Prepared for
Bhawesh Sah
By Team 4

CITY of BOSTON

BOS:311

A **DATA-DRIVEN** APPROACH TO ENHANCE CITIZEN SERVICES AND QUALITY OF LIFE

BOS :311

Table of Contents

BUSINESS PROBLEM

How can the city of Boston enhance the efficiency and efficacy of its city services, optimize the allocation of its resources, and **elevate the overall satisfaction of its residents**?

Our Approaches

01

Can we **predict the probability of issue resolution** when photographic evidence is submitted, while considering the impact of the **source** and **location** of the report?

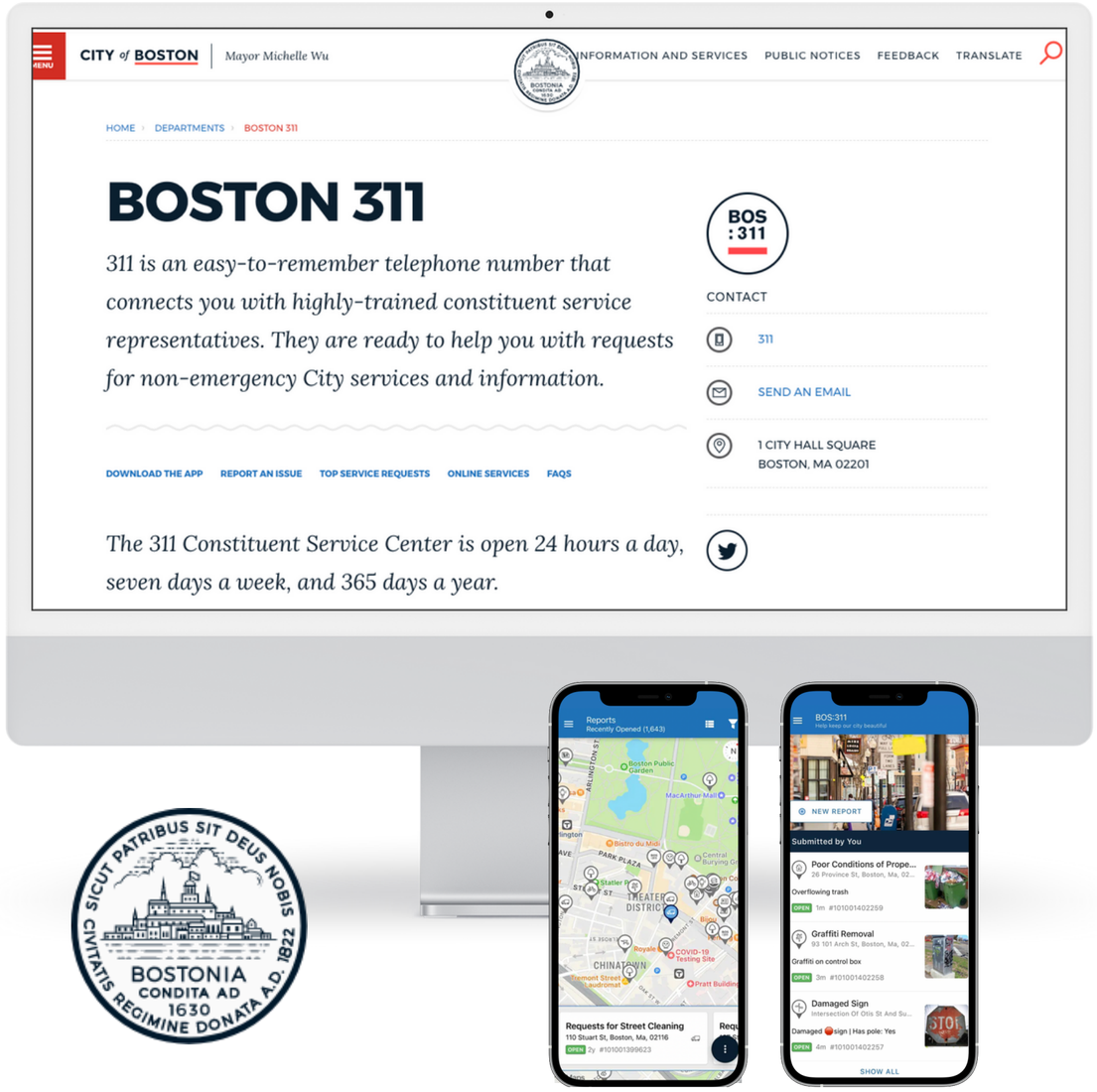
02

How can we leverage **seasonality patterns** from historical data to make accurate **predictions of the duration** it takes to **close a case?**

What is Boston 311?

The City of Boston's 311 Constituent Service Center

Is a vital tool **residents** can use to report a wide range of **non-emergency issues** by phone call, through the app or website. It is a 24/7 service, the 311 system has been instrumental in facilitating the reporting of these daily issues since **2011**.



Call 311

Download the App

Tweet @BOS311

File a Report Online

BOS:311

Data Profiling Report

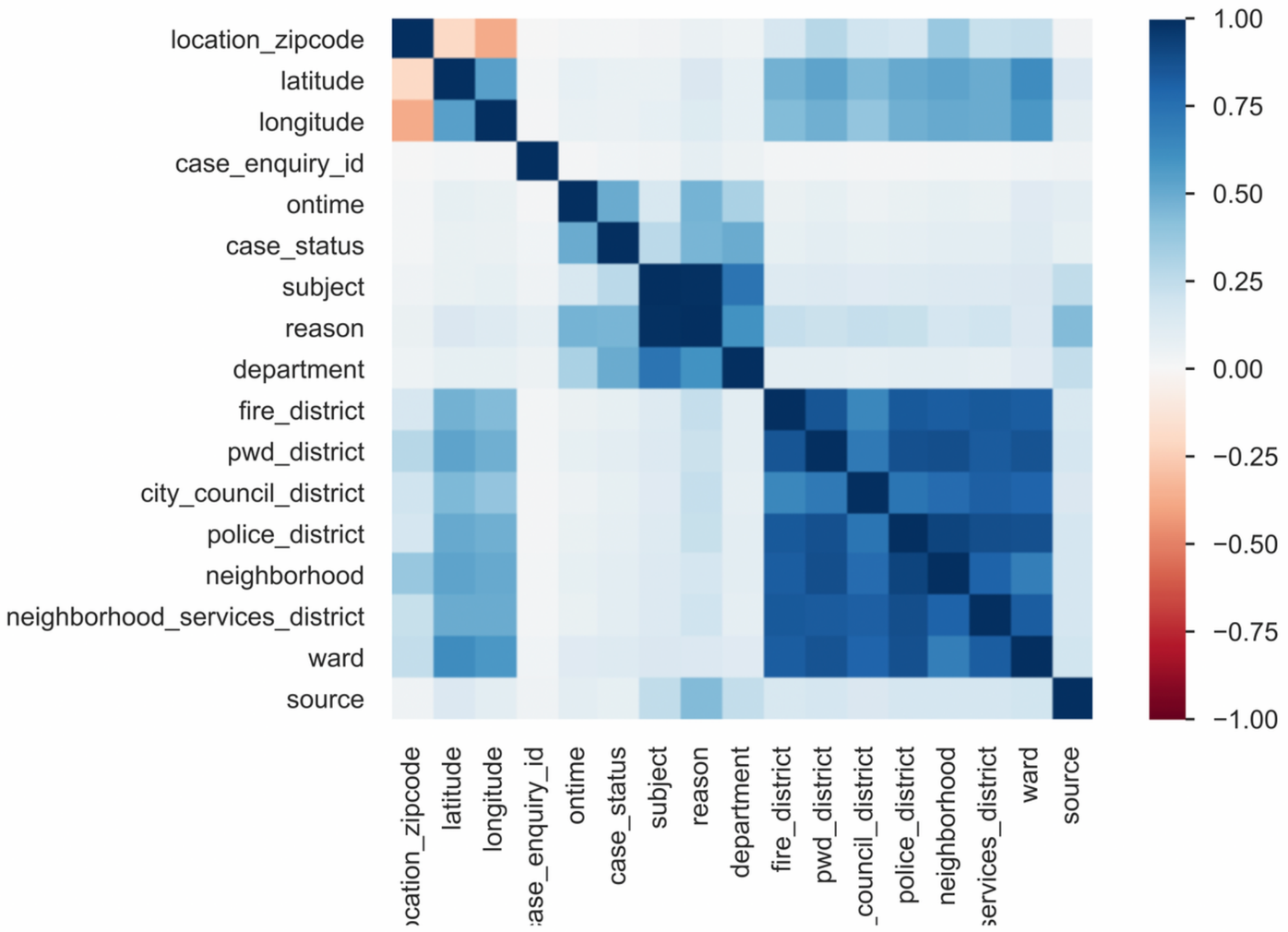
Number of variables	29
Number of observations	276723
Missing cells	534758
Missing cells (%)	6.7%
Duplicate rows	77
Duplicate rows (%)	< 0.1%
Total size in memory	61.2 MiB
Average record size in memory	232.0 B

Variable types

Categorical	26
Numeric	3

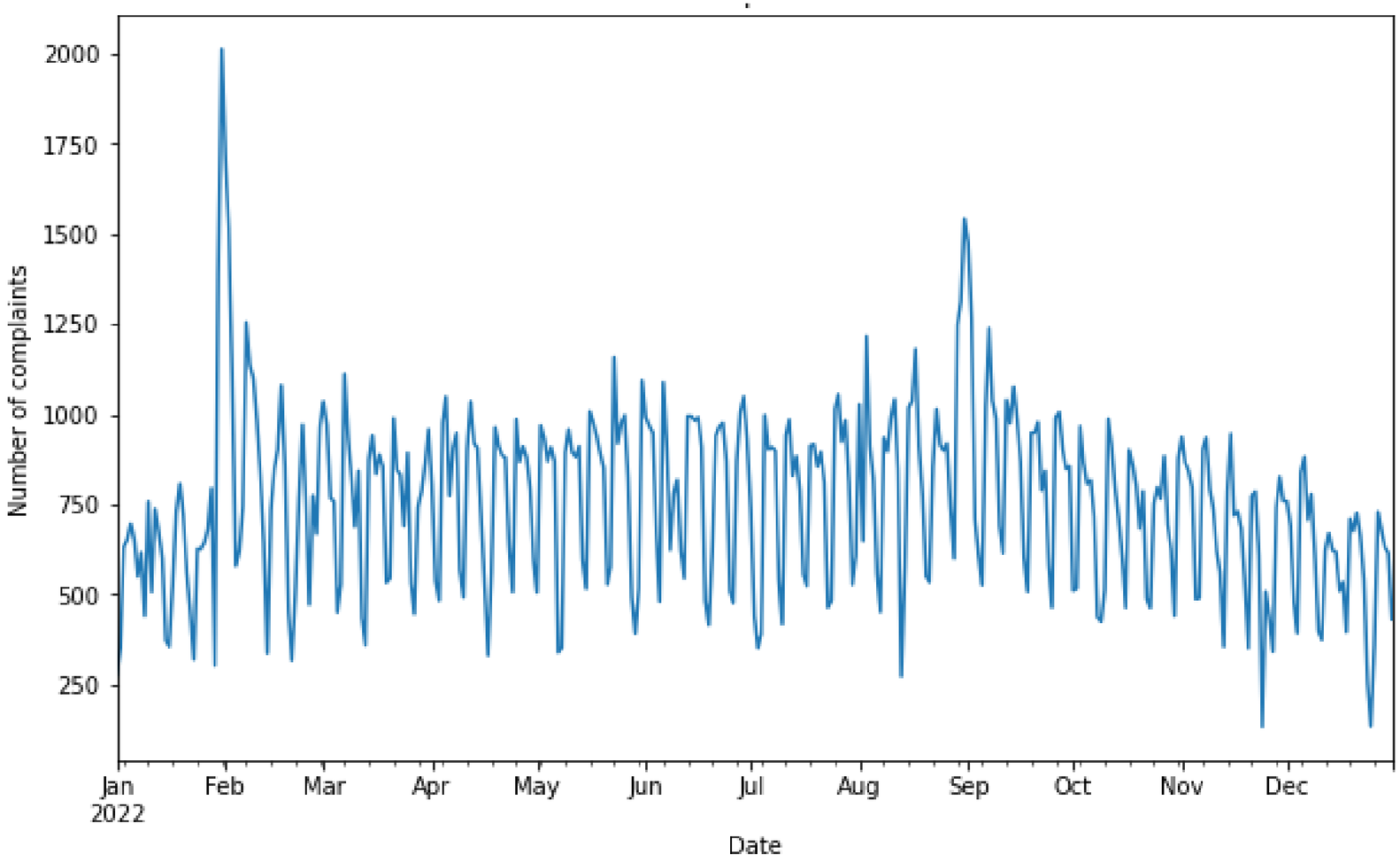
BOS:311

Correlation Matrix

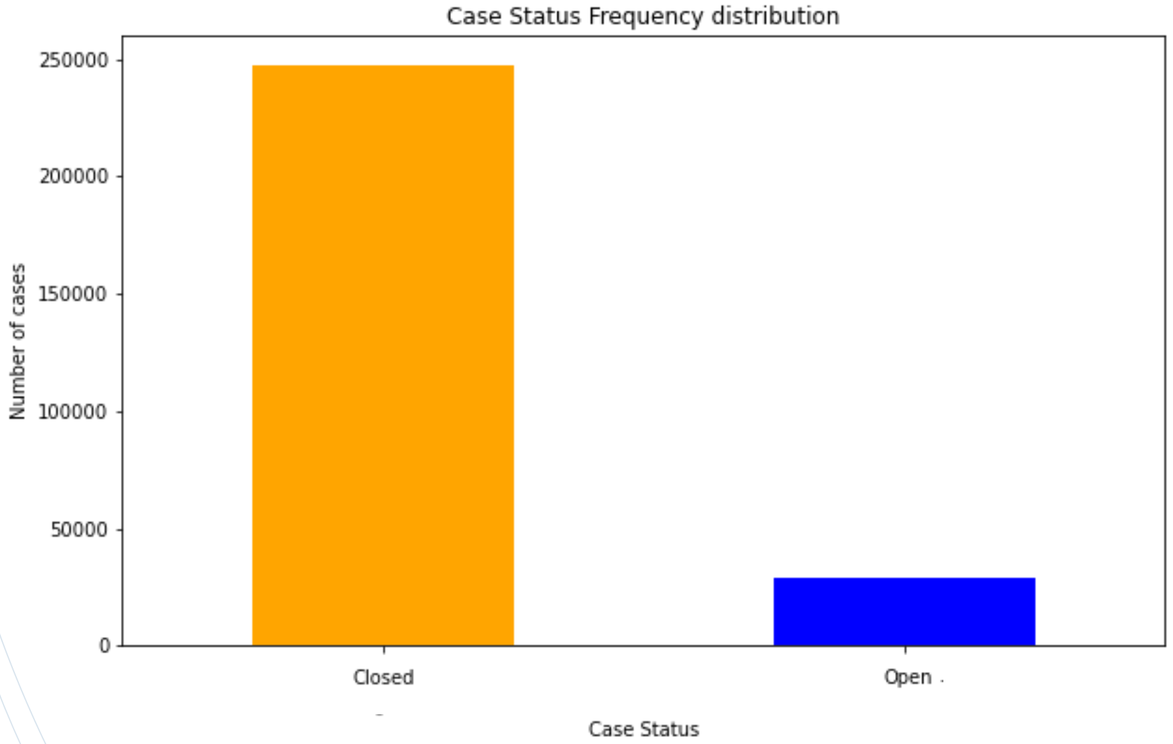


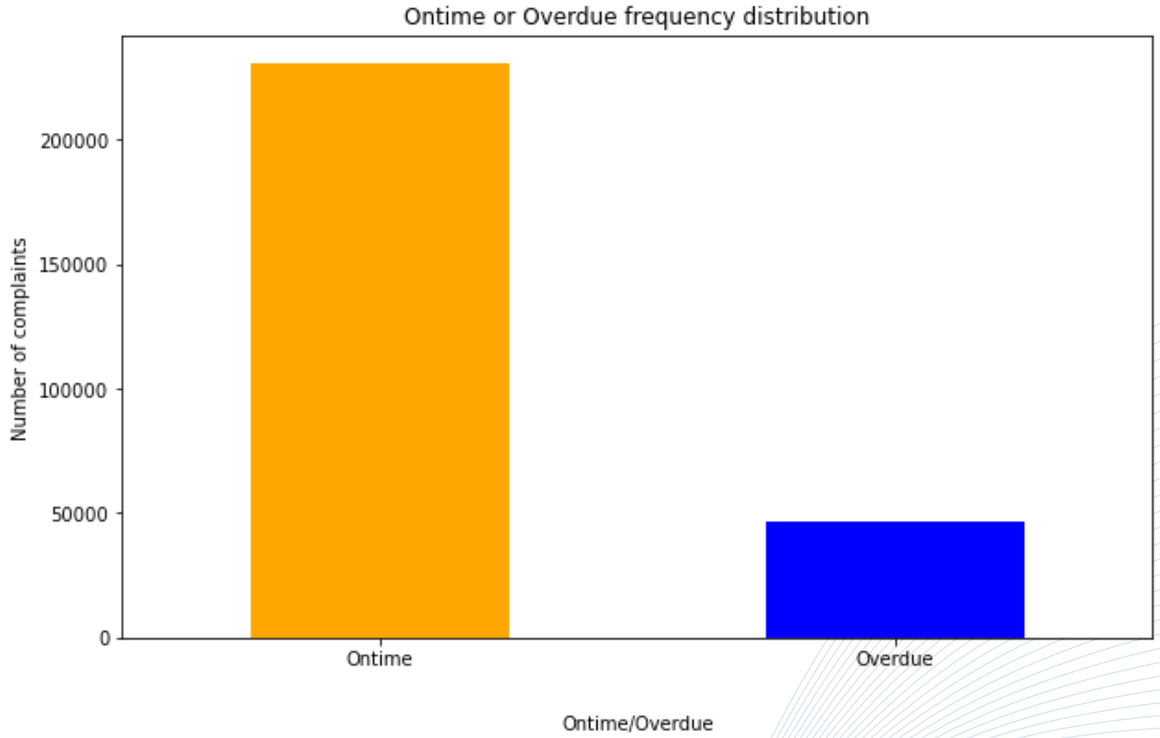
BOS
:311

Number of Complains Over Time



Status & Resolution Analysis





Data Cleaning & Transforation

BOS :311

Step 1

Remove:

- Case_enquiry_id

Split:

- Opendate column: open_date & open _time
- Targetdate: target_date & target_time
- Closing date: closing_date & closing_time

Rename:

- Subject - dept_responsible
- Reason - team_Responsible
- Type – complain_type

Step 2

Delete columns:

- Case_enquiry_id
- department
- queue
- ward
- precinct
- fire_district
- pwd_district
- city_council
- police_district
- location
- neighborhood_services_district
- open_dt
- target_dt
- closed_dt
- closure_reason

Create Boolean values: (1= photo & 0 = none)

a. 'Submitted photo'

b. 'Closed_photo'

Step 3

Convert columns to dummy variables:

- Source
- Ontime
- Case_status

Created New columns:

- Case_ID (Unique ID)
- Num_of Days(close_dt-open_dt)
- Season (Value Derived from open date)

Replace blank values with 'No Information provided for the following:

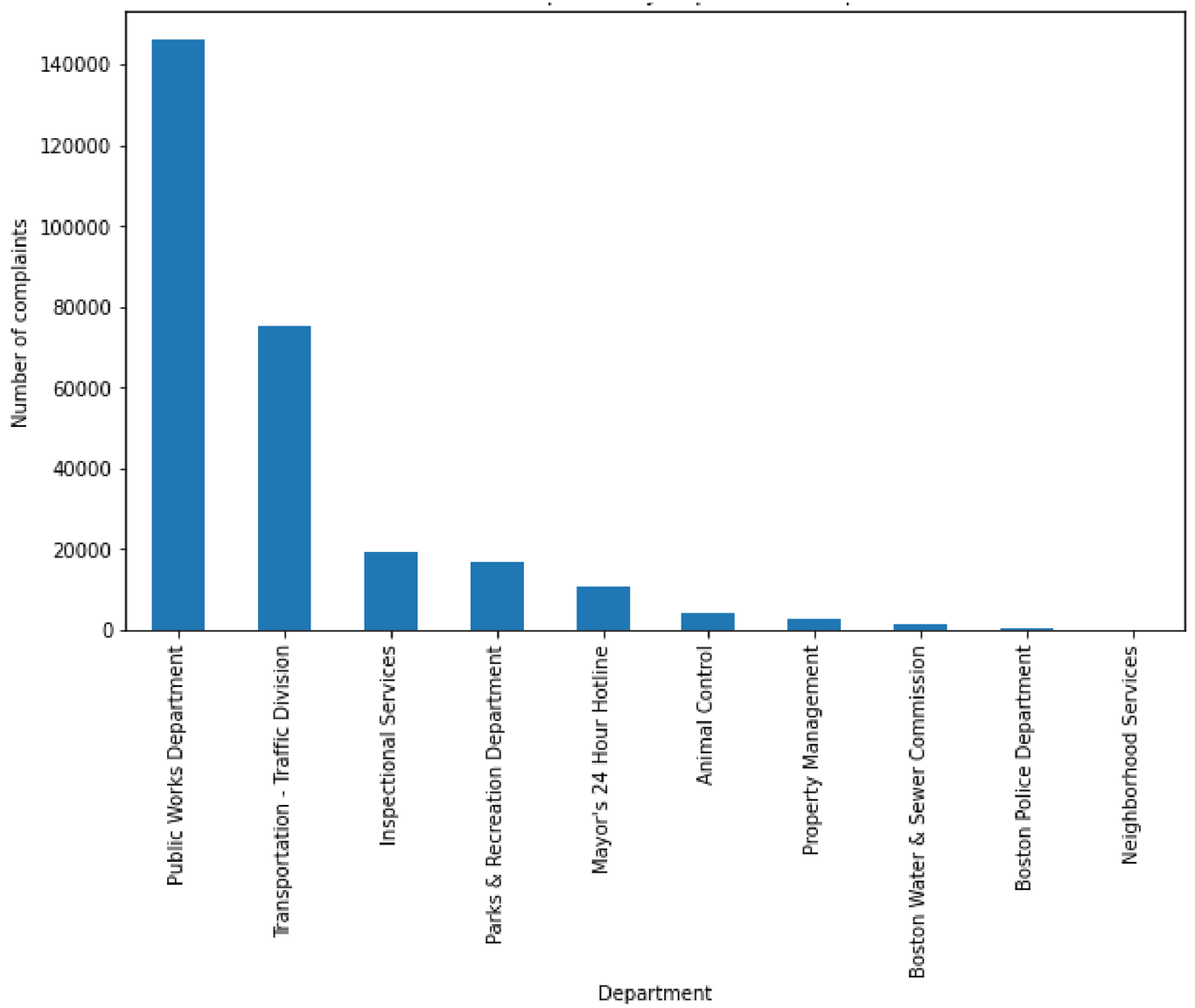
- Neighborhood
- Location_street
- Location_zipcode

Convert:

- Zip code to a 5-digit data value

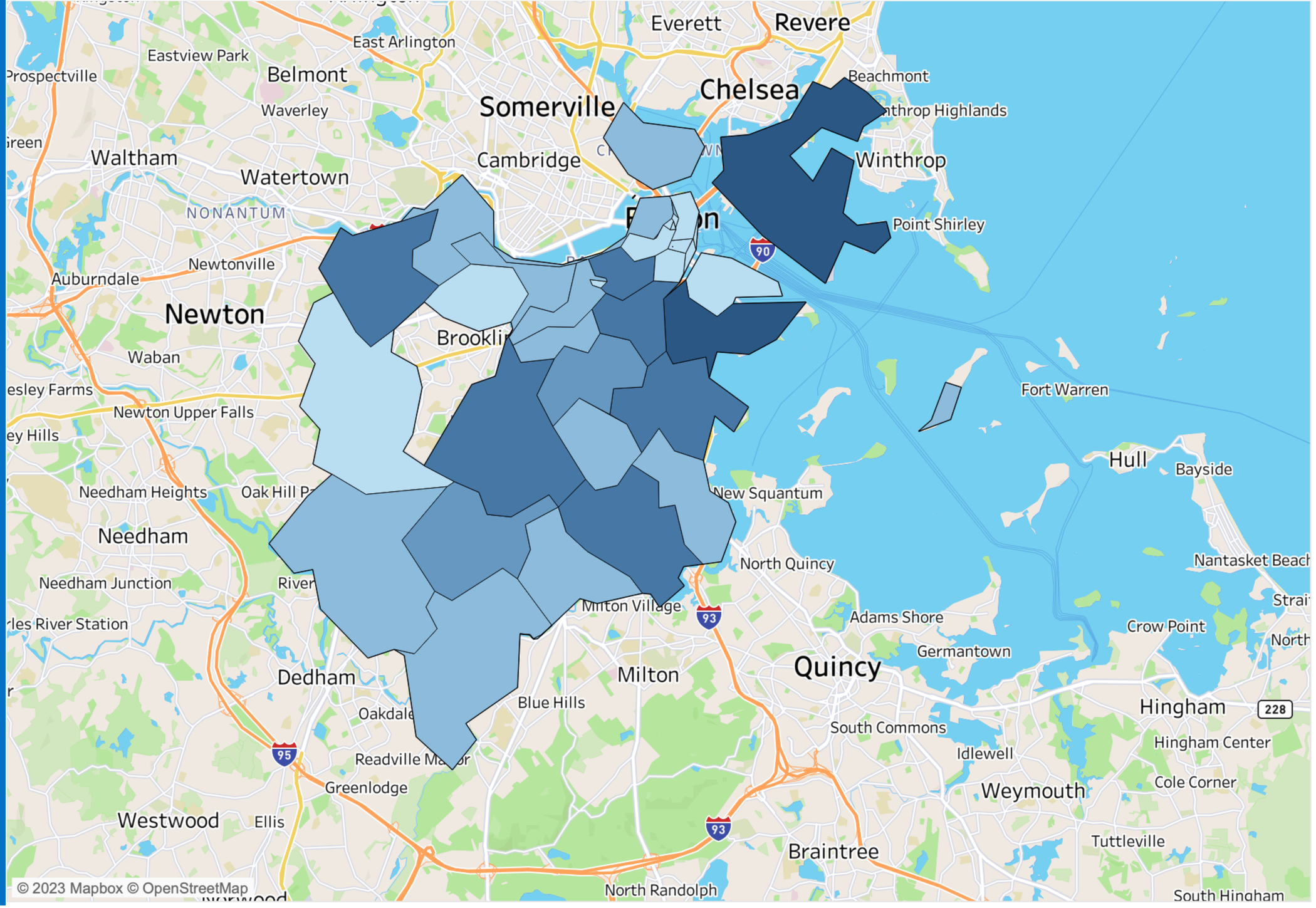
BOS :311

Number of Complains by Department Responsible




Maintenance
Employee
Signals
Building
Environmental
Cleaning
Comments
Signs
Measures
Bicycle
Traffic
Park
Noise
Requests
Management
Animal
Health
Abandoned
Vehicles
Enforcement
Housing
Highway
Weights
Administrative
Engineering
Catchbasin
Program
Trees
Safety
Lights
Generic
Street
Code
Needle
Services
Issues
Recycling
Disturbance
General


Count of Case Enquiry Id
2
19,000
Everett
Revere
East Arlington
Eastview Park
Belmont
Chelsea
Beachmont
Somerville
Waverley
Winthrop Highlands
CHARLESTOWN
Waltham
Cambridge
Winthrop
Watertown
NONANTUM
Boston
Point Shirley
ALLSTON
BACK BAY
Newtonville
Auburndale
SOUTH END
Newton
CHESTNUT HILL
Brookline
Waban
ROXBURY
Fort Warren
Newton Upper Falls
JAMAICA PLAIN
Hull
Bayside
Needham Heights
Oak Hill Park
DORCHESTER
New Squantum
Needham
WEST ROXBURY
North Quincy
Nantasket Beach
Needham Junction
Riverdale
Milton Village
Adams Shore
Crow Point
Germantown
HYDE PARK
Quincy
Dedham
Milton
Hingham
Blue Hills
Oakdale
South Commons
Hingham Center
Idlewell
Readville Manor
Weymouth
Cole Corner
Greenlodge
Westwood
Ellis
Braintree
Tuttleville
North Randolph
© 2023 Mapbox © OpenStreetMap


Count of Case Enquiry Id
2
19,000
Everett
Revere
East Arlington
Eastview Park
Belmont
Chelsea
Beachmont
Somerville
Waverley
Waltham
Cambridge
Winthrop
Watertown
NONANTUM
Point Shirley
Newtonville
Auburndale
Newton
Waban
Fort Warren
Newton Upper Falls
Hull
Bayside
Needham Heights
New Squantum
Needham
North Quincy
Nantasket Beach
Needham Junction
Milton Village
Adams Shore
Crow Point
Germantown
Quincy
Dedham
Milton
Blue Hills
Hingham
South Commons
Idlewell
Hingham Center
Readville
Greenlodge
Weymouth
Cole Corner
Westwood
Ellis
Braintree
Tuttleville
North Randolph
South Hingham
90
93
95
228
© 2023 Mapbox © OpenStreetMap

Cases closed within the target date were **flagged** as **1**; issue not closed as **0**

01

Can we **predict the probability of issue resolution** when photographic evidence is submitted, while considering the impact of the **source** and **location** of the report?

Impact variables considered

- Latitude
- Longitude
- Source_City_Worker_App
- Source_Constituent_Call
- Source_Employee_Generated
- Source_Self_Service

BOS :311

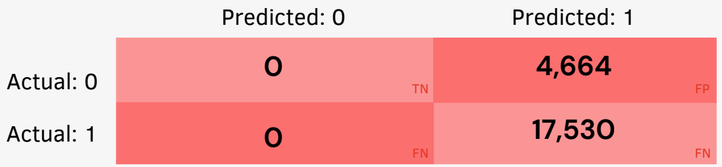
Logistic Regression

Attempt 01

Logistic regression imbalanced dataset

01

LOGISTIC REGRESSION ON IMBALANCED DATASET



Error: TN and FN not identified

F1 Score: 0.8825

Accuracy: 0.7898

Precision: 0.7898

Recall: 1.0

BOS : 311

Logistic Regression

Attempt 01

Logistic regression imbalanced dataset

01

Attempt 02

Logistic regression balanced dataset

02

LOGISTIC REGRESSION ON IMBALANCED DATASET

	Predicted: 0	Predicted: 1
Actual: 0	0 (TN)	4,664 (FP)
Actual: 1	0 (FN)	17,530 (FN)

Error: TN and FN not identified

F1 Score: 0.8825
Accuracy: 0.7898
Precision: 0.7898
Recall: 1.0

LOGISTIC REGRESSION ON BALANCED DATASET

	Predicted: 0	Predicted: 1
Actual: 0	6,915 (TN)	10,565 (FP)
Actual: 1	7,082 (FN)	10,670 (FN)

Improvement: TN and FN predicted
Error: F1 and accuracy scores compromised

F1 Score: 0.5473
Accuracy: 0.4991
Precision: 0.5024
Recall: 0.6010

BOS :311

Random Forest

RANDOM FOREST ON IMBALANCED DATASET

	Predicted: 0	Predicted: 1
Actual: 0	816 (TN)	3,848 (FP)
Actual: 1	988 (FN)	16,542 (FN)

F1 Score: 0.8724

Accuracy: 0.7821

Precision: 0.8112

Recall: 0.9436

Recommendation:

- Utilize the model for upcoming service requests
- Determine if request will meet target closure time
- Implement additional measures for requests predicted to be delayed
- Allocate special attention or resources to reduce delay

Attempt 01

Logistic regression
imbalanced dataset

01

Attempt 02

Logistic regression
balanced dataset

02

Attempt 03

Random Forest
imbalanced dataset

03

New column capturing the **season** was **created** which was further dummified.

02

How can we leverage **seasonality patterns** from historical data to make accurate **predictions of the duration** it takes to **close a case?**

Impact variables considered

Dept_Responsible	Latitude
Team_Responsible	Longitude
Neighborhood	Source_City_Worker_App
Submitted_Photo	Source_Constituent_Call
Season	Source_Employee_Generated
Source_Self_Service	

BOS :311

Multiple Regression

Observations:

- Dummification resulted in **65 predictor variables**
- Coefficients of determination: **0.21 (poor fit to data)**
- High RMSE of **375.54**

```
Coefficients: [ 7.83553324e-01 -2.50579632e+01  4.65345366e-01  8.9664665
9e-01
 -9.45334102e-02 -2.42340020e+00  2.70649667e-01  4.69345578e+12
  6.24927957e+12 -9.37483958e+11  6.20052296e+12  1.39798209e+13
 -4.12969610e+12  5.59843218e+12  5.59843218e+12 -1.89870542e+11
  7.15105859e+11 -1.89870542e+11  6.34604559e+12 -8.57125925e+12
 -1.89870542e+11 -7.91961320e+11 -1.89870542e+11  6.34604559e+12
 -7.91961320e+11  9.53825773e+12  6.34604559e+12 -1.89870542e+11
  6.34604559e+12 -8.40717937e+11 -1.89870542e+11 -8.57125925e+12
 -1.89870542e+11 -1.89870542e+11 -1.89870542e+11 -1.89870542e+11
 -1.89870542e+11 -1.89870542e+11 -1.89870542e+11 -8.57125925e+12
  6.34604559e+12  5.77247425e+11  5.77247425e+11  5.77247425e+11
  5.77247425e+11  5.77247425e+11  5.77247425e+11  5.77247425e+11
  5.77247425e+11  5.77247425e+11  5.77247425e+11  5.77247425e+11
  5.77247425e+11  5.77247425e+11  5.77247425e+11  5.77247425e+11
  5.77247425e+11  5.77247425e+11  5.77247425e+11  5.77247425e+11
  5.77247425e+11  5.77247425e+11  5.77247425e+11  5.77247425e+11
  5.77247425e+11  2.83491027e+11  2.83491027e+11  2.83491027e+11
  2.83491027e+11]
Mean squared error: 375.5409418774977
Coefficient of determination: 0.21459284129487244
```

Attempt 01

Multiple regression on impact variables

01

BOS :311

Multiple Regression

With forward feature selection

No	Feature	Coefficient
1	latitude	-16.88
2	dept_responsible_Inspectional Services	-1.97
3	dept_responsible_Parks & Recreation Department	3.95
4	dept_responsible_Property Management	80.99
5	dept_responsible_Public Works Department	-3.73
6	team_responsible_Building	15.06
7	team_responsible_Highway Maintenance	3.78
8	team_responsible_Housing	3.83
9	team_responsible_Recycling	11.37
10	team_responsible_Signs & Signals	8.95
11	team_responsible_Street Cleaning	1.1
12	team_responsible_Street Lights	19.25
13	team_responsible_Traffic Management & Engineering	9.54
14	team_responsible_Trees	28.47
15	neighborhood_Dorchester	1.48
16	neighborhood_Roxbury	1.09
17	neighborhood_South Boston / South Boston Water...	-1.03
18	season_Spring	3.53
19	season_Summer	1.75
20	season_Winter	5.7

Attempt 01

Multiple regression on impact variables

01

Attempt 02

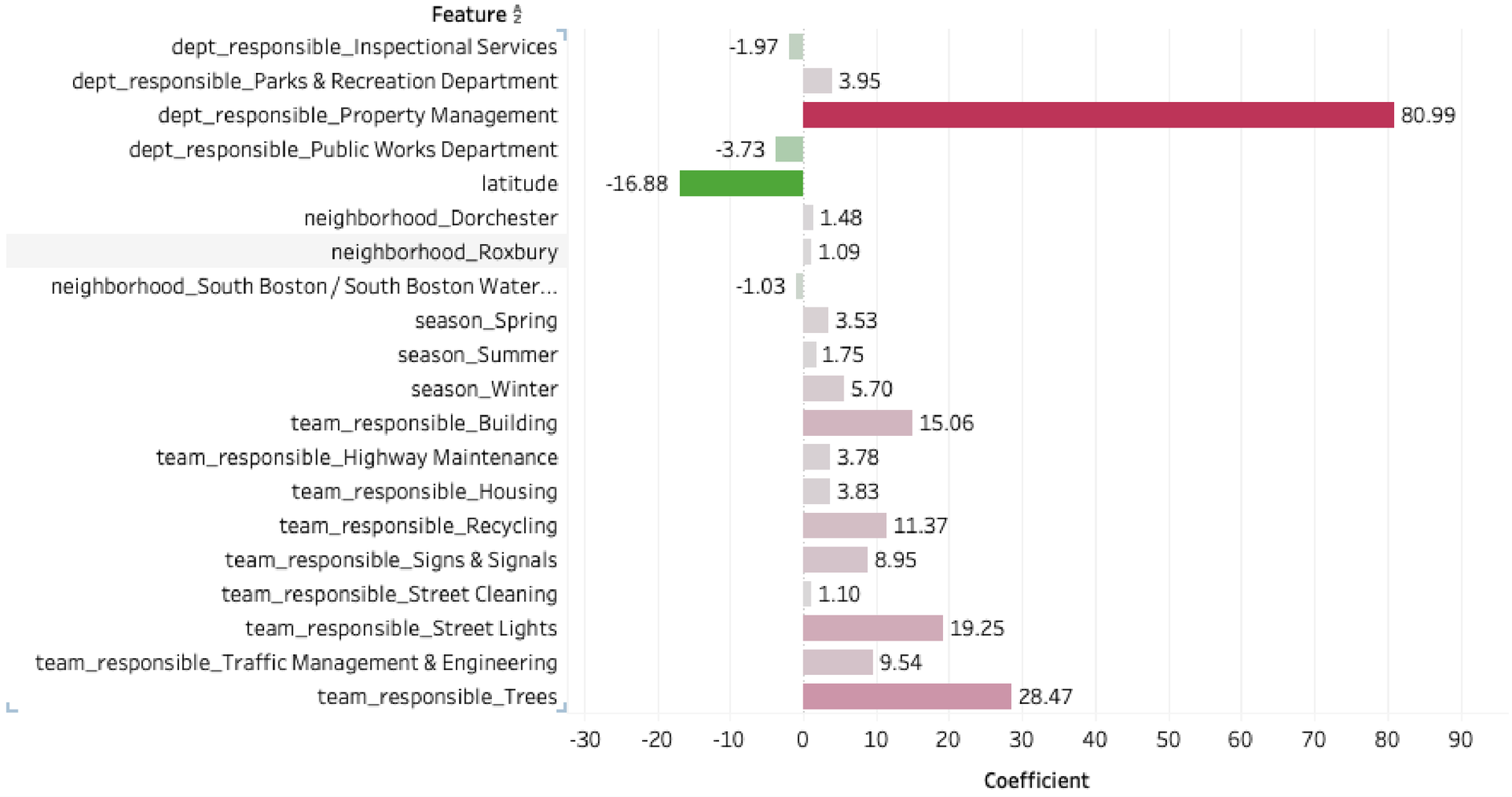
Multiple regression with forward feature selection

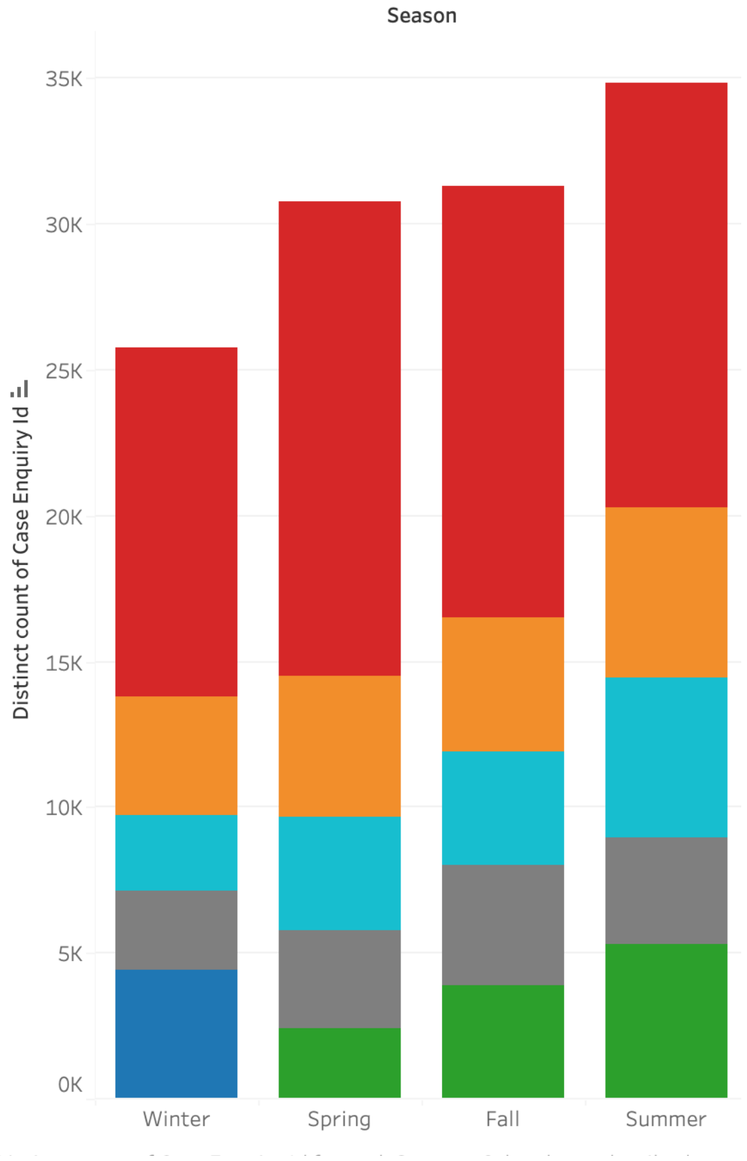
02

Observations:

- Features reduced to **20**
- RMSE improvement: **19.38**
- **Positive** coefficients → **increase** days
- **Negative** coefficients → **decrease** days

BOS:311 Impact of Features in Time Taken for Resolution





BOS:311

Top 5 Complains in Each Season

Complain Type

- Parking Enforcement
- Requests for Street Cleaning
- Improper Storage of Trash (Barrels)
- Missed Trash/Recycling/Yard Waste/Bulk Item
- CE Collection
- Request for Pothole Repair

Steps to Improve Service Efficiency

- Identify **root cause** of delay moving from west to east in Boston
- Review and improve **Property department**'s working system for faster case closure
- Implement **monitoring** for teams struggling with average closure time
- Study reasons for delay in **Dorchester** & **Roxbury** cases and improve resolution time
- Hire **more staff** and use special gear for **winter** season to avoid delays

THANK YOU.

Any questions?