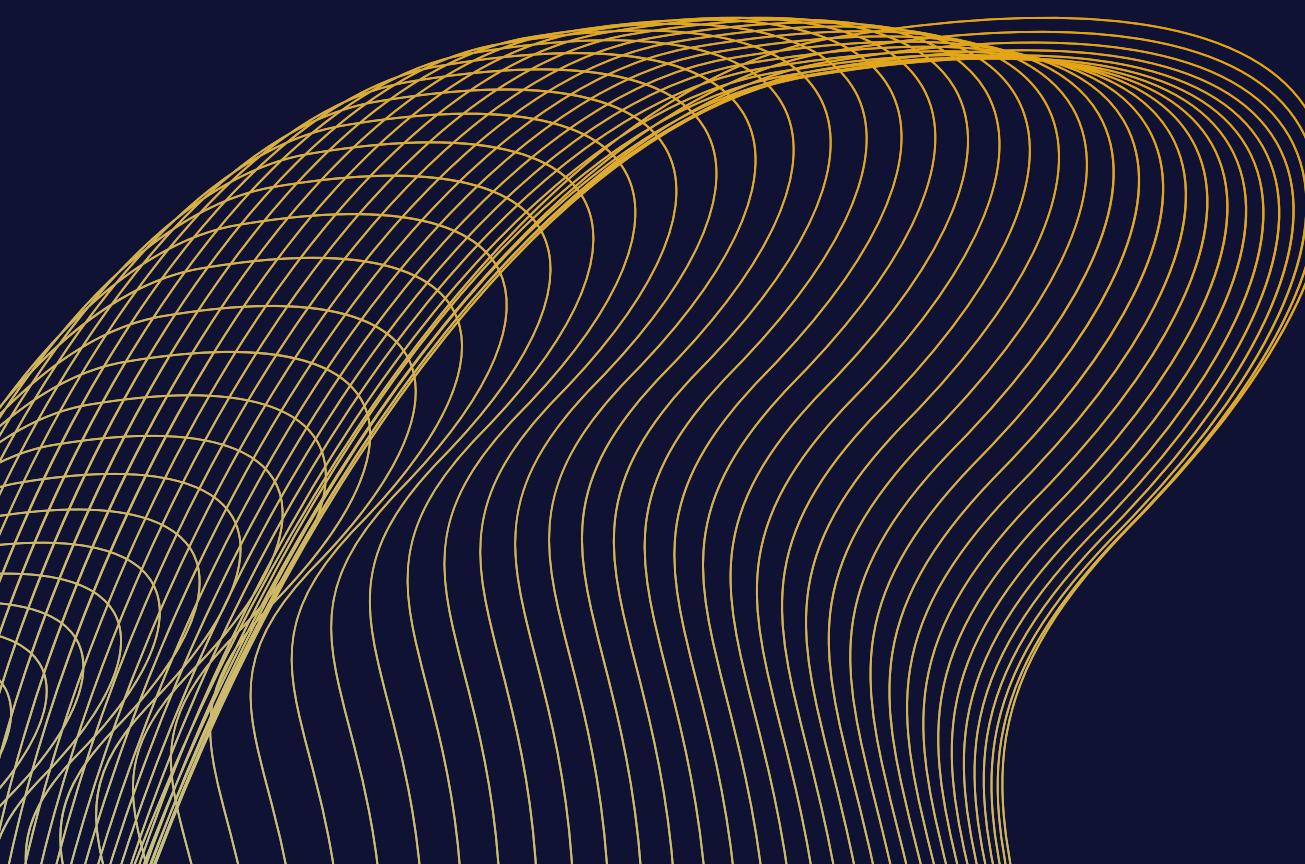


# 911 Call Volume Forecasting

Author: Connell Phillipps

---

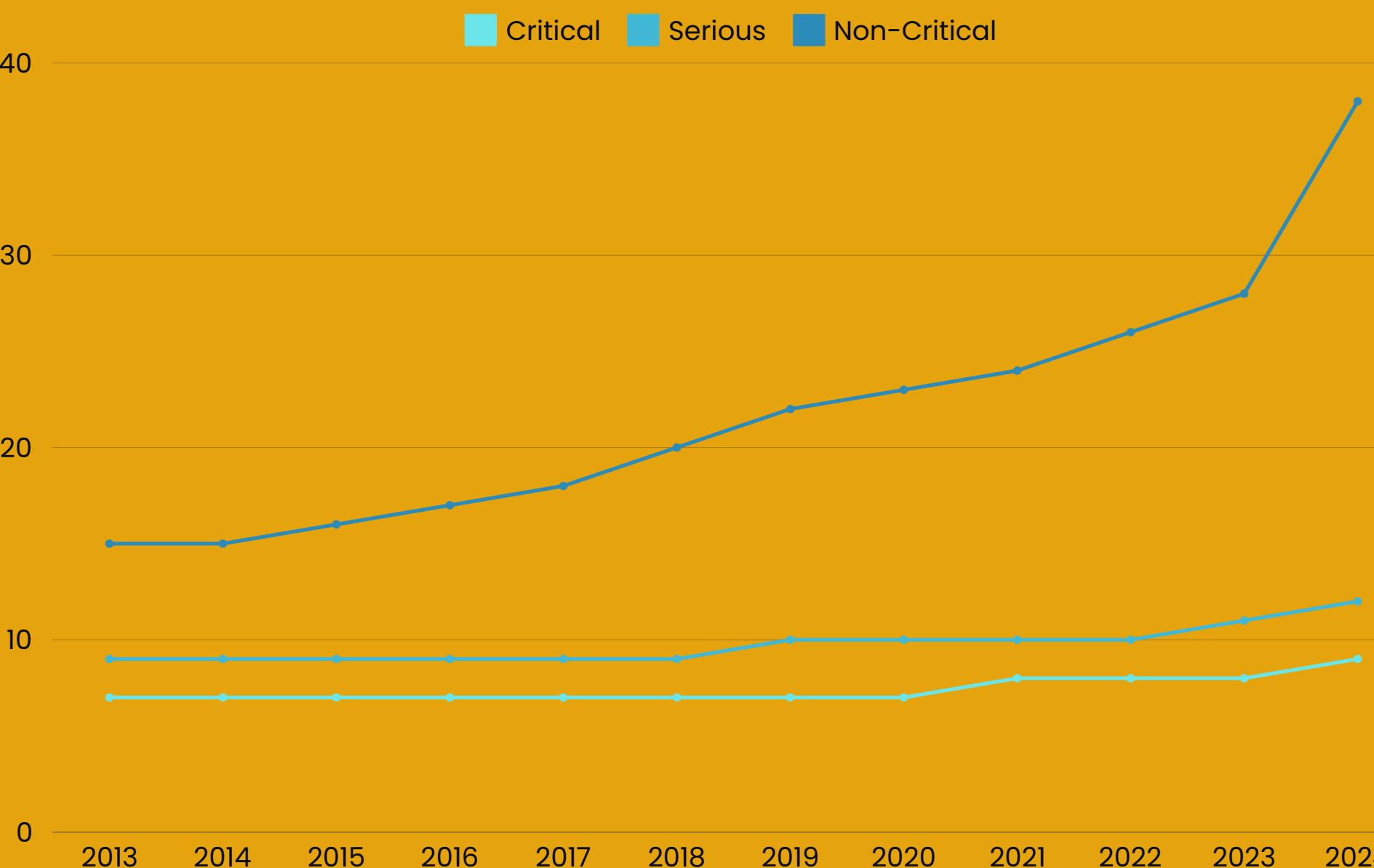




# Problem Statement

Is there any way to reduce 911 call response time by predicting call volume by location?

## Average Response Times - ([nyc.gov](https://nyc.gov))



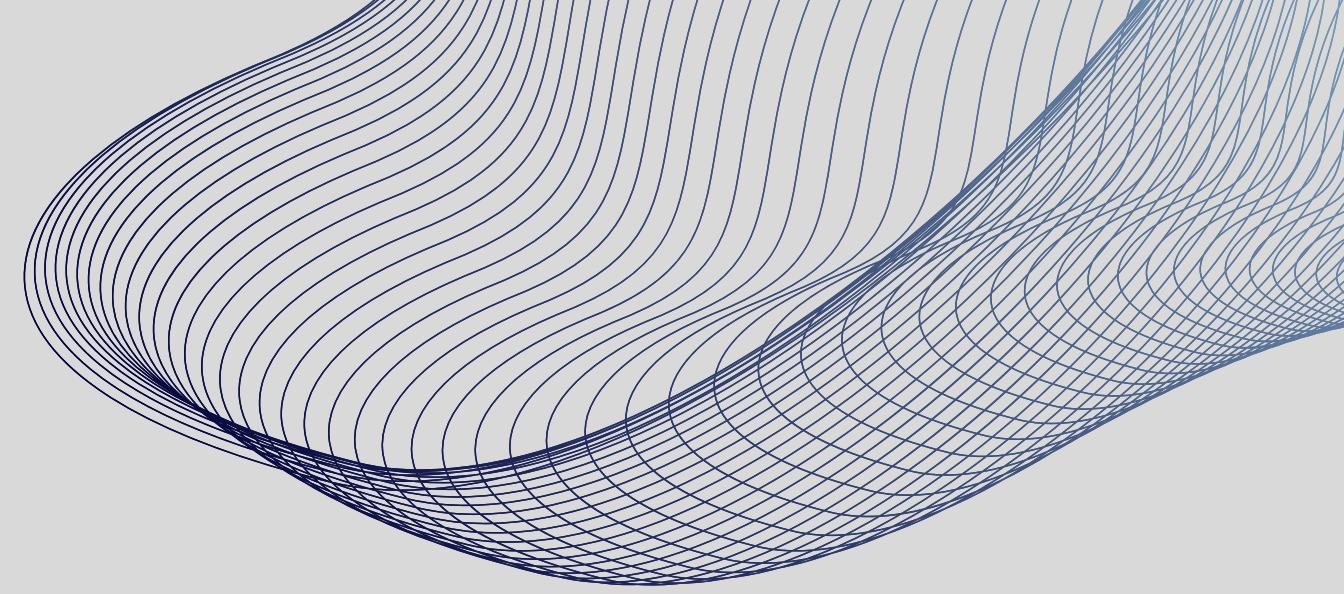
## Impact:

- Increased Public Opinion
- Improved Arrival scene
- Increase reporting



[Source](#)

# Data Cleaning



## Stage 1: Separated API Pulls

Rows	Columns
40.7M	20
Columns (20)	
Column Name	Description
# OBJECTID	objectid
# CAD_EVNT_ID	cad_evnt_id
CREATE_DATE	create_date
INCIDENT_DATE	incident_date
INCIDENT_TIME	incident_time
# NYPD_PCT_CD	nypd_pct_cd
BORO_NM	boro_nm
PATRL_BORO_NM	patrl_boro_nm
GEO_CD_X	geo_cd_x
GEO_CD_Y	geo_cd_y
RADIO_CODE	radio_code
TYP_DESC	typ_desc
CIP_JOBS	cip_jobs
ADD_TS	add_ts
DISP_TS	disp_ts
ARRIVD_TS	arrivd_ts
CLOSNG_TS	closng_ts
Latitude	latitude
Longitude	longitude
Location	location

[Source](#)

## Stage 2: Merged Dataset (12.3GB)

```
RangeIndex: 42473259 entries, 0 to 42473258
Data columns (total 20 columns):
 #   Column           Non-Null Count   Dtype  
 ---  -- 
 0   OBJECTID        6421740 non-null    float64
 1   CAD_EVNT_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  ARRIVD_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 
dtypes: float64(4), int64(3), object(13)
memory usage: 6.3+ GB
```

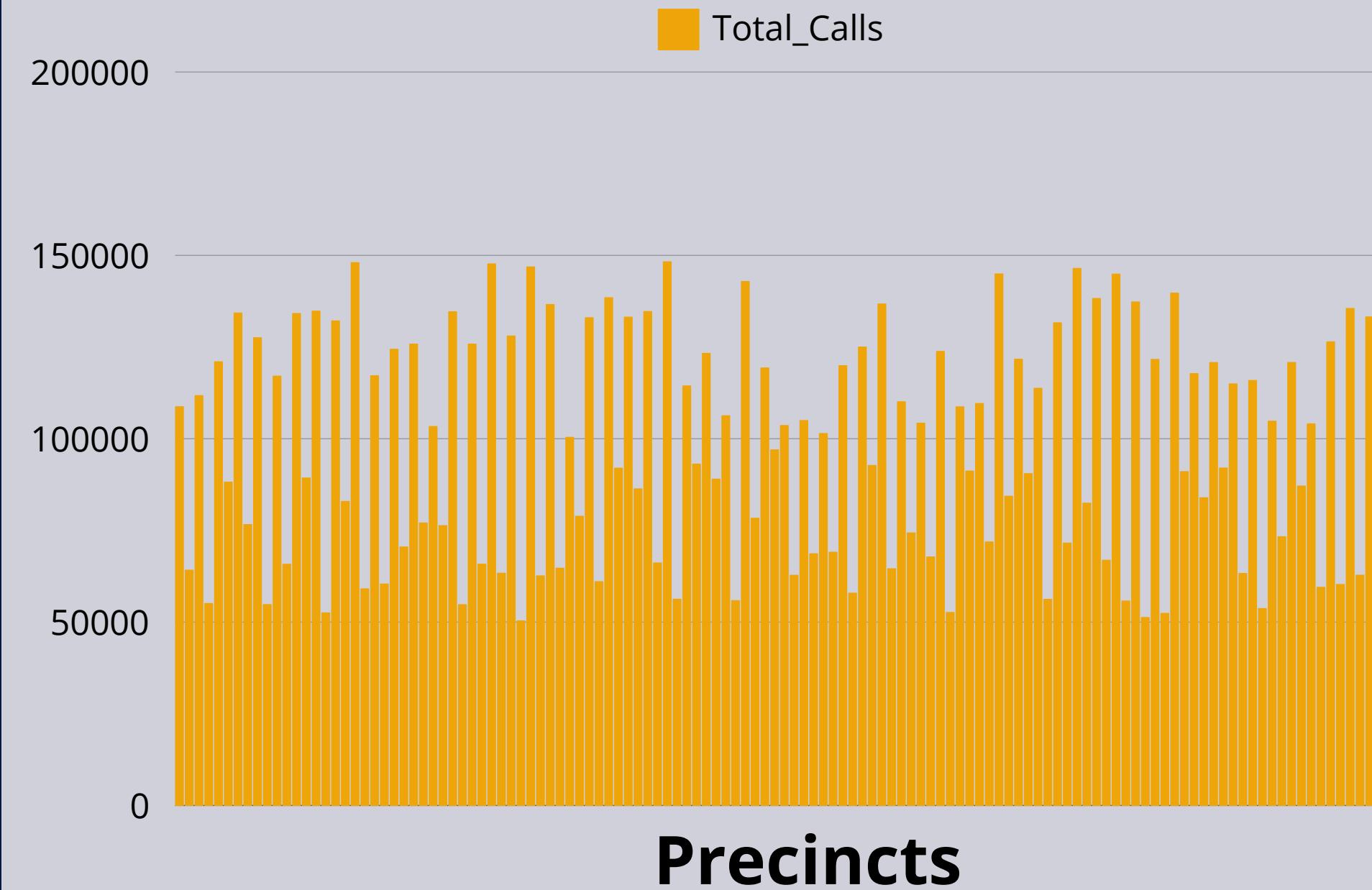
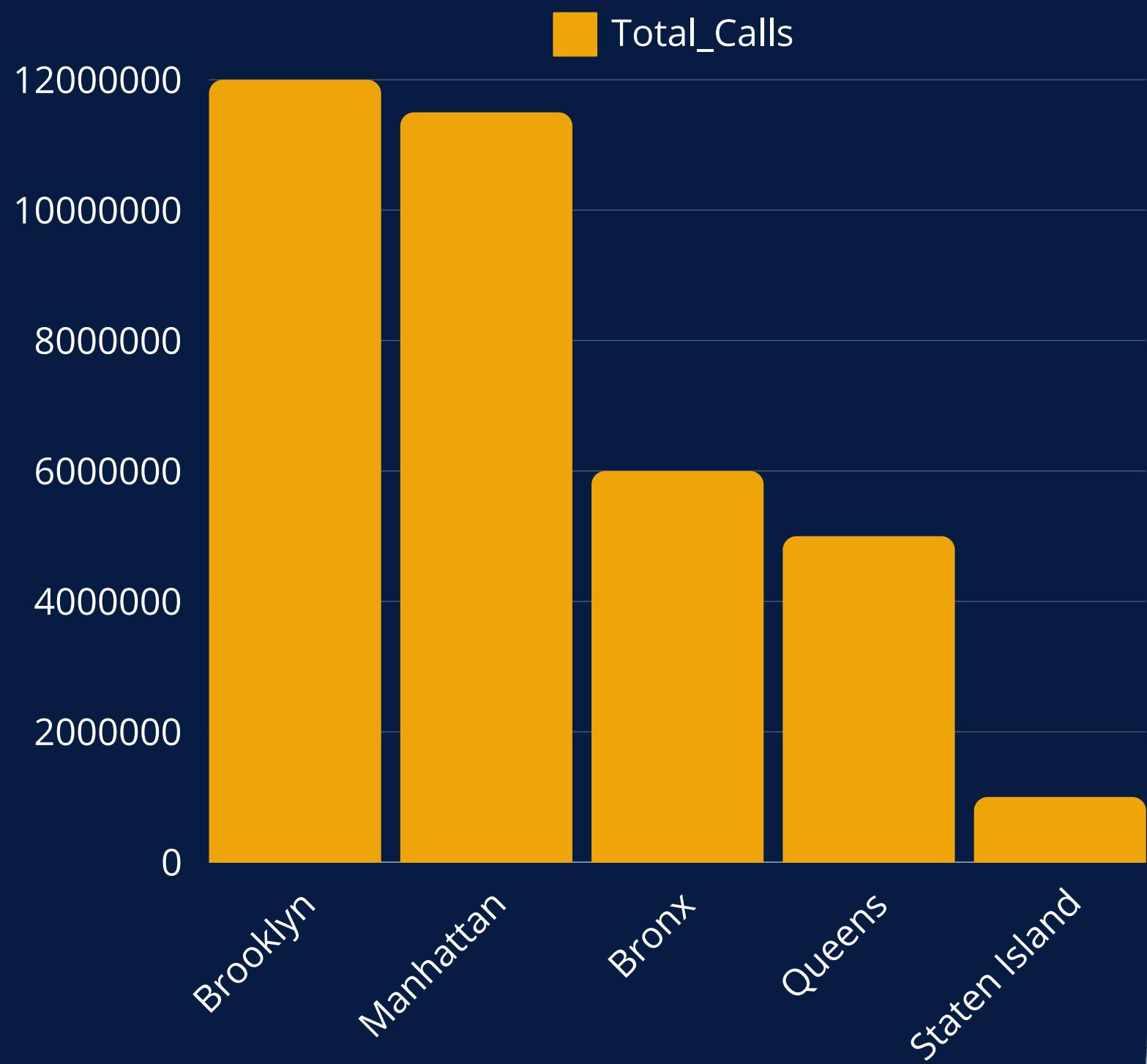
## Stage 3: Cleaned Data (3.2GB)

```
RangeIndex: 41021298 entries, 0 to 41021297
Data columns (total 8 columns):
 #   Column           Non-Null Count   Dtype  
 ---  -- 
 0   CAD_EVNT_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
dtypes: category(4), datetime64[ns](1), float64(2), int64(1)
memory usage: 1.4 GB
```



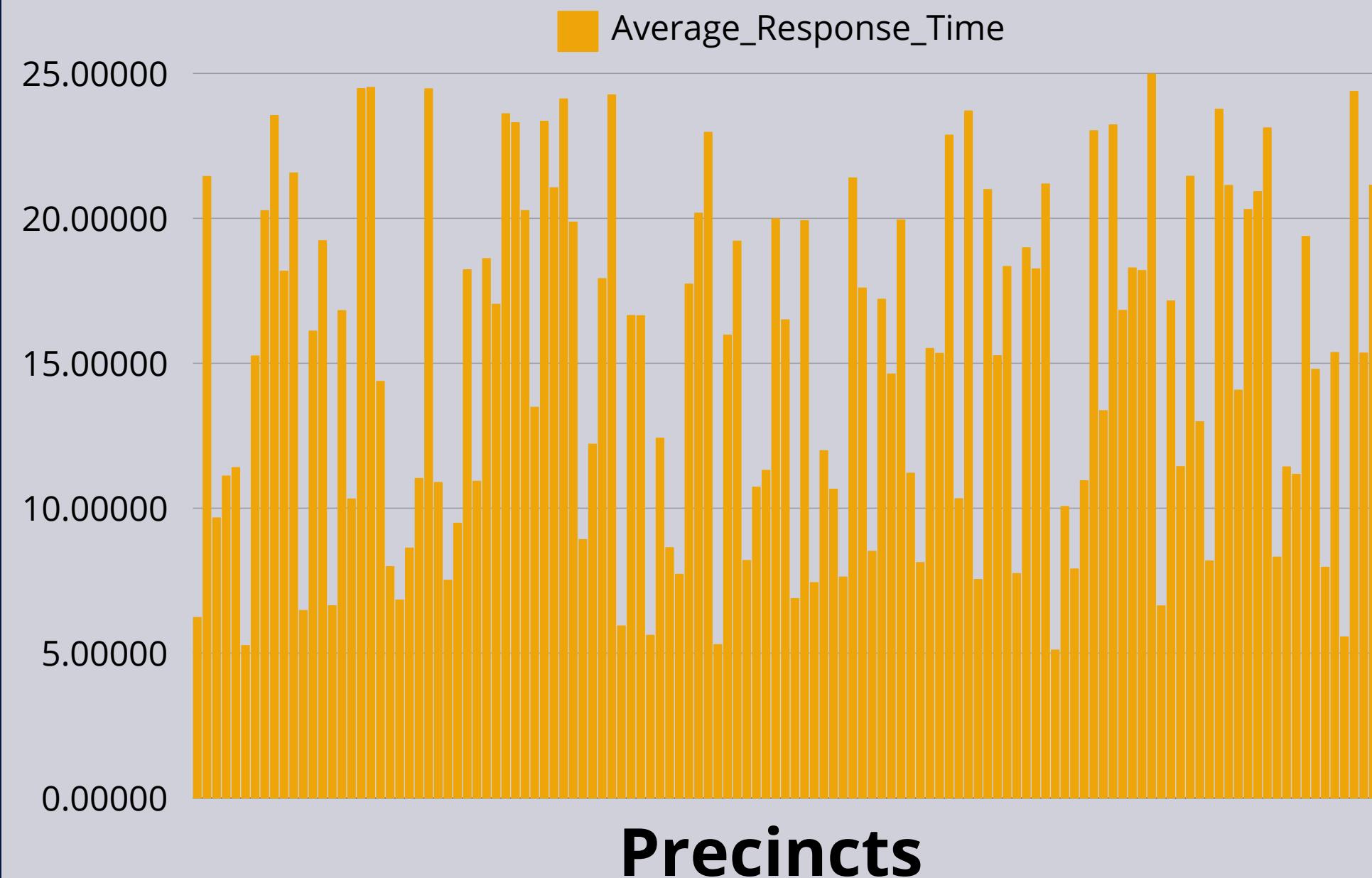
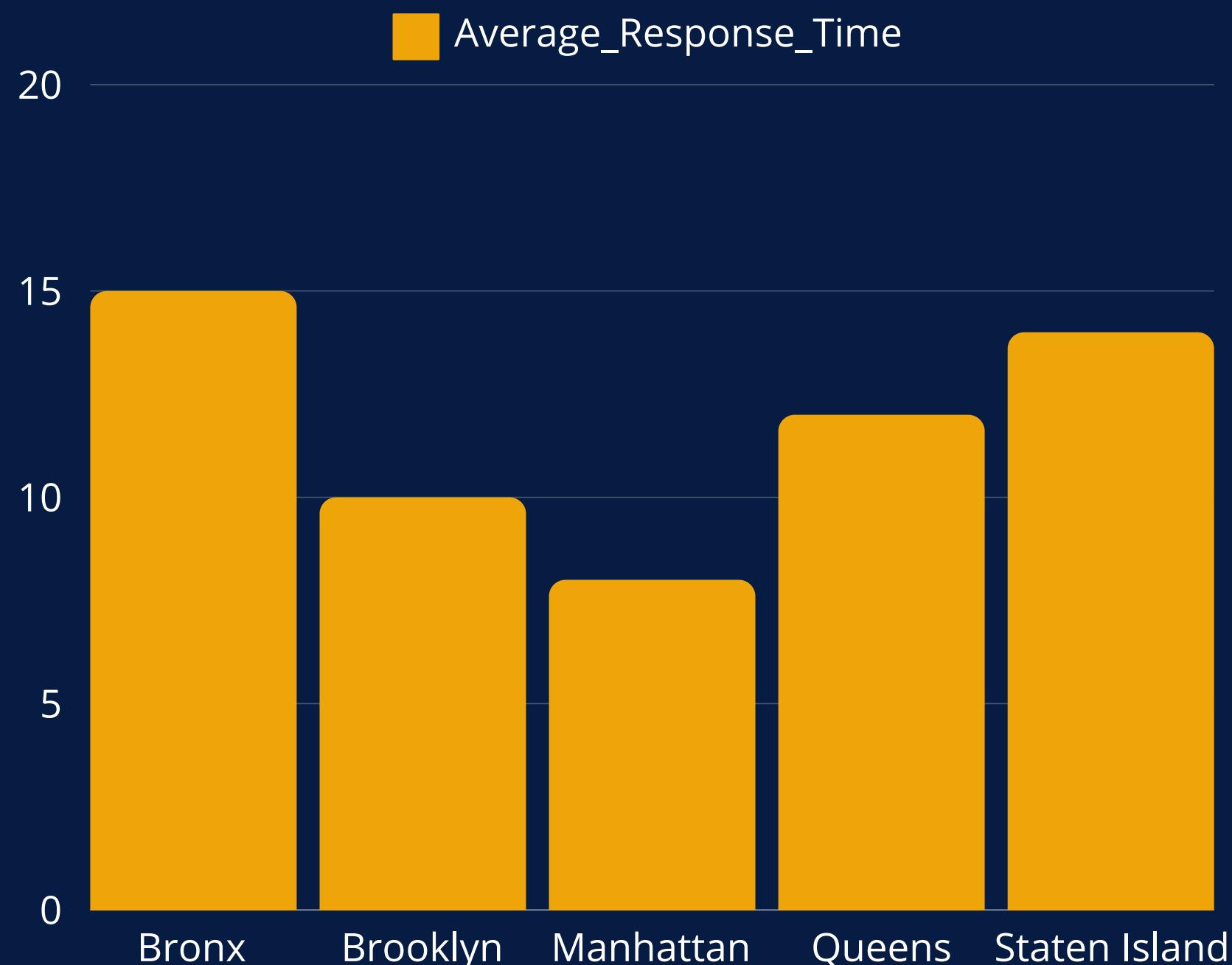
# Initial Findings

## Call Distributions

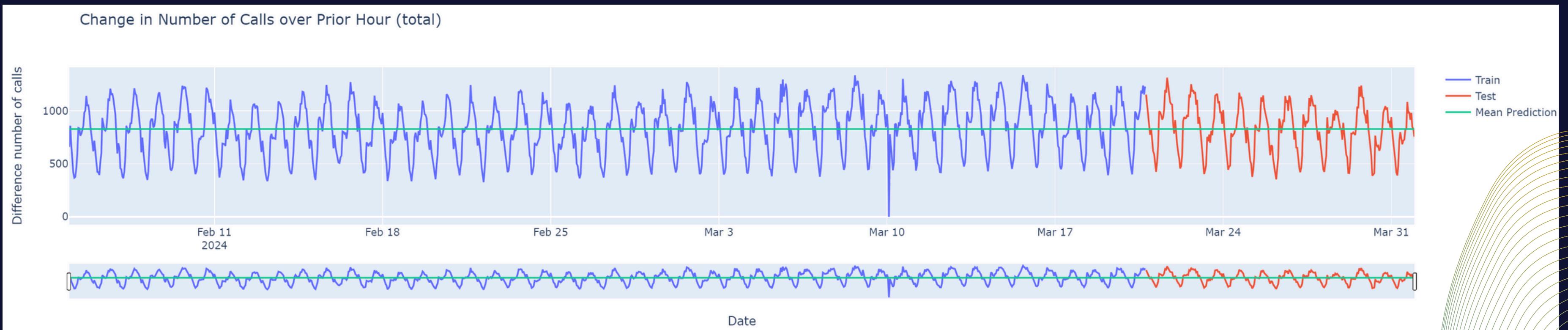


# Initial Findings

## Response Time Distribution



# Forecasting - Baseline



**Evaluation Metric:** Root Means Squared Error

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

	Baseline_train	Baseline_test
2_month	241.576298	220.590974
3_month	235.742804	234.115189
4_month	225.348839	251.276845
6_month	217.103911	256.616834
10_month	215.074416	250.217809
16_month	237.640749	236.249754





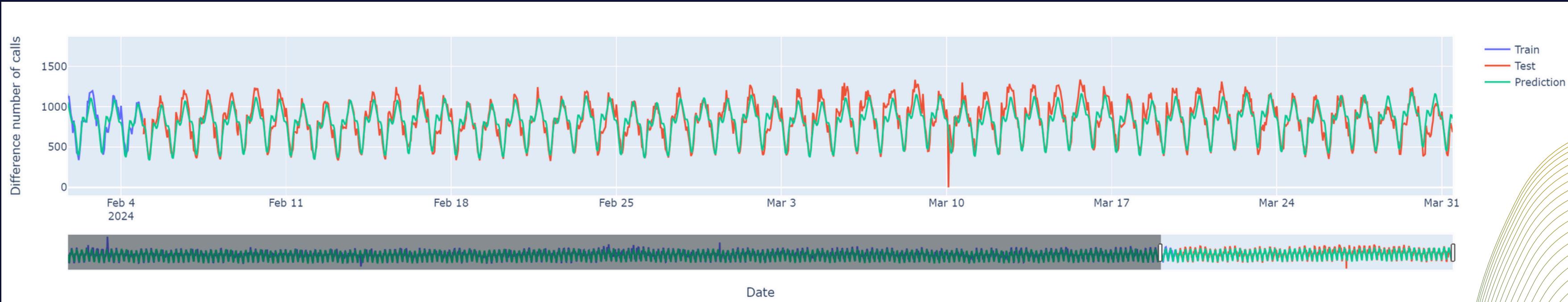
# Forecasting - Full Results

	Baseline_train	Baseline_test	SARIMAX_train_2	SARIMAX_test_2	SARIMAX_train_3	SARIMAX_test_3	SARIMAX_train_4	SARIMAX_test_4	SARIMAX_train_9	SARIMAX_test_9	SARIMAX_train_10	SARIMAX_test_10	SARIMAX_train_11	SARIMAX_test_11	SARIMAX_train_12
2_month	241.576298	220.590974	103.439	171.945	102.349	151.014	92.986	136.764	78.801	128.796	78.550	113.454	76.839	112.723	76.645
3_month	235.742804	234.115189	101.078	207.724	100.415	204.273	96.174	267.909	69.743	128.430	69.735	122.638	67.211	111.515	66.831
4_month	225.348839	251.276845	98.422	239.369	98.411	239.150	85.835	153.656	67.432	172.263	58.794	183.329	58.640	186.473	58.617
6_month	217.103911	256.616834	101.549	245.671	101.535	245.131	88.790	240.749	60.829	196.351	60.001	190.179	60.021	190.334	59.672
10_month	215.074416	250.217809	103.685	245.921	103.681	245.861	88.460	193.087	86.847	237.099	86.748	236.410	85.903	236.786	61.535
16_month	237.640749	236.249754	122.047	232.538	122.049	232.640	104.642	314.584	63.359	219.264	61.499	219.344	61.162	223.447	60.224
SARIMAX_test_12	SARIMAX_train_d2	SARIMAX_test_d2	SARIMAX_train_d3	SARIMAX_test_d3	SARIMAX_train_d4	SARIMAX_test_d4	SARIMAX_train_d9	SARIMAX_test_d9	SARIMAX_train_d10	SARIMAX_test_d10	SARIMAX_train_d11	SARIMAX_test_d11	SARIMAX_train_d12		
113.231	69.503	142.083	78.869	139.754	78.314	138.470	89.983	129.258	89.013	123.939	107.486	130.812	113.476		
112.062	76.835	180.999	85.229	168.980	85.403	167.811	97.602	133.215	99.479	127.786	99.518	129.203	145.993		
187.384	63.079	151.248	64.435	155.066	76.384	147.705	90.833	133.138	103.801	141.467	103.554	140.920	107.219		
204.009	63.812	165.982	75.731	165.850	75.230	165.173	90.250	169.066	101.161	162.373	103.635	165.021	117.836		
226.052	71.390	140.715	83.090	133.755	83.200	135.667	94.678	140.059	91.612	130.843	92.745	131.207	95.044		
215.527	103.990	257.711	134.052	250.355	134.494	256.523	176.622	264.426	180.623	257.895	184.277	260.761	194.969		

- **FaceBook Prophet 9 Month**
  - Best of over 200+ models

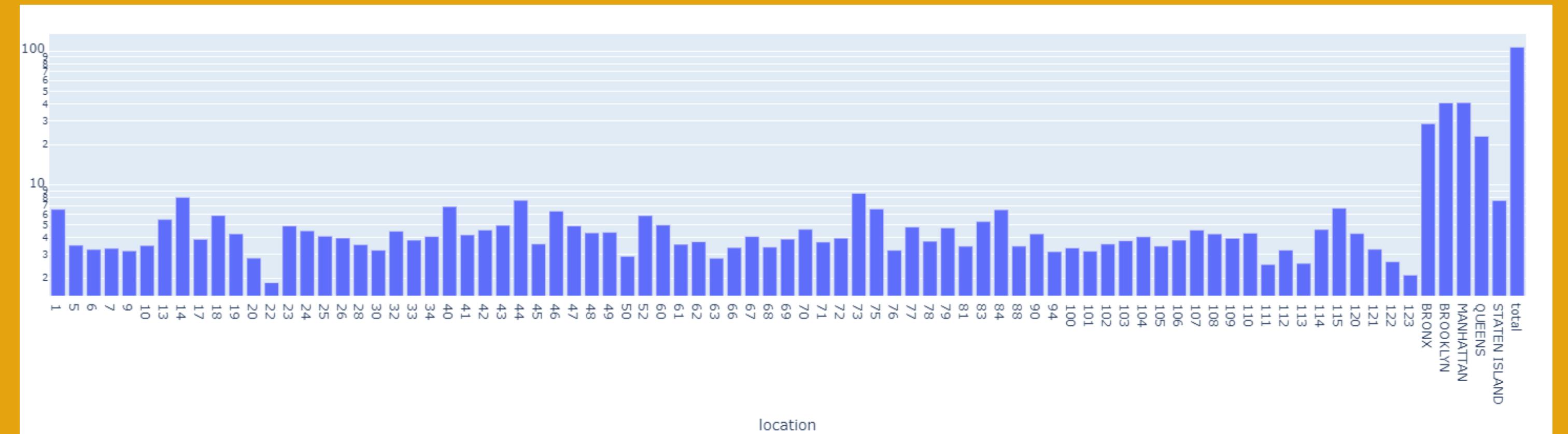
SARIMAX_test_d12	Prophet_train	Prophet_test
139.021	85.898	111.689
175.571	84.520	178.085
144.557	89.395	140.491
182.407	91.906	119.560
132.714	97.435	104.174
270.761	110.885	160.920

# Forecasting - Best Model



## Predictions

## Scores



# Heat Maps

## Borough



## Precinct

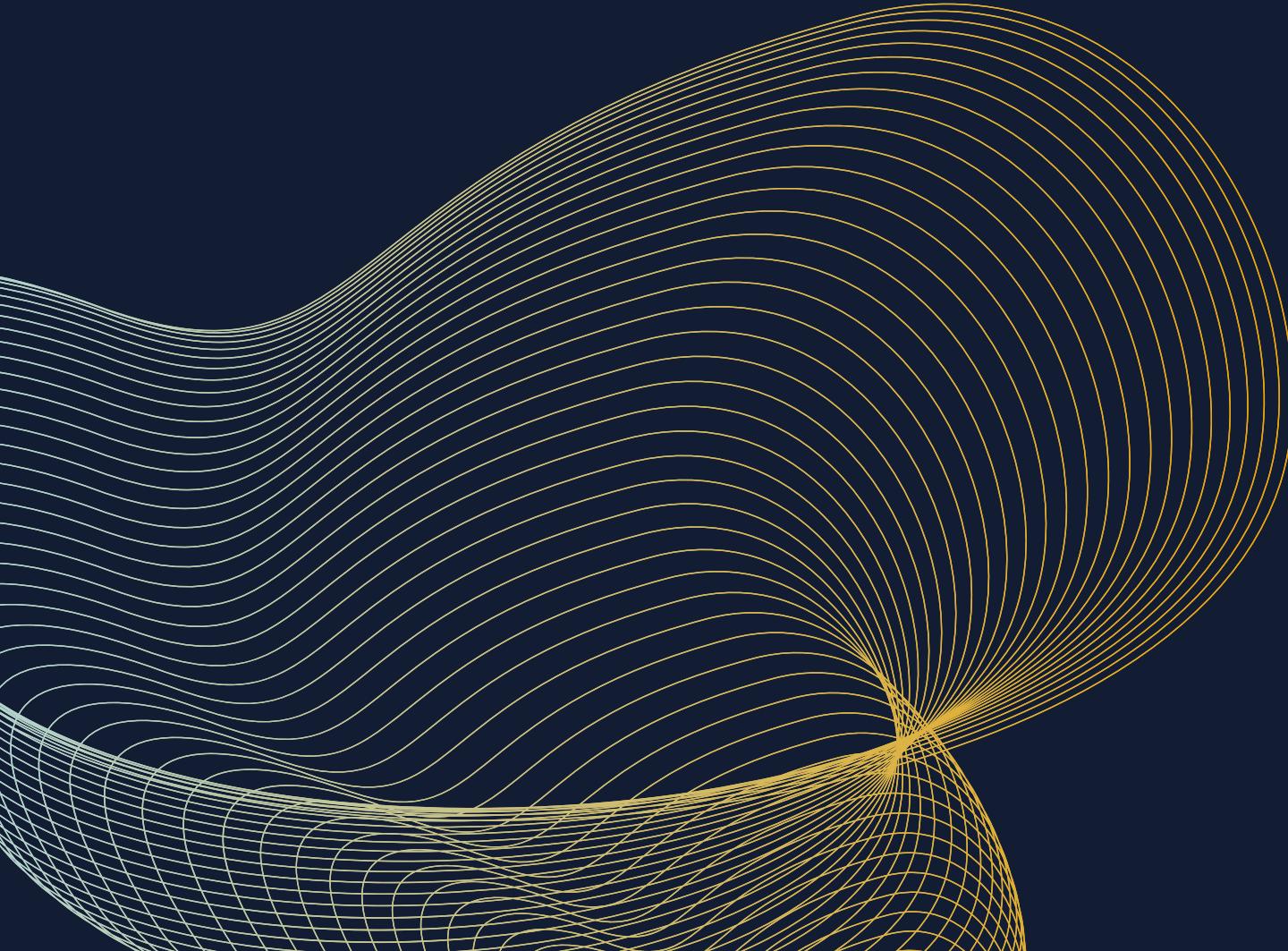


---

**DEMO**



# Whats Next!



## Priority 1 – Final Cleaning

- Orginize and clean github repo. I am a very orginized person and what to make it as professional as possible

## Priority 2 – Lat/Lon Groups

- Include lat lon location group to predict more granularly where calls are more likely to occur

## Priority 3 – Data Exploration

- The Data set has so much more information. I would love to take a deeper dive and have some fun exploring the data and trying to find other insights the data may hold

## Priority 4 – Update Date

- NYC Open data keeps data fairly up to date. Re-run notebooks with new data to make prediction further into future



# THANK YOU

**Github Project Link:**

[https://github.com/Connell-Phillipps/Capstone\\_911](https://github.com/Connell-Phillipps/Capstone_911)

**Learn More:**

Cmphilipps73@gmail.com

