

Estimating NYC 311 Service Request Wait Times

Danielle Medellin

Problem Statement



“A Service Request is your request for the City to provide you with assistance, perform an inspection, or address a problem.”

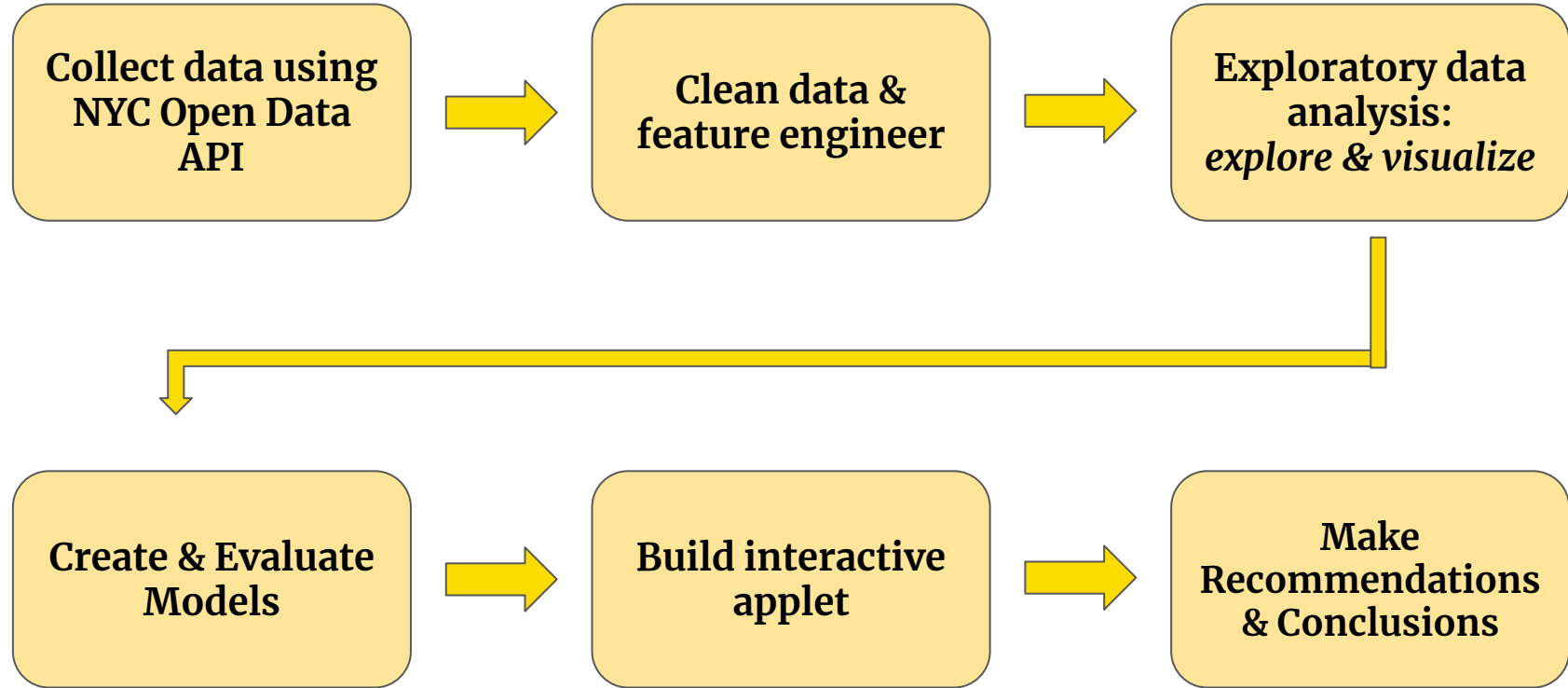
Due to the fact that there are so many different reasons for submitting service requests to 311, all with a wide range of urgency and importance, it can be difficult to gauge how long it will take for a request to actually get resolved.

Once a request is submitted, you receive a confirmation with the details of your request -- what's missing is an estimate of when your request will be resolved.

Our goal is to build a regression model that will predict the wait time a citizen can expect given the factors of their request. Our model's success will be measured with root-mean squared error (RMSE).

Additionally, we hope to build an applet that mimics what inputting a service request might look like, and in addition to just getting a description of the request, the app will also return an estimated wait time for the request to be closed.

Methodology & Workflow



Data Collection & Cleaning



Collected data using NYC Open Data API:

- Goal: collect 6 months worth of data
- Most recent starting from before Feb 1, 2020 due to COVID-19
- 600,000 observations
- Agency: NYPD, Status: Closed

Data cleaning:

- Cut down on unnecessary features
- Filled *nulls* for categorical features
- Eliminated negative wait times
- Limited wait time at 31 days

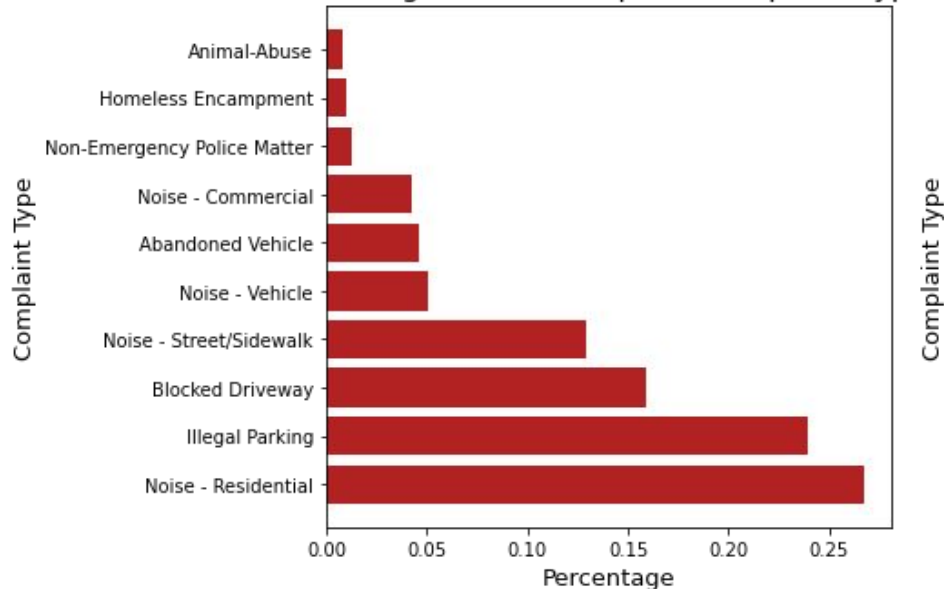
Final Features:

- Unique key
- Created date
- Closed date
- Agency (name & abbrev)
- Complaint type
- Descriptor
- Location type
- Status
- Borough
- Submission channel
- Resolution description & date
- Longitude
- Latitude
- Wait time

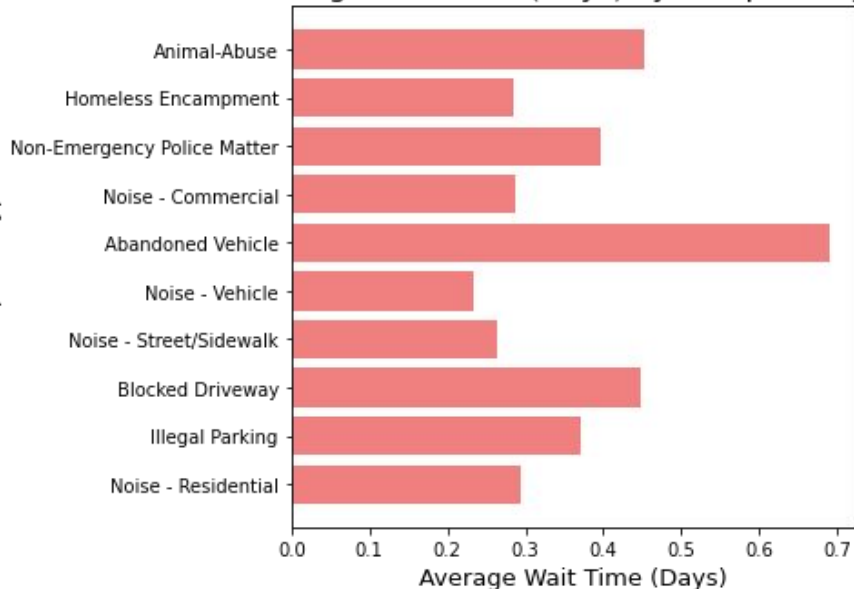
Exploratory Data Analysis



Percentage of Most Frequent Complaint Types

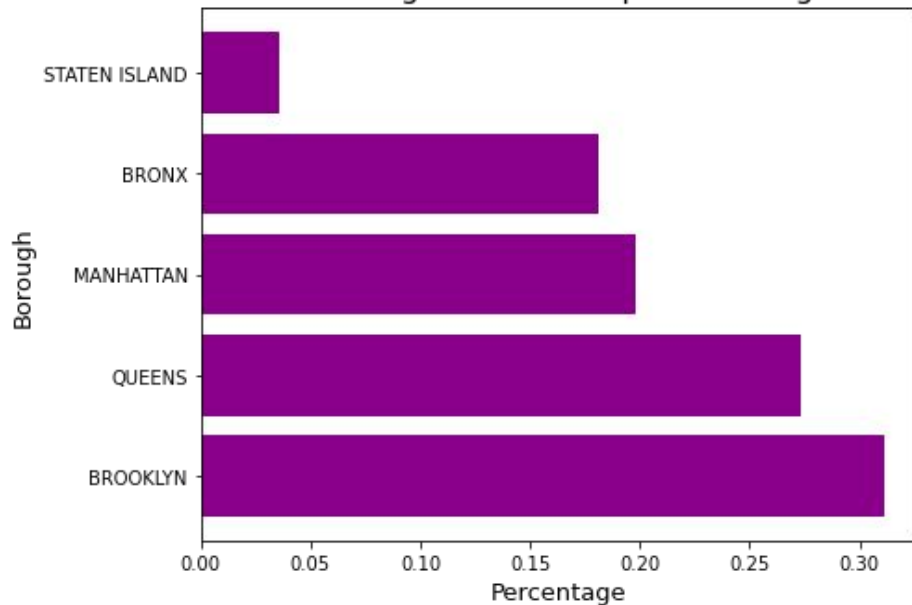


Average Wait Time (Days) by Complaint Type

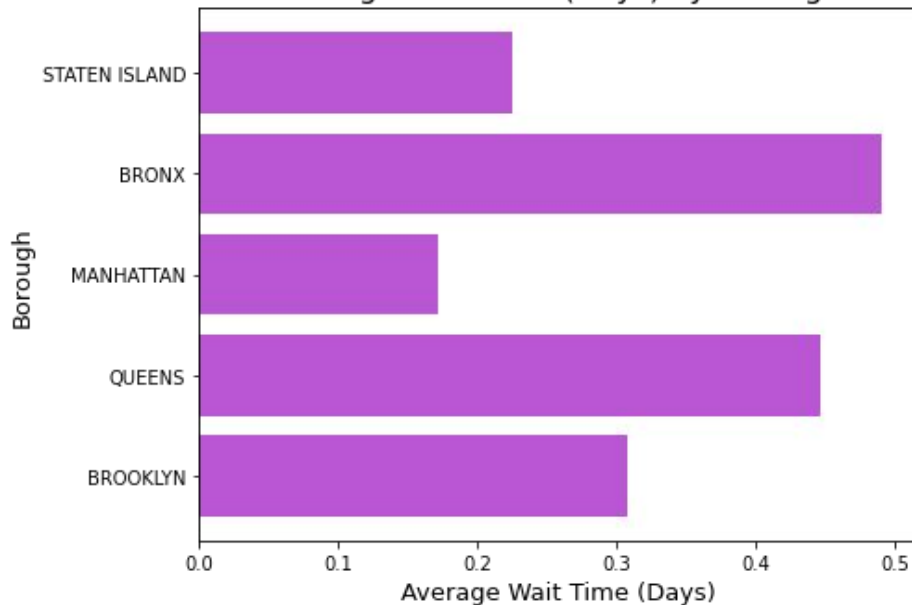


Exploratory Data Analysis

Percentage of Most Frequent Boroughs



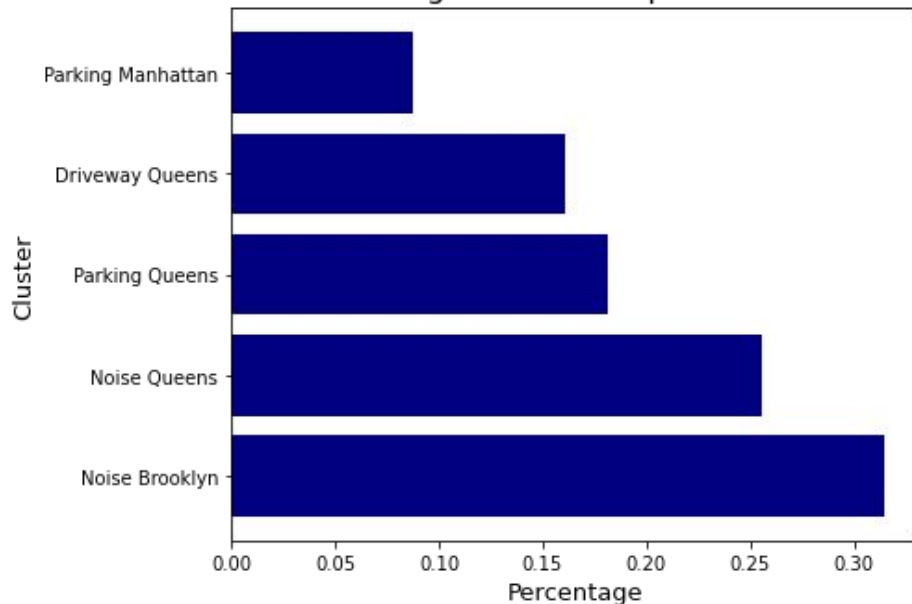
Average Wait Time (Days) by Borough



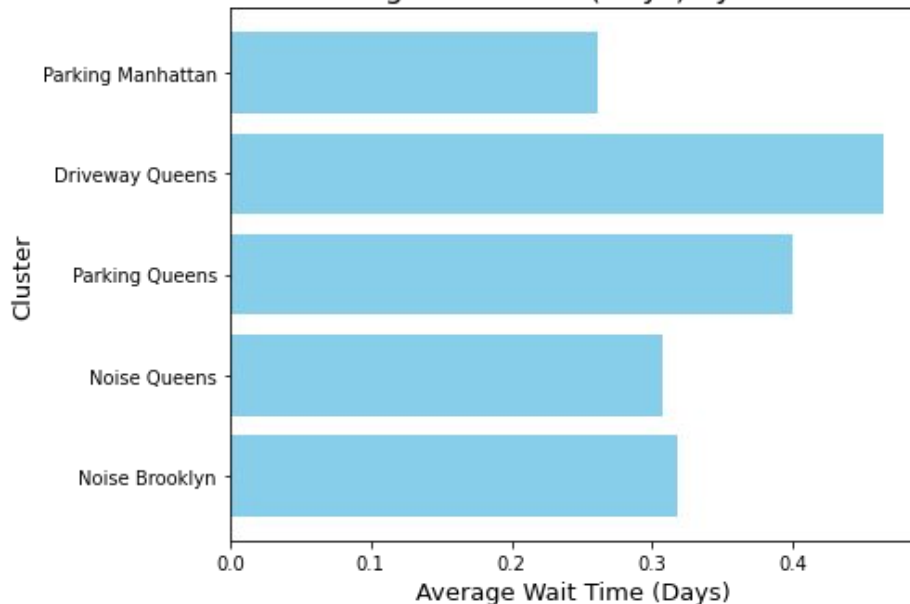
Exploratory Data Analysis



Percentage of Most Frequent Clusters



Average Wait Time (Days) by Cluster

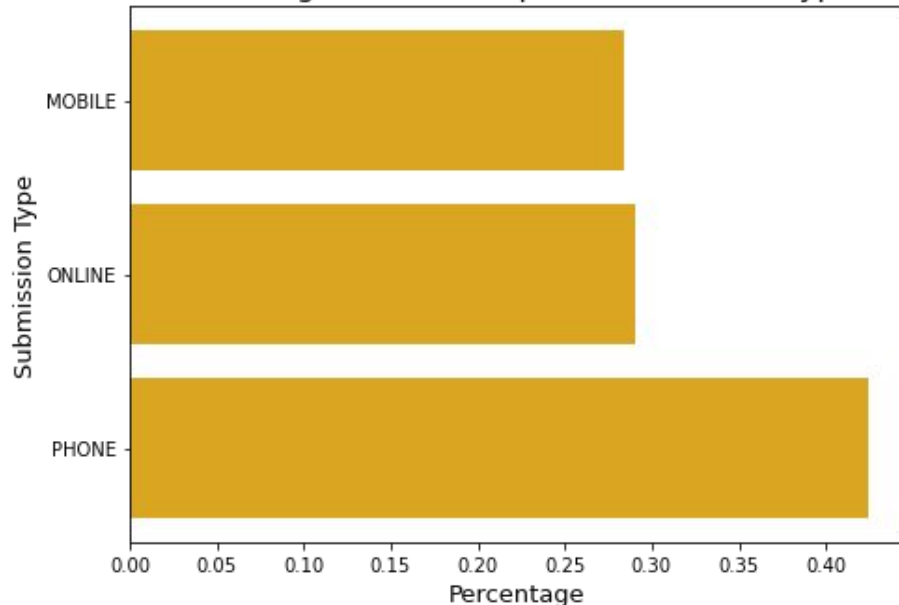


Clusters found using k-modes

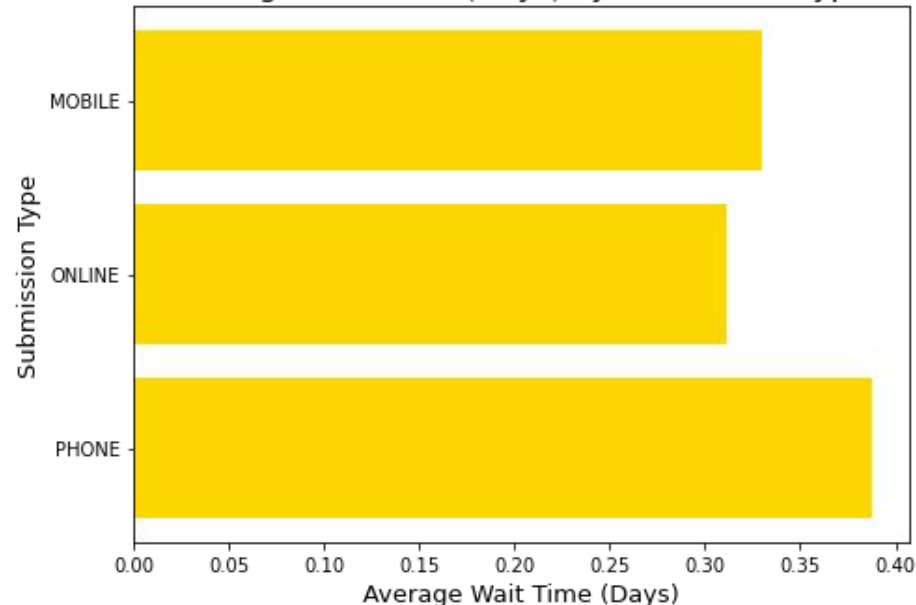
Exploratory Data Analysis



Percentage of Most Frequent Submission Types



Average Wait Time (Days) by Submission Type



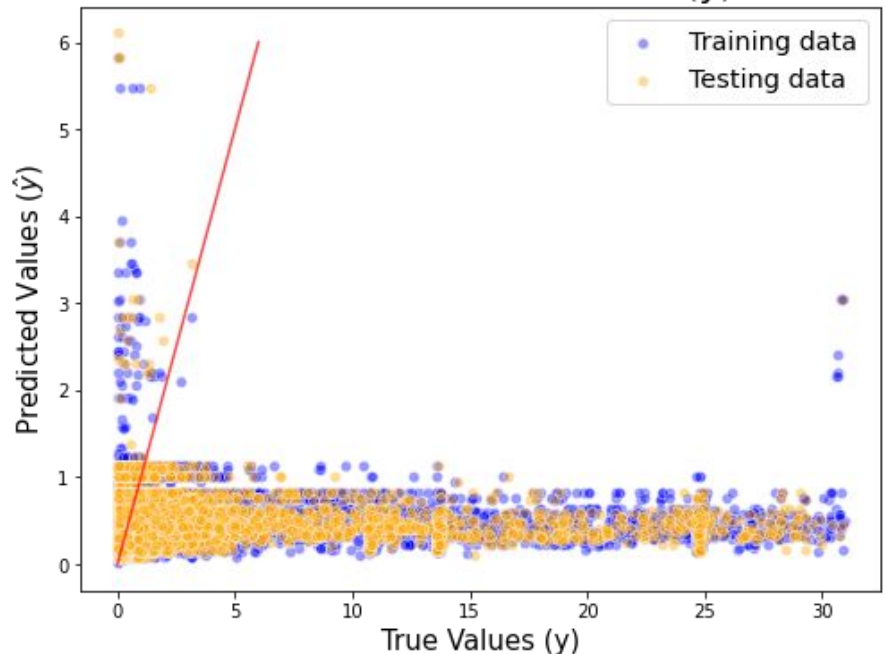
Model Features: complaint type, description, borough, submission channel, location type, predicted cluster

Model	Training RMSE	Testing RMSE	Difference (Test - Train)
Baseline	1.47	1.44	-0.039
Linear Regression	1.49	1.45	-0.039
Decision Tree	1.47	1.43	-0.039
Random Forest	1.47	1.43	-0.039
GLM - Gamma	1.47	1.43	-0.038
AdaBoost	1.47	1.43	-0.039
ARIMA	0.49	0.14	-0.35

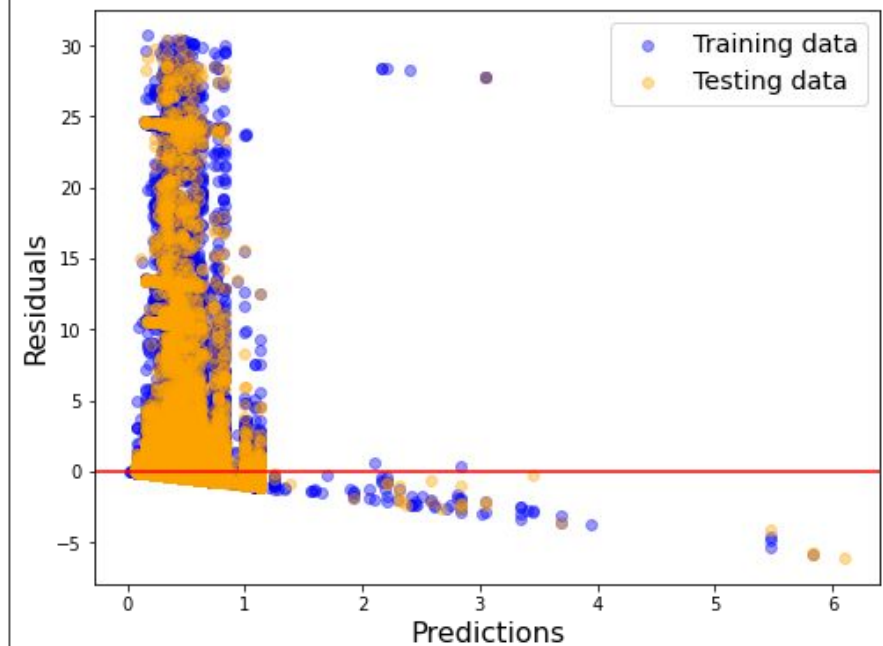
Gamma Regression: often used to measure waiting-time variables

Model Evaluation

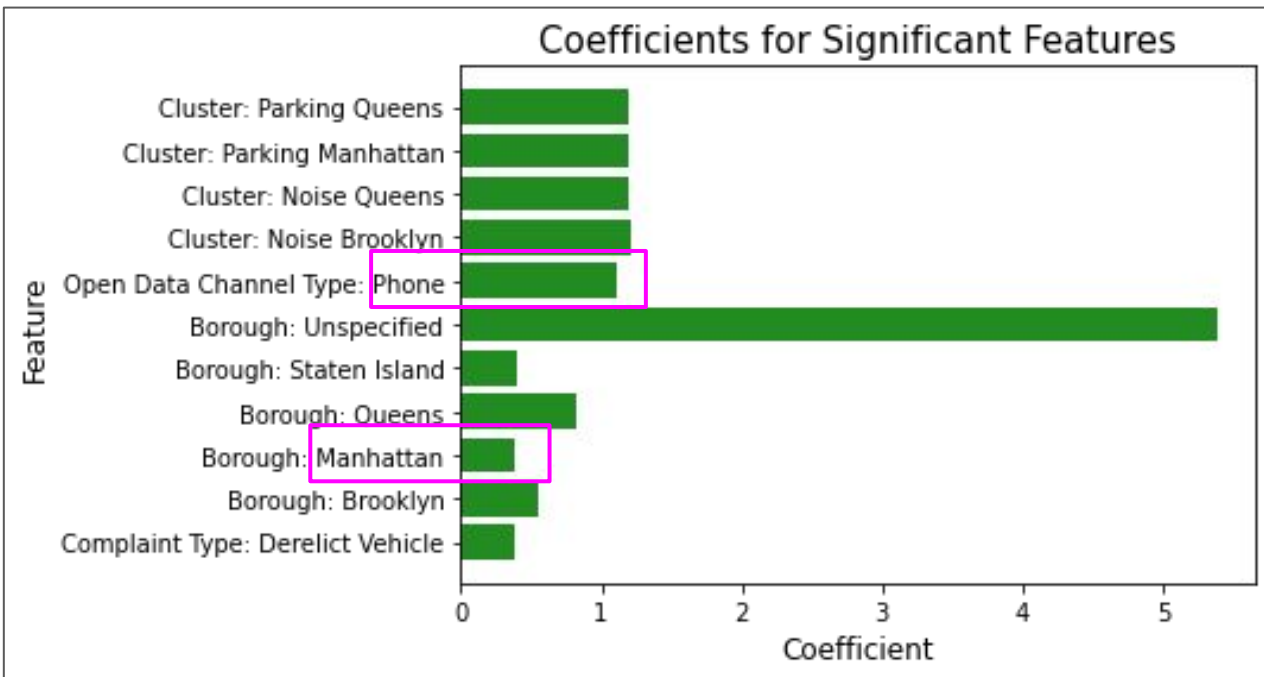
True Values v Predictions (y)



Residuals Plot

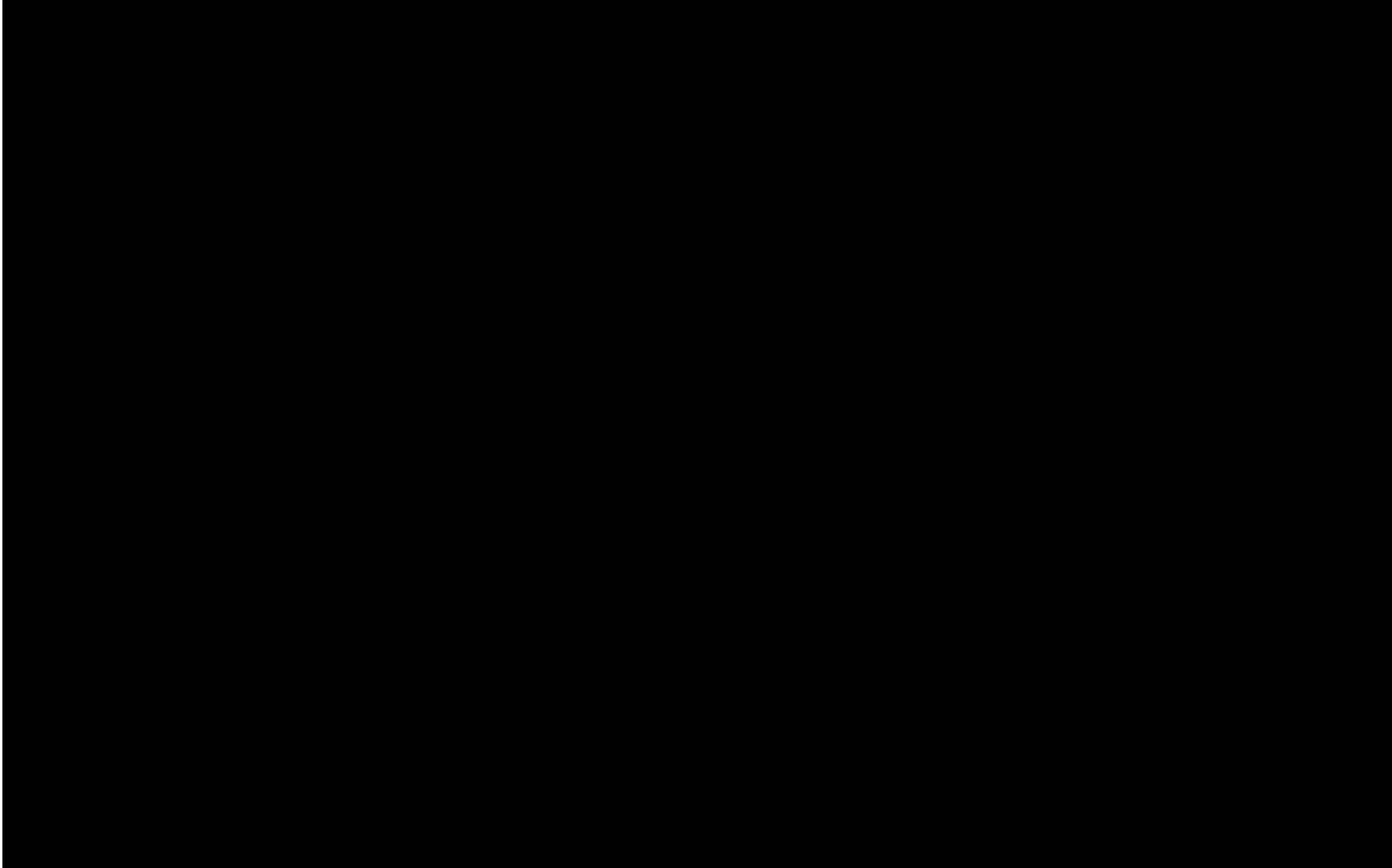


Model Evaluation



If the service request was made by *phone*, it will take **1.1 times** as much time to complete a 311 service request than those not submitted by phone, all else held constant.

If the service request was for *Manhattan*, it will take **.378 times** as much time than requests submitted not from Manhattan, all other factors held constant.



Thank you for your submission!
Your request is estimated to be resolved in 0.28 days.

Your Request:

Borough: BROOKLYN

Location: Street/Sidewalk

Complaint: Illegal Parking

Description (if applicable): Blocked Bike Lane

Submitted by: MOBILE

Submission Date & Time:

6-7-2020 21:10

If this is an emergency, please call 911.

Conclusions & Recommendations



- ❖ The features we explored are not good predictors of service request wait time
- ❖ Explore the time series model, add exogenous features
- ❖ Look into errors and extremes of the target variable, such as negative or large wait times
- ❖ Explore different types of requests sent to other agencies

References



Ashish. “Bank Customer Clustering (K-Modes Clustering).” *Kaggle*, Kaggle, 15 Mar. 2019, www.kaggle.com/ashydv/bank-customer-clustering-k-modes-clustering.

“Getting Started with NYC OpenData and the Socrata API.” Database and Programming Tutorials, holowczak.com/getting-started-with-nyc-opendata-and-the-socrata-api/5/.

Johnson, Jesse. “K-Modes.” *The Shape of Data*, 14 Sept. 2014, shapeofdata.wordpress.com/2014/03/04/k-modes/.

NYC OpenData. “311 Service Requests from 2010 to Present: NYC Open Data.” *311 Service Requests from 2010 to Present | NYC Open Data*, 7 June 2020, data.cityofnewyork.us/Social-Services/311-Service-Requests-from-2010-to-Present/erm2-nwe9.

NYC311. “About NYC311.” *311 Sets New Record with 44 Million Customer Interactions in 2018*, www1.nyc.gov/311/311-sets-new-record-in-2018.page.

Shadkam-Farrokhhi, Mahdi. “ARIMA Review.” *GitHub Enterprise*, 2020, git.generalassemb.ly/DSI-US-11/local_arima.

Shadkam-Farrokhhi, Mahdi. “Using cross_val_score with a Custom Metric.” *GitHub Enterprise*, 2020, git.generalassemb.ly/DSI-US-11/local_faq/blob/master/code/custom_cross_val_score.ipynb.

“Socrata Developer Portal.” *Socrata*, dev.socrata.com/foundry/data.cityofnewyork.us/erm2-nwe9.

Xmunoz. “Xmunoz/Sodapy.” *GitHub*, 2018, github.com/xmunoz/sodapy/blob/master/examples/soql_queries.ipynb.