Predicting Emergency Call Volume in NYC

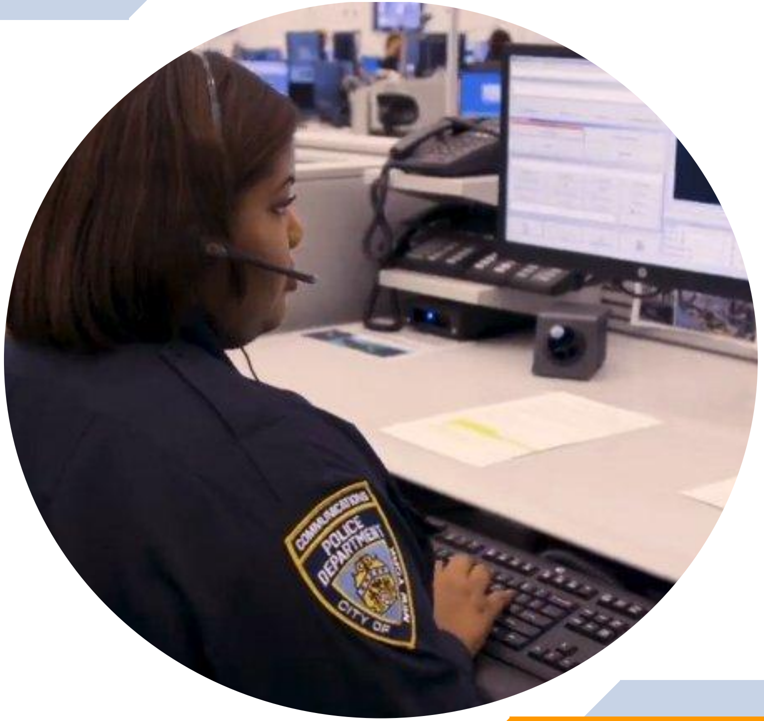
Ashley Poon, Chris Caress, Christina Holland, David Romo

PROBLEM AT HAND

Can we predict emergency call volume in NYC using time, weather, and traffic data?

While these features are beyond our control, the goal of this project is to increase preparedness for a surge in emergency services given specific conditions.

Using predictive modeling, we hope to be able to provide an indication of when the city should increase staff in emergency call centers and response teams as well as forecast potential staffing requirements.



Source: NYPD

DATA COLLECTION

EMS

Using NYC's open data source, we were able to collect EMS incident dispatch data from 2010 to 2016.

We then calculated the number of calls received on a given day and hour.

Weather

From NOAA, we requested daily weather data collected from the Central Park station - specifically, temperature and precipitation.

We then joined this weather data with our EMS data by day and hour.

Traffic Incidents

From NY State's open data source, we found historical traffic and transit events.

We calculated the number of traffic incidents on a given day and hour and joined this data with our EMS + weather data.

DATA DICTIONARY

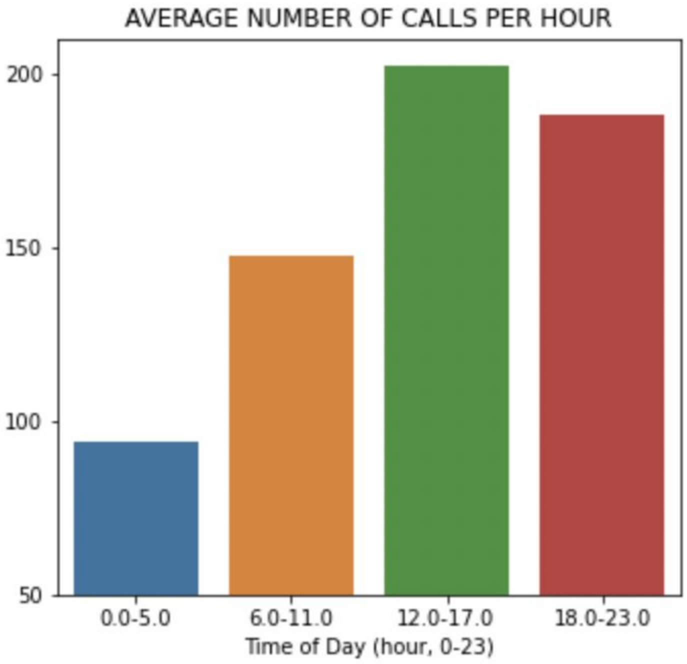
Column	Definition
num_calls	number of total NYC EMS calls within the hour
year	year (2010-2019)
month	month of the year (1-12)
day	day of the month (1-31)
hour	hour of the day (0-23)
PRCP	precipitation in inches
SNOW	snow (falling) in inches
SNWD	snow depth on ground in inches
TAVG_CALC	(max temp + min temp)/2 in Fahrenheit
Traffic Incidents	number of traffic incidents within the hour

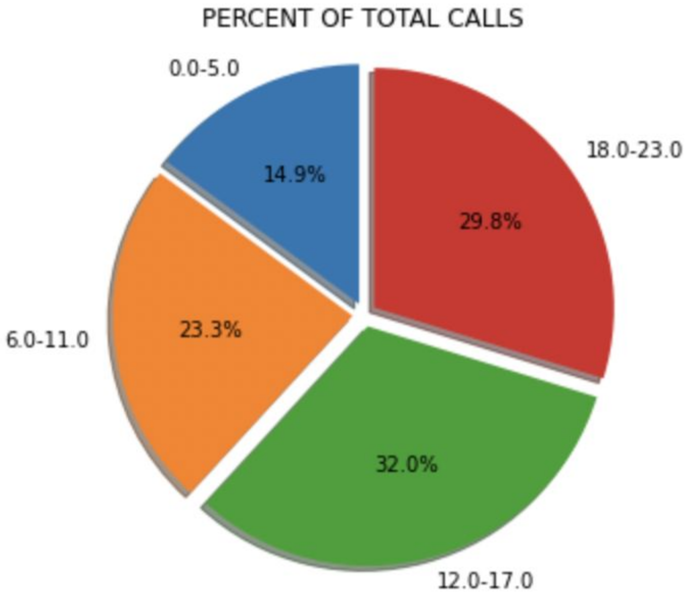
Exploratory Data Analysis

Averages for Each Quartile of Call Volume

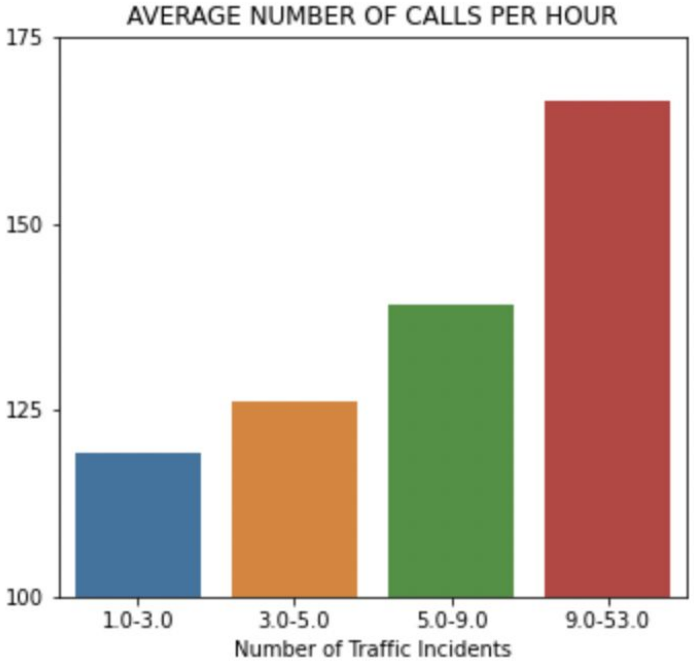
quartile	call volume	hour	temperature	traffic incidents	day	month	precipitation	snow depth	snowfall
1	73.295172	4.279836	56.067340	5.566657	15.877467	6.492104	0.144763	0.433556	0.088415
2	106.184019	4.732688	59.072881	4.718805	15.779500	6.614689	0.127019	0.356158	0.073801
3	157.022159	9.344096	57.874802	6.286800	15.829218	6.252770	0.137167	0.365654	0.092308
4	210.337575	12.408420	59.448790	10.153629	15.304430	5.989161	0.126588	0.453063	0.093230

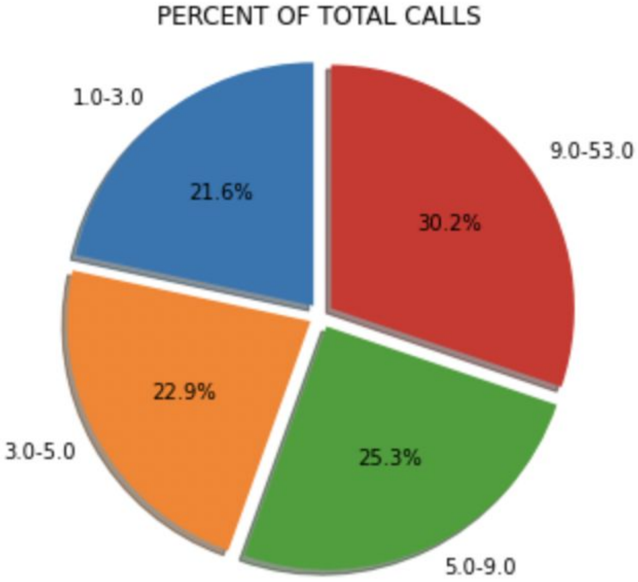
Impact of Time of Day



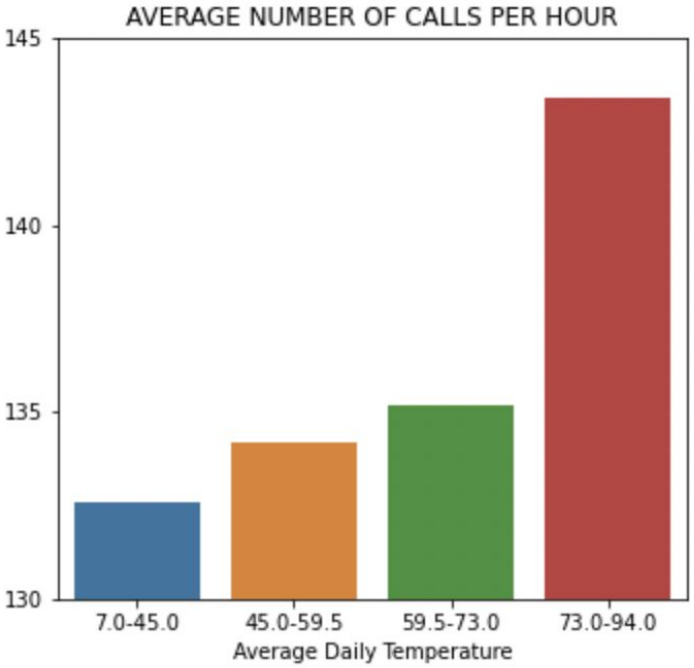


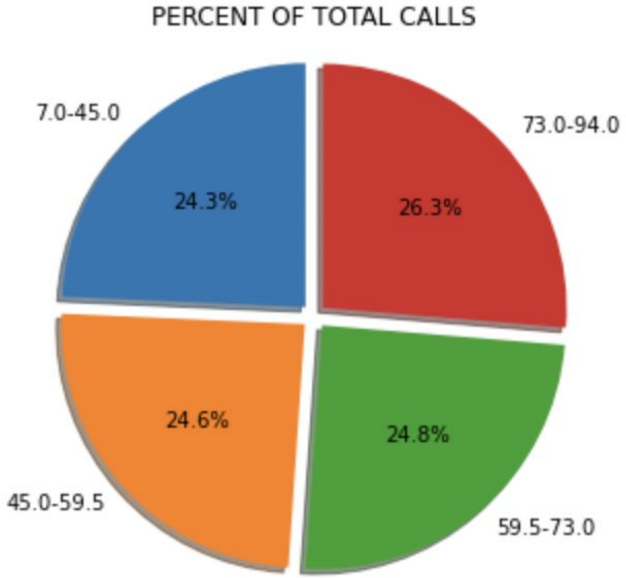
Impact of Traffic Incidents





Impact of Temperature





Modeling

MODELING PROCESS

Linear Regression

Random Forest

Neural Net

MODELING PROCESS

Linear Regression

Random Forest

Neural Net

LINEAR REGRESSION

- Default parameters
- R2 scores
  - Test: 0.52
  - Train: 0.51
- No overfitting but weak model

	coefs
hour	34.665246
Incidences	10.747473
TAVG_CALC	3.471342
SNWD	1.585477
month	0.925722
year	0.690116
SNOW	0.458211
PRCP	-0.606640
day	-1.185278

MODELING PROCESS

Linear Regression

Random Forest

Neural Net

MODELING PROCESS

Linear Regression

Random Forest

Neural Net

RANDOM FOREST

- Default parameters
  - Test score: 0.83
  - Train score: 0.98

overfit

- Best parameters (max_depth: 5, n_estimators: 250)
  - Test score: 0.81
  - Train score: 0.81
  - RMSE: 24 calls/hr

not overfit

	importance
hour	0.787306
TAVG_CALC	0.054226
Incidences	0.043073
day	0.040666
month	0.026652
year	0.024473
PRCP	0.017919
SNWD	0.004524
SNOW	0.001160

MODELING PROCESS

Linear Regression

Random Forest

Neural Net

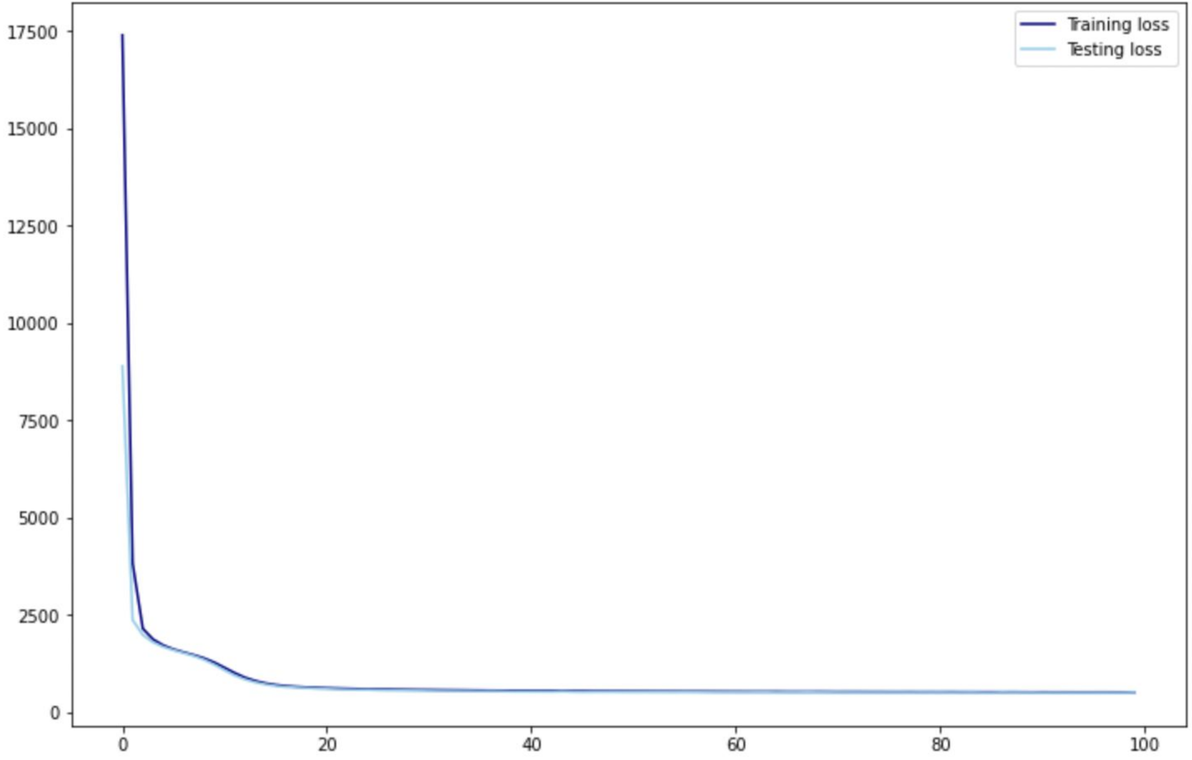
MODELING PROCESS

Linear Regression → Random Forest → **Neural Net**

NEURAL NET

- 3 dense layers (nodes: [128, 64, 64] , batch_size: 100 , epochs: 250)
  - Test score: 0.84
  - Train score: 0.86
  - RMSE: 22 calls
- Minimal overfitting and strongest model

NEURAL NET



- 3 dense layers

nodes: [128, 64, 64]

batch_size: 100

epochs: 250

loss: mean squared error

metric: mean average error

R-Squared Performance Summary

52% / 51%

Linear Regression

81% / 81%

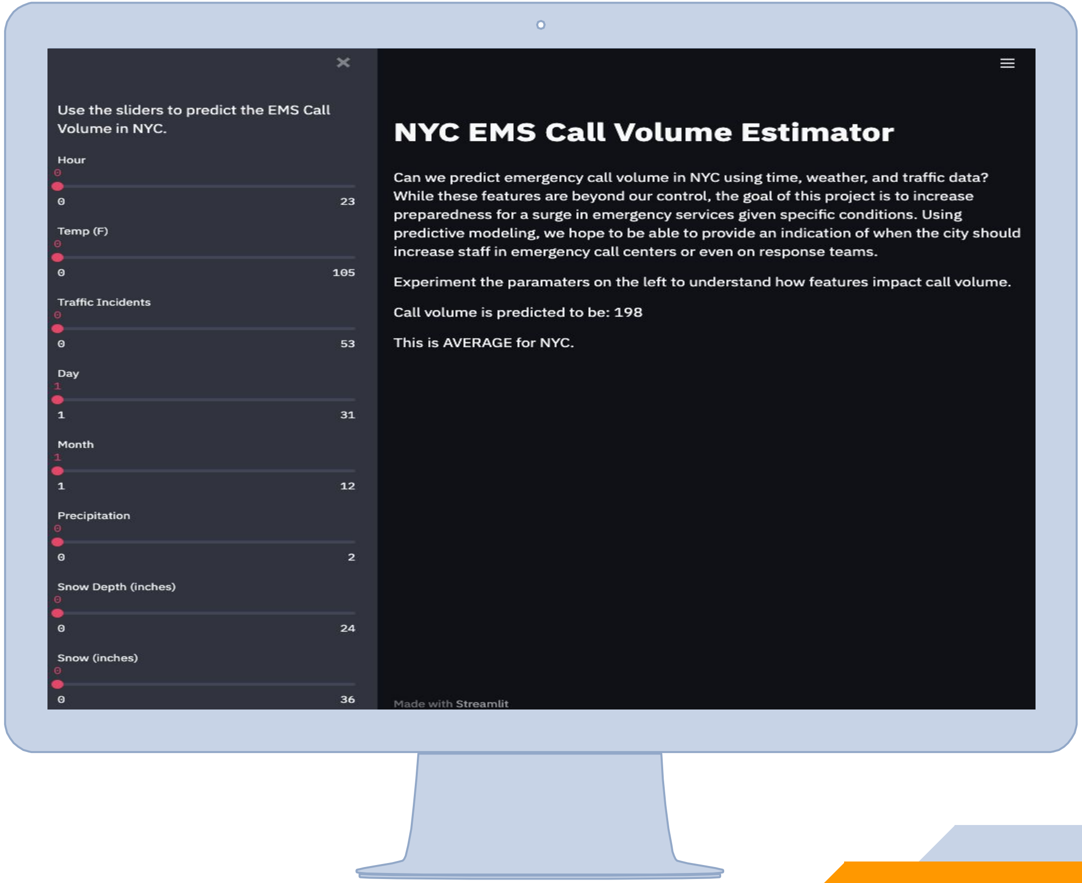
Random Forest

84% / 86%

Neural Net

Test / Train Score

DEMO


Use the sliders to predict the EMS Call Volume in NYC.
Hour
0
0
23
Temp (F)
0
0
105
Traffic Incidents
0
0
53
Day
1
1
31
Month
1
1
12
Precipitation
0
0
2
Snow Depth (inches)
0
0
24
Snow (inches)
0
0
36
NYC EMS Call Volume Estimator
Can we predict emergency call volume in NYC using time, weather, and traffic data? While these features are beyond our control, the goal of this project is to increase preparedness for a surge in emergency services given specific conditions. Using predictive modeling, we hope to be able to provide an indication of when the city should increase staff in emergency call centers or even on response teams.
Experiment the paramaters on the left to understand how features impact call volume.
Call volume is predicted to be: 198
This is AVERAGE for NYC.
Made with Streamlit

Conclusion

Summary

Our neural net model accounts for about 84% of variance. We start to see inaccurate predictions at the extreme high end of call volume. Other than that, the error is well within what we would expect.

The reason that this model is not designed for interpretability is because there is no control that can really be had over the inputs (weather, traffic, and time).

While there will always be emergency situations such as natural weather events, national emergency situations, etc. that cannot be properly accounted for, the goal of this project is to help predict and staff for more run-of-the-mill situations and make sure that the EMS team is allocated effectively.

	num_calls	preds	error
count	6367.000000	6367.000000	6367.000000
mean	136.157845	137.320007	1.162229
std	54.344129	49.156960	22.580943
min	36.000000	57.355877	-280.274490
25%	88.000000	92.681919	-11.937038
50%	127.000000	124.252640	2.716019
75%	185.000000	187.427269	15.743923
max	377.000000	252.046219	91.911209

Recommendation

Use weather forecasts and our model to create a ‘heatmap’ of staff needed and schedule people accordingly based on projected calls. This will help to allow efficient staffing while still preventing understaffing during times of high call volumes. If implemented well, this should help to avoid preventable deaths and account for stress on the system.

Next Steps

- Use historical traffic density and traffic incident type to improve our model predictability and decrease the RMSE. Currently, we know how many incidents happen but not the severity or the overall effect these have on traffic.
- We would also like to see research into ‘average handling times’ for EMS. This would ideally us better predict how much staff is needed.

THANKS!

Any questions?