Noise Complaints in NYC

Dakota Madden-Fong, Thomas Tuttle, Bailey Williams

Project Description and Goals

- Analyze data set of noise complaints in NYC from 12/31/2015 to 1/1/2017
- Attempt to find interesting patterns in data

- Create a model that, given a location, time and date, would give an idea of what parties should be near the user's current time and place

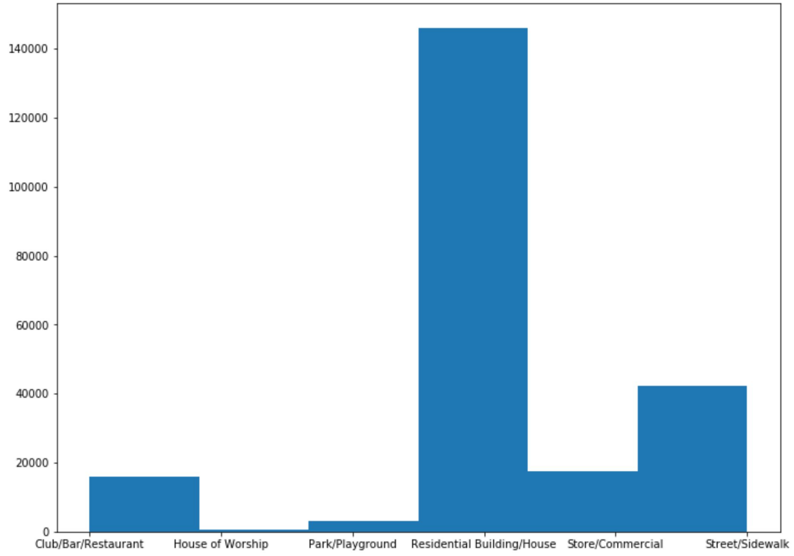
The Data Set

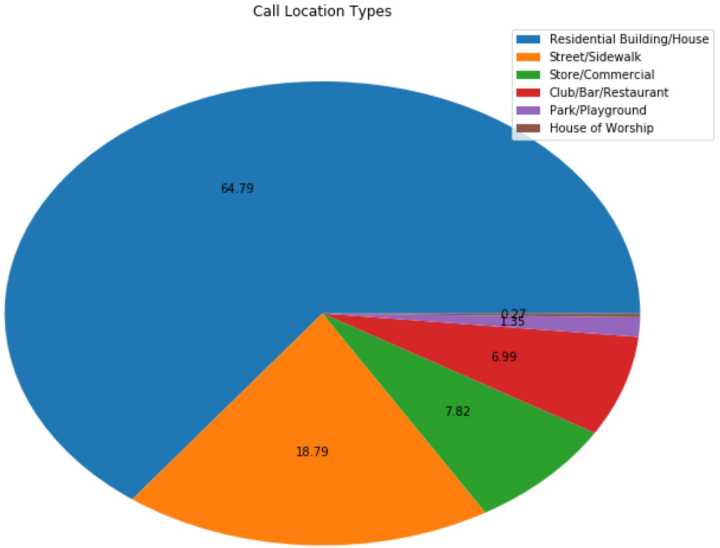
Created Date	Closed Date	Location Type	Incident Zip	City	Borough	Latitude	Longitude
2015-12-31 00:01:15	2015-12-31 03:48:04	Store/Commercial	10034.0	NEW YORK	MANHATTAN	40.86618344001468	-73.91893042945345
2015-12-31 00:02:48	2015-12-31 04:36:13	Store/Commercial	10040.0	NEW YORK	MANHATTAN	40.85932419390543	-73.93123733660876
2015-12-31 00:03:25	2015-12-31 00:40:15	Residential Building/House	10026.0	NEW YORK	MANHATTAN	40.79941540978025	-73.95337116858667
2015-12-31 00:03:26	2015-12-31 01:53:38	Residential Building/House	11231.0	BROOKLYN	BROOKLYN	40.6782851094981	-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	40.8587476839271	-73.86562454420242
2015-12-31 00:11:40	2015-12-31 06:24:00	Residential Building/House	11230.0	BROOKLYN	BROOKLYN	40.61700535900229	-73.95692046165364
2015-12-31 00:12:13	2015-12-31 00:38:09	Residential Building/House	11215.0	BROOKLYN	BROOKLYN	40.66505114462701	-73.98127790267175
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	40.71924726904146	-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	40.68532363531658	-73.92583881214411
2015-12-31 00:22:25	2015-12-31 16:24:02	Residential Building/House	11217.0	BROOKLYN	BROOKLYN	40.68423815261613	-73.9893526644669
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	40.83119590120817	-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	40.80234630821016	-73.9681096770604
2015-12-31 00:31:09	2015-12-31 01:44:19	Club/Bar/Restaurant	11103.0	ASTORIA	QUEENS	40.76577061799912	-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	-73.97080414491474

- Over 225,000 complaints
- Source: https://www.kaggle.com/somesnm/partynyc/data

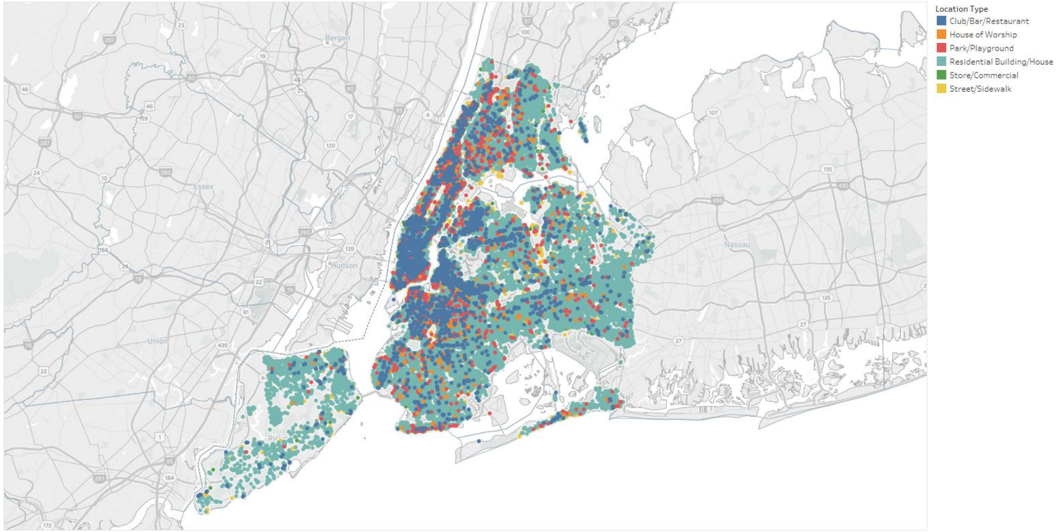
Exploratory Data Analysis

Calls By Location Type

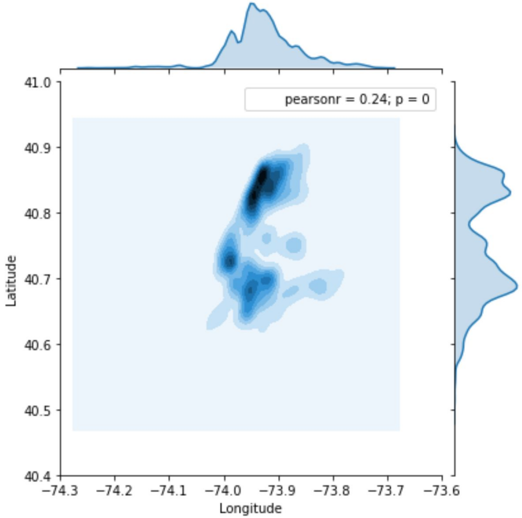




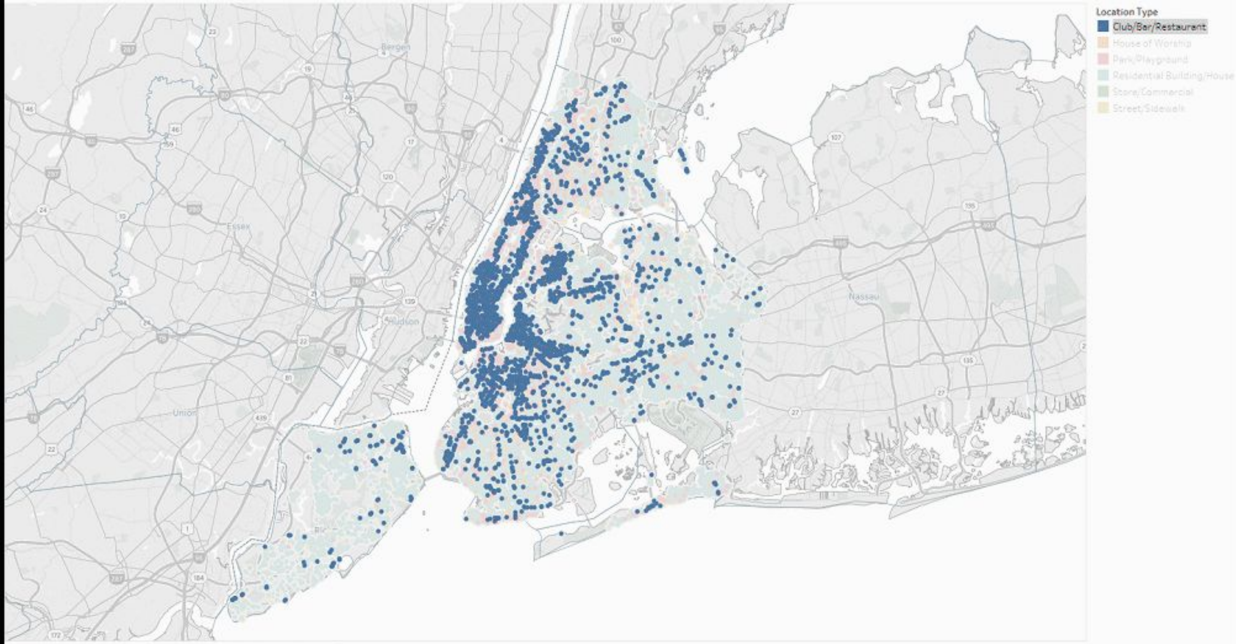
Calls by Location (Latitude/Longitude)



Map based on Longitude and Latitude. Color shows details about Location Type.

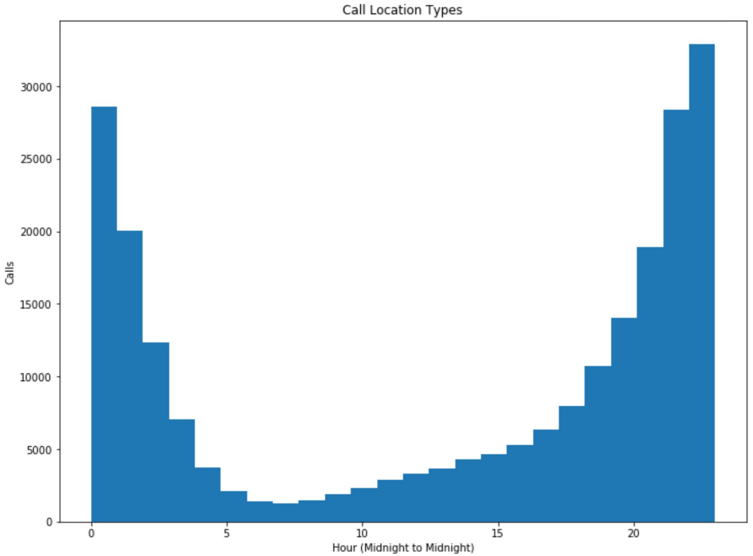


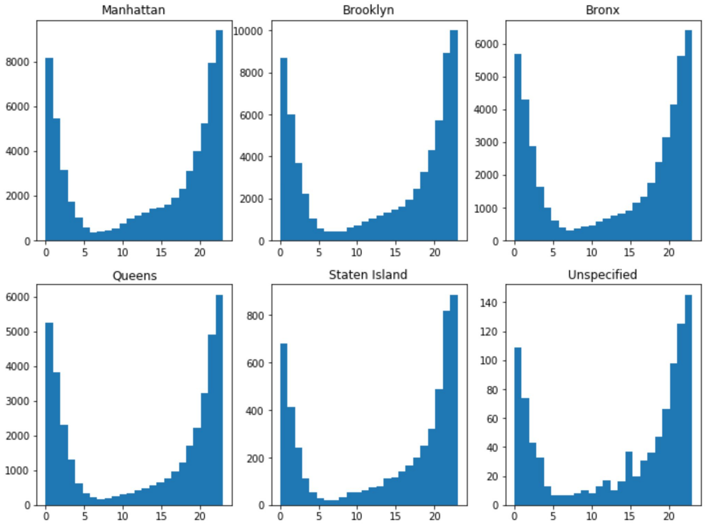
Sheet 1



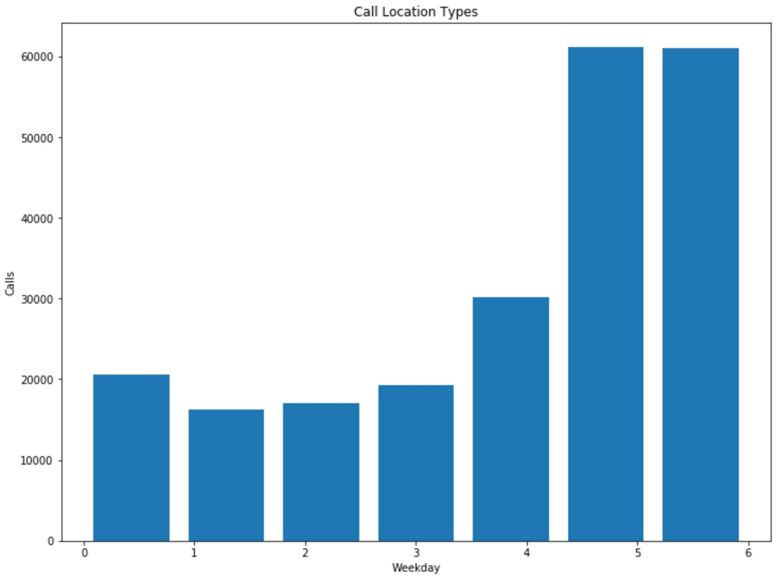
Map based on Longitude and Latitude. Color shows details about Location Type.

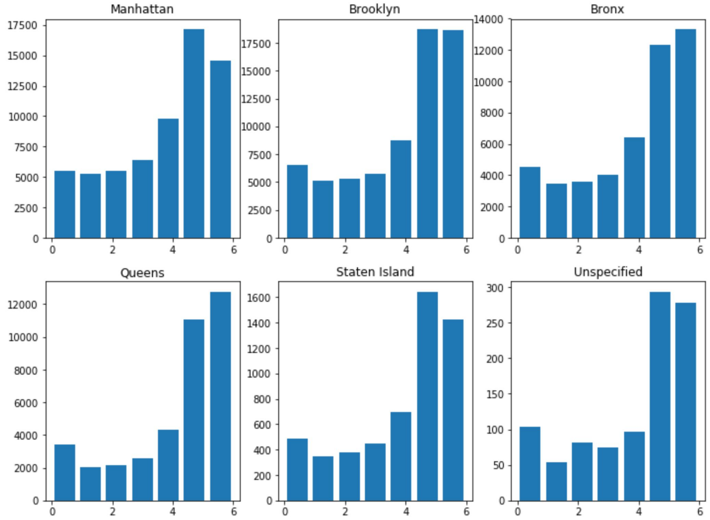
Calls By Hour (Total + By Borough)





Calls By Weekday (Total + By Borough)





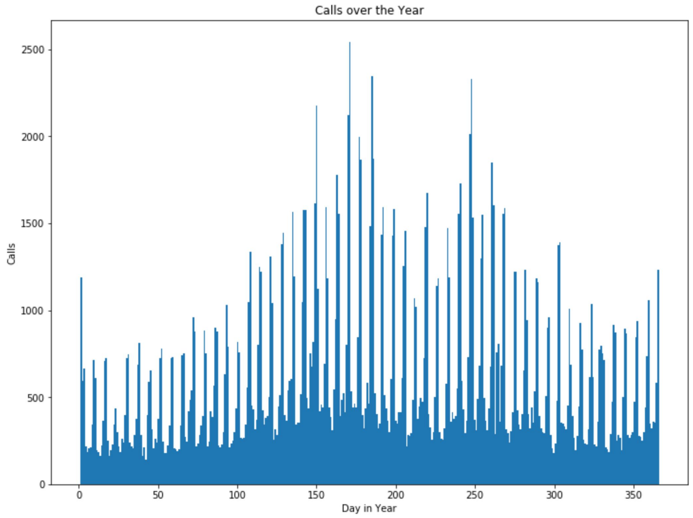
- Used python datetime library to determine weekday values
- 0 is Monday, 6 is Sunday

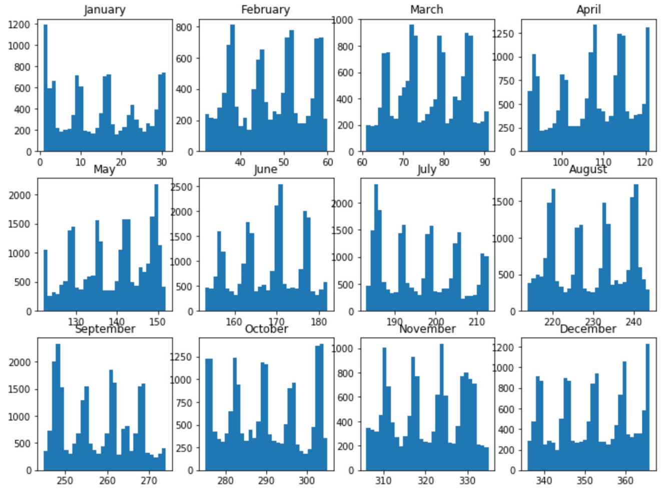
The Model

Developing the Model

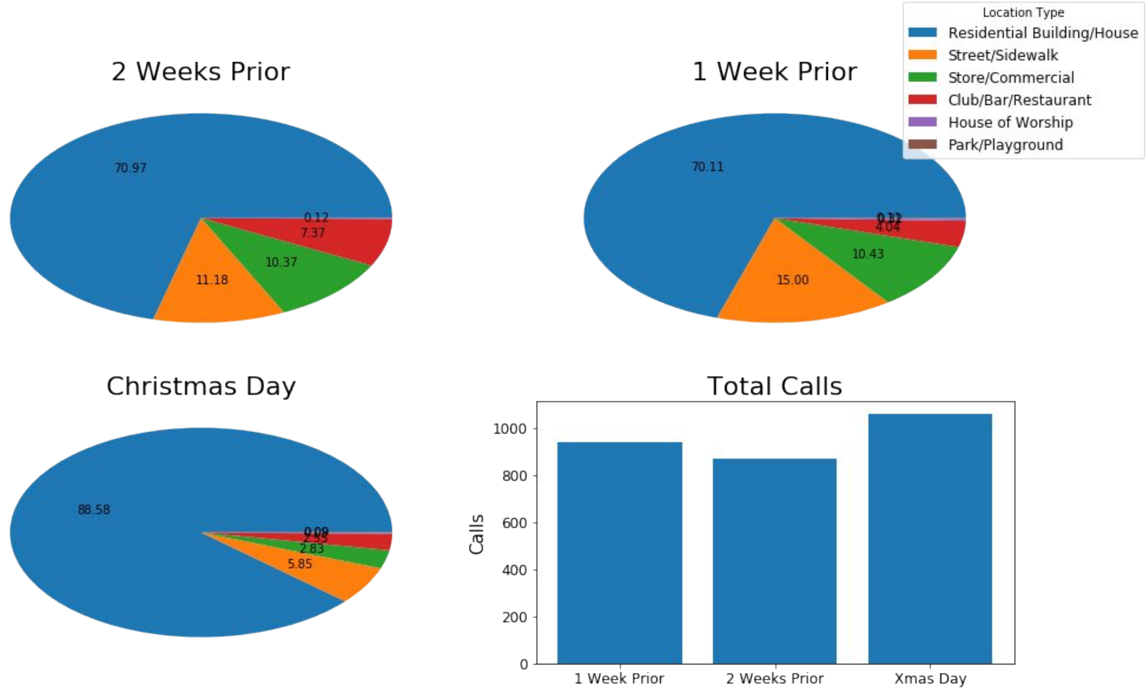
- Modeling Algorithm:
  - User inputs data
  - Data is used to “classify” the kind of parties in that area
  - Classification is done using KNN
- We need to figure out what variables to include:
  - Time (Hour, Weekday, Specific Date)
  - Location (Latitude/Longitude)
  - Location Type (Bar/Restaurant, House, etc.)

Importance of Specific Date

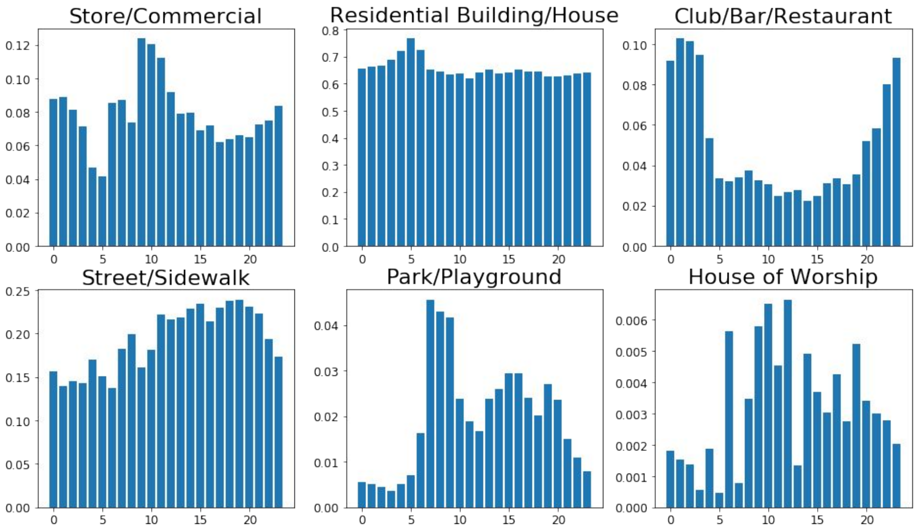




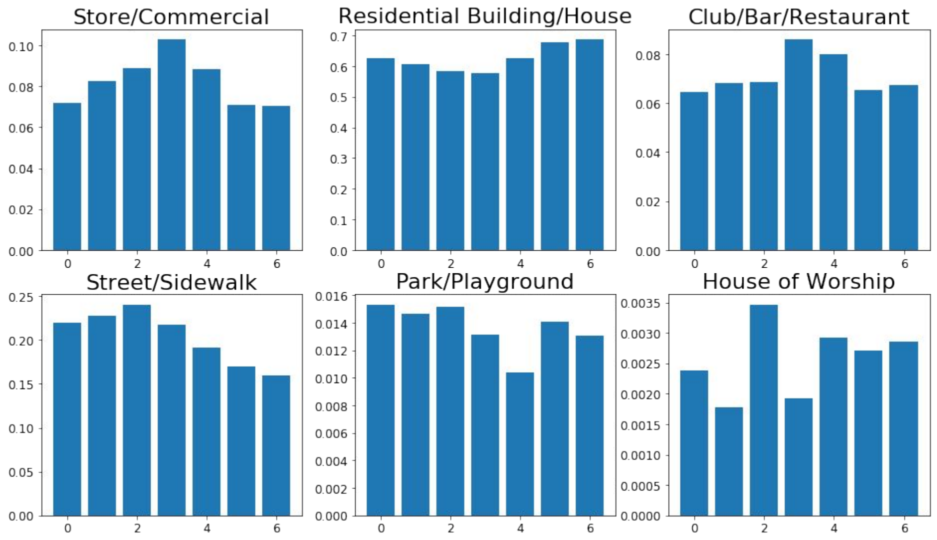
Importance of Specific Date



Importance of Hour and Weekday

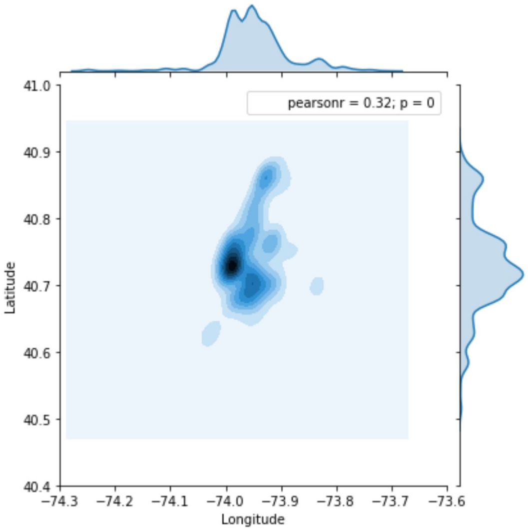


Proportion of Calls by Hour per Party Type

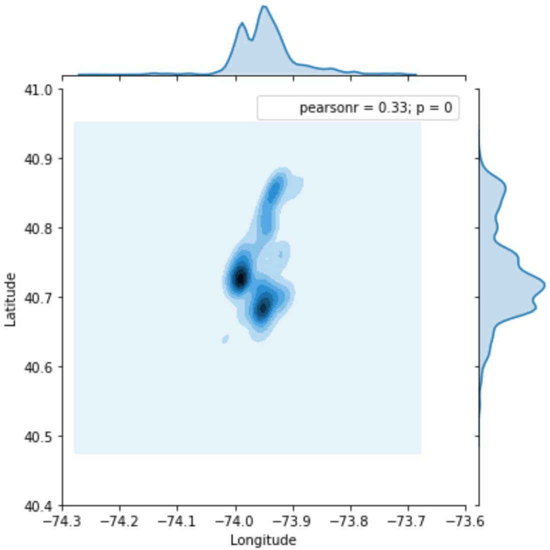


Proportion of Calls by Weekday per Party Type

Importance of Physical Location

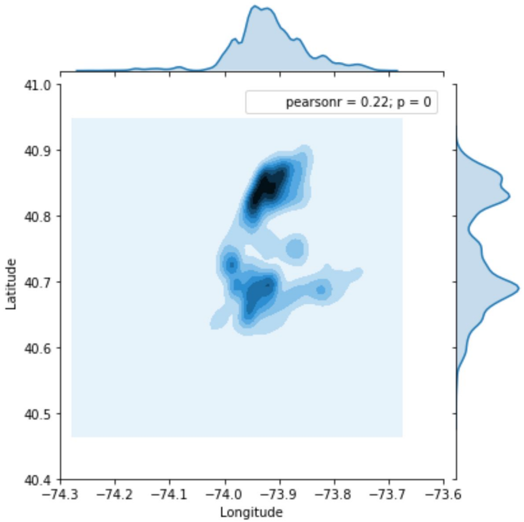


Store/Commercial

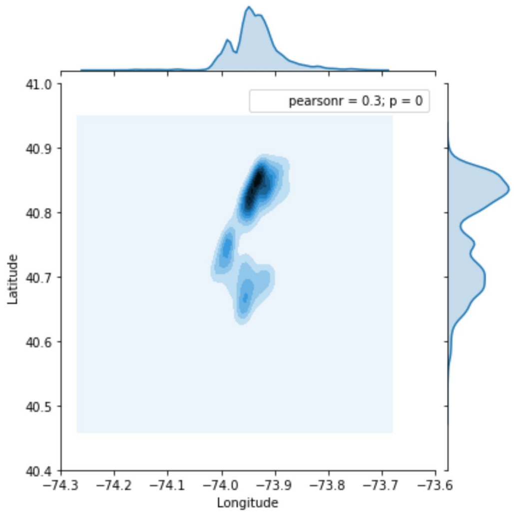


Club/Bar/Restaurant

Importance of Physical Location

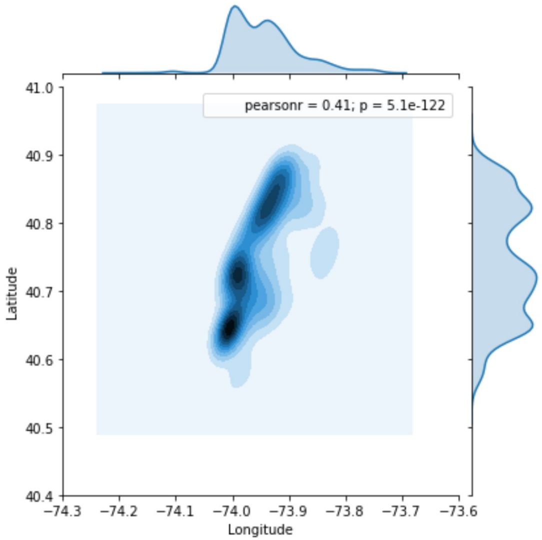


Residential Building/House

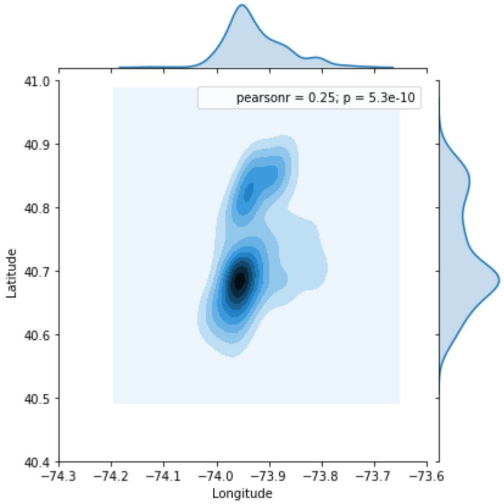


Street/Sidewalk

Importance of Physical Location



Park/Playground



House of Worship

Model Description

- Possible variables include: location (Latitude, Longitude), date, and time
- Model uses a k-nearest-neighbors algorithm to determine the type of party that should be currently available to the user based on noise complaints near that area
- Potential uses:
  - Wanting to find a cool party near your location at a certain time
  - You are moving and want to avoid an area that was a lot of a certain type of party

KNN Model

- Classification Problem
- Trying to accurately determine the party type of a given dataset

- Removed Residential Parties
    - Too many of them
    - Not useful

KNN Model

Location Type	Latitude	Longitude	Month	WeekdayInt	Hour	YearDay
Store/Commercial	40.866183	-73.918930	12	3	0	366
Store/Commercial	40.859324	-73.931237	12	3	0	366
Store/Commercial	40.755584	-73.885201	12	3	0	366
Store/Commercial	40.847007	-73.938193	12	3	0	366
Store/Commercial	40.865978	-73.919531	12	3	0	366

y

X

Training/Test

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)

KNC = KNeighborsClassifier(n_neighbors = K, weights = 'distance')

KNC.fit(X_train, y_train)

KNC.score(X_test, y_test)
```

Cross Fold Validation

KNC = **KNeighborsClassifier**(n_neighbors=**K**, weights = 'distance')

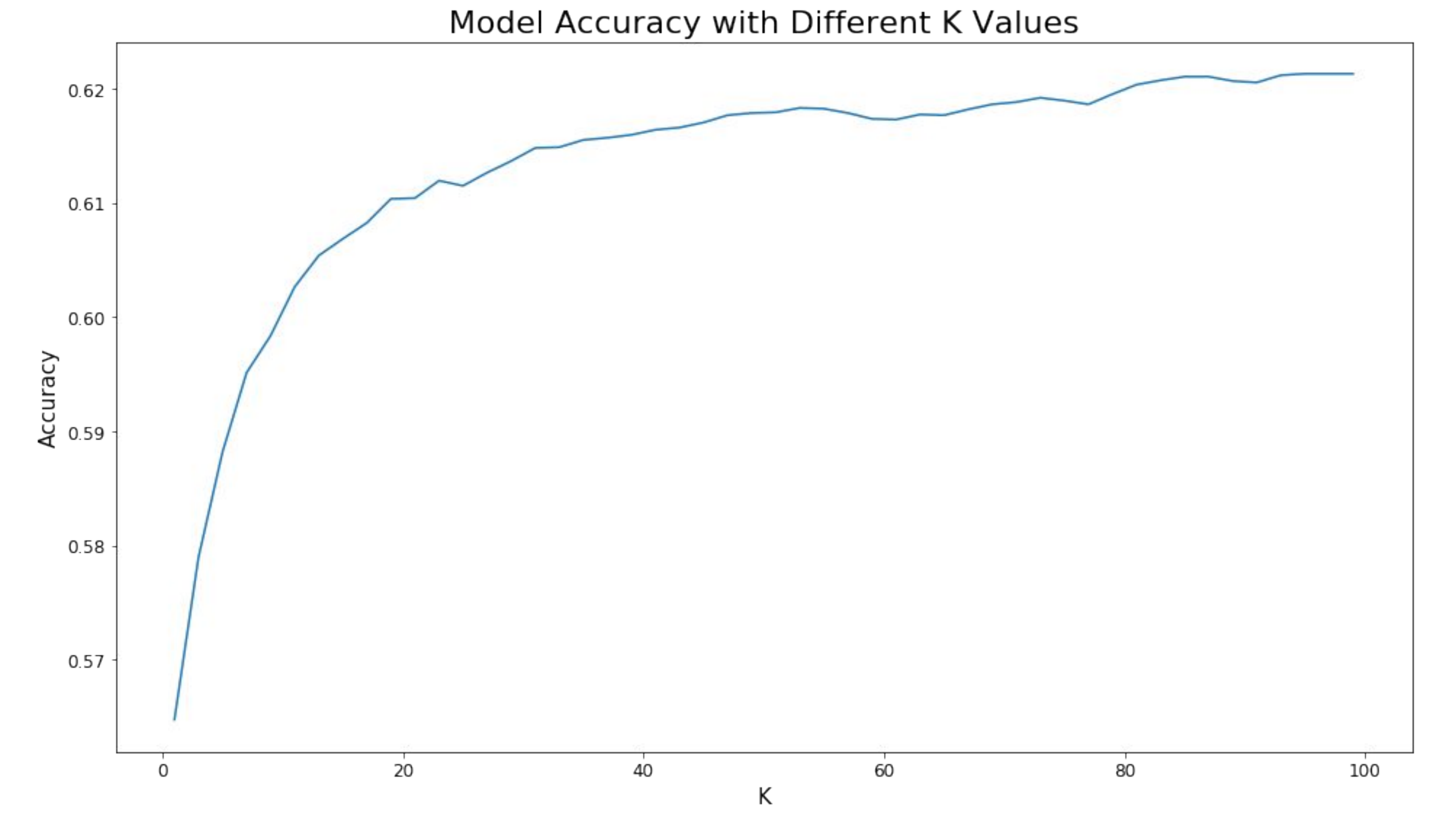
cross_val_score(KNC, X, y, cv=**N**, scoring='accuracy')

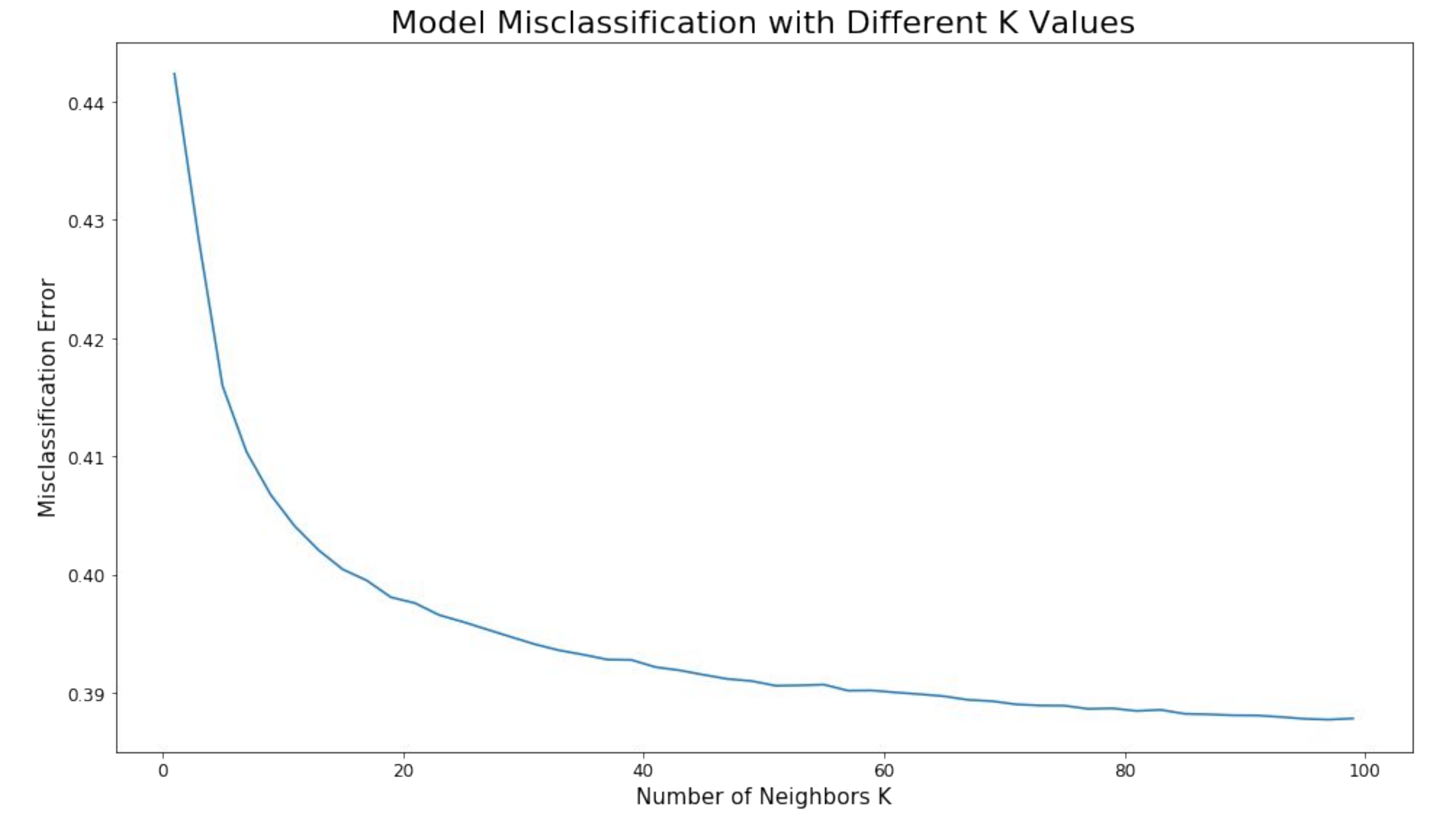
Break the data set randomly into **N** sets.

Each of the **N** sets takes a turn being the test set, while the others are the training set. **N** total tests.

The mean of all **N** tests is the accuracy at our specific **K** level.

Results


Model Accuracy with Different K Values
Accuracy
0.62
0.61
0.60
0.59
0.58
0.57
K
0
20
40
60
80
100



Model Misclassification with Different K Values
Misclassification Error
0.44
0.43
0.42
0.41
0.40
0.39
0
20
40
60
80
100
Number of Neighbors K

What did we learn?

- **Not good that all parameters were normalized, then weighted equally**
    - Could potentially get much better results by determining good relative weights

- **Using physical location as a parameter is problematic**
    - Similar businesses do not always exists in physical clusters

Closing Thoughts

- KNN may not be a great way to find a party.
- NYC, the world’s largest metropolitan area, has A LOT of parties. Who knew?

Questions?