Daily Call Volume Prediction

Using Historical Call Volume, Time, and Weather Data as parameters

Linear & Degree 2 Polynomial Regression Models

Glance at the dataset

*Only includes first 20 columns and 5 rows here.
There are 149 columns and 894 rows in total.

index	Date	Day of the Week	Total Inbound Calls	AWND	PRCP	SNOW	SNWD	TAVG	TMAX	TMIN	AWND_Lag1	AWND_Lag2	AWND_Lag3	AWND_Lag4	AWND_Lag5	AWND_Lag6	AWND_Lag7	AWND_Lag8	AWND_Lag9	AWND_Lag10
0	2022-01-01 00:00:00	Saturday	0.0	10.07	0.63	0.0	0.0	44	48.0	27.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	2022-01-02 00:00:00	Sunday	0.0	13.42	0.0	0.1	0.0	24	27.0	16.0	10.07	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	2022-01-03 00:00:00	Monday	78.0	5.59	0.0	0.0	0.0	21	34.0	13.0	13.42	10.07	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	2022-01-04 00:00:00	Tuesday	73.0	13.42	0.0	0.0	0.0	32	43.0	26.0	5.59	13.42	10.07	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	2022-01-05 00:00:00	Wednesday	70.0	15.88	0.0	0.0	0.0	34	42.0	20.0	13.42	5.59	13.42	10.07	NaN	NaN	NaN	NaN	NaN	NaN

NaN cells only exist in several top rows,
because there is no Lag data for those days.
I eliminated those rows in later analysis.

Parameters

2022-1 to 2024-6, excluding non-Sunday 0 call days

- **5 Historical Call Volume Parameters:**

  1-5 Lag Days

- **2 Time Parameters & 1 market index:**

  Year, Month (together as 1 column), S&P 500 index

- **140 (8 *20) Weather parameters:**

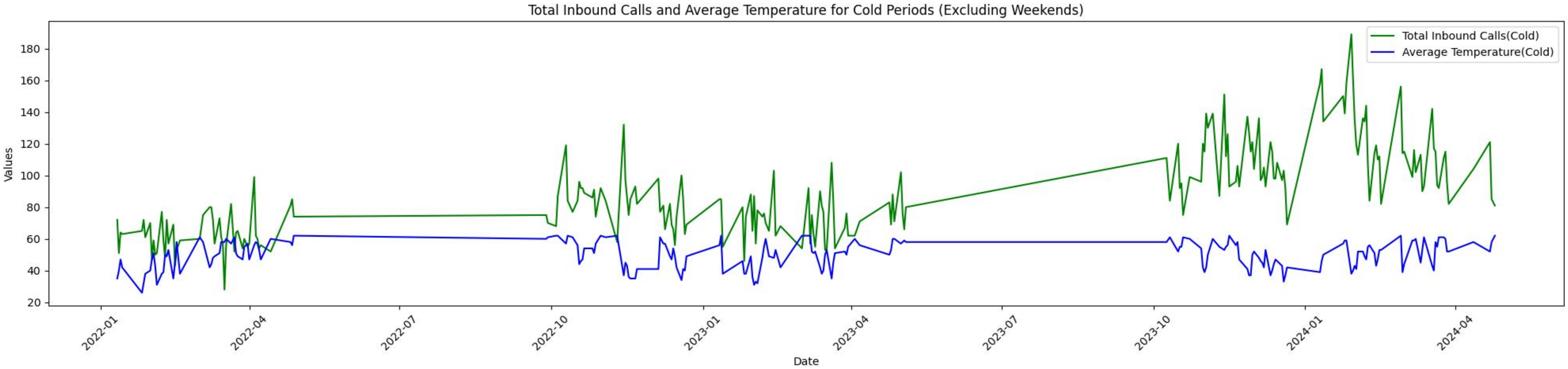
  0-19 Lag Days

  AWND - Average wind speed **SNOW -** Snowfall **HAIL -** Hail

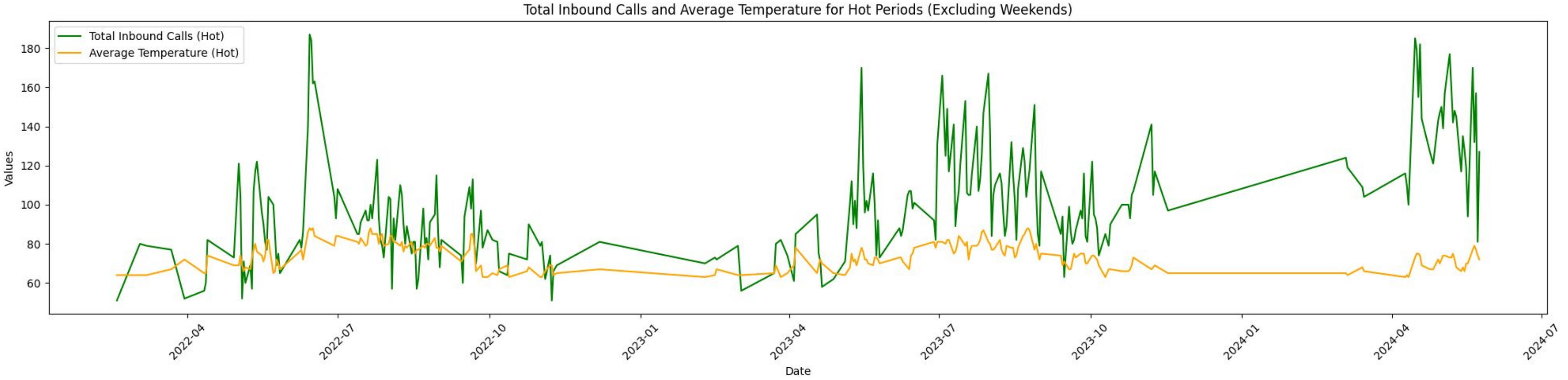
  TMAX - Maximum temperature **TAVG -** Average Temperature

  TMIN - Minimum temperature **PRCP -** Precipitation **SNWD** - Snow depth

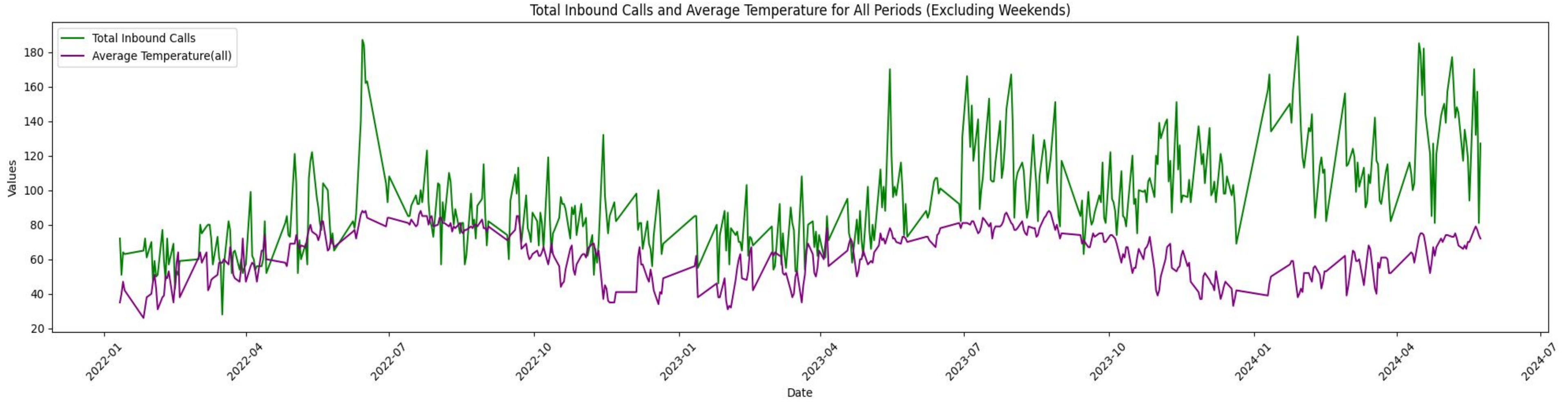
Cold Periods



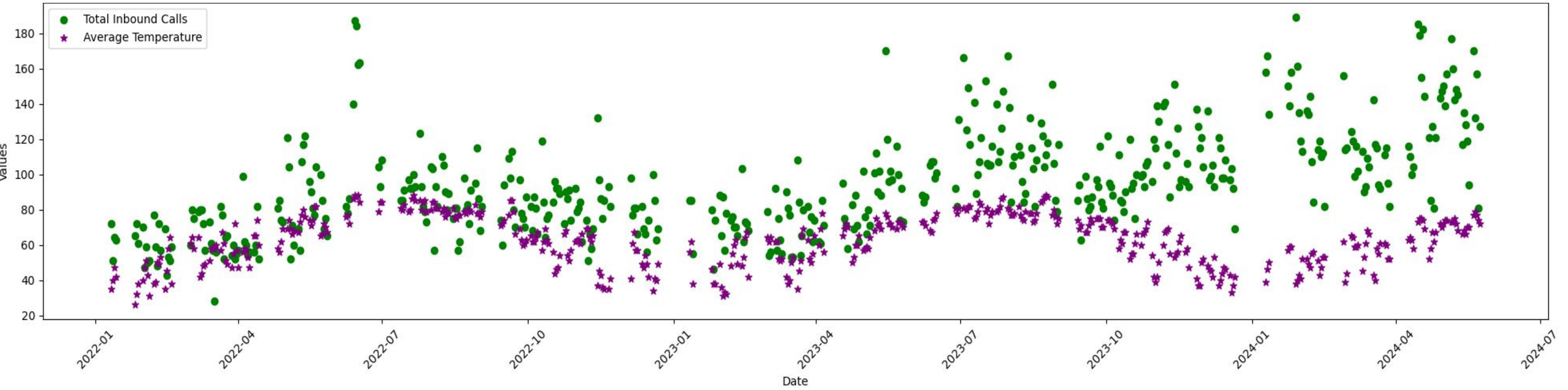
Hot Periods



All Periods



Total Inbound Calls and Average Temperature for All Periods (Excluding Weekends)
Total Inbound Calls
Average Temperature(all)
Values
Date



Total Inbound Calls
Average Temperature
Values
Date

Linear Regression for Cold Periods

```
                            OLS Regression Results
==============================================================================
Dep. Variable:     Total Inbound Calls   R-squared:                       0.913
Model:                             OLS   Adj. R-squared:                  0.879
Method:                  Least Squares   F-statistic:                     26.73
Date:                 Mon, 24 Jun 2024   Prob (F-statistic):           1.58e-80
Time:                         16:53:14   Log-Likelihood:                -1197.0
No. Observations:                  305   AIC:                             2568.
Df Residuals:                      218   BIC:                             2892.
Df Model:                           86
Covariance Type:             nonrobust
==============================================================================
```

Set **Friday** as the benchmark:
Monday: 54.51 more calls on average
Tuesday: 4.17 more calls on average*
Wednesday: 11.33 more calls on average*
Thursday: 5.75 more calls on average*
Saturday: 43.24 less calls on average
Sunday: 58.66 less calls on average

Set **Sunday** as the benchmark:
Monday: 103 calls on average
Tuesday: 62 calls on average
Wednesday: 70 calls on average
Thursday: 63 calls on average
Friday: 58.66 calls on average
Saturday: 15.42 calls on average

Important parameters interpretation

spending 0.0005 0.000

For every 1000 dollars spent on Google Ads, the inbounded call volume increases by 5 per day on average.

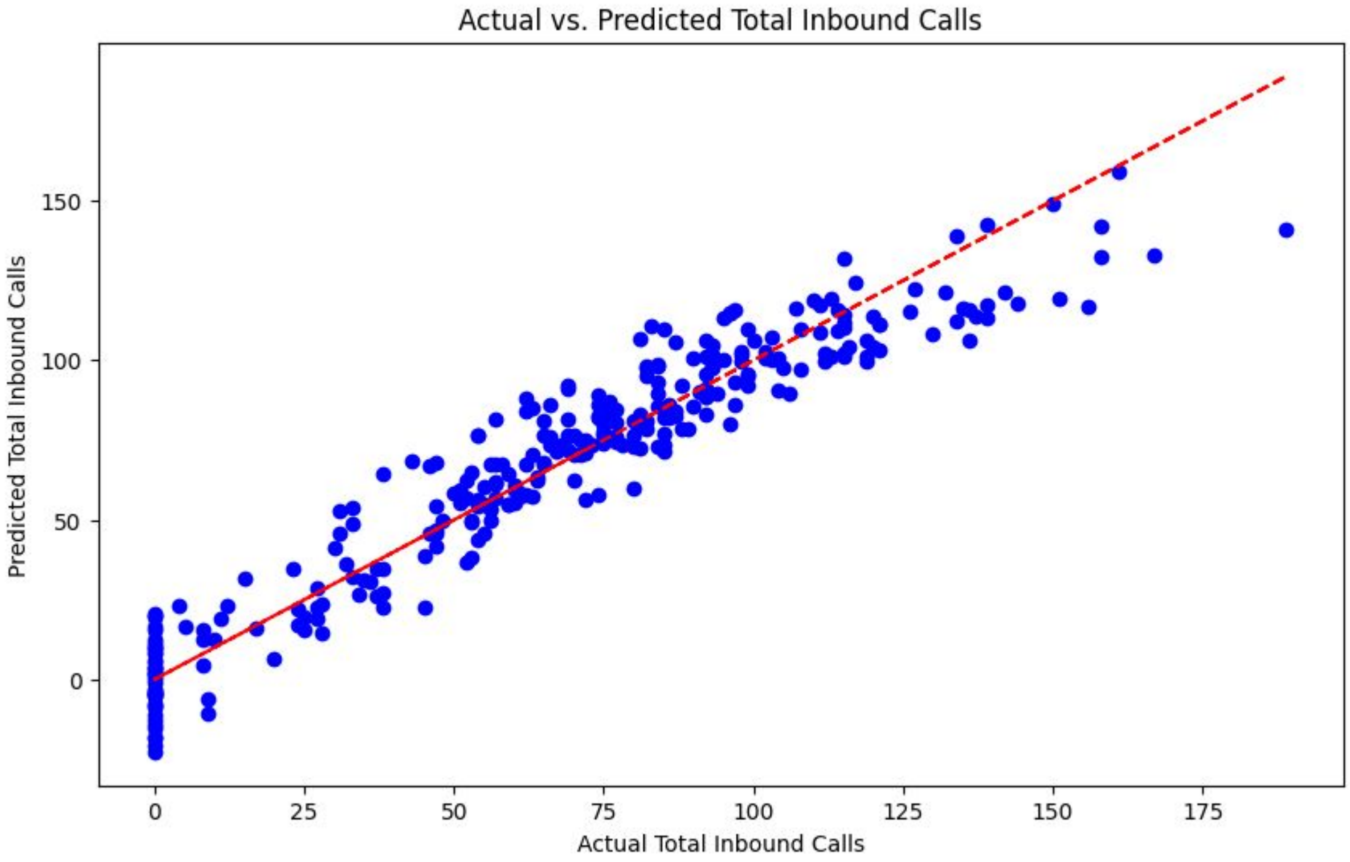
PRCP 5.0746 0.271 *

if precipitation increases by 1 inch, the inbounded call volume will increase by 5 per day on average.

Month_index_continous 1.1995 0.000

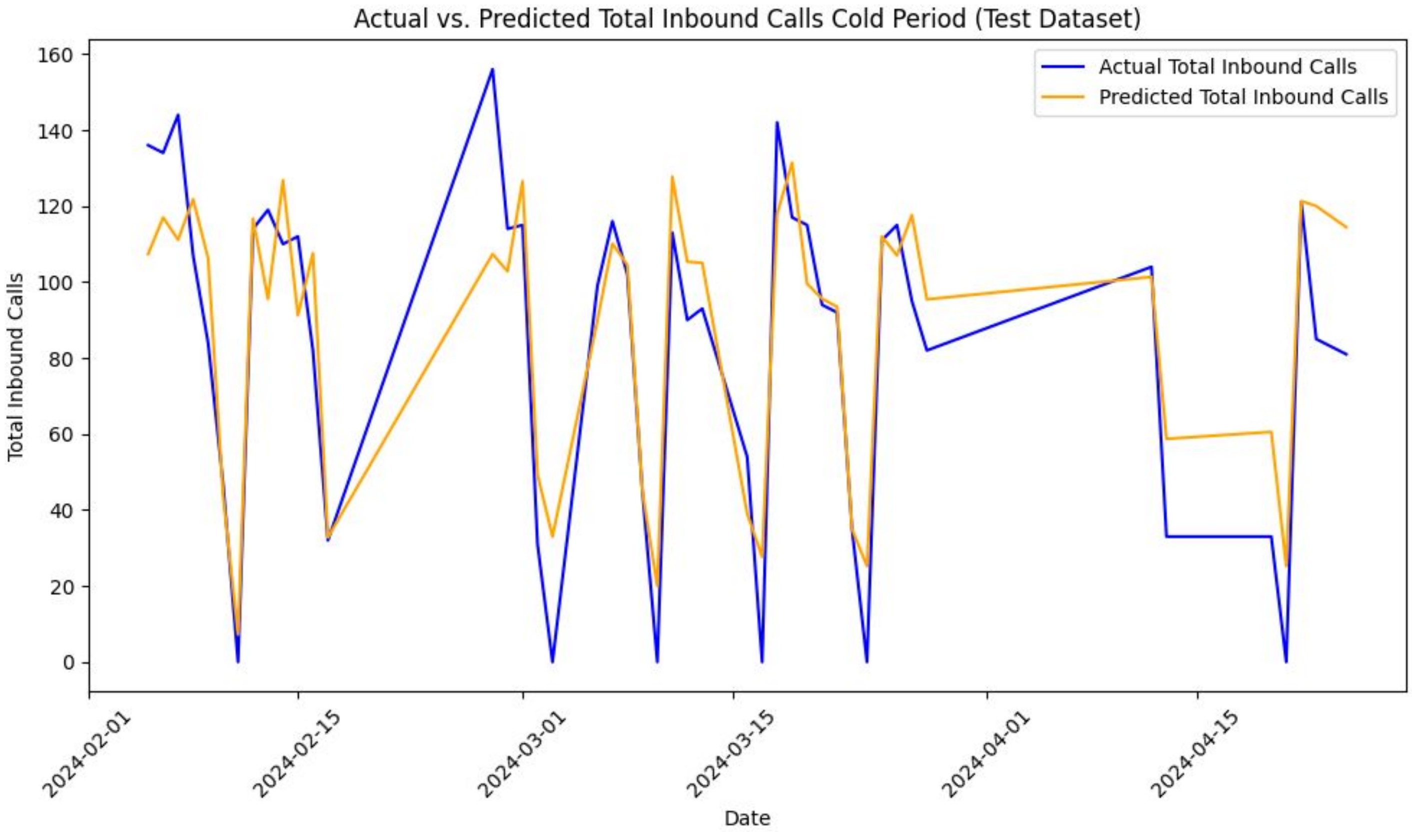
Every month later, the call volume will increase by 1 per day on average.

Linear Regression for Cold Periods



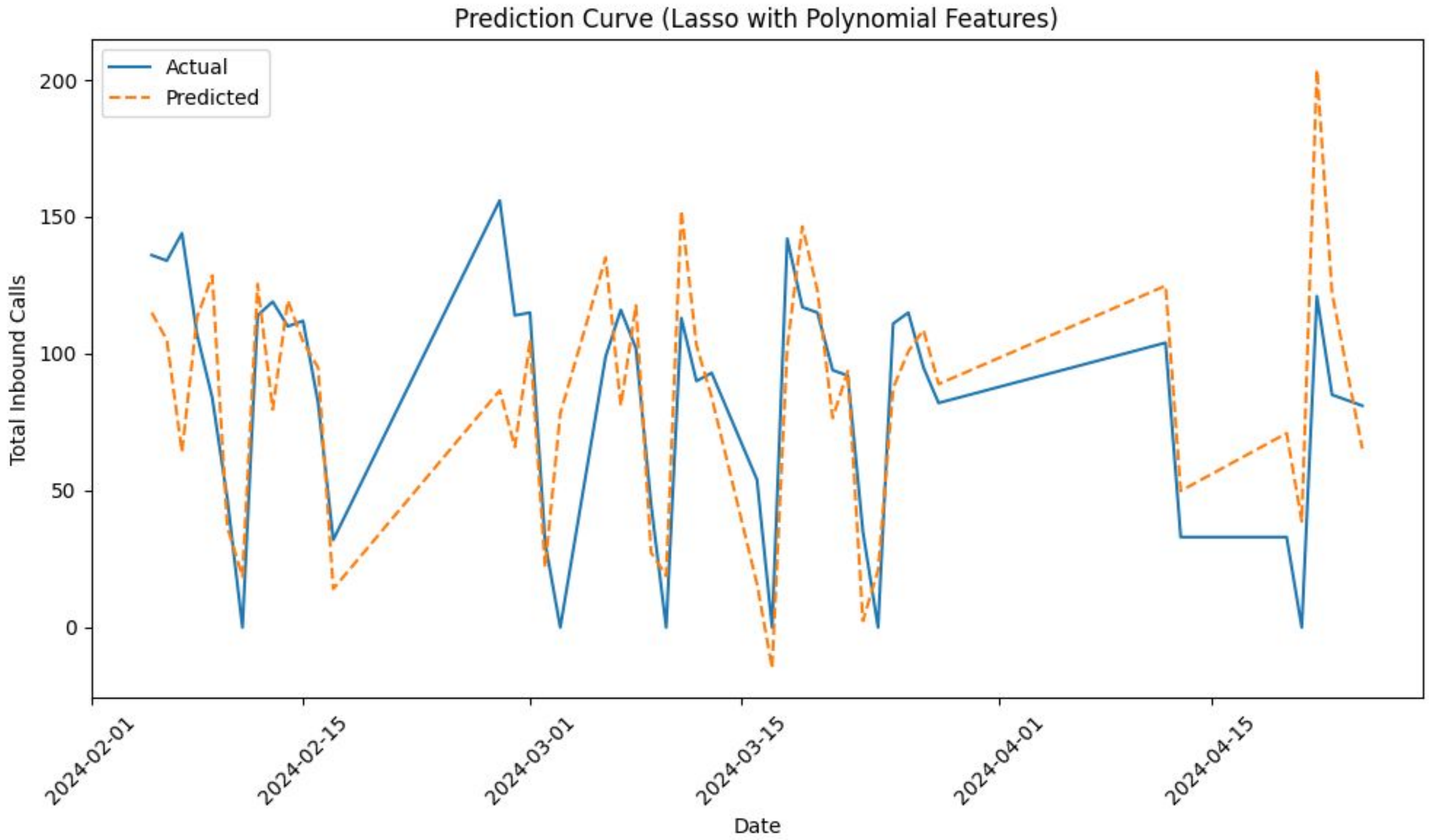
```
# Split the dataset into training (85%) and test (15%) sets
```

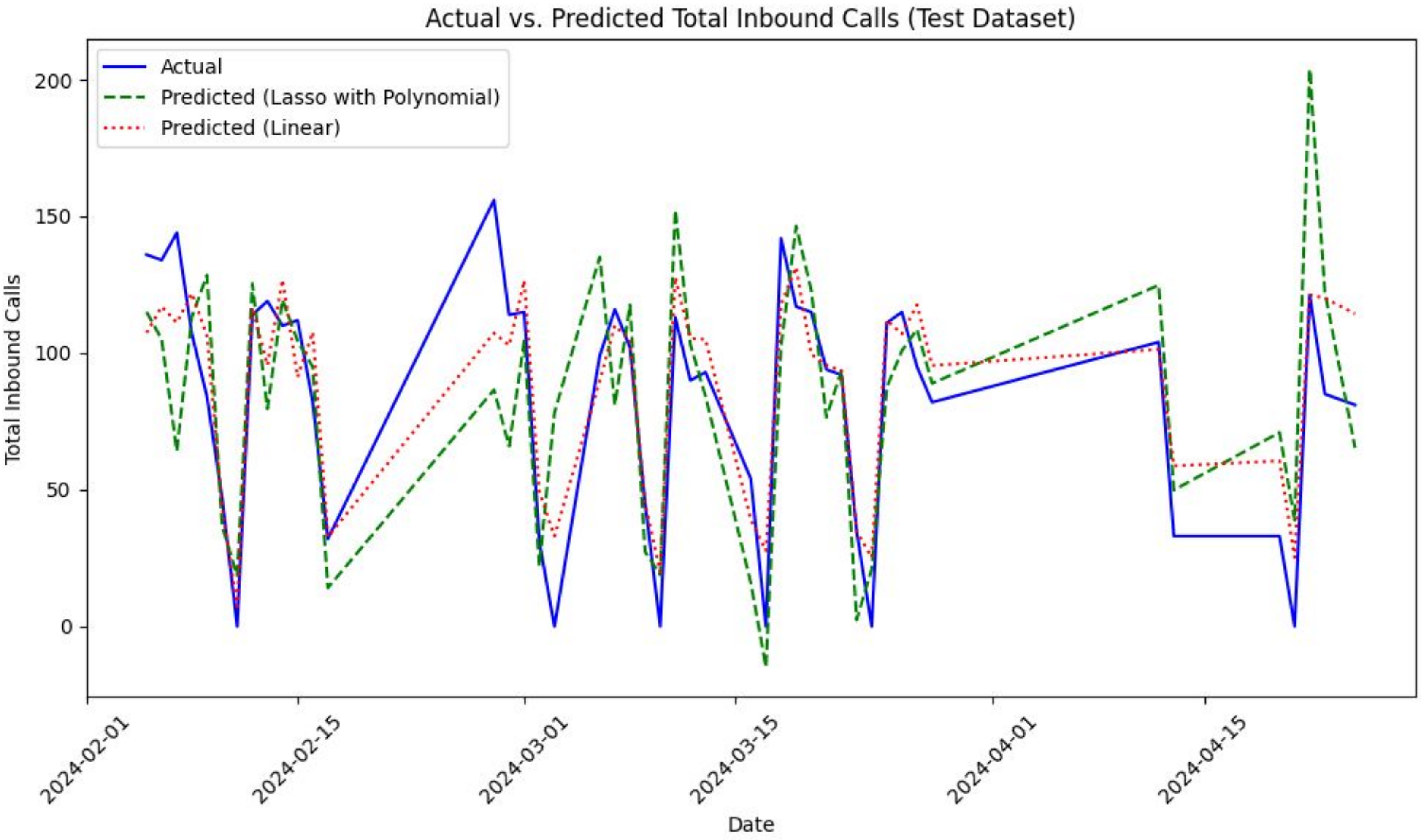
Mean Squared Error: **388.799**



Polynomial Regression for Cold Periods

Mean Squared Error: 1092.20





Actual vs. Predicted Total Inbound Calls (Test Dataset)
Actual
Predicted (Lasso with Polynomial)
Predicted (Linear)
Total Inbound Calls
200
150
100
50
0
2024-02-01
2024-02-15
2024-03-01
2024-03-15
2024-04-01
2024-04-15
Date

Linear Regression for Hot Periods

```
                            OLS Regression Results
==============================================================================
Dep. Variable:     Total Inbound Calls   R-squared:                       0.877
Model:                             OLS   Adj. R-squared:                  0.845
Method:                  Least Squares   F-statistic:                     27.14
Date:                 Mon, 24 Jun 2024   Prob (F-statistic):           8.72e-91
Time:                         15:27:25   Log-Likelihood:                -1451.5
No. Observations:                  347   AIC:                             3049.
Df Residuals:                      274   BIC:                             3330.
Df Model:                           72
Covariance Type:             nonrobust
==============================================================================
```

Set **Friday** as the benchmark:

Monday*: 42 more calls on average

Tuesday: 11 (-9.353, 28.943) 95%

Wednesday: 3 (-17.699, 24.227) 95%

Thursday: 0 (-18.485, 15.570) 95%

Saturday*: 69 less calls on average

Sunday*: 87 less calls on average

Set **Sunday** as the benchmark:

Monday: 129 calls on average

Tuesday: 98 calls on average

Wednesday: 90 calls on average

Thursday: 87 calls on average

Friday: 87 calls on average

Saturday: 18 calls on average

Important parameters interpretation

HAIL **22.2346** 14.289 1.556 **0.121** -5.897 50.366

Every inch increase in hail will result 22 more calls

spending **0.0003** 0.000 2.340 **0.020** 4.33e-05 0.001

For every 1000 dollars spent on Google Ads, the inbounded call volume increases by 3 per day on average.

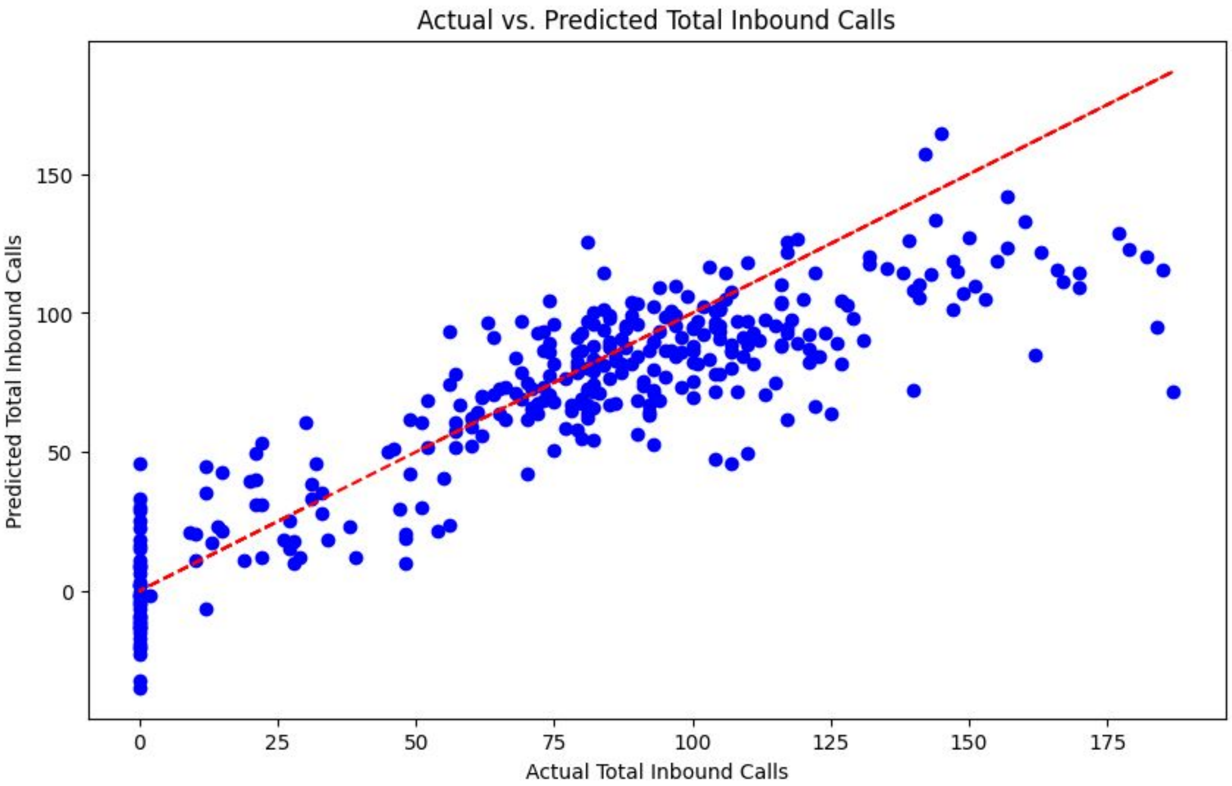
PRCP_Lag8 **6.5472** 3.703 1.768 **0.078** -0.742 13.837

if there is a precipitation 8 days ago, every 1 inch precipitation will increases the inbounded call volume by 6.5 per day on average.

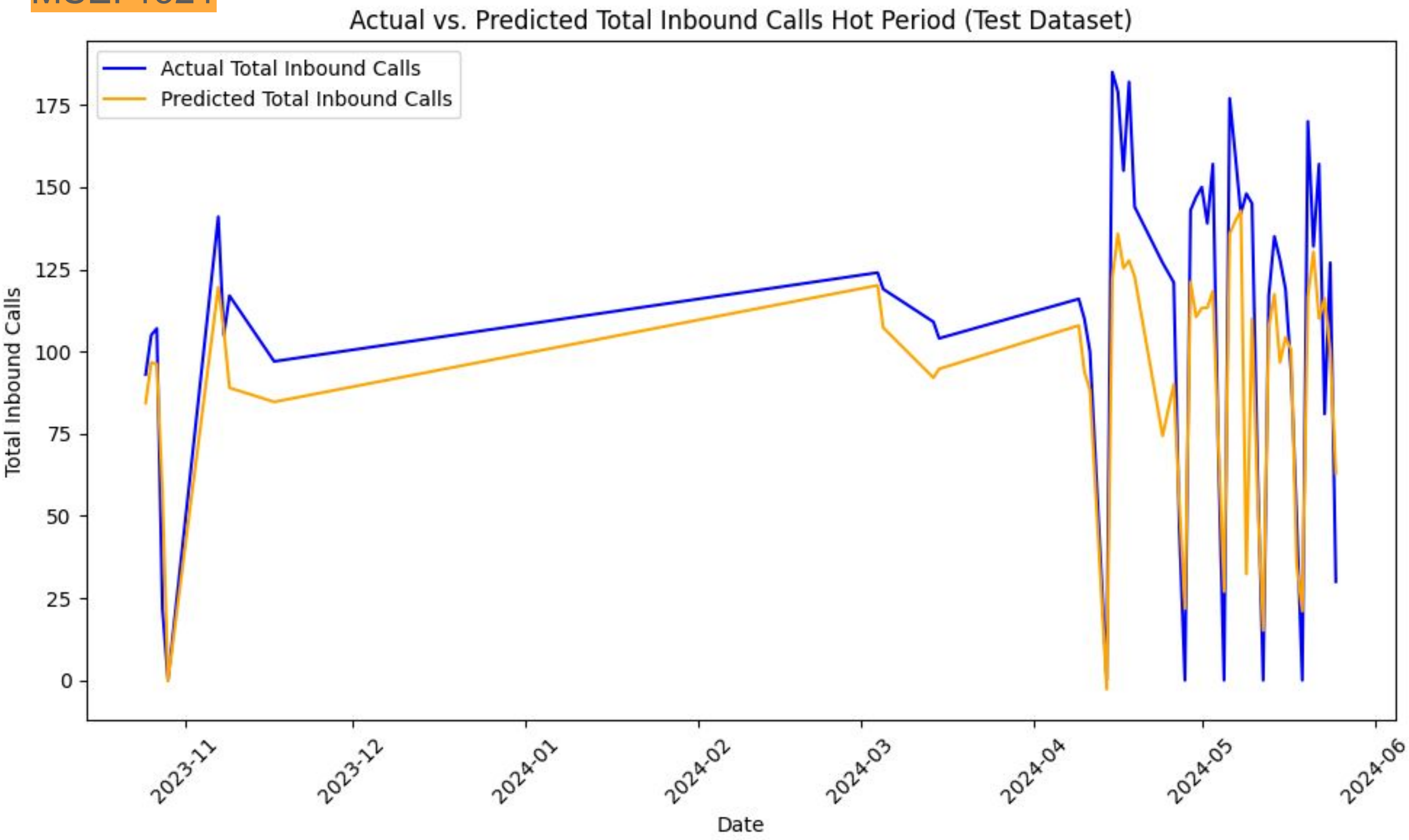
Month_index_continous **0.8187** 0.279 2.929 **0.004** **0.269** 1.369

Every month later, the call volume will increase by 0.8 per day on average.

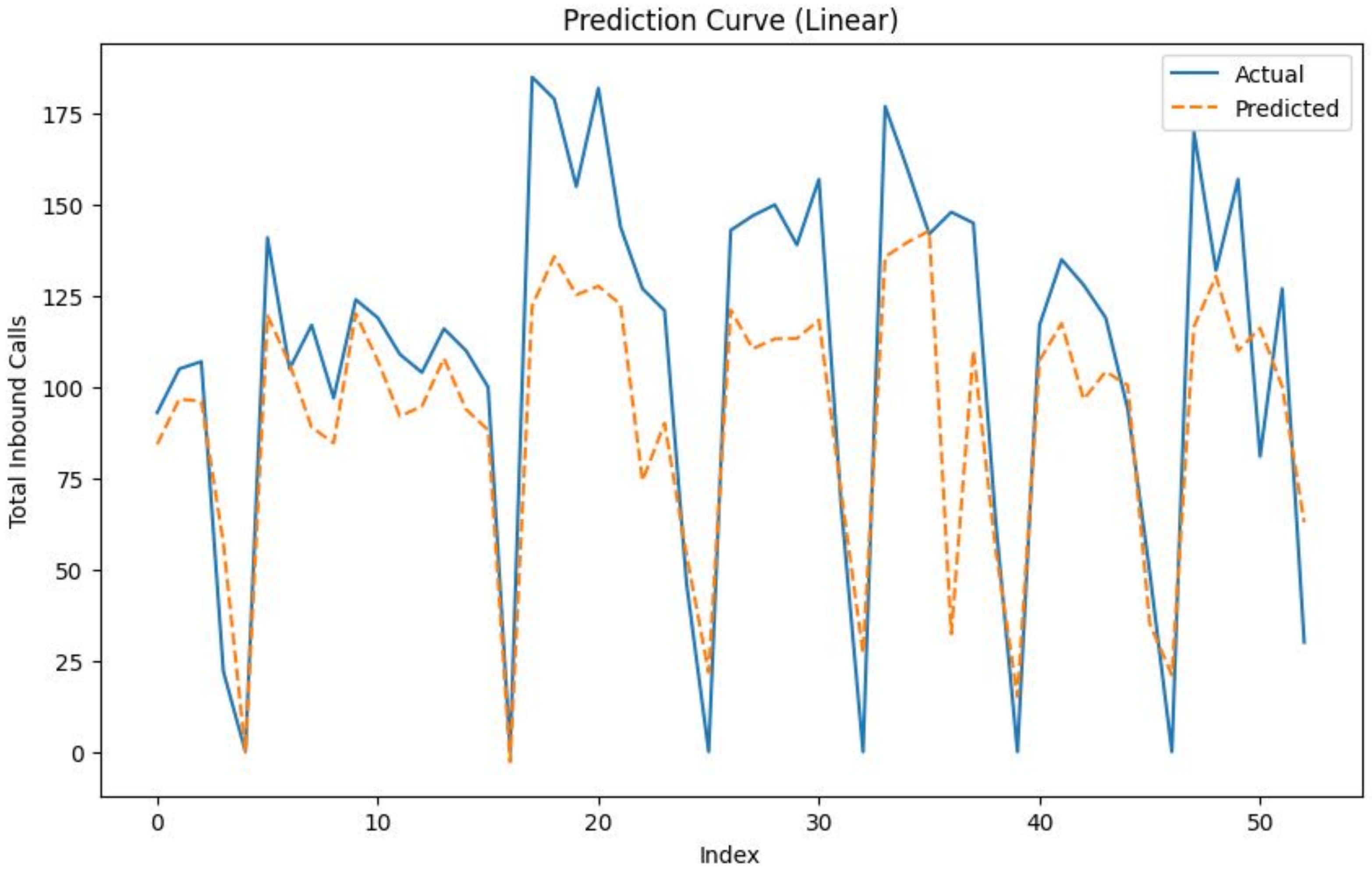
Linear Regression for Hot Periods



MSE: 1521

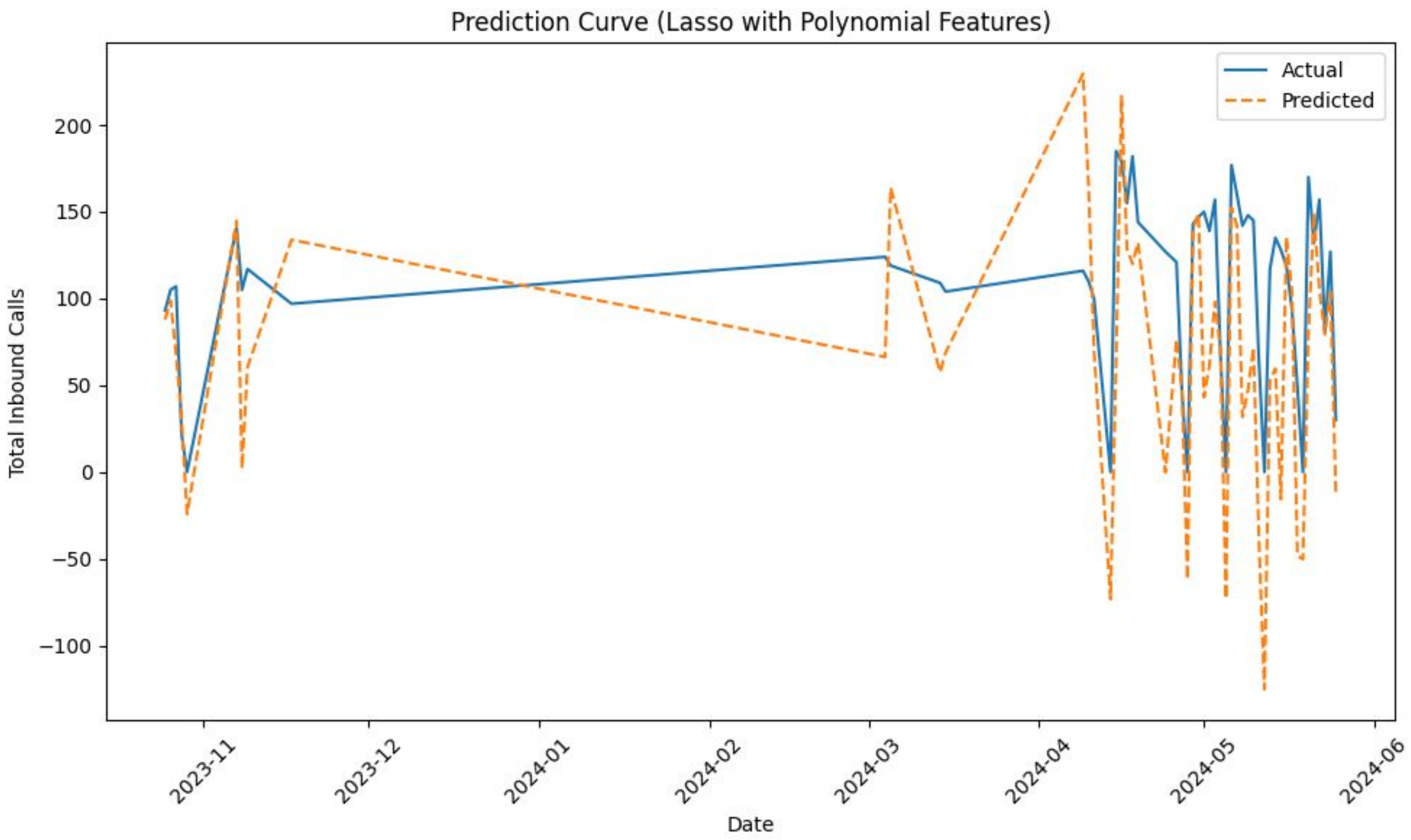


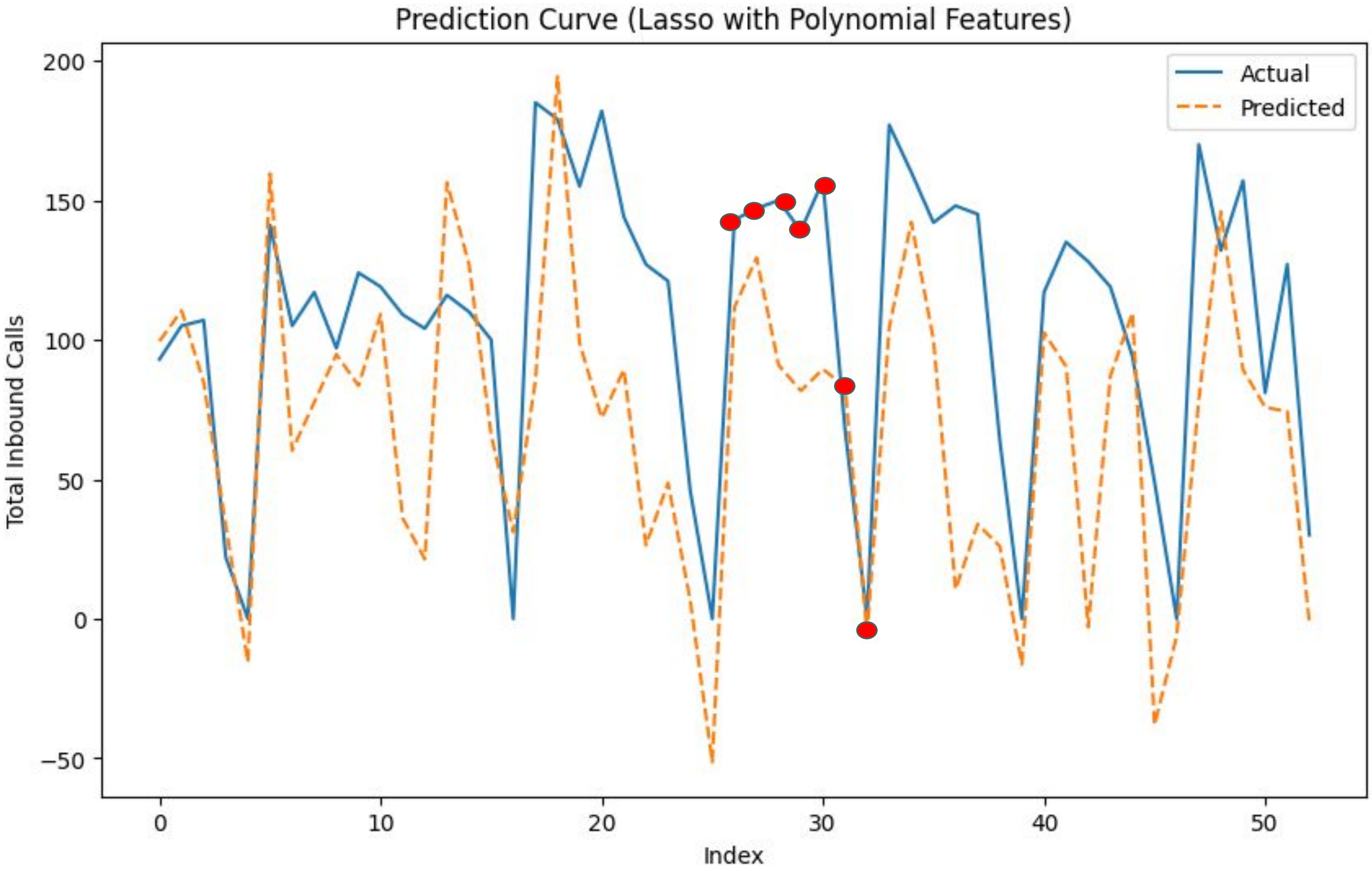
Actual vs. Predicted Total Inbound Calls Hot Period (Test Dataset)
Actual Total Inbound Calls
Predicted Total Inbound Calls
Total Inbound Calls
0
25
50
75
100
125
150
175
Date
2023-11
2023-12
2024-01
2024-02
2024-03
2024-04
2024-05
2024-06


Prediction Curve (Linear)
Actual
Predicted
Total Inbound Calls
Index
0
25
50
75
100
125
150
175
0
10
20
30
40
50

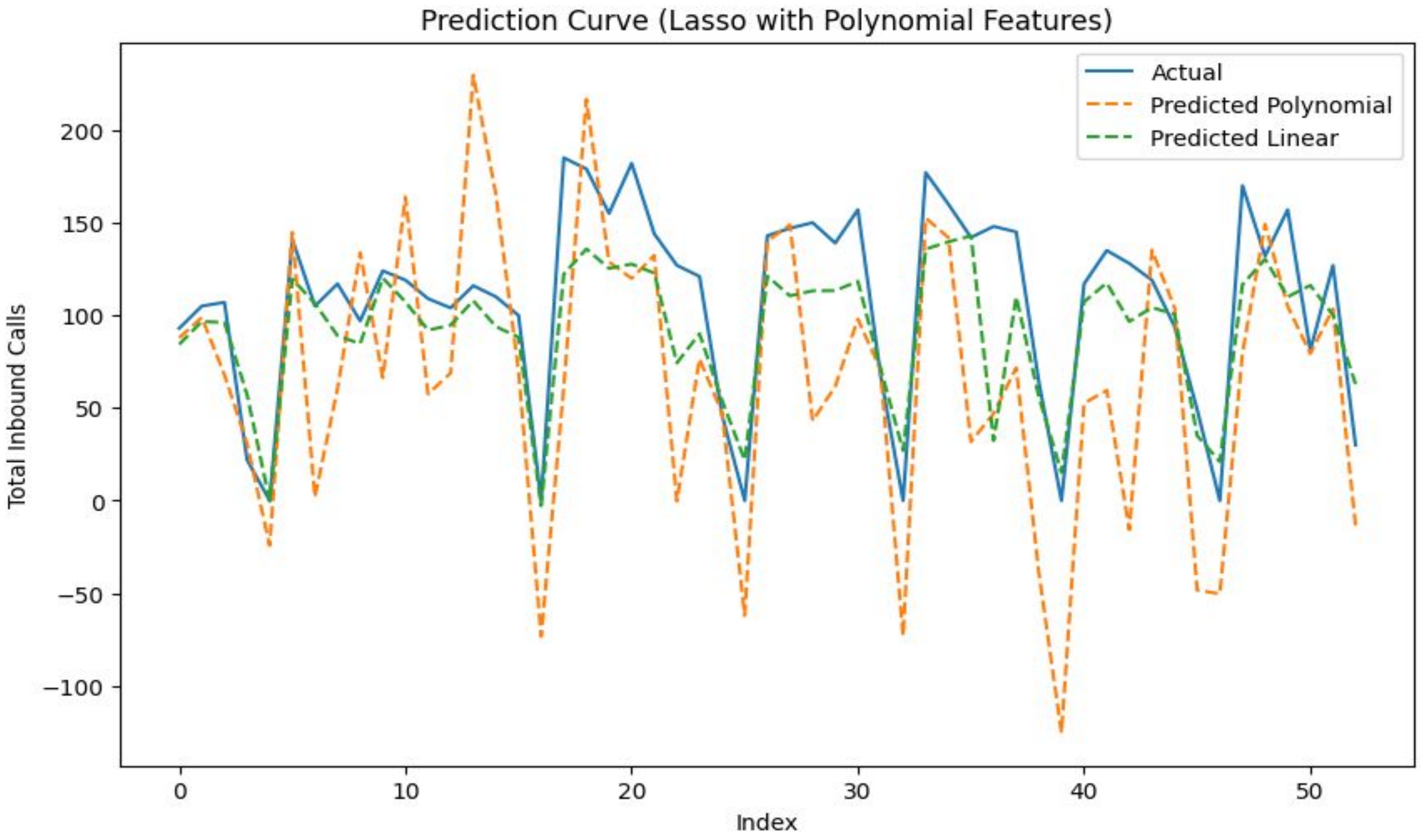
Polynomial Regression for Hot Periods

Mean Squared Error: 3189





Prediction Curve (Lasso with Polynomial Features)
Actual
Predicted
Total Inbound Calls
Index
200
150
100
50
0
−50
0
10
20
30
40
50


Prediction Curve (Lasso with Polynomial Features)
Actual
Predicted Polynomial
Predicted Linear
Total Inbound Calls
Index
200
150
100
50
0
−50
−100
0
10
20
30
40
50