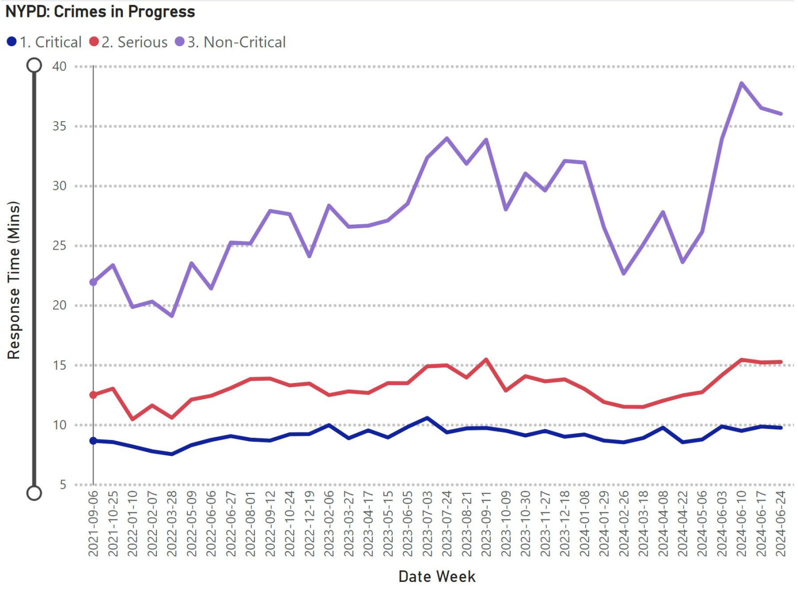
911 Hotspot Prediction

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 lifes, even if one life is saved with faster response is worth it!

The Data:

Data Dictionary:

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

12.3GB

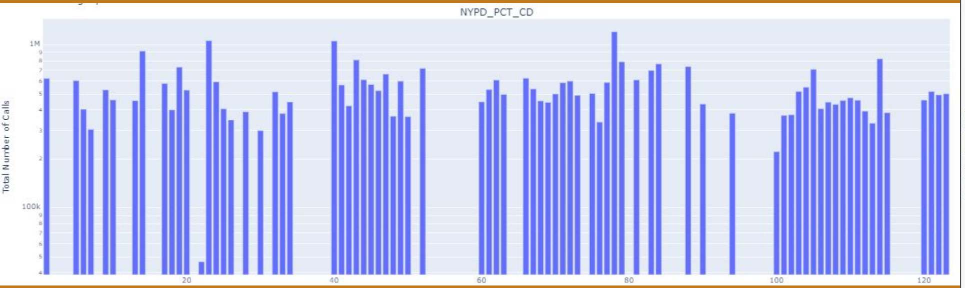
Preprocessed Dataset:

```
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
```

3.2GB

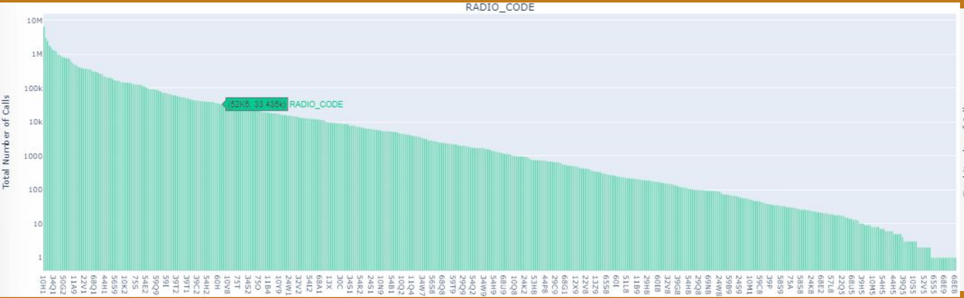
Initial Findings: Call Distribution by Category

NYPD Precinct



Boroughs

Radio Code



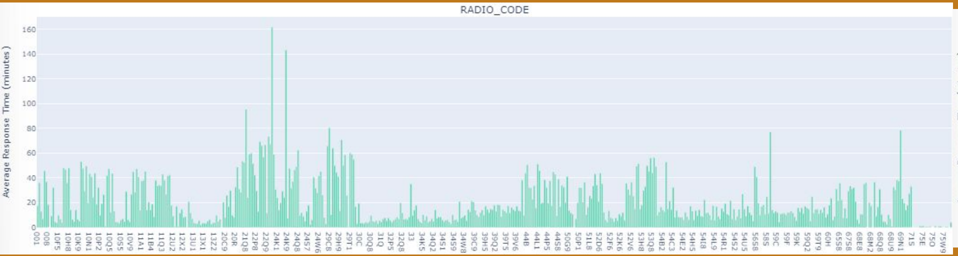
Crime in Progress

Initial Findings: Response Time by Category

NYPD Precinct

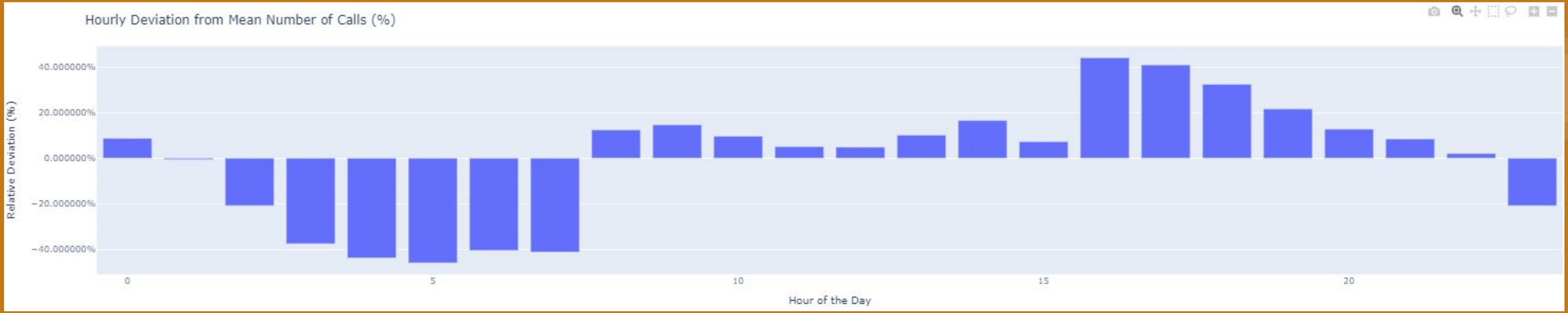
Boroughs

Radio Code

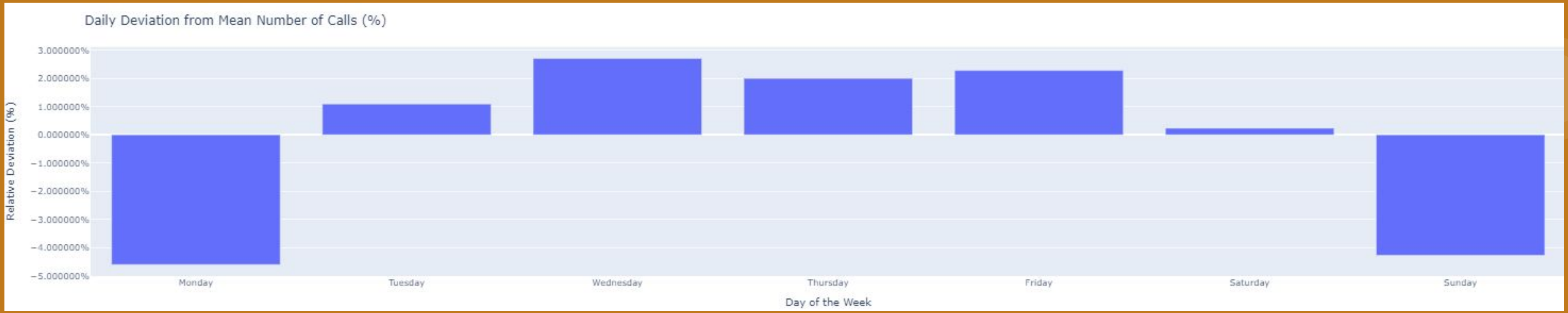


Crime in Progress

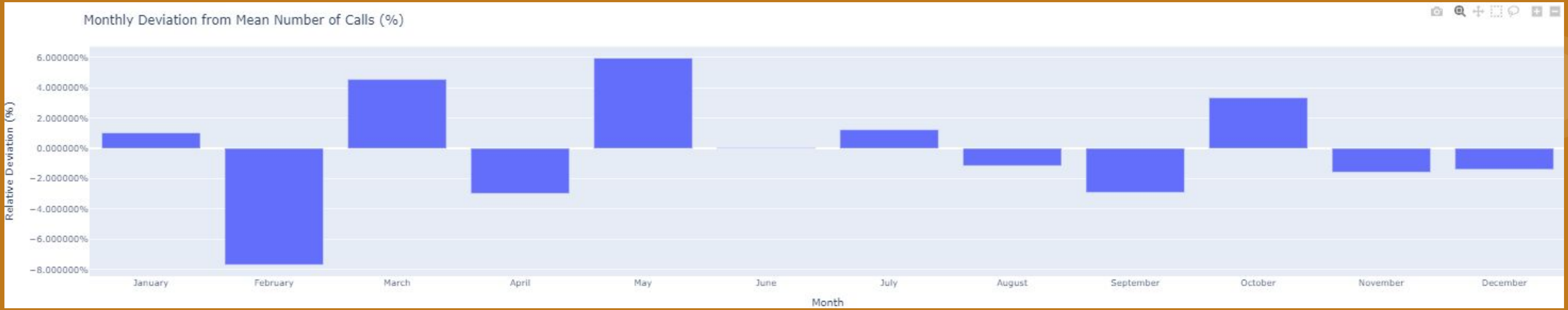
Initial Findings: Hourly Deviation from mean



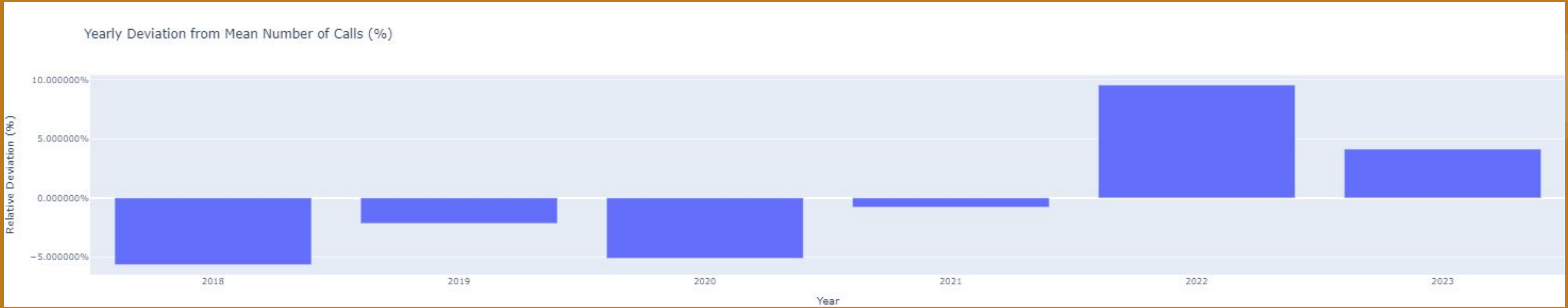
Initial Findings: Daily Deviation from mean



Initial Findings: Monthly Deviation from mean



Initial Findings: Yearly Deviation from mean

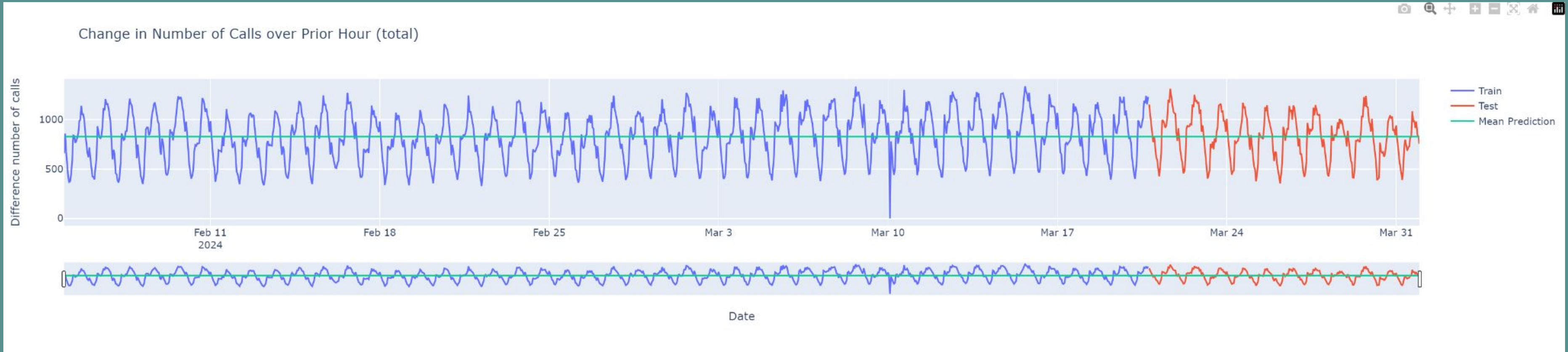


Data for Forecasting

12 MB

	1.0	5.0	6.0	7.0	9.0	10.0	13.0	14.0	17.0	18.0	...	120.0	121.0	122.0	123.0	BRONX	BROOKLYN	MANHATTAN	QUEENS	STATEN ISLAND	total
2018-01-01 00:00:00	7	6	10	11	17	9	13	27	5	20	...	10	11	14	16	147	285	209	146	51	838
2018-01-01 01:00:00	14	16	23	13	24	18	24	33	13	27	...	14	14	7	5	200	279	319	168	40	1006
2018-01-01 02:00:00	21	16	29	14	20	13	18	35	14	32	...	10	6	6	4	167	226	336	173	26	928
2018-01-01 03:00:00	16	8	12	15	14	18	14	28	13	26	...	11	7	9	4	148	213	254	155	31	801
2018-01-01 04:00:00	11	7	9	5	8	9	10	18	9	20	...	7	8	7	5	157	182	198	163	27	727

Forecasting - Baseline



Evaluation Metric:

- Root Means Squared Error

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

	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

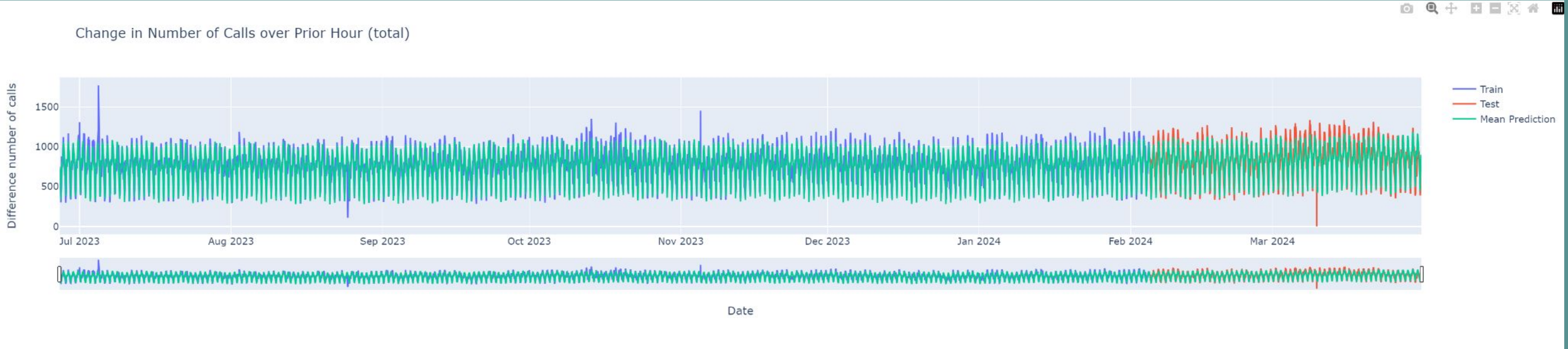
SARIMAX_test_12	SARIMAX_train_d2	SARIMAX_test_d2	SARIMAX_train_d3	SARIMAX_test_d3	SARIMAX_train_d4	SARIMAX_test_d4	SARIMAX_train_d9	SARIMAX_test_d9	SARIMAX_train_d10	SARIMAX_test_d10	SARIMAX_train_d11	SARIMAX_test_d11	SARIMAX_train_d12
113.231	69.503	142.083	78.869	139.754	78.314	138.470	89.983	129.258	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

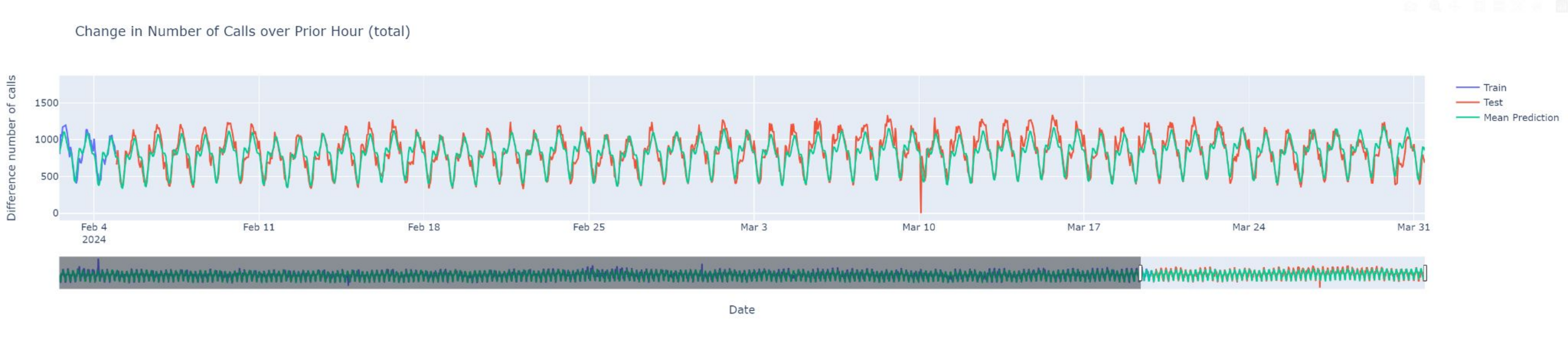
- Best run: **Facebook Prophet, 10 month** window!

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

Best score out of 90+ models run

Forecasting - Facebook Prophet Model





Forecasting - Each Location RMSE

Next Steps:

- Loop through all possible timeframes for Facebook Prophet model
    - Pick best model from that
- Loop through every column of original dataframe for each locational granularity.
    - Save predictions to another dataframe and send to visualizations notebook
- Graph predictions on a heat map of NYC
- Build Streamlit/Rshiny App
- Nice to have:
    - Create new aggregate database on lat/lon groups…