

911 Hotspot Prediction

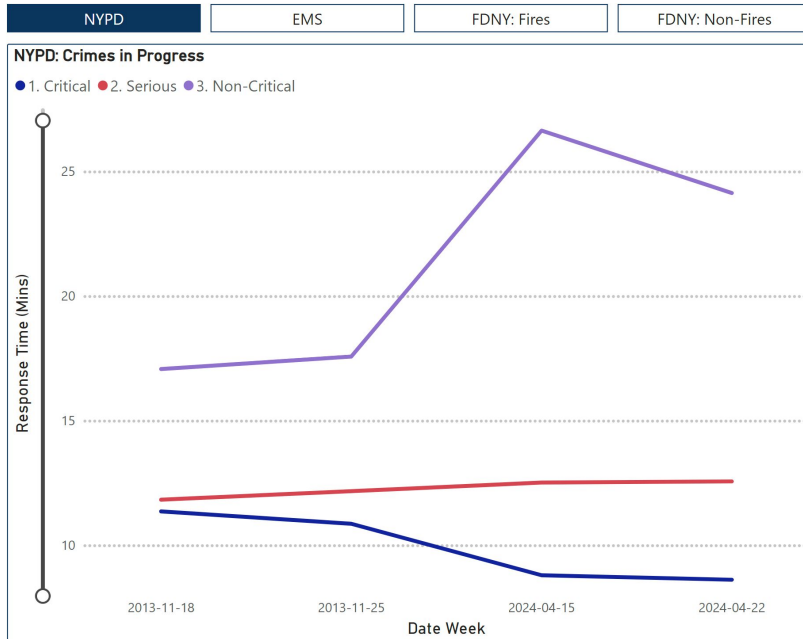
Author: Connell Phillipps





Problem Statement:

Is there any way to reduce 911 call response time by predicting services call locations?



<https://www.nyc.gov/site/911reporting/reports/response-time-trends.page>

Impact:

- Decreasing response time leads to increased public opinion.
- Increase public opinion leads to high crime reporting, more trust.
- Most crimes are not critical so decreasing response time doesn't really lead to less crime.
- Saving lives, even if one life is saved with faster response is worth it!



The Data:

Data Dictionary:

Column Name	Description	Type
CAD_EVT_ID	Unique identifier generated by the the IGAO 911 system	Plain Text T
CREATE_DATE	Date of call	Date & Time D
INCIDENT_DATE	Date of incident	Date & Time D
INCIDENT_TIME	Time of incident	Plain Text T
NYPD_PCT_CD	NYPD precinct call is in	Number #
BORO_NM	Borough call is in	Plain Text T
PATRL_BORO_NM	NYPD patrol Borough call is in	Plain Text T
GEO_CD_X	The X Coordinate of the midblock of the street segment where...	Plain Text T
GEO_CD_Y	The Y Coordinate of the midblock of the street segment where...	Plain Text T
RADIO_CODE	NYPD code used to inform NYPD member of service the nature...	Plain Text T
TYP_DESC	Description based on RADIO_CODE	Plain Text T
CIP_JOBS	Flag indicating if the call relates to a Crime In Progress (CIP)	Plain Text T
ADD_TS	Timestamp of when the call was added to the system	Date & Time D
DISP_TS	Timestamp of when the call was dispatched to a responding u...	Date & Time D
ARRIV_TS	Timestamp of when the responding unit arrived on the scene	Date & Time D
CLOSNG_TS	Timestamp of when the call was marked closed	Date & Time D
Latitude	The Latitude of the midblock of the street segment where the...	Number #
Longitude	The Longitude of the midblock of the street segment where th...	Number #



Merged Dataset:

```
RangeIndex: 42473259 entries, 0 to 42473258
Data columns (total 20 columns):
#   Column              Non-Null Count  Dtype
---  -
0   OBJECTID             6421740 non-null float64
1   CAD_EVT_ID           42473259 non-null int64
2   CREATE_DATE          42473259 non-null object
3   INCIDENT_DATE        42473259 non-null object
4   INCIDENT_TIME        41025977 non-null object
5   NYPD_PCT_CD         42473201 non-null float64
6   BORO_NM              42473259 non-null object
7   PATRL_BORO_NM        42473259 non-null object
8   GEO_CD_X             42473259 non-null int64
9   GEO_CD_Y             42473259 non-null int64
10  RADIO_CODE           42473259 non-null object
11  TYP_DESC             42473259 non-null object
12  CIP_JOBS             42473259 non-null object
13  ADD_TS               42473259 non-null object
14  DISP_TS              42473257 non-null object
15  ARRIV_TS             27602480 non-null object
16  CLOSNG_TS            42473130 non-null object
17  Latitude              42473259 non-null float64
18  Longitude             42473259 non-null float64
19  Location              40667946 non-null object
```

12.3GB

Preprocessed Dataset:

```
RangeIndex: 41021298 entries, 0 to 41021297
Data columns (total 8 columns):
#   Column              Non-Null Count  Dtype
---  -
0   CAD_EVT_ID           41021298 non-null int64
1   NYPD_PCT_CD         41021298 non-null category
2   BORO_NM              41021298 non-null category
3   RADIO_CODE           41021298 non-null category
4   CIP_JOBS             41021298 non-null category
5   ADD_TS               41021298 non-null datetime64[ns]
6   Latitude              41021298 non-null float64
7   Longitude             41021298 non-null float64
```

3.2GB

Initial Findings: Call Distribution by Category

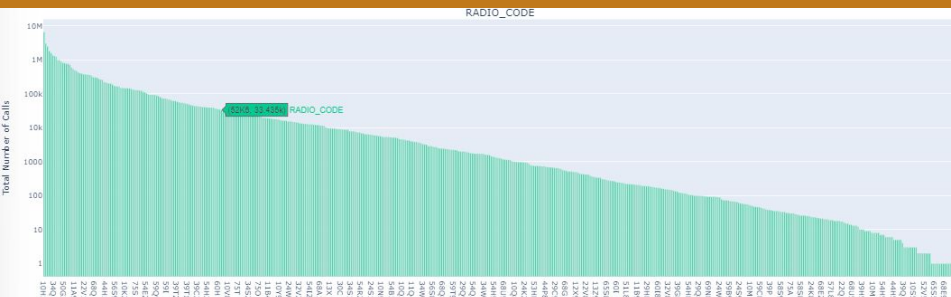
NYPD Precinct



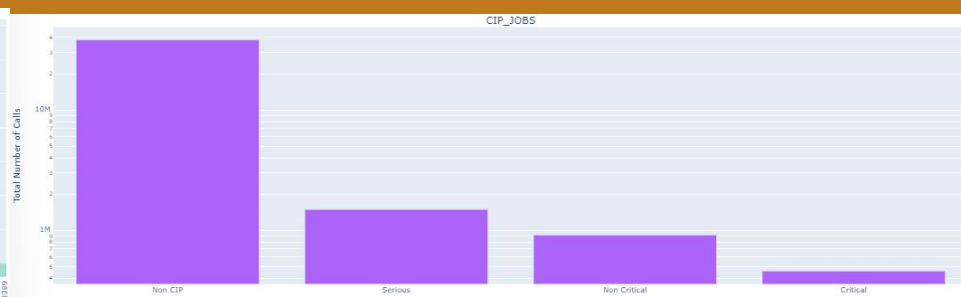
Boroughs



Radio Code



Crime in Progress

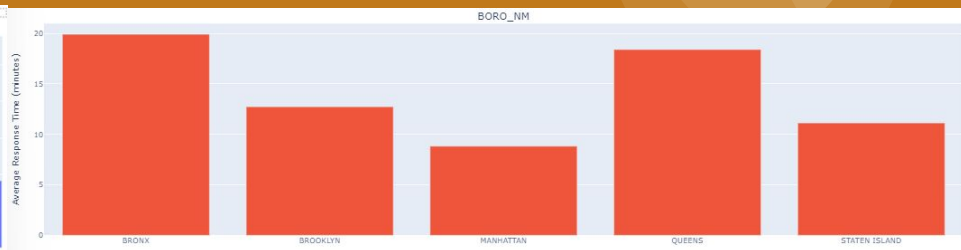


Initial Findings: Response Time by Category

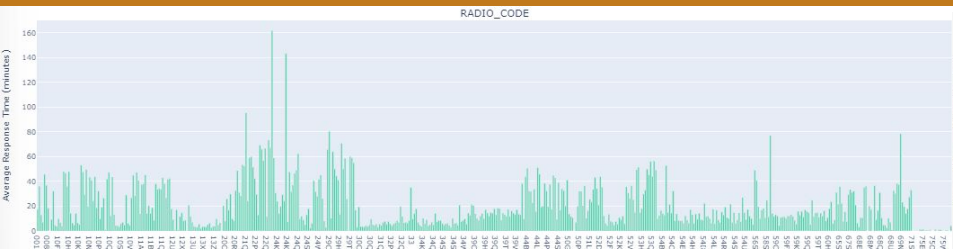
NYPD Precinct



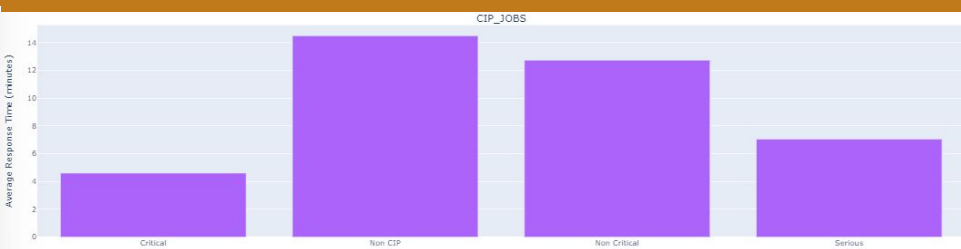
Boroughs



Radio Code



Crime in Progress



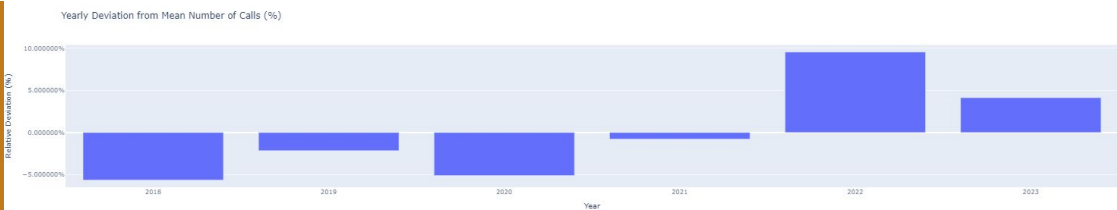
Initial Findings: Deviation from the Mean



Hourly

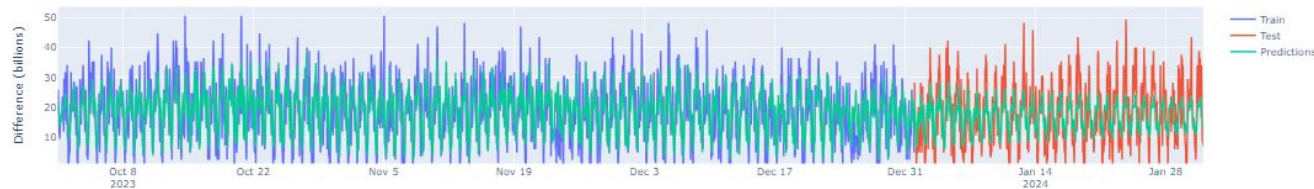


Monthly

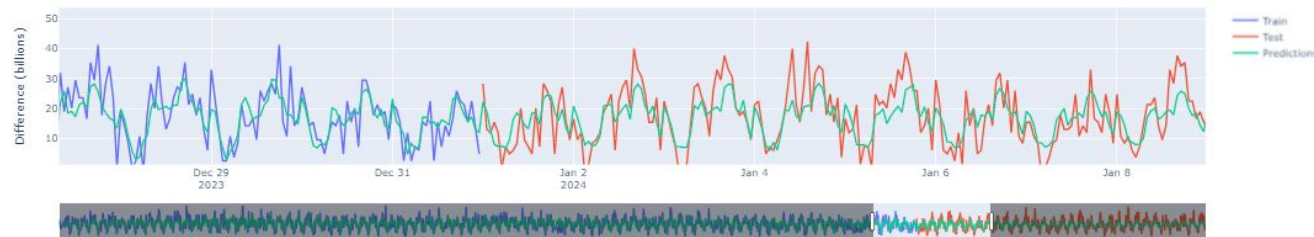


Forecasting

Change in Revenue Passenger Miles over Prior Year



Change in Revenue Passenger Miles over Prior Year



Evaluation Metric:

- Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|A_i - F_i|}{A_i}$$

A_i is the actual value

F_i is the forecast value

n is total number of observations

Baseline MAPE: 38.8% (prediction is just mean number of calls)

Model MAPE: 23.95% (without differencing added back in)



Next Steps:

- Add the trend and seasonality back into the predictions for an hourly timescale
- Create a function to loop through columns of a database and create predictive models for each column
- Graph predictions on a heat map of NYC
- Create new predictive model to forecast based on lat/lon
- Build streamlit/Rshiny App