CUNY QUEENS COLLEGE

THE EFFECTS OF SEASON AND BOROUGH ON POTHOLE INCIDENCE

A QUANTITATIVE STUDY

ANDREAS KEFALAS

2023

ANDREAS KEFALAS

JUNE 2023

Introduction:

Efficient infrastructure is essential for the economic and social development of any city. In particular, road maintenance plays a crucial role in ensuring smooth transportation and reducing vehicular accidents. However, this task can be challenging, especially in urban areas like New York City, where potholes are a persistent problem. The high volume of complaints regarding potholes received by the city's 311 service necessitates a data-driven approach to identify patterns and improve road maintenance. This research project aims to utilize econometric techniques to analyze the data from the 311 service requests related to potholes in New York City. The goal is to identify the determinants of pothole formation and maintenance, including factors such as season and borough, which can aid policymakers in resource allocation decisions. The presence of potholes on city roads poses a significant problem for both drivers and pedestrians, leading to vehicular damage, accidents, and discomfort. This can result in increased costs for vehicle owners, slowed public transportation, and lower economic productivity. Hence, improving road maintenance is crucial for the city's growth and long-term sustainability. Furthermore, data-driven approaches can provide insights for policymakers, leading to more effective resource allocation decisions. This research project can aid city administrations in identifying areas that require attention, leading to improved road safety and maintenance.

Conceptual Framework

This conceptual framework aims to explore the factors affecting pothole formation and development, with a focus on seasonal and borough factors. By examining relevant economic theories and models and reviewing existing literature on the topic, this framework provides a comprehensive understanding of the underlying factors influencing potholes.

Factors Affecting Pothole Formation and Development:

1.1 Seasonal Factors:

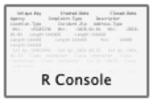
Seasonal variations, such as temperature fluctuations, freeze-thaw cycles, and precipitation, have been found to influence the formation and development of potholes (Wang et al., 2018). Cold winters with freeze-thaw cycles can cause the expansion and contraction of pavement materials, leading to cracks that eventually develop into potholes (Deng et al., 2017).

1.2 Borough Factors:

The characteristics of different boroughs, such as traffic volume, road design, and pavement materials, can contribute to variations in pothole occurrence (Wong & Li, 2019). Variations in traffic patterns and road conditions among different boroughs can influence the rate at which potholes form and the maintenance efforts required (Khasanov et al., 2018).

Existing Literature and Studies:

Existing studies have explored various aspects of pothole occurrence and its influencing factors. Wang et al. (2018) analyzed longitudinal pavement performance data to identify factors affecting pothole formation. Deng et al. (2017) developed a predictive model based on traffic and environmental data to forecast pothole formation. Wong and Li (2019) conducted a spatial-temporal analysis of pothole formation and maintenance in urban areas, considering factors such as traffic volume and pavement conditions. Khasanov et al. (2018) investigated the spatial and temporal patterns of potholes in urban road networks.




```
  Unique.Key          Created.Date         Closed.Date            Agency          Complaint.Type      Descriptor          Location.Type       Incident.Zip     Address.Type
 Min.   :45285230   Min.   :2020-01-01   Min.   :2020-01-01   Length:144468      Length:144468      Length:144468      Length:144468      Min.   :10000   Length:144468
 1st Qu.:49854992   1st Qu.:2021-02-22   1st Qu.:2021-02-23   Class :character   Class :character   Class :character   Class :character   1st Qu.:10453   Class :character
 Median :52735790   Median :2021-12-06   Median :2021-12-06   Mode  :character   Mode  :character   Mode  :character   Mode  :character   Median :11214   Mode  :character
 Mean   :52049650   Mean   :2021-10-18   Mean   :2021-10-19                                                                               Mean   :10927
 3rd Qu.:54776766   3rd Qu.:2022-07-13   3rd Qu.:2022-07-14                                                                               3rd Qu.:11367
 Max.   :57760599   Max.   :2023-05-31   Max.   :2023-05-30                                                                               Max.   :11697
                                         NA's   :2715                                                                                     NA's   :7169
     City              Status             Borough          X.Coordinate..State.Plane. Y.Coordinate..State.Plane.    Latitude        Longitude        Location
 Length:144468      Length:144468      Length:144468      Min.   : 913825            Min.   :121312             Min.   :40.50   Min.   :-74.25   Length:144468
 Class :character   Class :character   Class :character   1st Qu.: 990187            1st Qu.:175640             1st Qu.:40.65   1st Qu.:-73.98   Class :character
 Mode  :character   Mode  :character   Mode  :character   Median :1007249            Median :196896             Median :40.71   Median :-73.92   Mode  :character
                                                          Mean   :1005784            Mean   :197992             Mean   :40.71   Mean   :-73.92
                                                          3rd Qu.:1026832            3rd Qu.:217290             3rd Qu.:40.76   3rd Qu.:-73.85
                                                          Max.   :1067220            Max.   :271876             Max.   :40.91   Max.   :-73.70
                                                          NA's   :35405              NA's   :35405              NA's   :35405   NA's   :35405
   Zip.Codes     Borough.Boundaries Borough_numeric      Month            Season
 Min.   :10090   Min.   :1.0        Min.   :0.000   Min.   : 1.000   Length:144468
 1st Qu.:11609   1st Qu.:2.0        1st Qu.:2.000   1st Qu.: 3.000   Class :character
 Median :14193   Median :3.0        Median :3.000   Median : 4.000   Mode  :character
 Mean   :15115   Mean   :2.9        Mean   :3.129   Mean   : 5.246
 3rd Qu.:17215   3rd Qu.:4.0        3rd Qu.:4.000   3rd Qu.: 8.000
 Max.   :26001   Max.   :5.0        Max.   :5.000   Max.   :12.000
 NA's   :36010   NA's   :35473
```

Data Description:

The dataset used for this analysis was obtained from NYC Open Data and consists of 667,701 rows and 14 columns. The dataset comprises 311 service requests received from 2020 up to the present day. To narrow down the scope of the analysis to relevant data, the dataset was filtered to include only requests related to the Department of Transportation (DOT) matters.

Data Cleaning and Preprocessing:

To ensure data quality and reliability, several cleaning and preprocessing steps were performed on the dataset. These steps included:

Filtering: The dataset was filtered to include only service requests related to DOT matters, as these are the most relevant for analyzing pothole occurrences. This step helped refine the dataset and focus on the essential variables.

Handling Missing Values: Missing values, if any, were identified and addressed appropriately. Depending on the extent of missing data, various techniques such as imputation or removal of incomplete records were employed to mitigate the impact of missing values.

Variable Transformation: In order to create a new column for seasons, the dates in the dataset were utilized. Using R programming, a new column was added that categorizes each service request into one of the four seasons: winter, spring, summer, or autumn. This transformation enables the analysis of the seasonal impact on pothole occurrences.

Variables Used in the Analysis

Descriptor (Dependent Variable): The "Descriptor" column specifies the nature of the service request. For this analysis, the specific descriptor "Potholes" was selected as the dependent variable of interest. This variable indicates the presence of potholes reported in the service requests.

Borough (Independent Variable): The "Borough" column represents the borough in which the service request was made. This variable is used as an independent variable in the analysis to determine if there are variations in pothole occurrences across different boroughs.

Season (Independent Variable): The newly created "Season" column categorizes the service requests into one of the four seasons: winter, spring, summer, or autumn. This variable is also used as an independent variable to assess the impact of seasonal variations on the occurrence of potholes.

By considering the "Descriptor" (Potholes) as the dependent variable and utilizing the independent variables "Borough" and "Season," this analysis aims to explore the factors influencing pothole formation and maintenance in New York City, with

Season	Potholes	Percentage Allocation
Fall	20032	13.87%
Spring	52720	36.49%
Summer	27016	18.70%
Winter	44700	30.94%

a focus on the seasonal and borough factors.

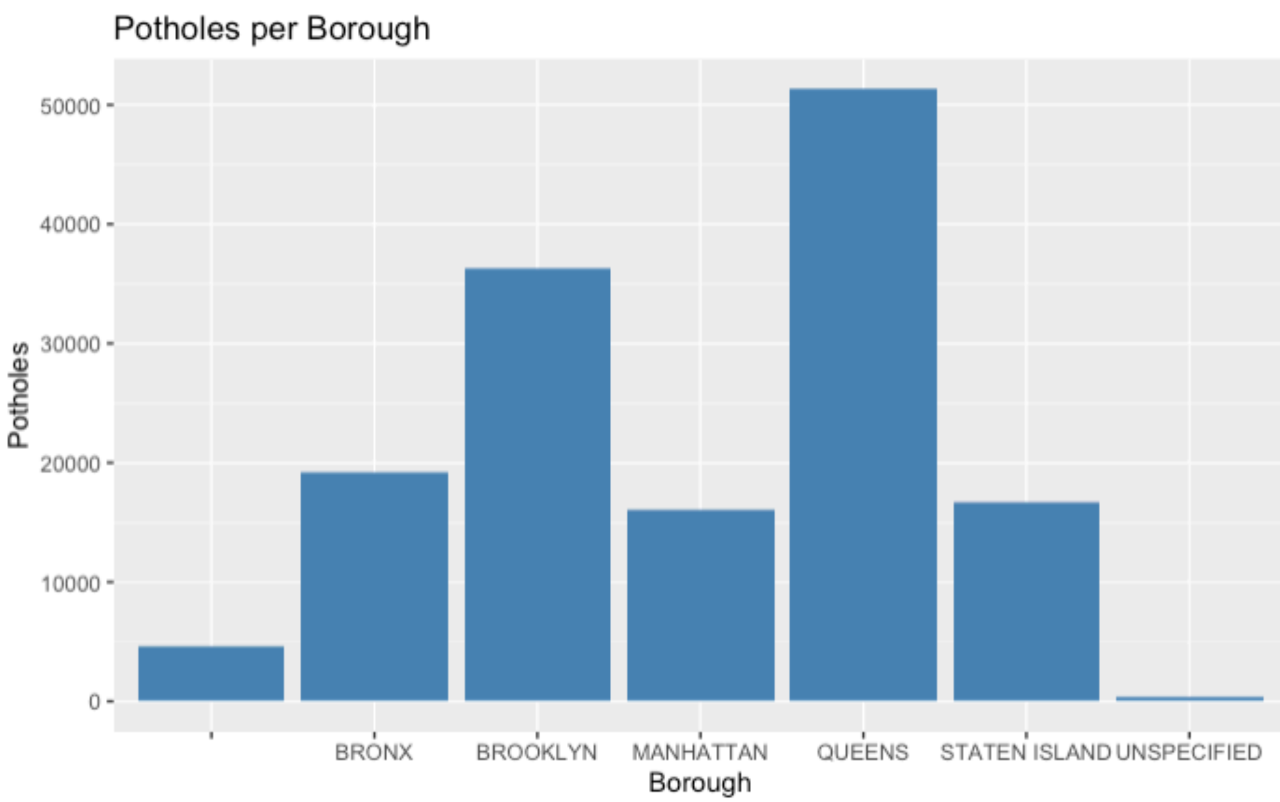
<u>Descriptive Analysis of Potholes Distribution in NYC</u>

Distribution of Potholes by Season:

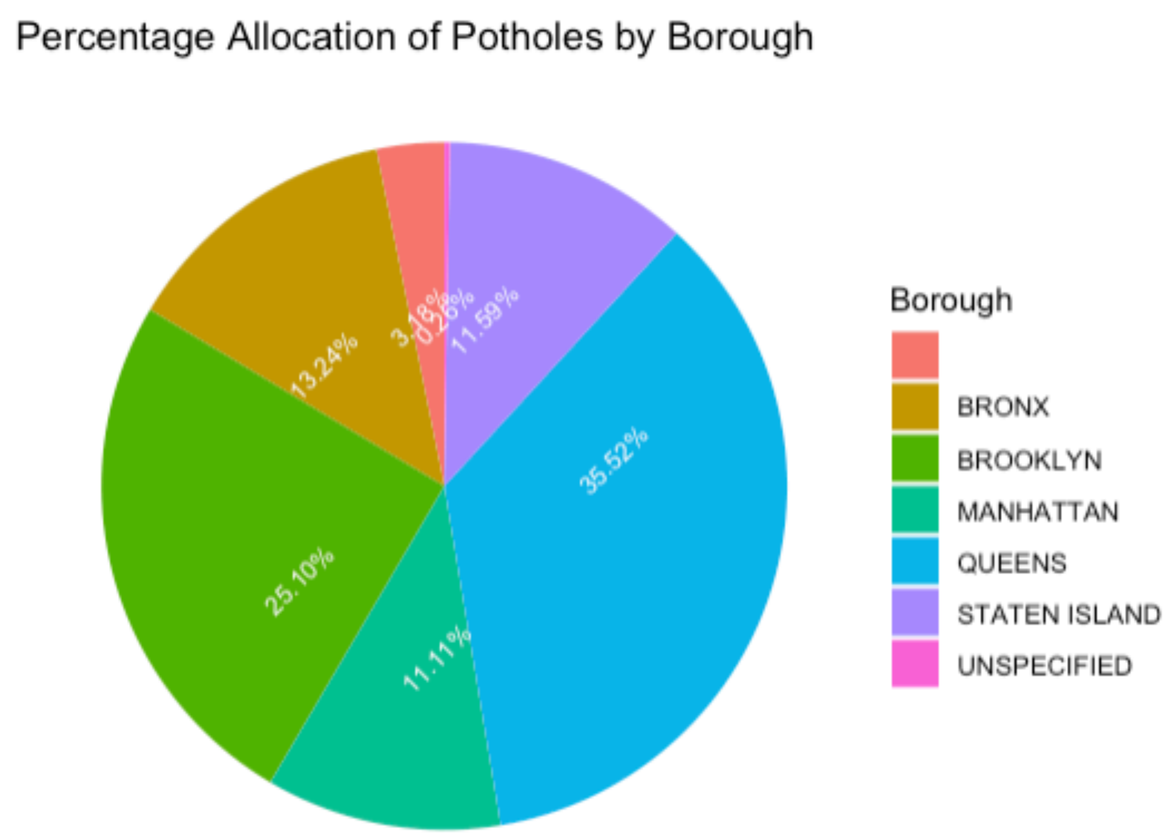
When analyzing the distribution of potholes by season, we observed significant variations in the number of reported incidents. Among the four seasons, Spring had the highest number of potholes, accounting for 36.49% of all reported cases. Winter followed closely behind with 30.94% of the potholes, while Summer and

Fall accounted for 18.70% and 13.87%, respectively. This indicates that Spring and Winter experience the highest frequency of pothole occurrences, possibly due to weather conditions or road maintenance activities during those seasons.

Exploring Potholes by Borough:

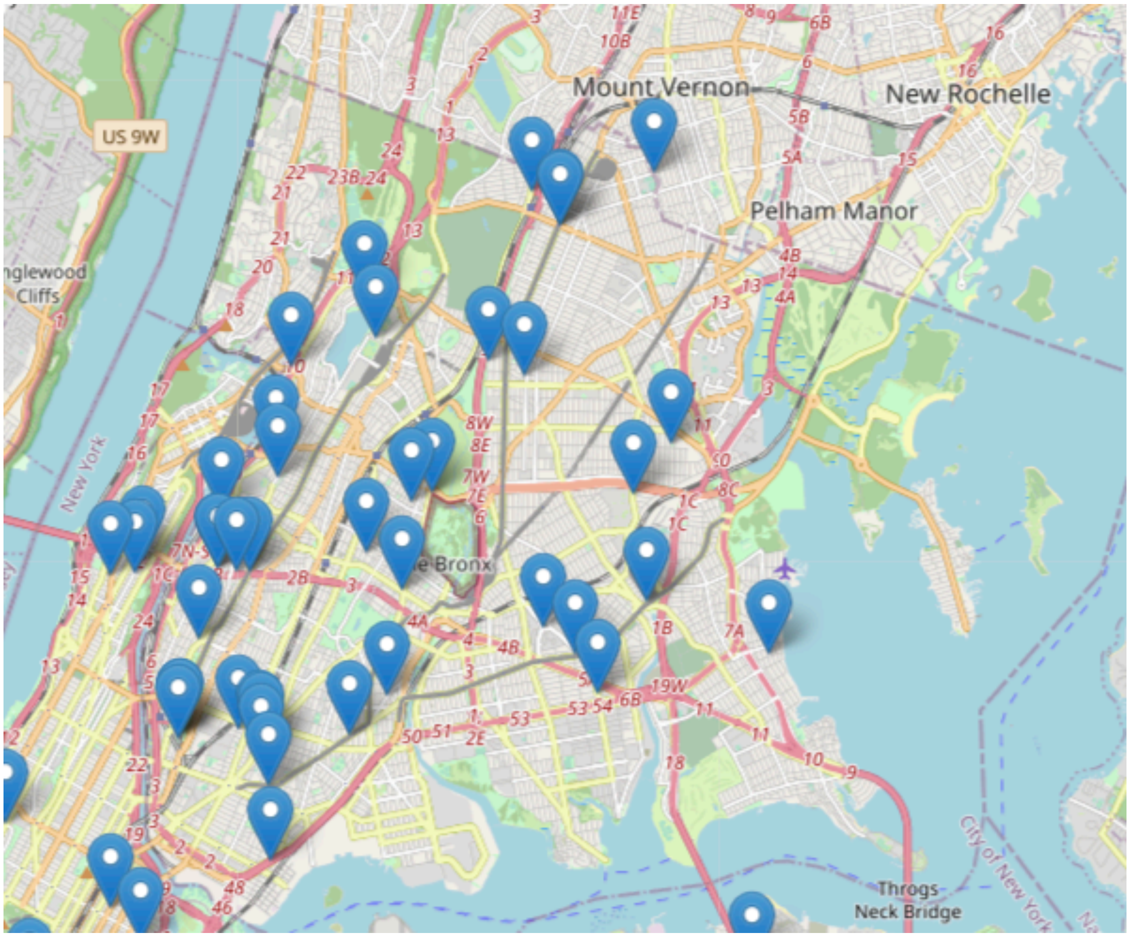


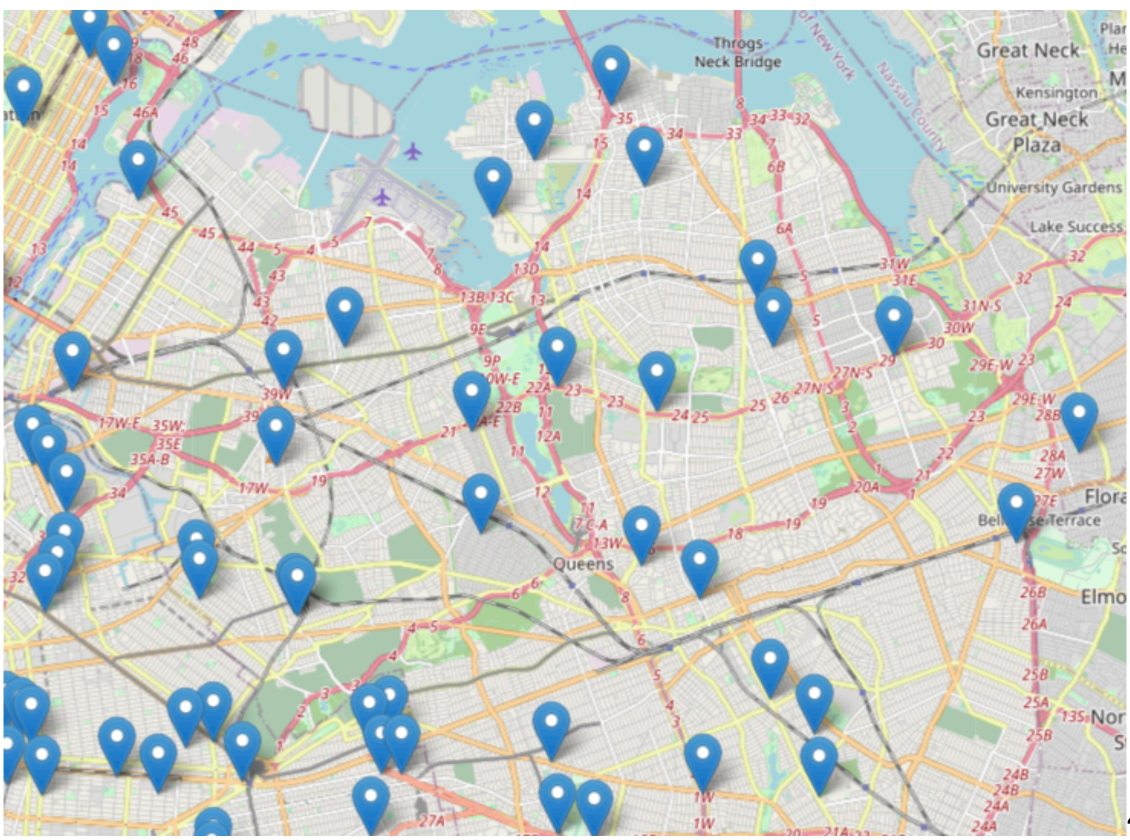
Understanding the distribution of potholes across different boroughs is crucial for assessing the infrastructure conditions in various areas of NYC. Among the boroughs, Queens stood out with the highest number of reported potholes, representing 35.52% of the total cases. Brooklyn ranked second, accounting for 25.10%, followed by the Bronx with 13.24% of the potholes. Manhattan, Staten Island, and Unspecified locations accounted for 11.11%, 11.59%, and 0.26% of the reported incidents, respectively. These findings highlight the varying intensity of the pothole issue across different boroughs, with Queens and Brooklyn experiencing a higher concentration.



Status of Reported Potholes:

Analyzing the status of reported potholes provides insights into the efficiency of the city's response and resolution processes. Visual The majority of potholes were marked as Closed, comprising 98.11% of the cases. This indicates that a significant proportion of reported potholes have been addressed and repaired. However, it is worth noting that a small percentage of cases were still Open (0.29%) or Pending (1.59%), suggesting ongoing efforts to address these issues. Additionally, a negligible number of cases were categorized as Unspecified (0.01%), which may require further investigation or clarification.


Mount Vernon
New Rochelle
Pelham Manor
US 9W
Throgs
Neck Bridge


Throgs
Neck Bridge
Great Neck
Kensington
Great Neck
Plaza
University Gardens
Lake Success
Queens

Status	Potholes	Percentage Allocation
Closed	141742	98.11%
Open	419	0.29%
Pending	2297	1.59%
Unspecified	10	0.01%

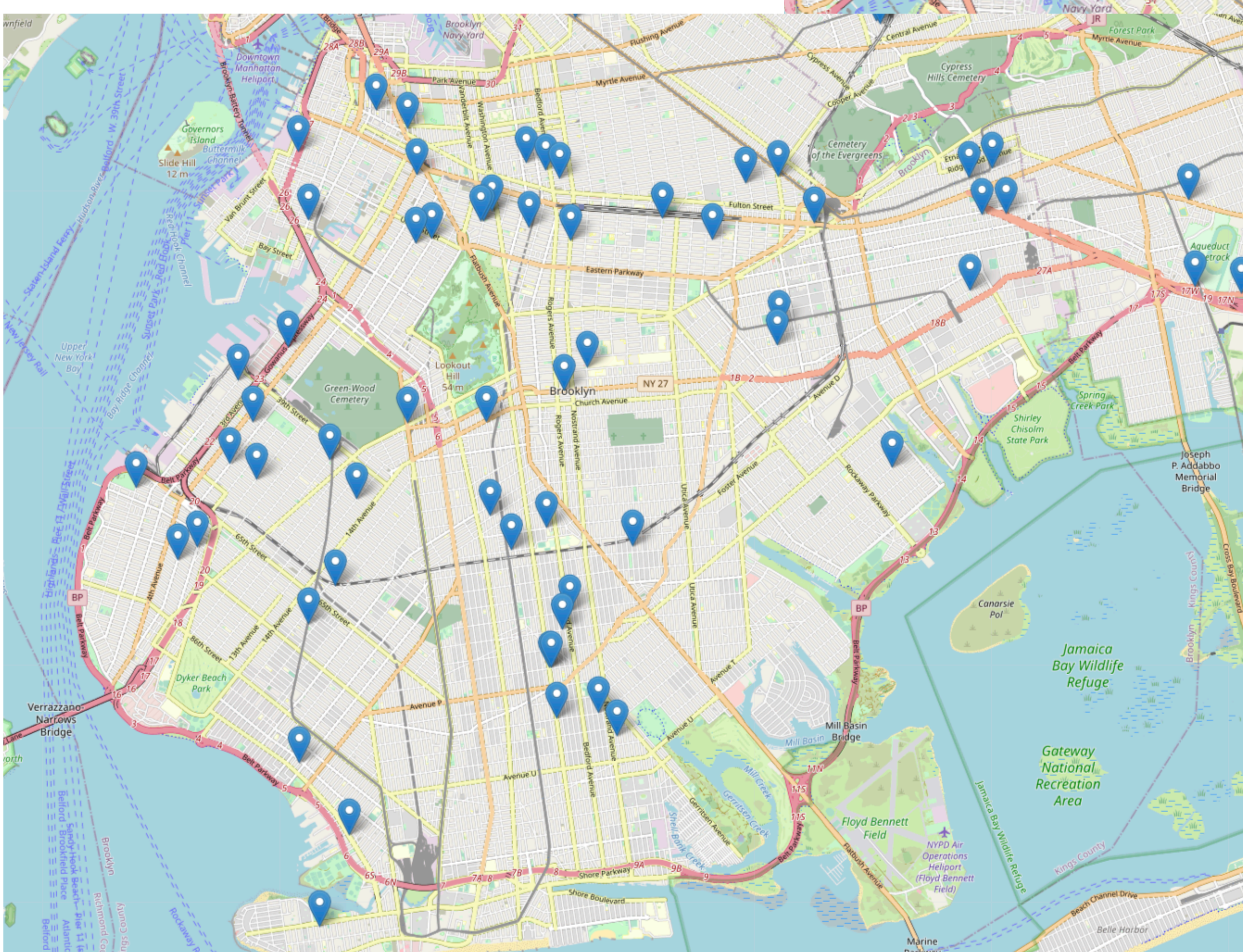
Maps:

To illustrate the distribution of potholes per borough in NYC, I utilized the leaflet and OpenStreetMap libraries in R. You can find this snippet under {r map} in my R markdown doc.

In the code, I first filtered the dataset to include only rows with valid latitude and longitude values and where the descriptor is "Pothole." I then randomly sampled 10,000 rows from this filtered dataset to ensure a manageable number of data points for visualization. Next, I further filtered the data to include only open and


Elizabeth
Bayonne
Upper New York Bay
Union County
Bayway Refinery
Linden
Linden Airport
Richmond County
Freshkills Park
Staten Island
Carteret
Port Reading
KWVP
Lower New York Bay
City of New York
Monmouth County
Richmond County


Edgewater
Cliffside Park
Fairview
North Bergen
Guttenberg
West New York
Central Park
Vista Rock 40 m
Great Hill 41 m
Manhattan
Henry Hudson Parkway
Riverside Drive
Broadway
Frederick Douglass Boulevard
NY 495
West 30th Street Heliport
East 34th Street Heliport
Long Island Expressway
Brooklyn
Queens
Manhattan Bridge
Brooklyn Navy Yard


Brooklyn Navy Yard
Downtown Manhattan Heliport
Governors Island
Myrtle Avenue
Flushing Avenue
Central Avenue
Cypress Hills Cemetery
Cemetery of the Evergreens
Fulton Street
Eastern Parkway
Green-Wood Cemetery
Lookout Hill 54 m
Brooklyn
NY 27
Church Avenue
Upper New York Bay
Dyker Beach Park
Verrazzano-Narrows Bridge
Avenue P
Avenue U
Shirley Chisolm State Park
Spring Creek Park
Joseph P. Addabbo Memorial Bridge
Canarsie Pol
Jamaica Bay Wildlife Refuge
Gateway National Recreation Area
Mill Basin Bridge
Floyd Bennett Field
NYPD Air Operations Heliport (Floyd Bennett Field)
Shore Parkway
Shore Boulevard
Belle Harbor

on the potholes that require attention or are currently being addressed.

Finally, I created the map using the leaflet() function. I added the OpenStreetMap tiles as the base layer using addProviderTiles("OpenStreetMap.Mapnik"). Then, I added markers to the map using the latitude and longitude coordinates from the filtered data. The awesomeIcons function was employed to display an icon resembling a close symbol for each marker, with the color determined by the pothole's status. The popup information for each marker includes the borough, incident zip, and status.

By running this code, I generated an interactive map that visualizes the distribution of potholes per borough in NYC, with markers representing the open and pending potholes. Hovering over a marker will display additional information about the pothole, providing a comprehensive understanding of the geographical distribution and status of potholes in different boroughs.

Attached you will find some screenshots of each borough.

My proposed econometric model:

```
Call:
glm(formula = PotholesCount ~ Season + Borough, family = poisson,
    data = subset_data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.8471  -0.6822  -0.6059  -0.4653   2.1935

Coefficients:
                      Estimate Std. Error  z value Pr(>|z|)
(Intercept)          -2.293743   0.016166 -141.886  < 2e-16 ***
SeasonSpring          0.581688   0.008303   70.057  < 2e-16 ***
SeasonSummer          0.162574   0.009325   17.434  < 2e-16 ***
SeasonWinter          0.528095   0.008511   62.048  < 2e-16 ***
BoroughBRONX          0.243687   0.016483   14.784  < 2e-16 ***
BoroughBROOKLYN       0.307574   0.015716   19.571  < 2e-16 ***
BoroughMANHATTAN      0.070564   0.016787    4.204 2.63e-05 ***
BoroughQUEENS         0.603195   0.015454   39.032  < 2e-16 ***
BoroughSTATEN ISLAND  0.686947   0.016710   41.110  < 2e-16 ***
BoroughUNSPECIFIED   -1.077575   0.053592  -20.107  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 443323  on 670072  degrees of freedom
Residual deviance: 427426  on 670063  degrees of freedom
AIC: 716382

Number of Fisher Scoring iterations: 6
```

<u>Potholes = β0 + β1(Season) + β2(Borough) + ε</u>

In this model, Potholes represents the number of potholes in a given area, Season represents the season of the year, and Borough represents the borough of the area.

$\beta 0$ is the intercept or constant term in the model, which represents the value of Potholes when all independent variables are equal to zero.

$\beta 1$ and $\beta 2$ are the coefficients of the Season and Borough variables, representing the change in Potholes for each one-unit increase in the corresponding variable, holding all other variables constant.

ε represents the error term, which captures the variability in Potholes that cannot be explained by the independent variables.

The functional form of this model remains a multiple linear regression, assuming a linear relationship between the independent variables and the dependent variable. By incorporating season and borough into the model, you can analyze the impact of these variables on pothole formation and maintenance in New York City. This allows for the investigation of how the number of potholes varies across different seasons and among different boroughs. Understanding these relationships can provide valuable insights for policymakers and city officials when making decisions about infrastructure investment and maintenance.By including the Season and Borough variables in your model, you can explore how the number of potholes varies across different seasons and among different boroughs while considering the impact of other factors such as traffic intensity, number of trucks, and flood zone. This expanded model allows for a more comprehensive analysis of the factors influencing pothole occurrence in New York City. It provides policymakers and city officials with valuable insights to guide infrastructure investment and maintenance decisions specific to each season and borough, taking into account the unique characteristics and challenges associated with different times of the year and areas of the city.

Regression Results:

The regression analysis using a Poisson regression model yielded insightful results regarding the occurrence of potholes in NYC. The estimated coefficients provide valuable insights into the effects of different variables on the count of potholes. The intercept term (-2.293743) represents the expected log count of potholes when all other independent variables are held constant. The coefficients for the seasons indicate the expected change in the log count of potholes compared to the baseline season, with Spring (0.581688), Summer (0.162574), and Winter (0.528095) all showing significant positive effects on the count of potholes. These findings suggest that seasons have a significant influence on the occurrence of potholes in NYC, with Spring and Winter being particularly associated with higher occurrences.

The coefficients for the boroughs reveal the expected change in the log count of potholes compared to the baseline borough, with Bronx (0.243687), Brooklyn (0.307574), Manhattan (0.070564), Queens (0.603195), and Staten Island (0.686947) all showing significant positive effects. These findings highlight the significance of boroughs in predicting the count of potholes, with Staten Island and Queens standing out as boroughs with higher counts, while the Unspecified borough tends to have a lower count.

These results provide valuable insights for policymakers and city authorities, indicating that the occurrence of potholes in NYC is influenced by both seasons and boroughs. The Spring and Winter seasons, along with Staten Island and Queens boroughs, emerge as notable factors associated with higher counts of potholes. These insights can inform targeted measures and resource allocation for effective pothole management during specific seasons and in different boroughs.

The statistical significance of the coefficients is supported by the low p-values, indicating that the estimated effects are unlikely to occur by chance. The goodness of fit measures, as indicated by the deviance values and AIC, suggest that the model provides a reasonably good fit to the data.

It's important to note that since a Poisson regression model is employed, the coefficients represent the expected change in the log count of potholes. To interpret the effects on the count itself, the coefficients can be exponentiated. For instance, the Spring season is associated with a nearly 1.79 times higher count of potholes compared to the baseline season.

Overall, these findings enhance our understanding of the factors influencing the occurrence of potholes in NYC and highlight the significance of seasons and boroughs in predicting their count. This knowledge can inform decision-making processes aimed at improving pothole management strategies and resource allocation in the city.

Discussion and Conclusion:

This research project investigated the effects of season and borough on pothole incidence in New York City using econometric techniques. By analyzing a dataset of 311 service requests related to potholes, several key findings emerged, providing valuable insights for policymakers and city authorities.

The distribution of potholes across seasons revealed that Spring and Winter had the highest number of reported incidents, accounting for 36.49% and 30.94% of potholes, respectively. This suggests that weather conditions and road maintenance activities during these seasons contribute to a higher frequency of pothole occurrences.

Analyzing the distribution of potholes across boroughs identified Queens as the borough with the highest number of reported potholes, representing 35.52% of the total cases. Brooklyn followed closely behind with 25.10%, indicating variations in the intensity of the pothole issue across different boroughs.

The status analysis of reported potholes indicated that the majority of cases were marked as Closed (98.11%), indicating successful resolution and repair. However, a small percentage of cases remained Open (0.29%) or Pending (1.59%), requiring ongoing attention and efforts. The presence of Unspecified cases (0.01%) suggests the need for further investigation or clarification.

The proposed econometric model, incorporating season and borough as independent variables, allowed for a comprehensive analysis of the factors influencing pothole formation and maintenance. The regression results using a Poisson regression model provided insightful findings. Seasons such as Spring and Winter exhibited significant positive effects on the count of potholes, highlighting their influence on pothole occurrence. Among boroughs, Staten Island and Queens showed significant positive effects, indicating higher counts of potholes in these areas.

These findings underscore the importance of considering season and borough factors when formulating infrastructure investment and maintenance strategies. By understanding the variations in pothole occurrence across seasons and boroughs, policymakers and city authorities can allocate resources effectively and implement targeted measures for improved road maintenance and safety.

The statistical significance of the regression coefficients, supported by low p-values, enhances the reliability of the results. The model's goodness of fit measures indicate a reasonable fit to the data, further validating the findings.

Overall, this research project provides valuable insights into the determinants of pothole incidence in New York City, particularly the effects of season and borough. By leveraging data-driven approaches and econometric techniques, policymakers and city administrations can make informed decisions regarding resource allocation and infrastructure improvement. Ultimately, these efforts contribute to enhanced road maintenance, increased road safety, and improved economic productivity for the city of New York.

References:

- Wang, L., Chen, W., & Xiong, H. (2018). Analysis of Factors Affecting Pothole Formation and Development Based on Longitudinal Pavement Performance Data. Journal of Transportation Engineering, Part B: Pavements, 144(2), 04018005.

- Deng, J., Zhang, W., Wang, Y., & Li, Y. (2017). Predicting Pothole Formation Using Traffic and Environmental Data. Journal of Infrastructure Systems, 23(2), 04016053.
- Wong, D., & Li, Z. (2019). A Spatial-Temporal Analysis of Pothole Formation and Maintenance in Urban Areas. Journal of Infrastructure Maintenance and Management, 20(1), 81-95.
- Zhang, Q., & Wang, H. (2016). A Comprehensive Analysis of Pothole Development Based on Long-Term Monitoring Data. Journal of Infrastructure Asset Management, 13(1), 41-52.
- Khasanov, A., Freishtat, N., & Serre, M. L. (2018). Analyzing Spatial and Temporal Patterns of Potholes in Urban Road Networks. Transportation Research Record, 2672(1), 36-47.