FINDINGS

- Trends in motor accidents in Queens, Brooklyn and Manhattan largely follow the same pattern pre and post-pandemic whereas the Bronx and Staten Island especially are less volatile
- Manhattan has gotten 'comparatively safer' having motor accident counts closely matching Queens prior to the pandemic to now having counts closely matching the Bronx
- Staten Island by far is expected to have the lowest number of motor accidents whereas Brooklyn is expected to have the highest number of accidents

LIMITATIONS

- Data related to average or median annual income and population for each bororugh would have been helpful in contextualizing variations in crash counts across boroughs
- Visual verification of gap sizes following an exponential distribution was done without an exact algorithm for determining fair bin sizes. Either a more exact method for specifying bin sizes and measuring differences or verifying the validity of a poisson model should be used
- Due to the large lambda parameters, the poisson models were difficult to make probabilistic predictions with, this was circumvented via using a normal approximation

CONCLUSION AND KEY RECOMMENDATIONS

In conclusion trends in motor accident counts across boroughs in New York City follow a general pattern but differ upon closer analysis. The poisson process model works well for modeling predicted counts however, due to the large lambda parameter, falls short at predicting probabilities.

Policy recommendations include:

- Efforts on reducing accidents should be focused on Brooklyn
- Closer study of Manhattan to see how it as kept it's accident counts down relative to where it was pre-pandemic
- Closer study of the Bronx to see why the trend in accidents has been going down

Steps for further research:

- Plotting motor accident counts in boroughs with respect to population or average annual income
- Improving binning when verifying that gap sizes follow an exponential distribution
- Using data from years further back given stable and reasonable trends for poisson prediction
- Using more advanced models like Negative Binomial or Poisson-Lognormal[2, 3, 4]

ACKNOWLEDGEMENTS

Thank you to Northeast Big Data Innovation Hub, National Student Data Corp and the U.S. Department of Transportation and Federal Highway Administration for their services and support. Thank you to Emily Rothenberg, TDSP staff and the TDSP community for administering this project, providing support and offering their time to bounce my ideas off of. This project has been a great opportunity to work with real world data in all it's complexities and build predictive models. Special thanks to the UC Berkeley Data Science Program, in particular Data140 and Data100 staff for providing me with the statistical foundation and technical skills to employ the analysis I did.

1. http://prob140.org/textbook/content/Chapter_18/03_The_Gamma_Family.html
2. http://prob140.org/textbook/content/Chapter_15/04_Exponential_Distribution.html
3. http://prob140.org/textbook/content/Chapter_07/03_Poissonizing_the_Multinomial.html
4. https://www.sciencedirect.com/science/article/abs/pii/S0001457507001996?casa_token=1XZHKSmL36MAAAAA:PacLyHC9IAxVzr0I9_11D4ehVKsw1e-aNXjRSnVrEv3UVhXekPBnUaze3XDaqigFQ3gI_nQtwA
5. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0197338
6. https://pdfs.semanticscholar.org/937e/7178711f97841e435d051a23e5f07ffb834a.pdf