NYC311 Housing and Buildings Complaint Analysis

Purwadhika Final Project

Azka Nur Afifah

Azka Nur Afifah

Education: Chemical Engineering

Past experience: Graduate Research Assistant

Interest: Renewable energy, education, data analysis

Email: azkanuraf@gmail.com

github.com/azukacchi

Introduction

NYC 311

- Provides government and non-emergency services.
- Complaints will be forwarded to various agencies
  - local governments, public utilities, and other public services
- Agencies:
  - NYC Police Department: noise, illegal parking, blocked driveway
  - Dept of Housing Preservation and Dev: heat/hot water, unsanitary condition, plumbing
  - Dept of Transportation: street, street light, traffic signal condition

NYC 311

- Provides government and non-emergency services.
- Complaints will be forwarded to various agencies
  - local governments, public utilities, and other public services
- Agencies:
  - NYC Police Department: noise, illegal parking, blocked driveway
  - Dept of Housing Preservation and Dev: heat/hot water, unsanitary condition, plumbing
  - Dept of Transportation: street, street light, traffic signal condition

Problem Statement

- Dept of Housing Preservation and Development (HPD) is the agency responsible to process the complaints related to housing and building
- Receives 2nd highest number of calls of all agencies[1]
- Manage large volume of complaints by focusing on the most pressing complaint and determine the possibility of the complaints.

Dataset

1. NYC 311 Service Request

   Complaint details such as complaint type, location and address, complaint response, status.

   Raw dataset: >10M rows, 16 columns

2. PLUTO Dataset

   The Primary Land Use Tax Lot Output by NYC Department of City Planning. Contains land use and geographic data at tax lot level.

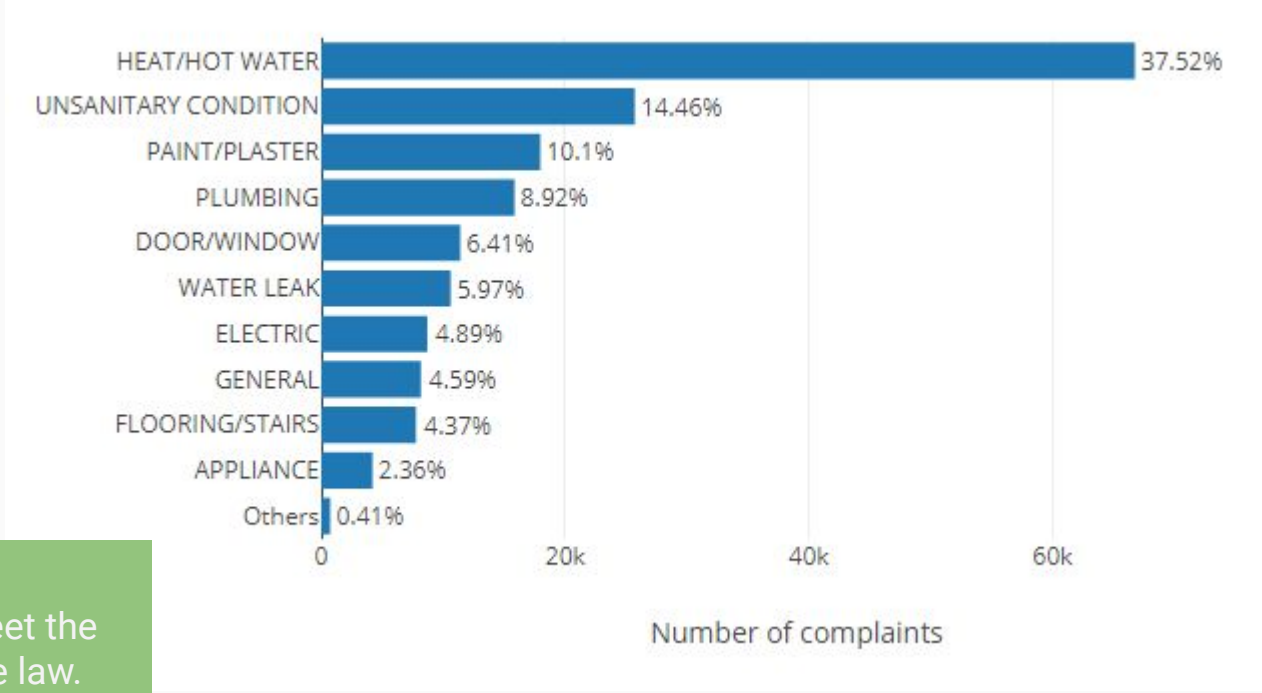
   Raw dataset: 89k rows, 87 columns

Exploratory Data Analysis

Top 10 Complaints in 2015 - 2019

>⅓

of housing and buildings complaints are HEAT/HOT WATER complaints

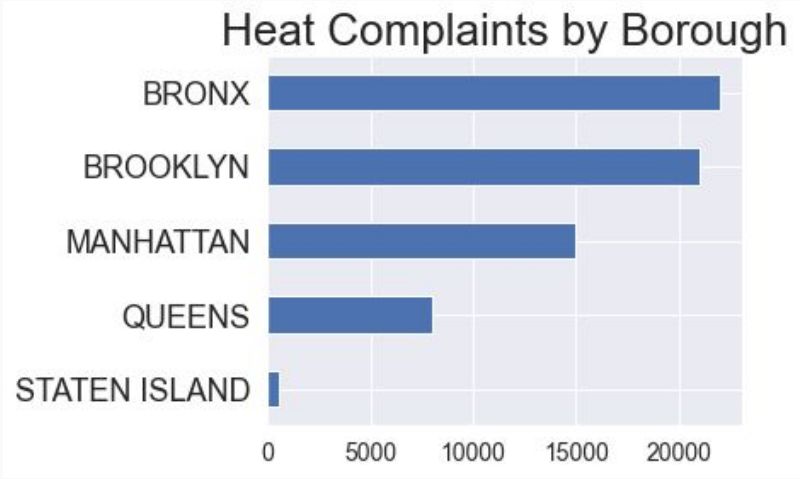


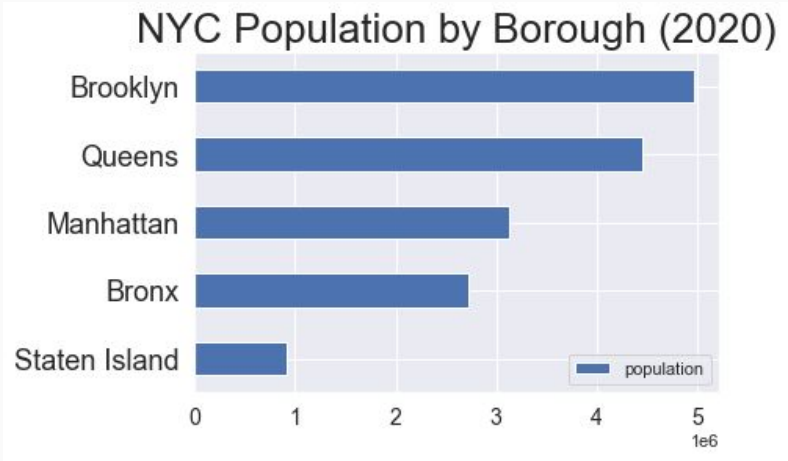
HEAT/HOT WATER Complaints?
Building owners are required to meet the heat requirements according to the law. This includes residential building, commercial building, and other public facilities.

In each year, HEAT/HOT WATER has the highest number of complaints

Heat Complaints at Borough Level

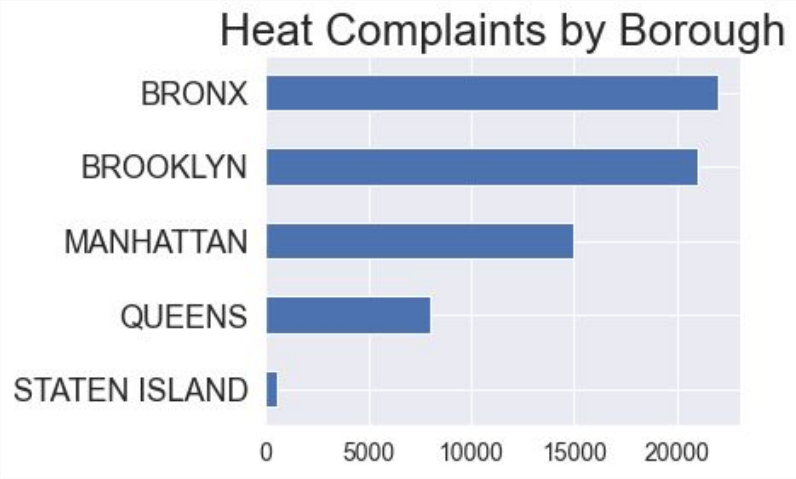
- **Bronx** has the highest number of HEAT/HOT WATER complaints
- Bronx is only the **4th** most-populated borough
- Bronx has the **highest heat complaints density**

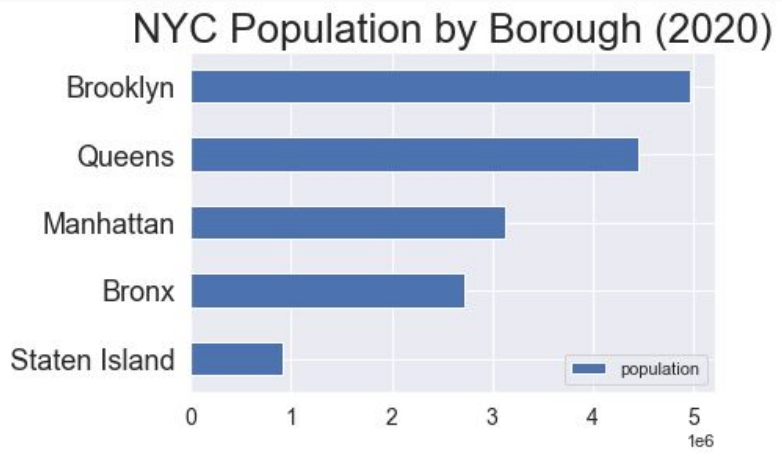




Heat Complaints at Borough Level

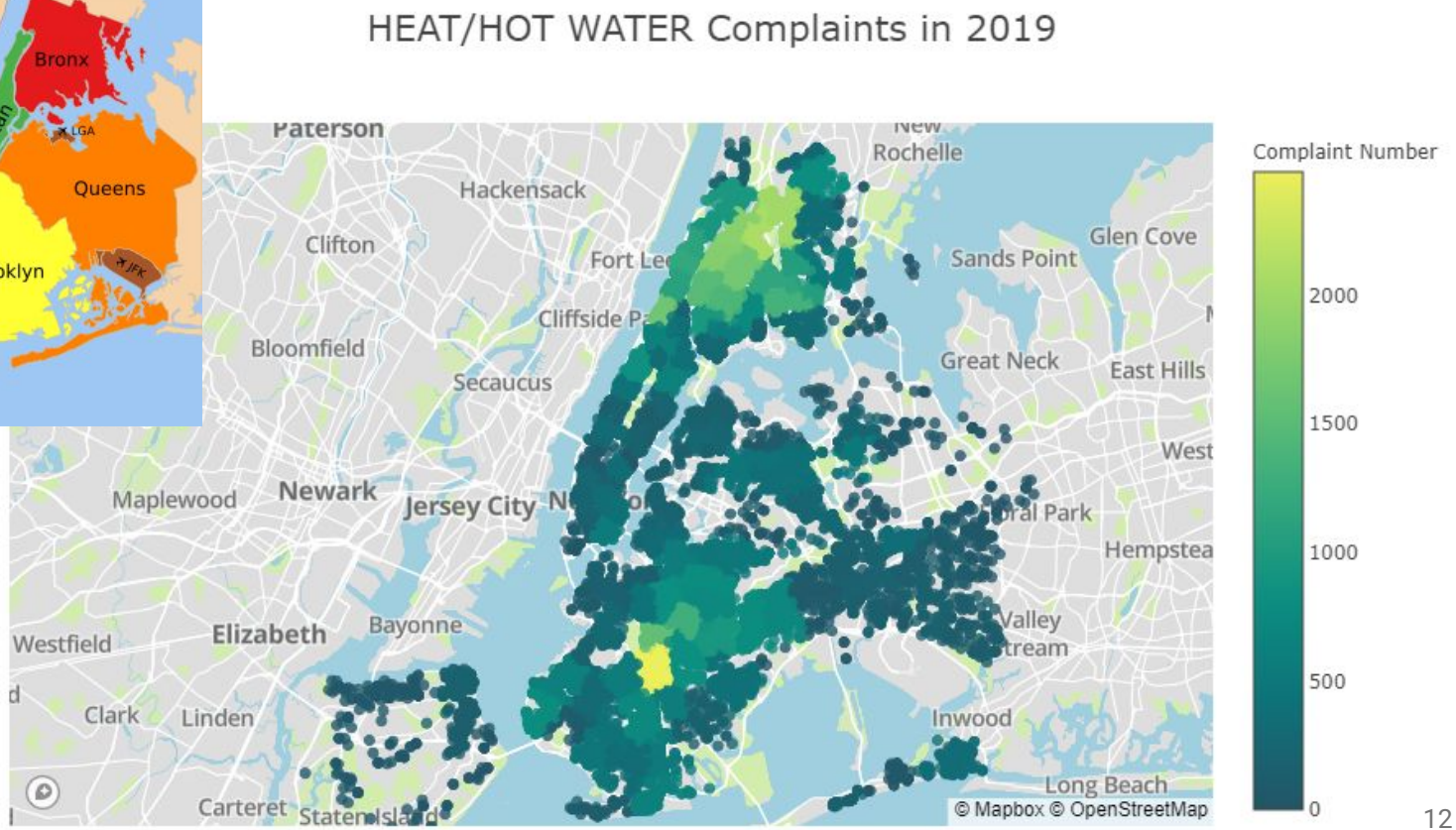
- **Bronx** has the highest number of HEAT/HOT WATER complaints
- Bronx is only the **4th** most-populated borough[2]
- Bronx has the **highest heat complaints density**
- **Focus on HEAT/HOT WATER complaints in Bronx**





Heat Complaints Map at Zip Code Level

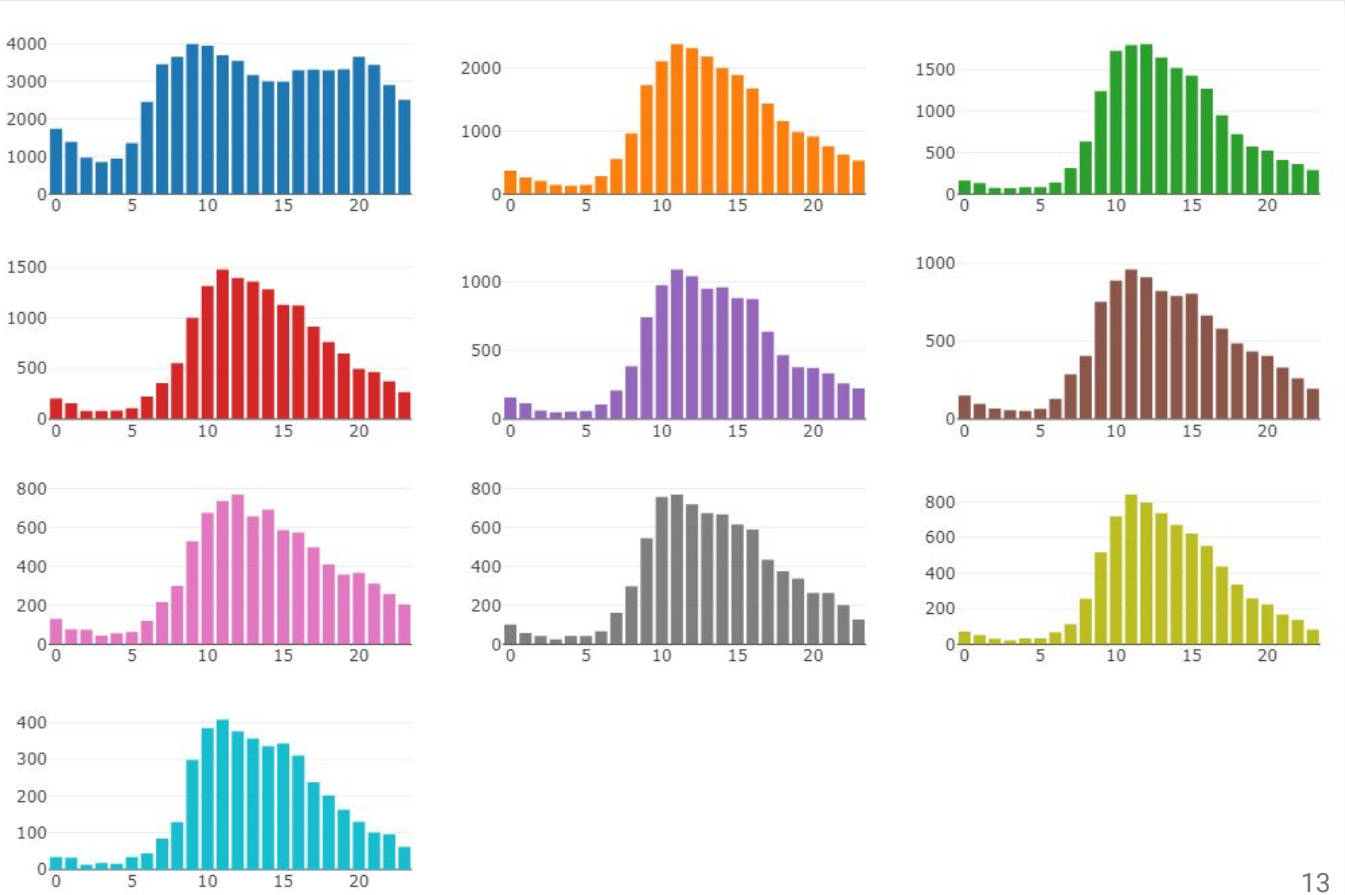
NYC Borough Map
Source: Wikipedia[3]



Heat Complaints Pattern

Complaint Type Throughout the Day

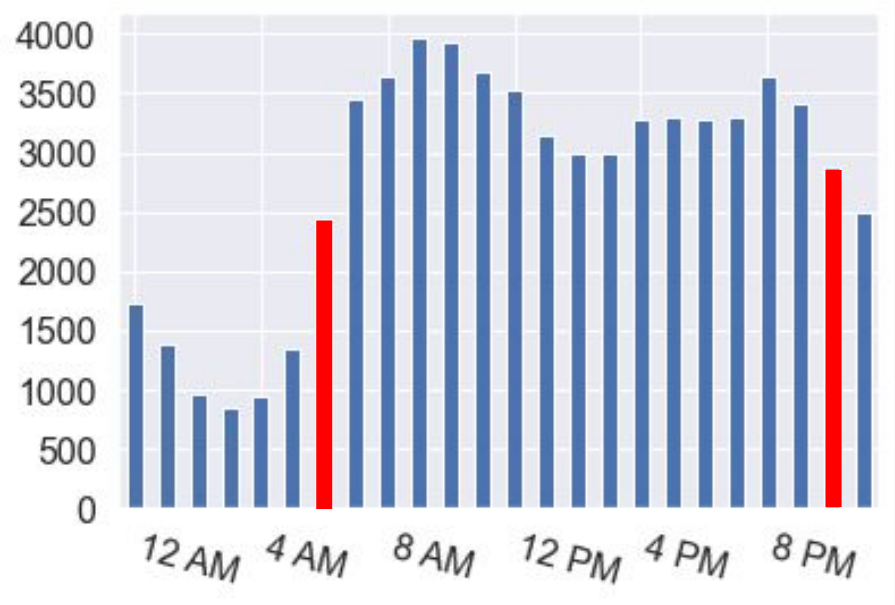
- HEAT/HOT WATER
- UNSANITARY CONDITION
- PAINT/PLASTER
- PLUMBING
- DOOR/WINDOW
- WATER LEAK
- ELECTRIC
- GENERAL
- FLOORING/STAIRS
- APPLIANCE



Heat Complaints Pattern

Heat Complaint Throughout the Day

- 6 AM - 10 PM
  - Minimum inside temperature requirement (20 C) only if outside temperature falls below certain point (12.7 C)
- 10 PM - 6 AM
  - Minimum inside temperature requirement (16.6 C) should always be met



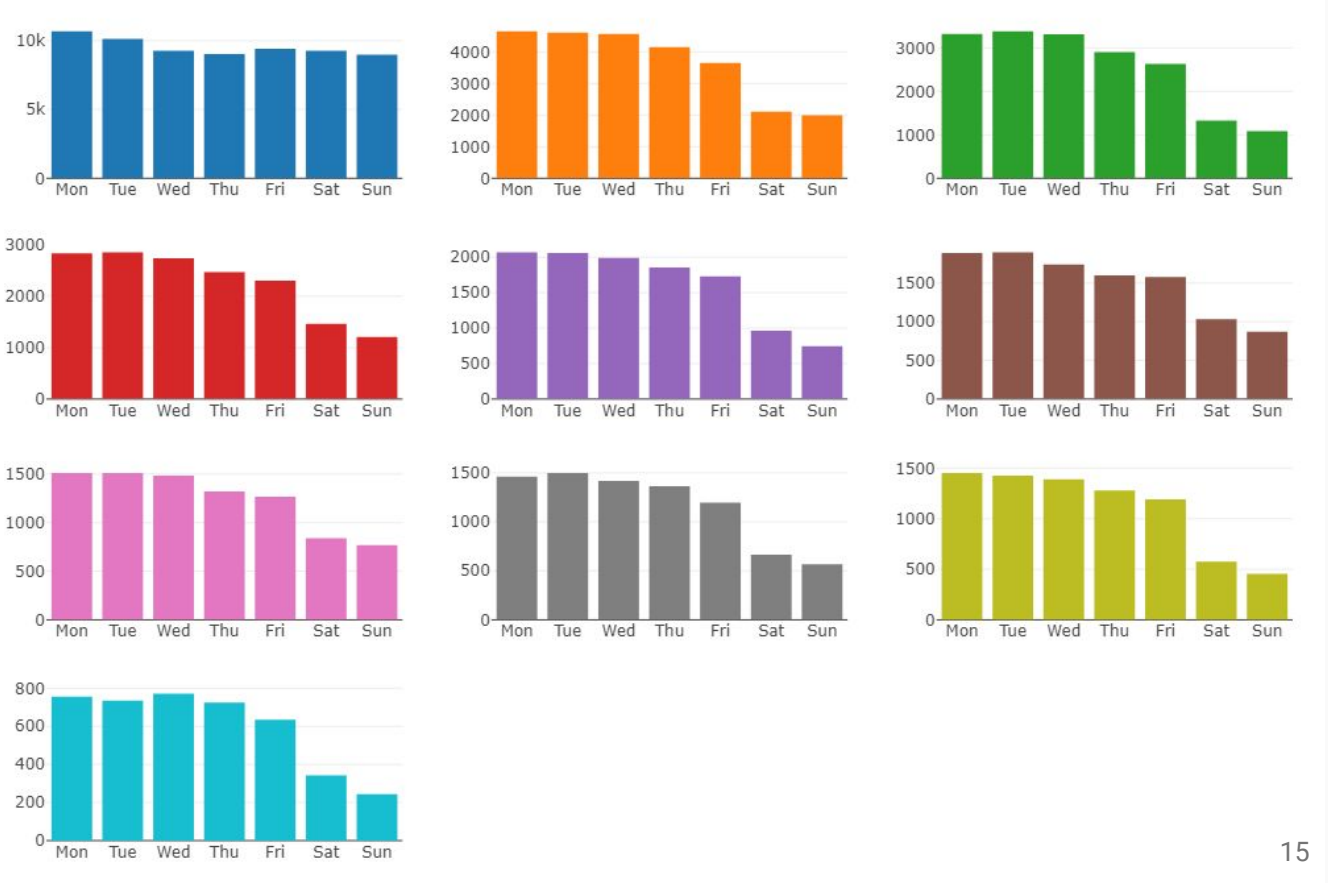
Complaint number started to rise at 6 AM

Number stayed high even after work hour ends (~4 PM)

Heat Complaints Pattern

Complaint Type Throughout the Week

- HEAT/HOT WATER
- UNSANITARY CONDITION
- PAINT/PLASTER
- PLUMBING
- DOOR/WINDOW
- WATER LEAK
- ELECTRIC
- GENERAL
- FLOORING/STAIRS
- APPLIANCE

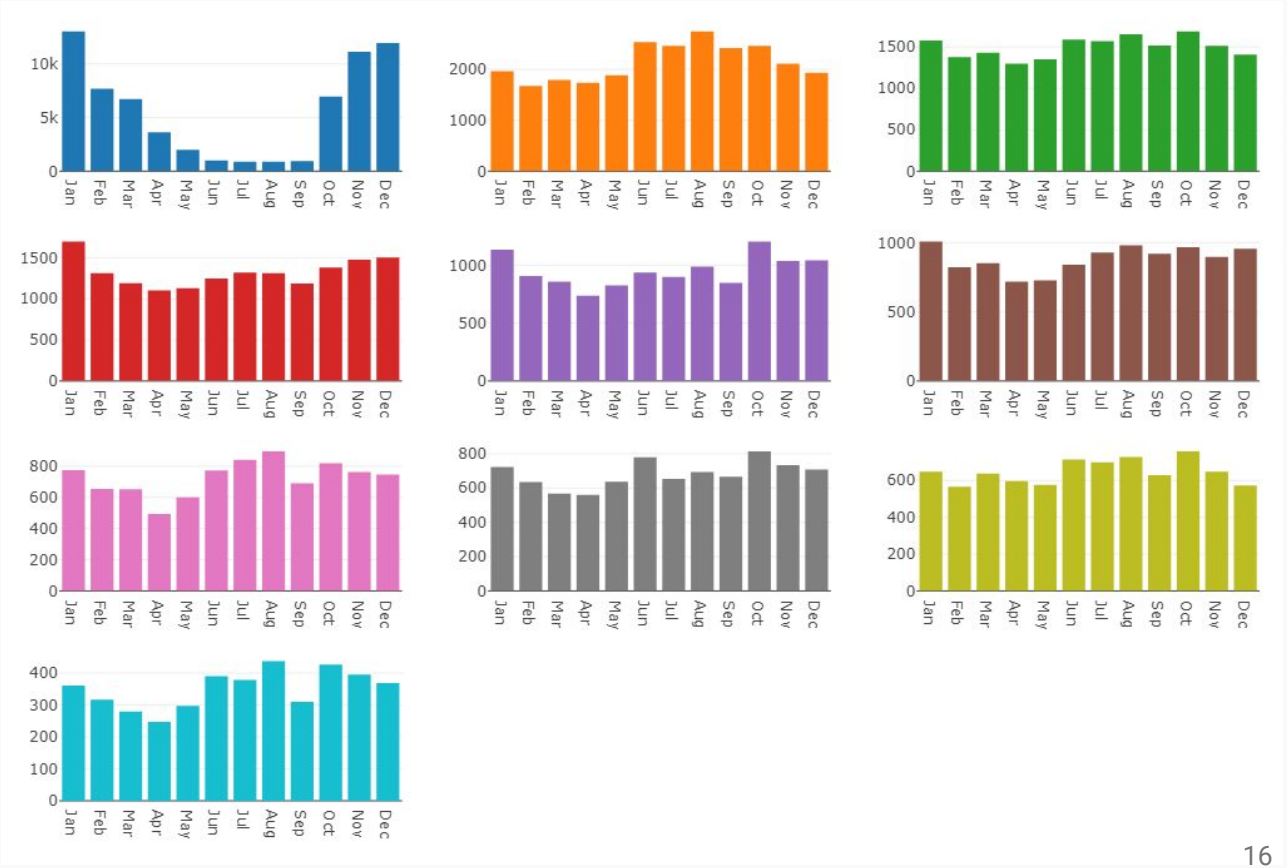


Steady number of heat complaints throughout the week suggests that most of the complaints came from residential building

Heat Complaints Pattern

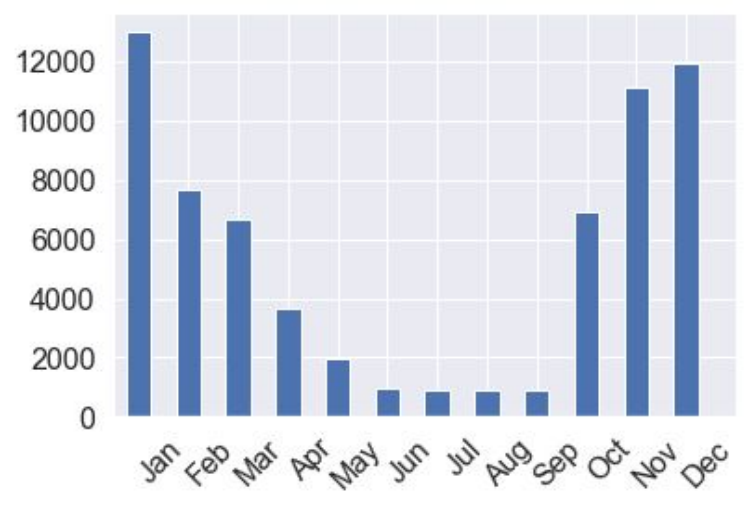
Complaint Type Throughout the Year

- HEAT/HOT WATER
- UNSANITARY CONDITION
- PAINT/PLASTER
- PLUMBING
- DOOR/WINDOW
- WATER LEAK
- ELECTRIC
- GENERAL
- FLOORING/STAIRS
- APPLIANCE

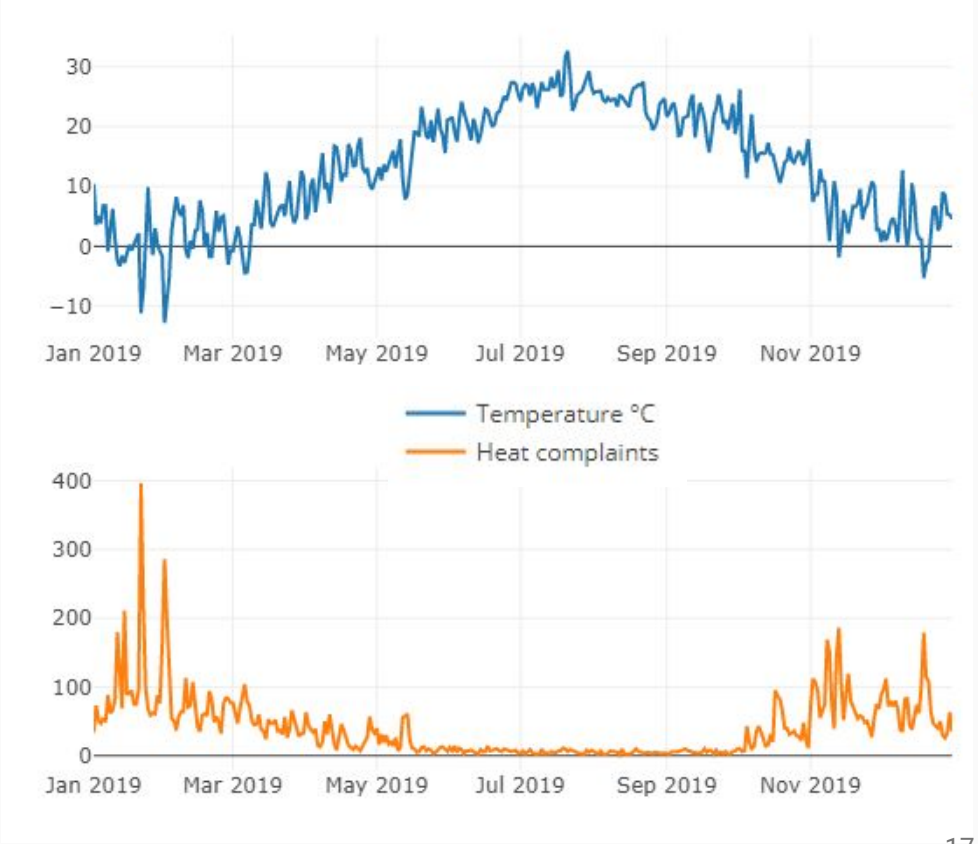


Heat Complaints Pattern

Heat Complaint Throughout the Year



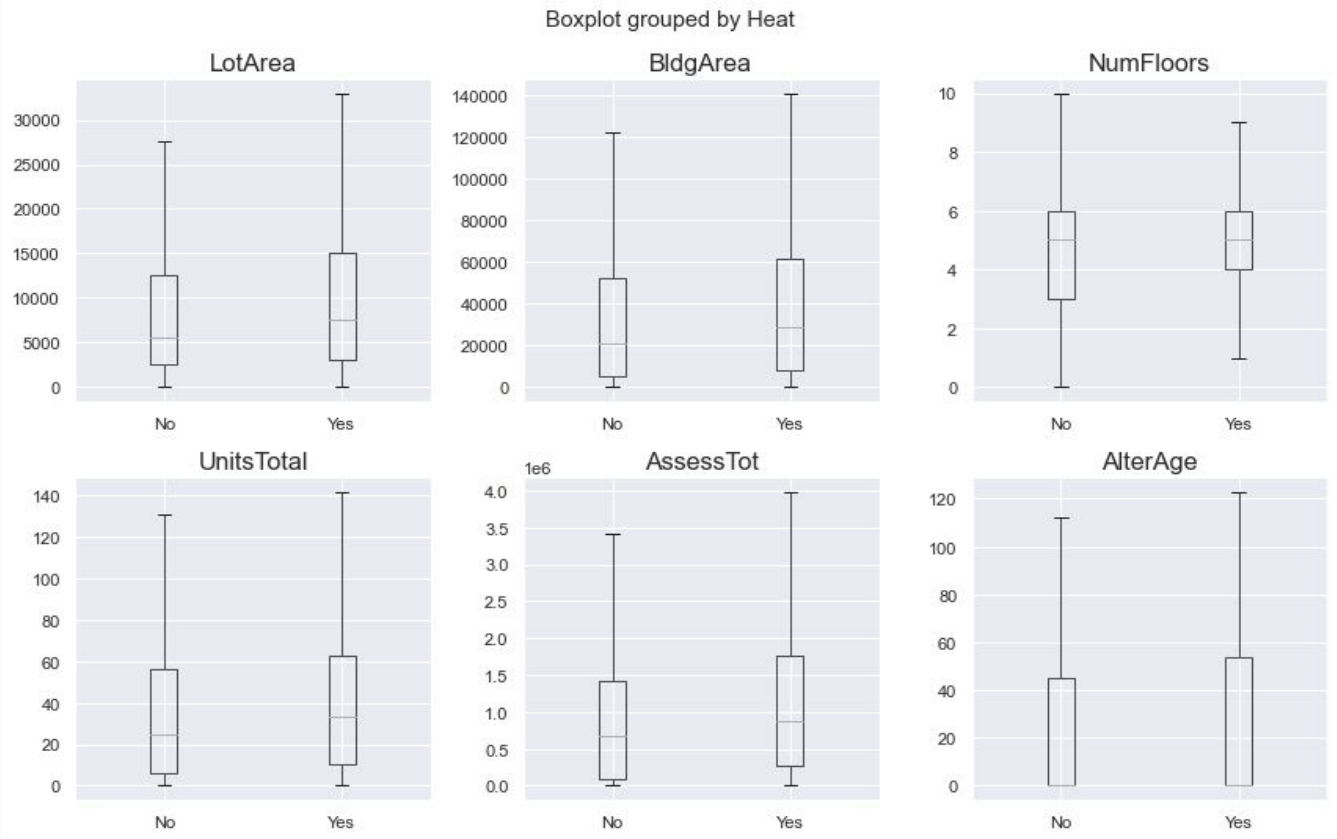
Regulation period: Oct 1 - May 31
Complaints about excessive heat: opens Jun-Sep



Building Characteristics for Heat and Non-Heat Complaints

No difference in distribution observed between the building with and without heat complaints

This might happen because each building can file a heat as well as a non-heat complaint, therefore the building characteristics are not inclusive to either type of complaint.



Predictive Modelling

Goal

- Predict whether a building will have a heat complaint at a certain time during a day on a certain month, given its characteristics.
- When the prediction is done for buildings at a certain area (for example zip code), we can tell the total heat complaint in that area and decide which area needs more attention

Features Used

Dataset used was the NYC 311 Service Request and PLUTO Dataset merged on the incident address. There are 12 final features used:

No	Feature	Description
1	month	the month the complaint happens
2	OfficeArea	an estimate of the exterior dimensions of the portion of the structure(s) allocated for office use
3	GarageArea	an estimate of the exterior dimensions of the portion of the structure(s) allocated for garage use
4	StrgeArea	an estimate of the exterior dimensions of the portion of the structure(s) allocated for storage use
5	ResArea	an estimate of the exterior dimensions of the portion of the structure(s) allocated for residential use
6	BldgArea	the total gross area in square feet
7	OtherArea	allocated area other than for Residential, Office, Retail, Garage, Storage, Loft or Factory use
8	NumFloors	in the tallest building on the tax lot, the number of full and partial stories starting from the ground floor
9	UnitsTotal	the sum of residential and non-residential (offices, retail stores, etc.) units in all buildings on the tax lot
10	AssessTot	the tentative assessed total value for Fiscal Year 2018
11	hourbin	the hour time the complaint happens, mapped into early morning, morning, afternoon, and evening
12	bldgperlot	building area per lot area

Baseline Model

Majority classifier

Always predict the majority class, in this case 0 or "No" (not heat complaint), for the target value

F1 score: 0
Accuracy: 0.62

Single feature

The class is predicted by only one variable. We will make a decision tree with only one decision level with the most-informative feature (decision stump).

F1 score: 0.42
Accuracy: 0.64

Benchmark Model

Cross -validate using combination of several models and strategies on train and evaluation set

the models used are:

1. Logistic Regression
2. Decision Tree Classifier
3. Stochastic Gradient Descent
4. Naive Bayes

the strategy used for these pipelines are:

1. Oversampling with SMOTE
2. Scaling with MinMaxScaler
3. Feature selection with RFE
4. Oversampling with RandomOverSampler
5. Combined oversampling (SMOTE) and undersampling (TomekLinks) with SMOTETomek
6. **Combined oversampling (SMOTE) and undersampling (ENN - Edited Nearest Neighbours)with SMOTEENN**

Best model: Decision Tree Classifier

steps	Test set F1
model	0.584766
SMOTE, model	0.587515
Scaling, SMOTE, model	0.595282
RFE, SMOTE, model	0.600187
Scaling, RandOverSamp, model	0.598615
Scaling, SMOTETomek, model	0.598376
Scaling, SMOTEENN, model	**0.639501**

Model Improvement

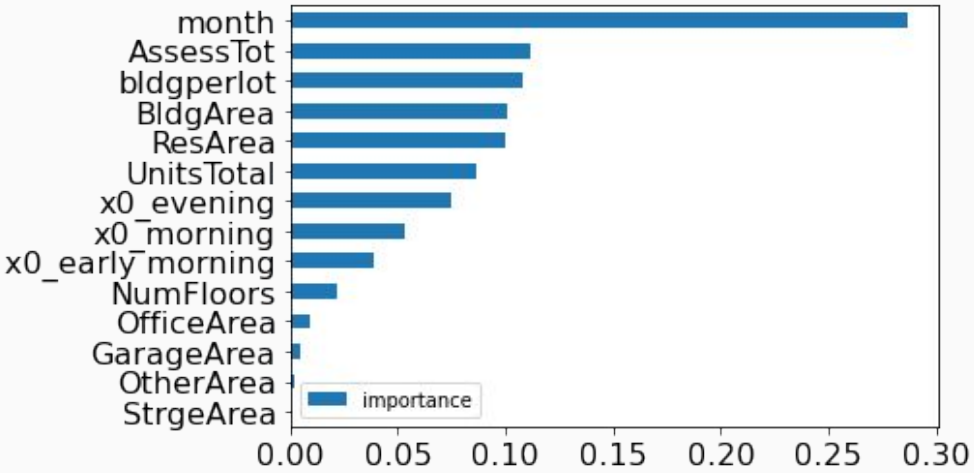
Improve the benchmark model (DecisionTreeClassifier with scaling and SMOTE-ENN) with these strategies:

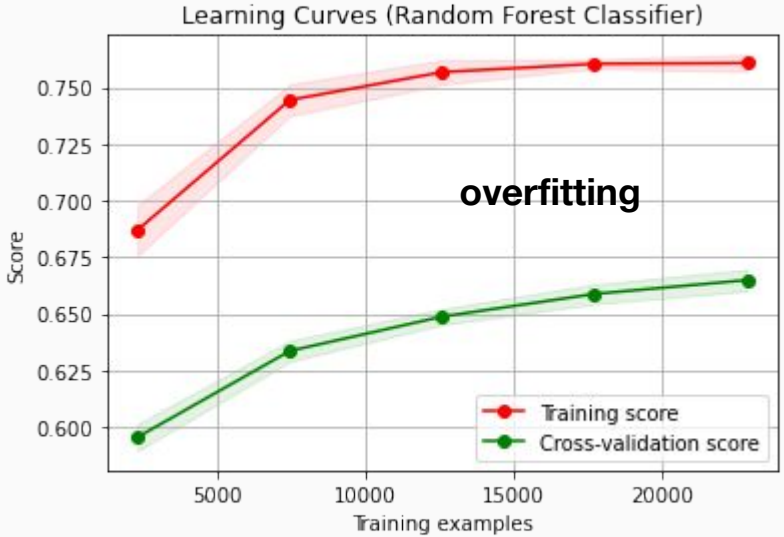
1. Boosting with AdaBoostClassifier
2. Bagging with RandomTreeClassifier
3. Hyperparameter Tuning

Fit model with train and validation dataset, then evaluate on test dataset

Best Model

Random Tree Classifier with 50 estimators and maximum depth of 20 has the best result with F1 score and accuracy of 0.67





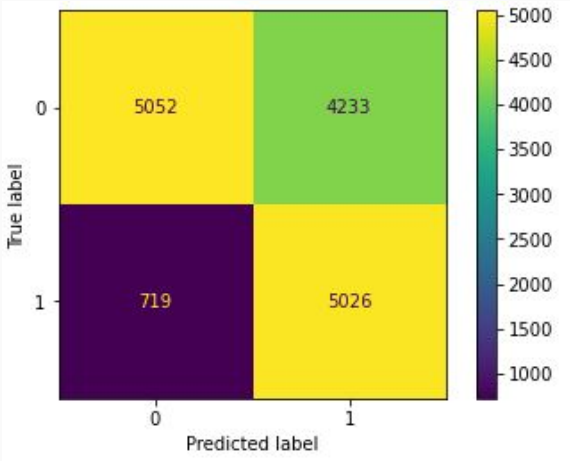
Making Use of the Model

1. **Predict** heat complaint at building/tax lot level
2. **Count** the total heat complaints in a certain area (e.g. zip code)
3. **Choose** the area with the highest number of complaints to focus on

Thus, we want to be as confident as possible when predicting whether a building will have a heat complaint or not -> **increase the threshold**, for example, to 0.7 to increase the precision

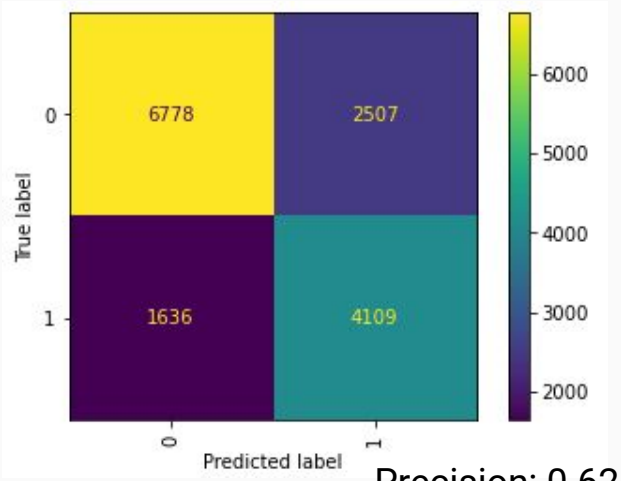
Comparing Thresholds

Threshold = 0.5



Precision: 0.54
F1 score: 0.67
Accuracy: 0.67

Threshold = 0.7



Precision: 0.62
F1 score: 0.66
Accuracy: 0.72

Prediction Example

zipcode

10452

hour

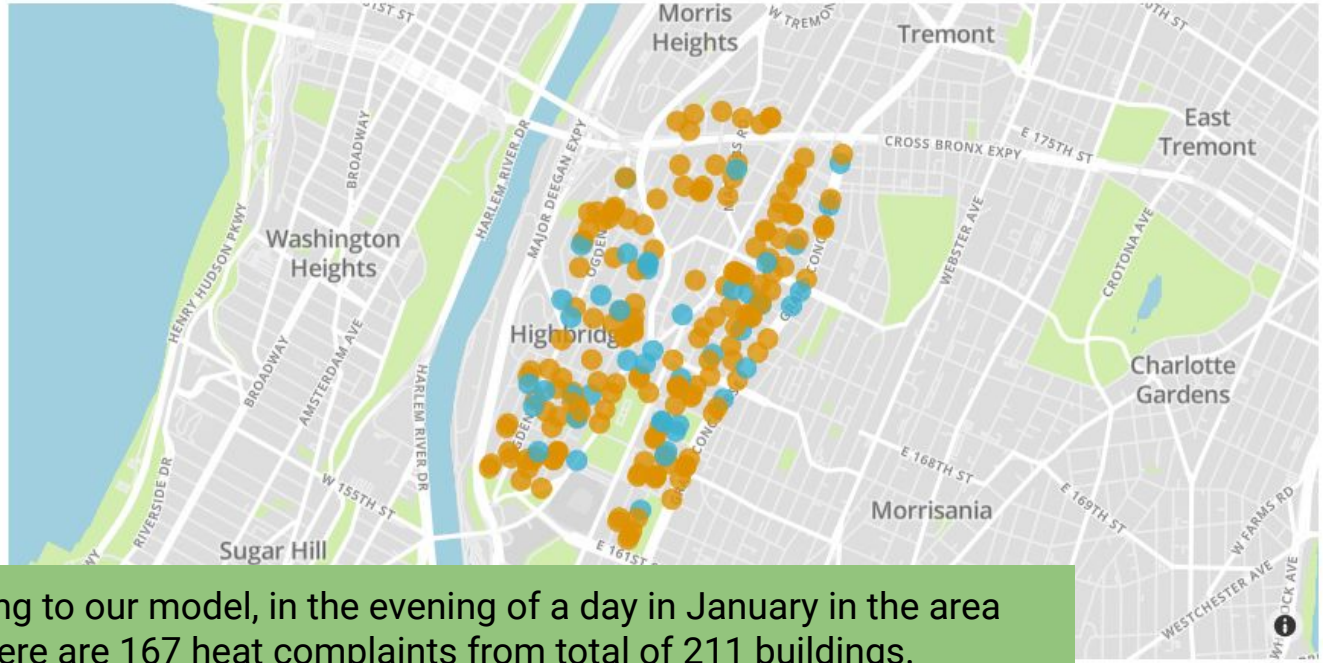
evening

month

Jan

Submit

167 Heat Complaints in Zipcode 10452 Area, Total Houses 211



Interpretation: According to our model, in the evening of a day in January in the area with zip code 10452 there are 167 heat complaints from total of 211 buildings.

Conclusion

- We have analyzed housing and buildings complaints and chose heating problems as the complaint we should focus on, more specifically in Bronx.
- We have analyzed the pattern related to heat complaints in comparison with non-heat complaints.
- We have built a model that predicts the type of complaint sent to the NYC 311 Service Call based on several available tax lot information.
- The model suffers from overfitting, but with threshold adjustment we can get a more confident prediction.

Thank You

References

[1] NYC City Council Data Team. (n.d.). *Are City Agencies Responding to 311?* New York City Council. Retrieved February 24, 2021, from https://council.nyc.gov/data/311-agency/

[2] NYC OpenData. (n.d.). *NYC Population by Borough*. Retrieved February 24, 2021, from https://data.cityofnewyork.us/City-Government/NYC-Population-by-Borough/h2bk-zmw6/data

[3] Wikipedia contributors. (2021, February 9). *Boroughs of New York City*. Wikipedia. https://en.wikipedia.org/wiki/Boroughs_of_New_York_City

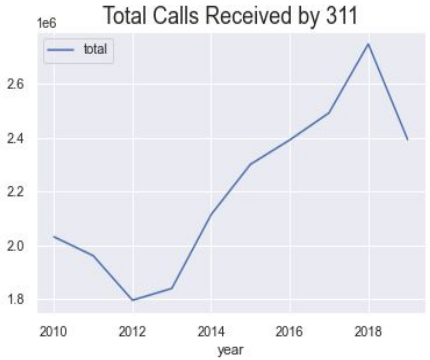
Dashboard


Final Project - Azka Nur Afifah
127.0.0.1:5000/#intro
Incognito
Update
40.90
40.85
40.80
40.75
40.70
40.65
40.60
40.55
1000
800
600
400
200
NYC311 HOUSING AND BUILDINGS
COMPLAINT ANALYSIS
PURWADHIKA JCDS-0506 FINAL PROJECT
AZKA NUR AFIFAH

PROBLEM STATEMENT

About NYC311

NYC311 by the New York City government provides the public with government and non-emergency services. Complaints will be forwarded to various agencies such as local governments, public utilities, and other public services. Citizens can file a complaint about various problem, such as illegal parking, noise, heating, and apartment maintenance. The complaints are coming from five boroughs, which are Queens, Brooklyn, Bronx, Staten Island, and Manhattan.



Housing and Buildings Complaints

Department of Housing Preservation and Development (HPD) is the agency responsible to process the complaints related to housing and building. HPD fields most reports, second only to the NYC Police Department.

HPD is seeking help to manage the large volume of complaints by focusing on the most pressing complaint and determine the possibility of the complaints. This analysis provides insight to the trends of the complaint received by HPD.

Datasets Used

NYC 311 SERVICE REQUESTS

original source from NYC OpenData

data only related to HPD, compiled and uploaded to server by edX Data Science and Machine Learning Capstone Project course (2.37 GB)

PLUTO DATASET

The Primary Land Use Tax Lot Output by NYC Department of City Planning (download here)

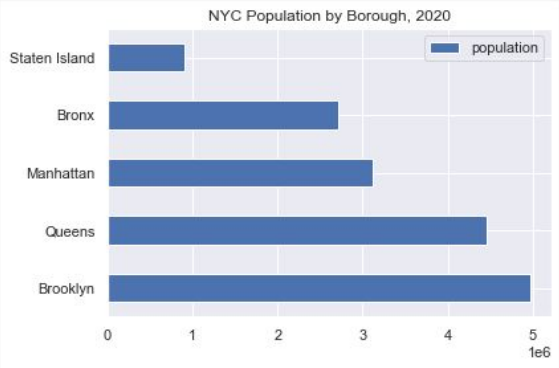
EXPLORATORY DATA ANALYSIS

The visualization in this page is made based on a smaller dataset of 180k rows, sampled randomly, from 2015 to 2019.

There are 14 types of complaints under Housing and Buildings category, with HEAT/HOT WATER having the highest number of complaints. This type of complaint comprises 37.9% of the total complaints on housing and buildings.

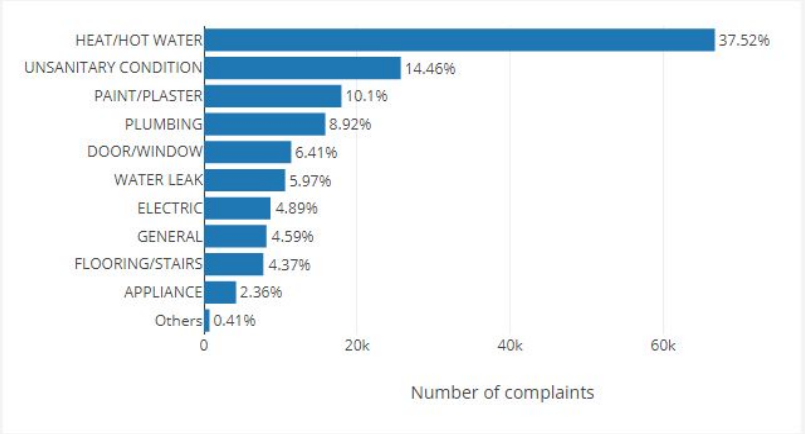
Building owners are required to meet the heat requirements according to the law. This includes residential building, commercial building, and other public facilities such as child care, nursing home, homeless shelter, and public school. When the heat requirements are not met, citizens can file a complaint. The NYC311 Service does not serve complaints coming from private homes, apartments, and other residential spaces.

Bronx has the highest number of HEAT/HOT WATER complaints from 2015 - 2019, followed closely with Brooklyn. On the other hand, Bronx only has population of 2.7 million, the fourth highest after Brooklyn, Queens, and Manhattan. This means the heat complaint in Bronx has the highest density compared in the other borough.



Top 10 Complaints in 2015 - 2019

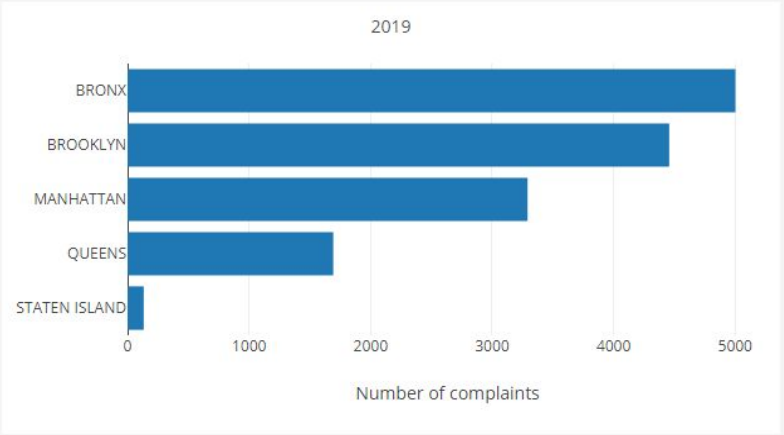
HEAT/HOT WATER complaints made up 37.5% of total houses and buildings complaints.



Heat Complaints in Each Borough

year

Bronx has the highest HEAT/HOT WATER complaints almost every year between 2015 and 2019, except in 2018.



HOME PROBLEM STATEMENT EDA HEAT PATTERN PREDICT

Complaints Map

The map below shows the distribution of total and HEAT/HOT WATER complaints in NYC at zipcode level.

year

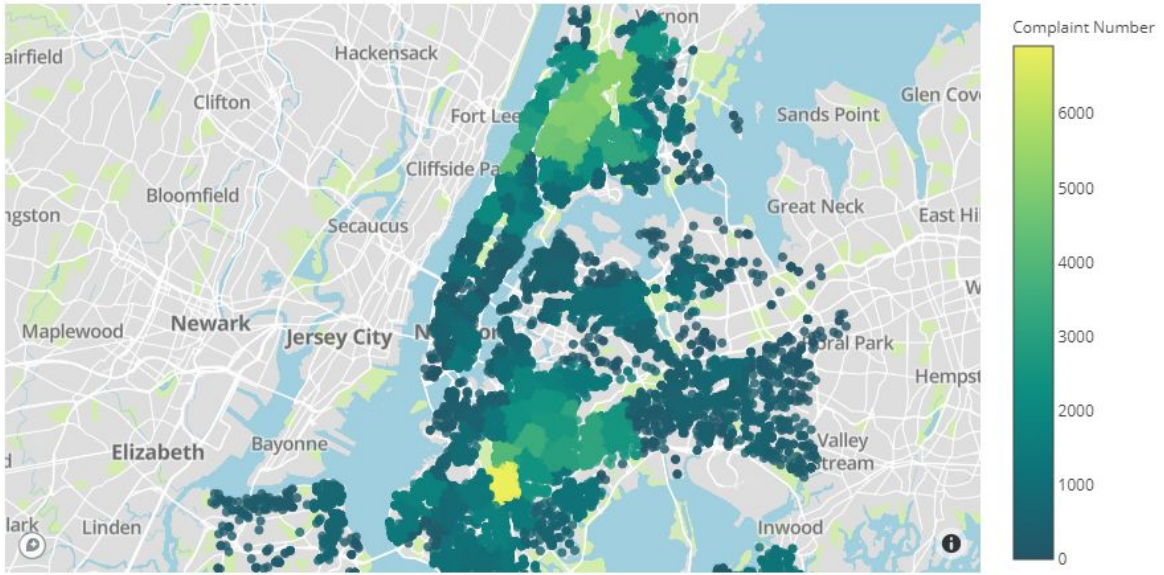
2019

column

Total

Submit

Total Complaints in 2019



HEAT PATTERN

We can observe patterns in heat complaints throughout the year, week, and day. The city regulation demands building owners to provide tenants with minimum heat requirements between Oct 1st and May 31. The heat requirements vary based on the hour and outside temperature. During work hours (6 AM - 10 PM) the heat requirement (minimum 68 degrees) is applied only if the outside temperature falls below 55 degrees. Outside that time window (during nighttime) heat should always meet the required temperature (minimum 62 degrees) without no specification about outside temperature.

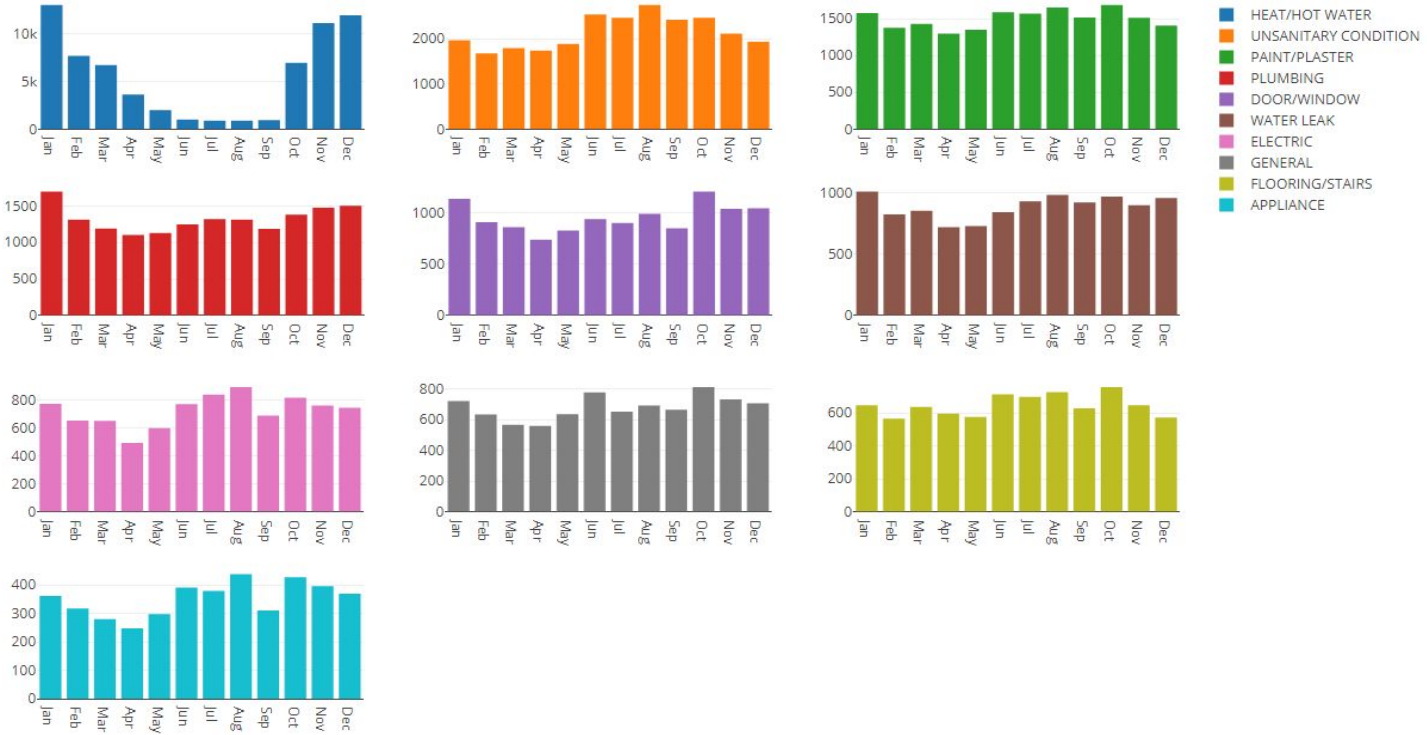
The higher minimum temperature requirement during work hours (starts at 6 AM) explains the daily HEAT/HOT WATER complaint pattern which was very low during nighttime (0 - 5 AM) and started to rise at 6 AM time. This also explains the reason the number of HEAT/HOT WATER complaints stayed high even after normal work hour ends (~4 PM), while the other complaint types kept decresing until it reached nighttime.

The regulation specifies heat requirement during the day, which applies throughout the week during Oct 1 - May 31. Since there is no specific regulation on heat requirement during the week, there is no difference in number of HEAT/HOT WATER complaints throughout the week, even during weekend. This suggests that most HEAT/HOT WATER complaints come from residential building, unlike the other complaints that have lower number during the weekend.

While the other type of complaints have a relatively steady number of complaints throughout the year, HEAT/HOT WATER complaint has lowest number from late spring (June) until early fall (Sep). The HEAT/HOT WATER complaints about excessive heat are accepted only during this period.

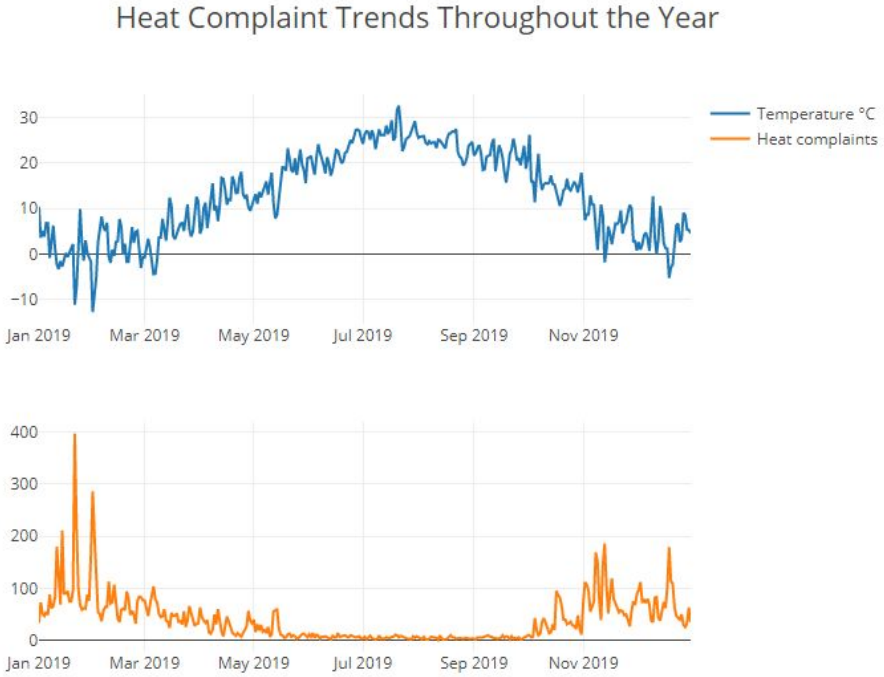
interval monthly

Complaint Type Throughout the Year



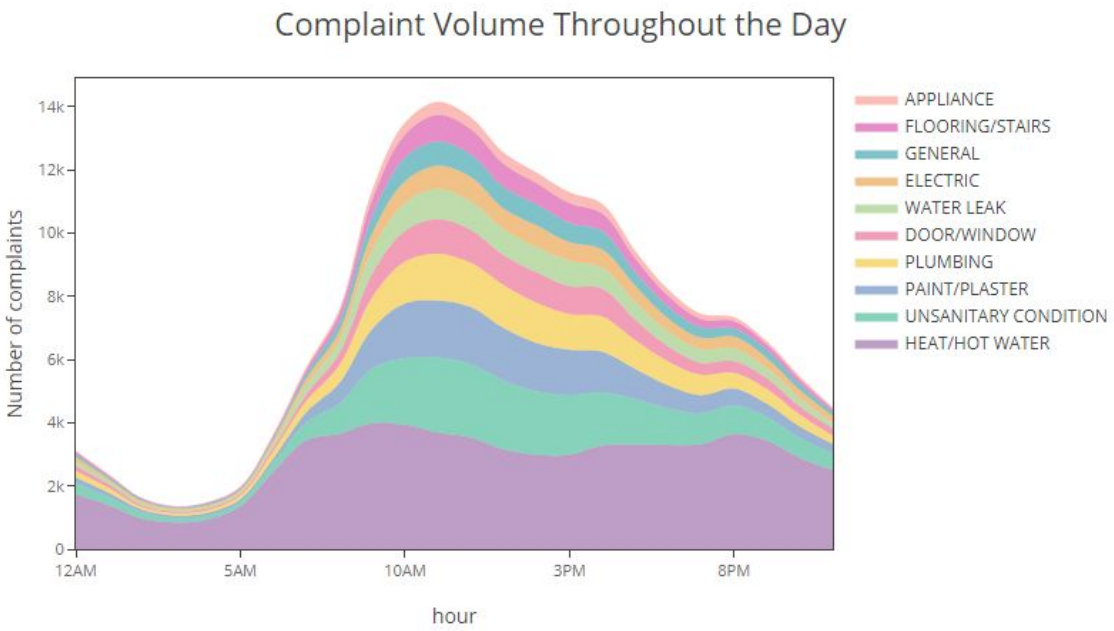
Trends Throughout the Year

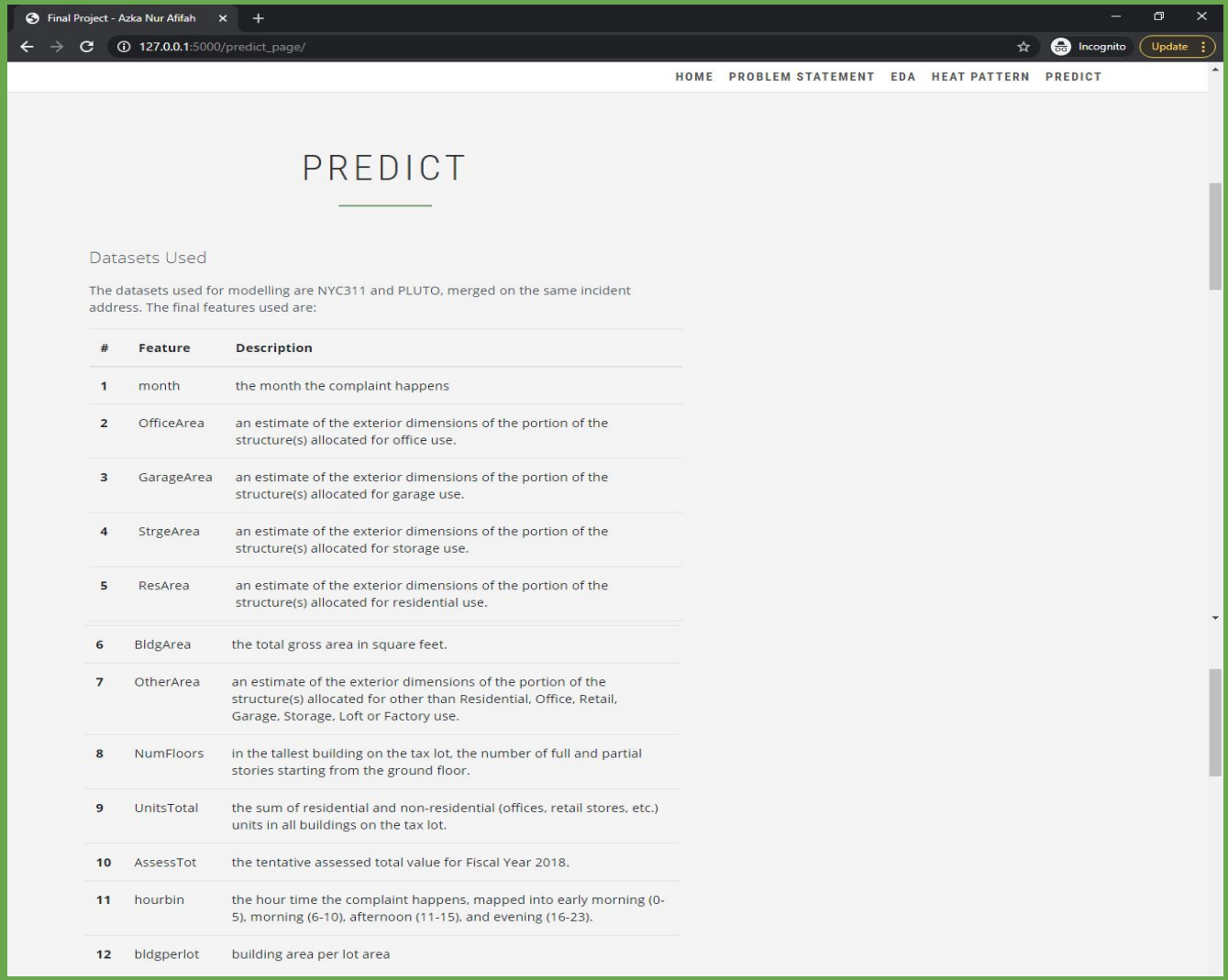
We can look into heat complaint trends throughout the year by comparing the number of complaints and the average temperature in NYC in 2019. As expected, HEAT/HOT WATER complaints started happening in fall and peaked in January when the temperature hit the lowest. The highest peak of heat complaints happened in the same day during the lowest peak of temperature, which were Jan 21st and 31st. (dataset from NOAA)



Complaint Volume Throughout the Day

At any given hour, HEAT/HOT WATER has the highest number of complaints among the other under houses and buildings complaints.



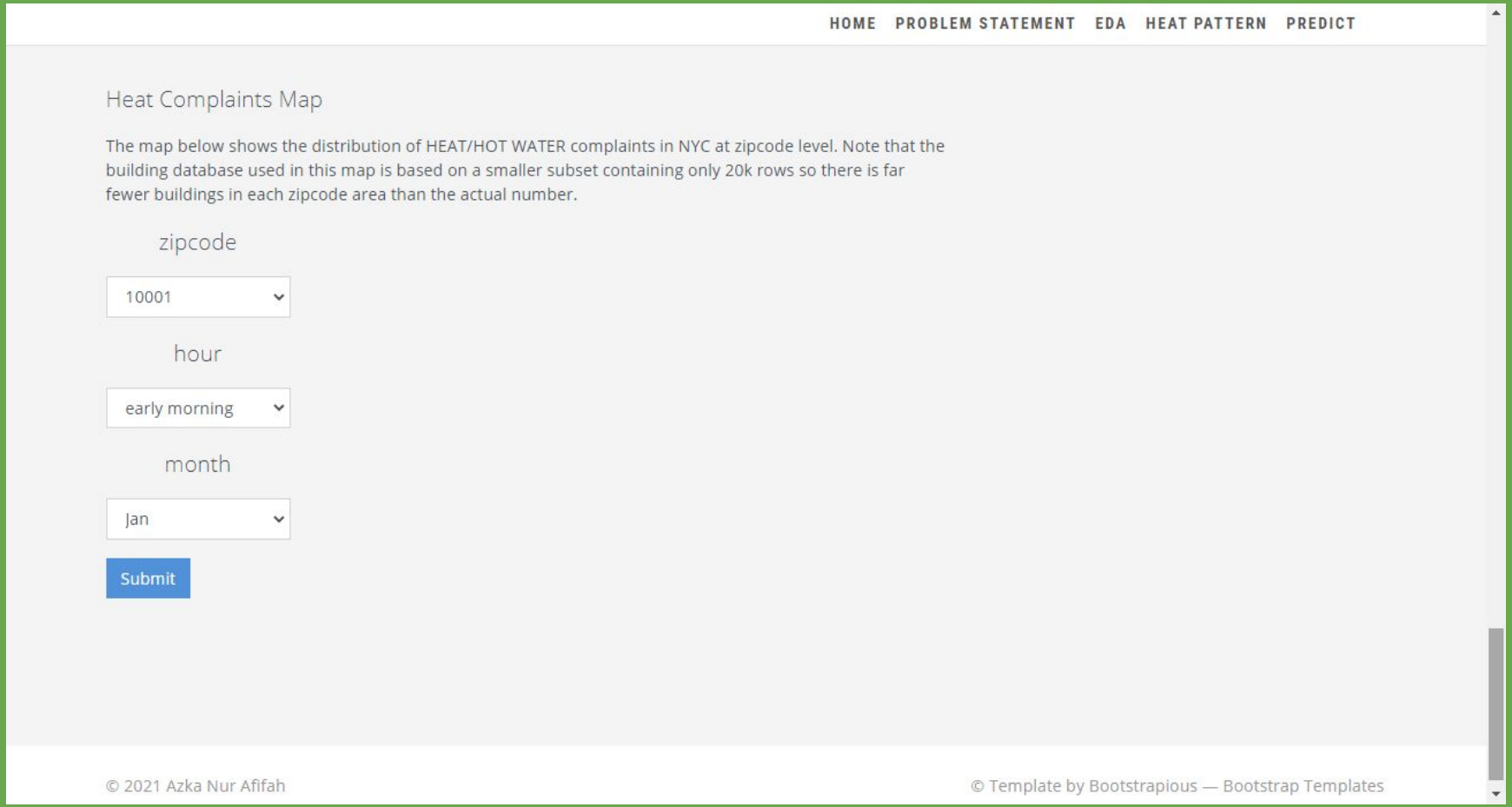

Final Project - Azka Nur Afifah
127.0.0.1:5000/predict_page/
Incognito
Update
HOME
PROBLEM STATEMENT
EDA
HEAT PATTERN
PREDICT
PREDICT
Datasets Used
The datasets used for modelling are NYC311 and PLUTO, merged on the same incident address. The final features used are:
Feature Description
1 month the month the complaint happens
2 OfficeArea an estimate of the exterior dimensions of the portion of the structure(s) allocated for office use.
3 GarageArea an estimate of the exterior dimensions of the portion of the structure(s) allocated for garage use.
4 StrgeArea an estimate of the exterior dimensions of the portion of the structure(s) allocated for storage use.
5 ResArea an estimate of the exterior dimensions of the portion of the structure(s) allocated for residential use.
6 BldgArea the total gross area in square feet.
7 OtherArea an estimate of the exterior dimensions of the portion of the structure(s) allocated for other than Residential, Office, Retail, Garage, Storage, Loft or Factory use.
8 NumFloors in the tallest building on the tax lot, the number of full and partial stories starting from the ground floor.
9 UnitsTotal the sum of residential and non-residential (offices, retail stores, etc.) units in all buildings on the tax lot.
10 AssessTot the tentative assessed total value for Fiscal Year 2018.
11 hourbin the hour time the complaint happens, mapped into early morning (0-5), morning (6-10), afternoon (11-15), and evening (16-23).
12 bldgperlot building area per lot area

NYC 311 Service Request Dataset (Cleaned)

Unique Key	Created Date	Complaint Type	Incident Zip	Incident Address	Street Name	City	Latitude	Longitude
37426123	2017-10-13 07:04:11	HEAT/HOT WATER	10453	1815 DAVIDSON AVENUE	DAVIDSON AVENUE	BRONX	40.84946333006640	-73.9122638766879
32006264	2015-11-17 16:18:50	PAINT/PLASTER	10456	1290 GRAND CONCOURSE	GRAND CONCOURSE	BRONX	40.8364999249858	-73.9157502711184
41706758	2019-02-13 20:31:50	HEAT/HOT WATER	10459	823 HOME STREET	HOME STREET	BRONX	40.82778307140349	-73.8991000134166
32348872	2016-01-05 06:39:39	HEAT/HOT WATER	10463	3605 SEDGWICK AVENUE	SEDGWICK AVENUE	BRONX	40.88154126381984	-73.8964833998230
30920087	2015-06-23 09:09:48	ELECTRIC	10458	235 EAST 203 STREET	EAST 203 STREET	BRONX	40.87368371214091	-73.8853985120936
42876054	[illegible]	WATER LEAK	10472	[illegible]	BEACH AVENUE	BRONX	40.82906101434141	-73.8657242460269

PLUTO Dataset (Cleaned)

Incident Address	LotArea	BldgArea	ComArea	ResArea	OfficeAr	RetailAre	GarageA	StrgeAre	FactryAr	OtherAr	NumBld	NumFloo	UnitsTot	BldgFron	BldgDep	AssessTo	YearBuil	YearAlte
1815 DAVIDSO AVENUE	8190	37600	0	37600	0	0	0	0	0	0	1	6	40	75	88	589950	1930	1991
1290 GRAND CONCOU	10705	48000	0	48000	0	0	0	0	0	0	1	6	56	101	105	1213650	1923	1991
[illegible]	6189	16800	0	16800	0	0	0	0	0	0	1	5	15	44.25	94	181350	1910	1991


HOME PROBLEM STATEMENT EDA HEAT PATTERN PREDICT
Heat Complaints Map
The map below shows the distribution of HEAT/HOT WATER complaints in NYC at zipcode level. Note that the building database used in this map is based on a smaller subset containing only 20k rows so there is far fewer buildings in each zipcode area than the actual number.
zipcode
10001
hour
early morning
month
Jan
Submit
© 2021 Azka Nur Afifah
© Template by Bootstrapious — Bootstrap Templates

HOME PROBLEM STATEMENT EDA HEAT PATTERN PREDICT

Heat Complaints Map

The map below shows the distribution of HEAT/HOT WATER complaints in NYC at zipcode level. Note that the building database used in this map is based on a smaller subset containing only 20k rows so there is far fewer buildings in each zipcode area than the actual number.

Interpretation: According to our model, in the early morning of a day in January in the area with zip code 10019 there are 53 heat complaints from total of 58 buildings.

zipcode

10019

hour

early morning

month

Jan

Submit

53 Heat Complaints in Zipcode 10019 Area, Total Houses 58