Analysis of Noise Complaints in New York City

Willamette University CS 429 Data Science Spring 2018
Dakota Madden-Fong, Thomas Tuttle, Bailey Williams

Project Description

The goal of this project was to analyze a data set containing noise complaints in the New York City area and to create a model that would describe the type of “parties” in a given area at a given time and place.

The data set contained over 225,000 noise complaints from 12/31/2015 to 1/1/2017.

The Data Set

Created Date	Closed Date	Location Type	Incident Zip	City	Borough	Latitude	Longitude
2015-12-31 00:01:15	2015-12-31 03:48:04	Store/Commercial	10034.0	NEW YORK	MANHATTAN	40.86618344001468	-73.91893042945345
2015-12-31 00:02:48	2015-12-31 04:36:13	Store/Commercial	10040.0	NEW YORK	MANHATTAN	40.85932419390543	-73.93123733660876
2015-12-31 00:03:25	2015-12-31 00:40:15	Residential Building/House	10026.0	NEW YORK	MANHATTAN	40.79941540978025	-73.953371716858667
2015-12-31 00:03:26	2015-12-31 01:53:38	Residential Building/House	11231.0	BROOKLYN	BROOKLYN	40.678285109498 1	-73.99466779426595
2015-12-31 00:05:10	2015-12-31 03:49:10	Residential Building/House	10033.0	NEW YORK	MANHATTAN	40.85030372032608	-73.93851562699031
2015-12-31 00:08:05	2015-12-31 01:59:12	Residential Building/House	10467.0	BRONX	BRONX	40.858747683927 1	-73.86562454420242
2015-12-31 00:11:40	2015-12-31 06:24:00	Residential Building/House	11230.0	BROOKLYN	BROOKLYN	40.617005359002 29	-73.95692046165364
2015-12-31 00:12:13	2015-12-31 00:38:09	Residential Building/House	11215.0	BROOKLYN	BROOKLYN	40.66505114462701	-73.98127790267 175
2015-12-31 00:12:37	2015-12-31 05:03:39	Residential Building/House	10463.0	BRONX	BRONX	40.875894942376384	-73.91247127084895
2015-12-31 00:14:13	2015-12-31 06:25:40	Store/Commercial	11372.0	JACKSON HEIGHTS	QUEENS	40.75558360239671	-73.88520104800678
2015-12-31 00:15:36	2015-12-31 02:58:09	Residential Building/House	11213.0	BROOKLYN	BROOKLYN	40.66785694336881	-73.93199345800588
2015-12-31 00:16:28	2015-12-31 00:59:23	Store/Commercial	10033.0	NEW YORK	MANHATTAN	40.84700715687885	-73.93819337596233
2015-12-31 00:18:04	2015-12-31 03:49:12	Store/Commercial	10034.0	NEW YORK	MANHATTAN	40.86597800798226	-73.9195308453907
2015-12-31 00:19:23	2015-12-31 02:18:42	Club/Bar/Restaurant	11375.0	FOREST HILLS	QUEENS	40.7192472690414 6	-73.84237048590964
2015-12-31 00:19:43	2015-12-31 00:29:37	Club/Bar/Restaurant	10027.0	NEW YORK	MANHATTAN	40.81453554135292	-73.95914444206745
2015-12-31 00:19:49	2015-12-31 05:04:42	Residential Building/House	10468.0	BRONX	BRONX	40.873034748443594	-73.90195294442927
2015-12-31 00:20:05	2015-12-31 01:40:22	Residential Building/House	11417.0	OZONE PARK	QUEENS	40.67771291666382	-73.83232708056931
2015-12-31 00:20:45	2015-12-31 04:51:45	Residential Building/House	11233.0	BROOKLYN	BROOKLYN	40.685323653531658	-73.92583881214411
2015-12-31 00:22:25	2015-12-31 01:40:22	Residential Building/House	11217.0	BROOKLYN	BROOKLYN	40.68423815261613	-73.98935266446669
2015-12-31 00:22:38	2015-12-31 06:06:17	Residential Building/House	11355.0	FLUSHING	QUEENS	40.758123914829945	-73.82884714328377
2015-12-31 00:23:23	2015-12-31 16:28:20	Residential Building/House	11234.0	BROOKLYN	BROOKLYN	40.633336809850135	-73.92071208195789
2015-12-31 00:25:43	2015-12-31 08:42:51	Residential Building/House	10452.0	BRONX	BRONX	40.831195901208 17	-73.93034856129522
2015-12-31 00:26:25	2015-12-31 06:53:33	Residential Building/House	10468.0	BRONX	BRONX	40.86950166576785	-73.90443848773681
2015-12-31 00:28:24	2015-12-31 05:19:00	Club/Bar/Restaurant	10025.0	NEW YORK	MANHATTAN	40.802346308210 16	-73.9681096770604
2015-12-31 00:31:09	2015-12-31 01:44:19	Club/Bar/Restaurant	11103.0	ASTORIA	QUEENS	40.765770617999 12	-73.91898767632166
2015-12-31 00:31:20	2015-12-31 06:34:54	Residential Building/House	10458.0	BRONX	BRONX	40.86905269304852	-73.89030927857495
2015-12-31 00:31:59	2015-12-31 02:34:08	Residential Building/House	10021.0	NEW YORK	MANHATTAN	40.76872682443117	-73.95408580869662
2015-12-31 00:33:25	2015-12-31 01:13:26	Residential Building/House	11238.0	BROOKLYN	BROOKLYN	40.67900614917824	-73.9708041449147 4

Above are the first few rows of the data set. The set gave us the date and time the complaints were made and closed, as well as the location of the incident. Note the location data is given in latitude and longitude.

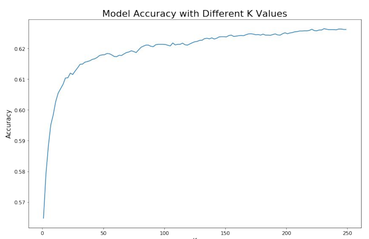
Data from:
https://www.kaggle.com/somesnm/partynyc/data

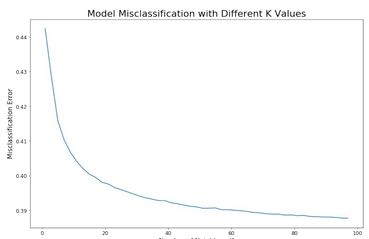
Building the KNN Model

We determined that physical location, as well as date, month, day of the week and time of day are important to include in the model that determines location type. We built a **KNN model** using these parameters and the sci-kit-learn package. We remove residential locations as we want to focus on non-residential parties.

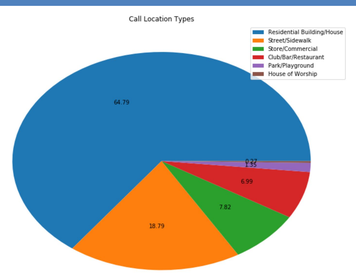
Results

To test the accuracy of our model, we split our data into training and testing data, then compared the known classifications of the testing data to the classifications returned by the model. Even at optimal K values, our accuracy was not very high.

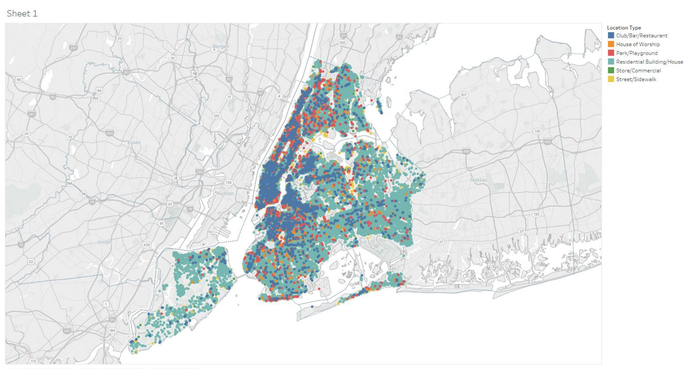




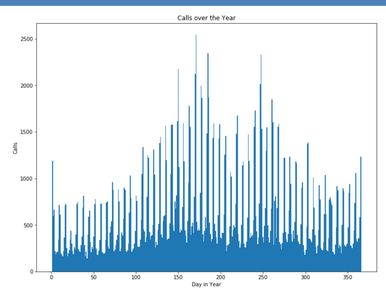
Exploratory Data Analysis



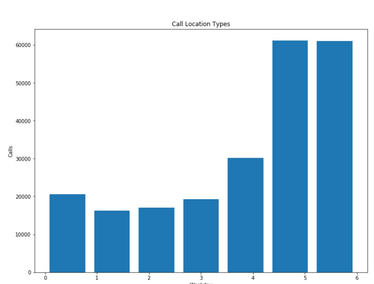
Proportion of Call Location Types. Observe the high proportion of residential and street complaints.

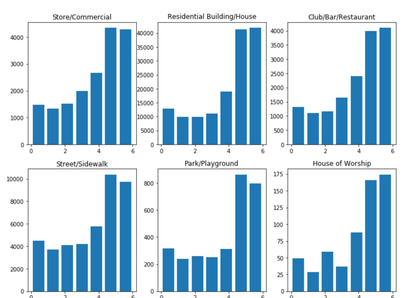


Call Locations by Latitude and Longitude.

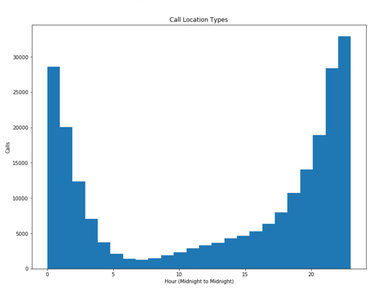


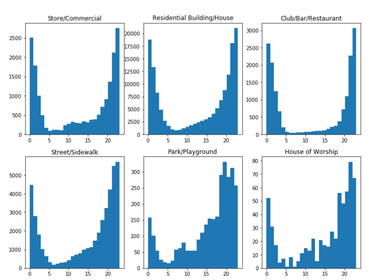
Total Number of Calls each Day. Observe that there is not a consistent weekly pattern.





Number of Calls each Weekday (Total + by Location Type) (Starting at Monday and Ending at Sunday).





Number of Calls each Hour (Total + by Location Type). Observe how the hourly activity changes based on location type.

Conclusions

Multiple issues plagued our model.

First, physical location is not a good metric to use when trying to categorize location type, as parties of similar type are not usually physically clustered together.

Second, all parameters the model operates on were weighted equally. We believe that if we were able to determine good relative weights for our parameters, our model would perform significantly better.