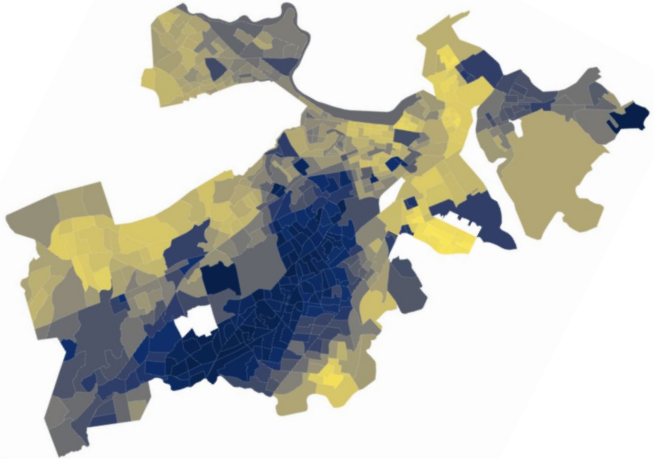
CITY *of* BOSTON
PERMITTING

Deliverable 3

CS 506 - Team D
David E. Kim (Team Lead)
Lukas Werk
Richard Hao
Efim Sokolov
Jackson Fisk



Analyze approved and rejected permits, including those assessed under Article 80, to **uncover insights linked to broader societal, political, and environmental concerns**.

Boston Permitting Process

Application
Apply and pay underlying fees for the permit

→ **Plan Review & Zoning**
City examines the plan for compliance with the zoning laws

→ **Rejected** → **Appeal**
ZBA

→ **Approved**
Valid for six months

Current Progress

- Data cleaning & visualizations
- Preliminary Analysis
- Census data merging
- Geospatial visualizations
- Project base questions answered
- Basic Exploratory Analysis
- Extension project

Datasets

Approved Permits

Approved Building Permits

Article 80 Permits

Approved Permits subject to Article 80

Zoning Board of Appeal

ZBA decisions

Census Data

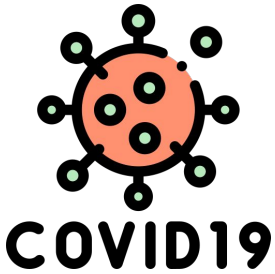
For additional analysis

COVID19 Data

Extension Project

Extension Project

Questions



- How did the **pandemic outbreak** affect the permitting process?
- What **factors can be predicted** by machine learning to influence permit approvals?

Datasets

- Dataset of Massachusetts **COVID-19 data** including dates and confirmed case counts.
- **Combined dataset** of approved and rejected permits.

Information

- **Mathematical methods** to discover trends in COVID-19 and permitting processes over time.
- Implement a **decision tree model** to illustrate the factors influencing permit approvals and rejections.
- **Objective: Explore insights related to societal, political, and environmental factors.**

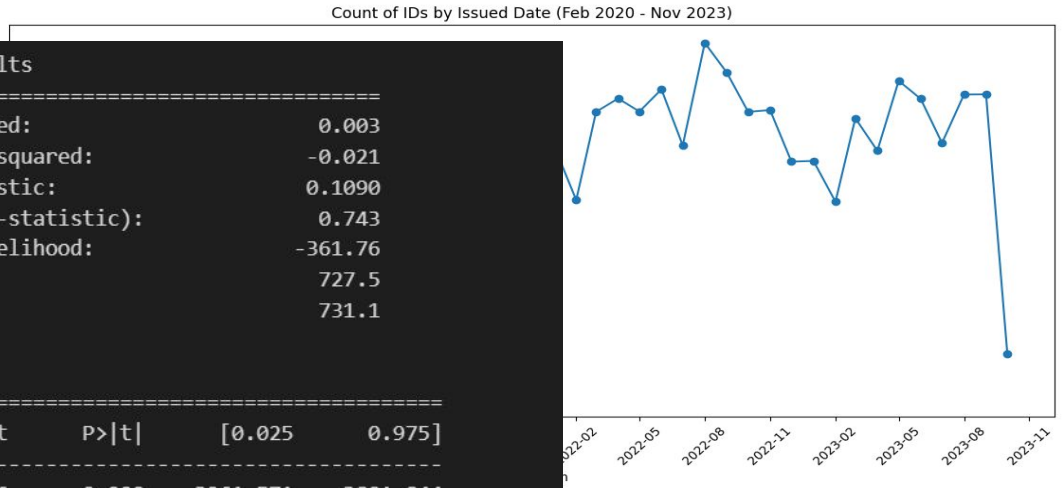
OLS Regression Results

Dep. Variable:	project_count	R-squared:	0.003
Model:	OLS	Adj. R-squared:	-0.021
Method:	Least Squares	F-statistic:	0.1090
Date:	Mon, 11 Dec 2023	Prob (F-statistic):	0.743
Time:	20:10:45	Log-Likelihood:	-361.76
No. Observations:	45	AIC:	727.5
Df Residuals:	43	BIC:	731.1
Df Model:	1		
Covariance Type:	nonrobust		

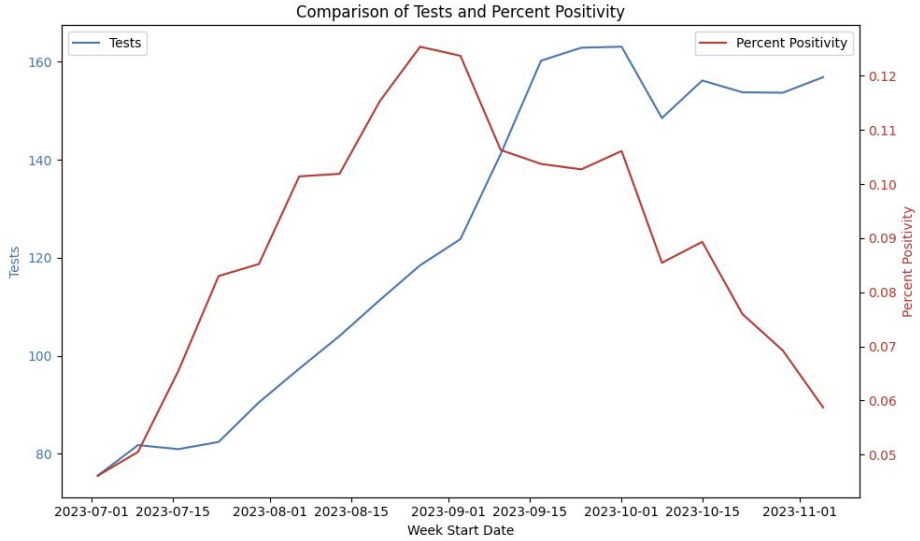
	coef	std err	t	P>|t|	[0.025	0.975]
const	3531.7075	133.950	26.366	0.000	3261.571	3801.844
Confirmed cases	-0.0005	0.002	-0.330	0.743	-0.004	0.003

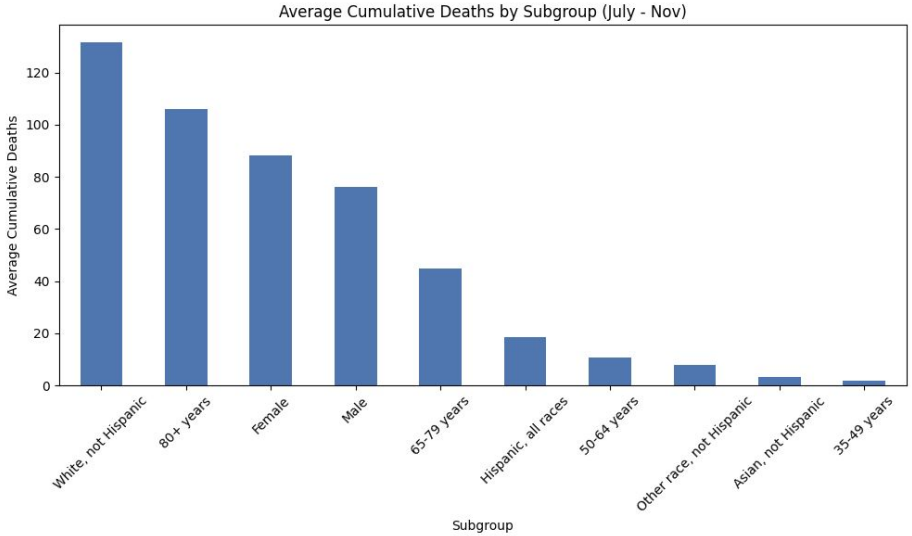
Omnibus:	27.946	Durbin-Watson:	0.921
Prob(Omnibus):	0.000	Jarque-Bera (JB):	51.222
Skew:	-1.834	Prob(JB):	7.54e-12
Kurtosis:	6.723	Cond. No.	1.04e+05

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.04e+05. This might indicate that there are strong multicollinearity or other numerical problems.

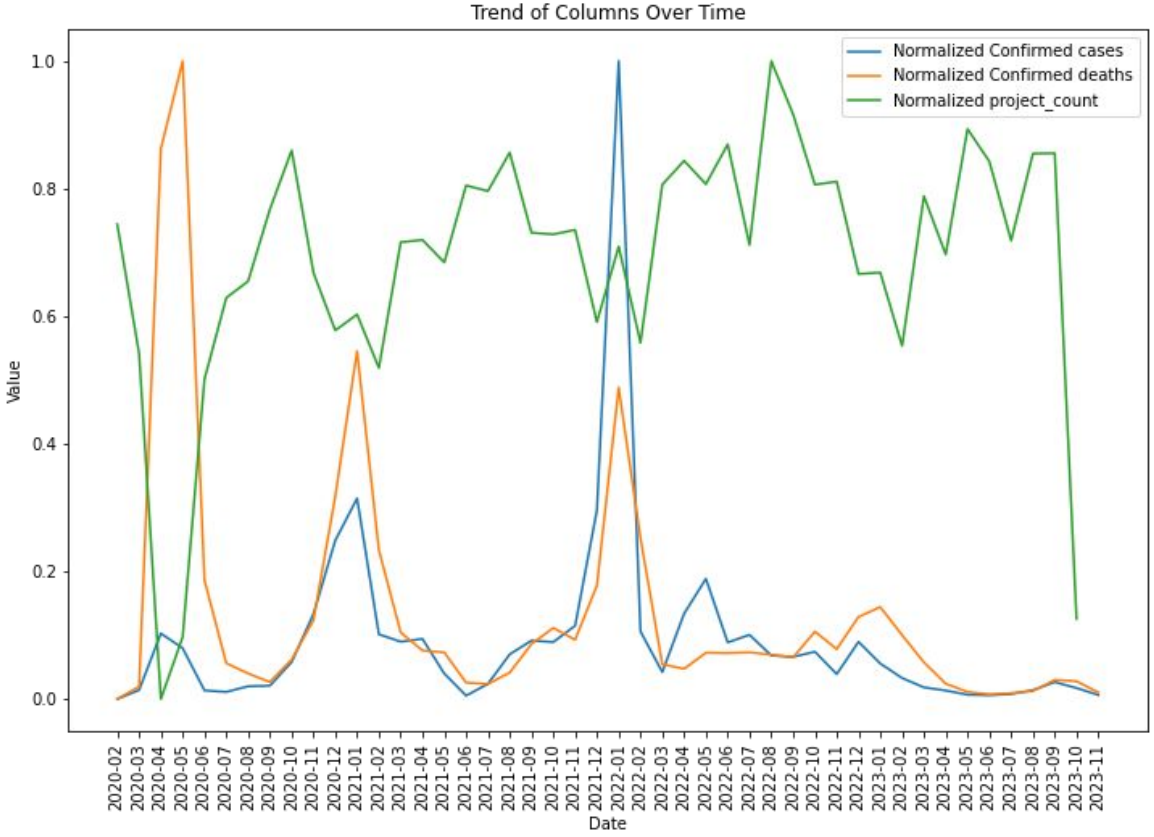


COVID19

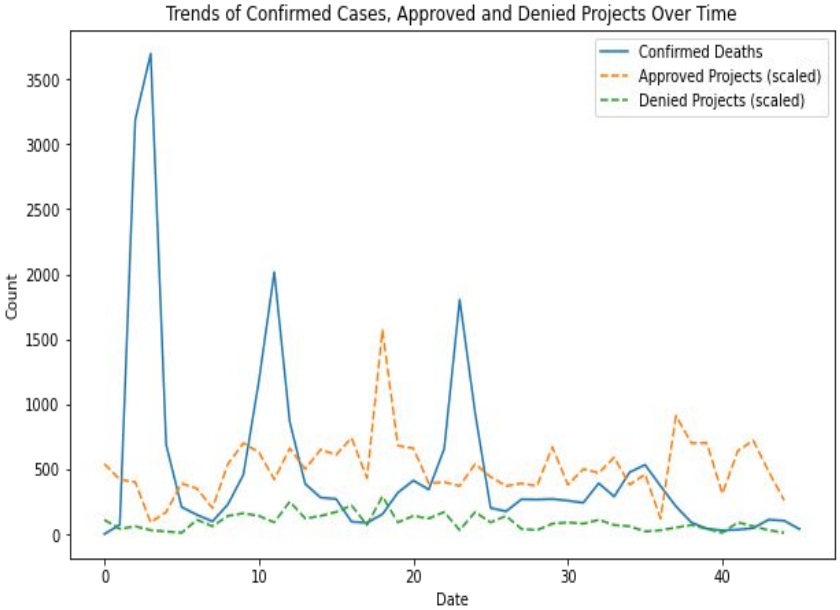

Comparison of Tests and Percent Positivity
Tests
Percent Positivity
Tests
Percent Positivity
Week Start Date
160
140
120
100
80
0.12
0.11
0.10
0.09
0.08
0.07
0.06
0.05
2023-07-01
2023-07-15
2023-08-01
2023-08-15
2023-09-01
2023-09-15
2023-10-01
2023-10-15
2023-11-01

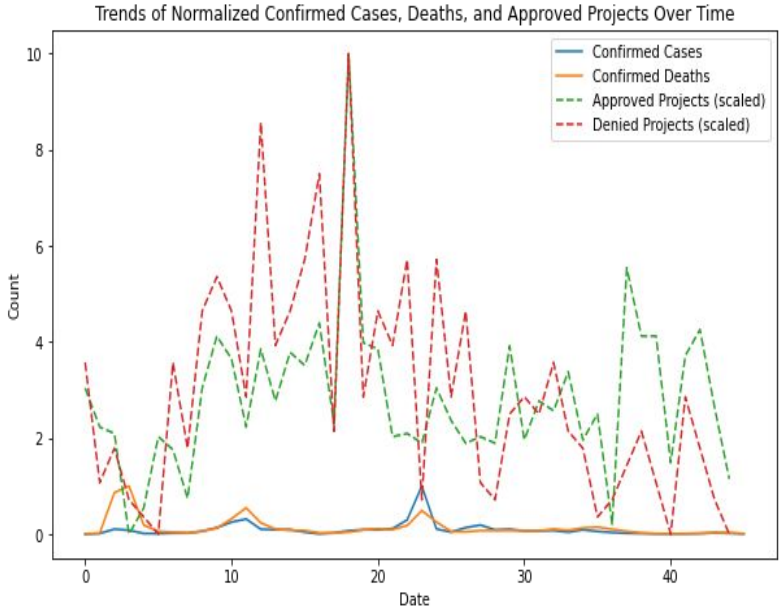

Average Cumulative Deaths by Subgroup (July - Nov)
Average Cumulative Deaths
Subgroup
120
100
80
60
40
20
0
White, not Hispanic
80+ years
Female
Male
65-79 years
Hispanic, all races
50-64 years
Other race, not Hispanic
Asian, not Hispanic
35-49 years

COVID19

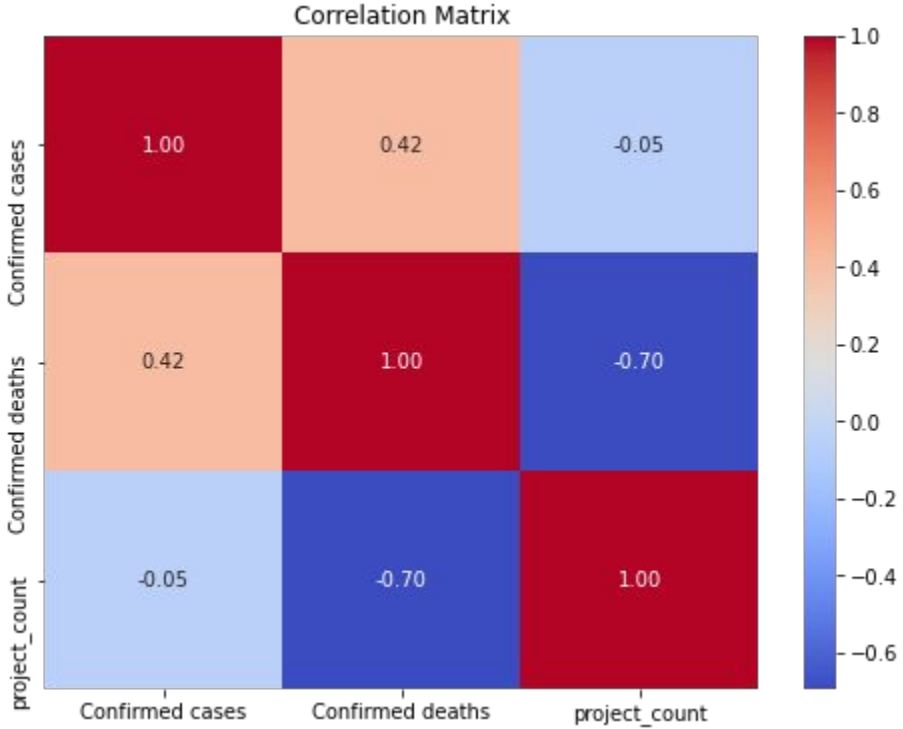

Trend of Columns Over Time
Normalized Confirmed cases
Normalized Confirmed deaths
Normalized project_count
Value
Date
1.0
0.8
0.6
0.4
0.2
0.0
2020-02
2020-03
2020-04
2020-05
2020-06
2020-07
2020-08
2020-09
2020-10
2020-11
2020-12
2021-01
2021-02
2021-03
2021-04
2021-05
2021-06
2021-07
2021-08
2021-09
2021-10
2021-11
2021-12
2022-01
2022-02
2022-03
2022-04
2022-05
2022-06
2022-07
2022-08
2022-09
2022-10
2022-11
2022-12
2023-01
2023-02
2023-03
2023-04
2023-05
2023-06
2023-07
2023-08
2023-09
2023-10
2023-11

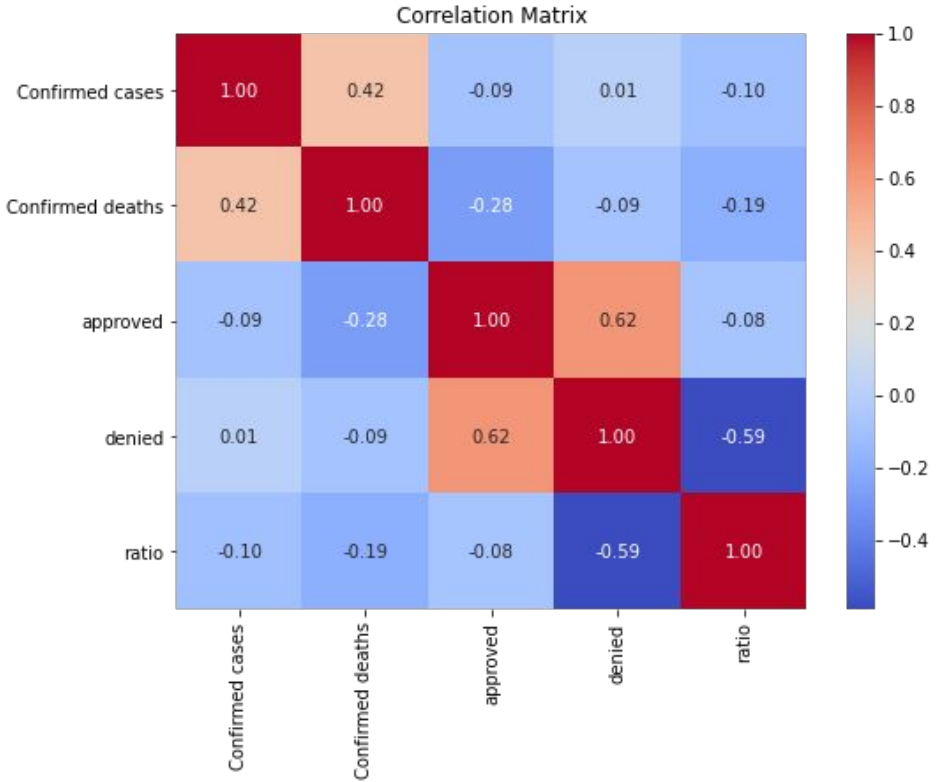
COVID19


Trends of Confirmed Cases, Approved and Denied Projects Over Time
Confirmed Deaths
Approved Projects (scaled)
Denied Projects (scaled)
Count
Date
0
500
1000
1500
2000
2500
3000
3500
0
10
20
30
40

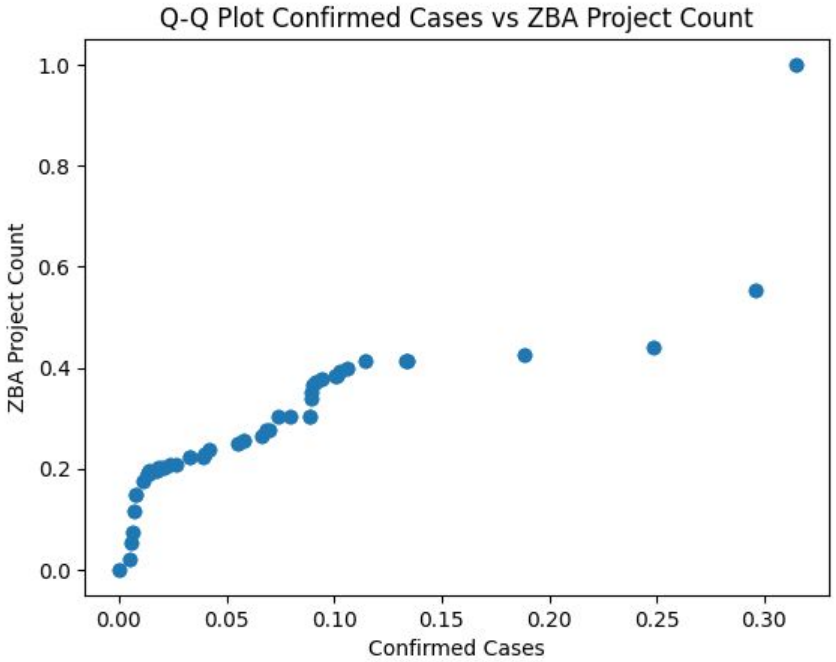

Trends of Normalized Confirmed Cases, Deaths, and Approved Projects Over Time
Confirmed Cases
Confirmed Deaths
Approved Projects (scaled)
Denied Projects (scaled)
Count
Date
0
2
4
6
8
10
0
10
20
30
40

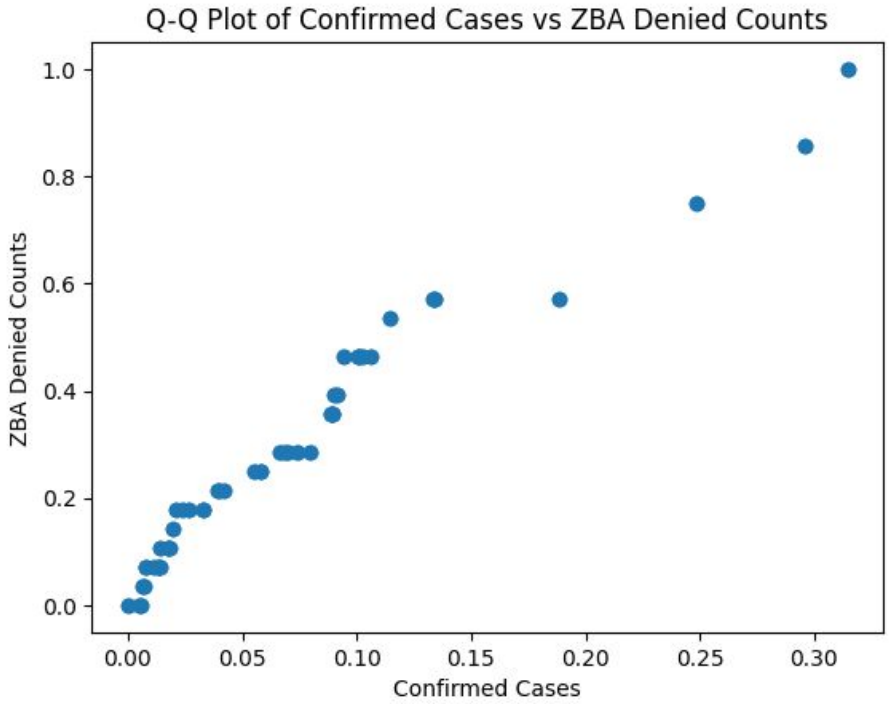
COVID19


Correlation Matrix
Confirmed cases
Confirmed deaths
project_count
1.00
0.42
-0.05
0.42
1.00
-0.70
-0.05
-0.70
1.00
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6


Correlation Matrix
Confirmed cases
Confirmed deaths
approved
denied
ratio
1.00
0.42
-0.09
0.01
-0.10
0.42
1.00
-0.28
-0.09
-0.19
-0.09
-0.28
1.00
0.62
-0.08
0.01
-0.09
0.62
1.00
-0.59
-0.10
-0.19
-0.08
-0.59
1.00
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4

Q-Q Plots


Q-Q Plot Confirmed Cases vs ZBA Project Count
ZBA Project Count
Confirmed Cases
0.0
0.2
0.4
0.6
0.8
1.0
0.00
0.05
0.10
0.15
0.20
0.25
0.30


Q-Q Plot of Confirmed Cases vs ZBA Denied Counts
ZBA Denied Counts
Confirmed Cases
0.0
0.2
0.4
0.6
0.8
1.0
0.00
0.05
0.10
0.15
0.20
0.25
0.30

COVID19

OLS Regression Results

Dep. Variable:	normalized_project_count	R-squared:	0.484
Model:	OLS	Adj. R-squared:	0.472
Method:	Least Squares	F-statistic:	40.27
Date:	Wed, 06 Dec 2023	Prob (F-statistic):	1.15e-07
Time:	22:36:43	Log-Likelihood:	23.554
No. Observations:	45	AIC:	-43.11
Df Residuals:	43	BIC:	-39.49
Df Model:	1		
Covariance Type:	nonrobust		

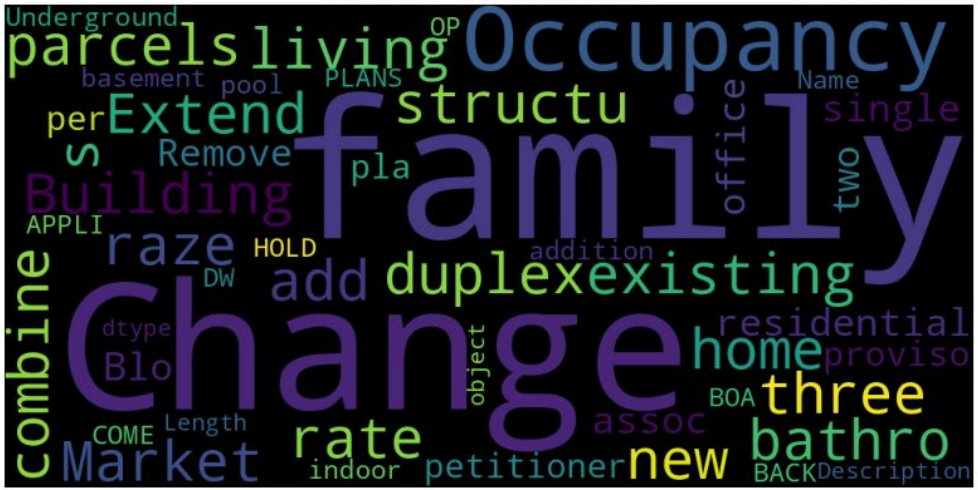
	coef	std err	t	P>|t|	[0.025	0.975]
const	0.7832	0.026	29.695	0.000	0.730	0.836
normalized_Confirmed deaths	-0.6777	0.107	-6.346	0.000	-0.893	-0.462

Omnibus:	33.031	Durbin-Watson:	1.458
Prob(Omnibus):	0.000	Jarque-Bera (JB):	108.580
Skew:	-1.767	Prob(JB):	2.64e-24
Kurtosis:	9.739	Cond. No.	4.98

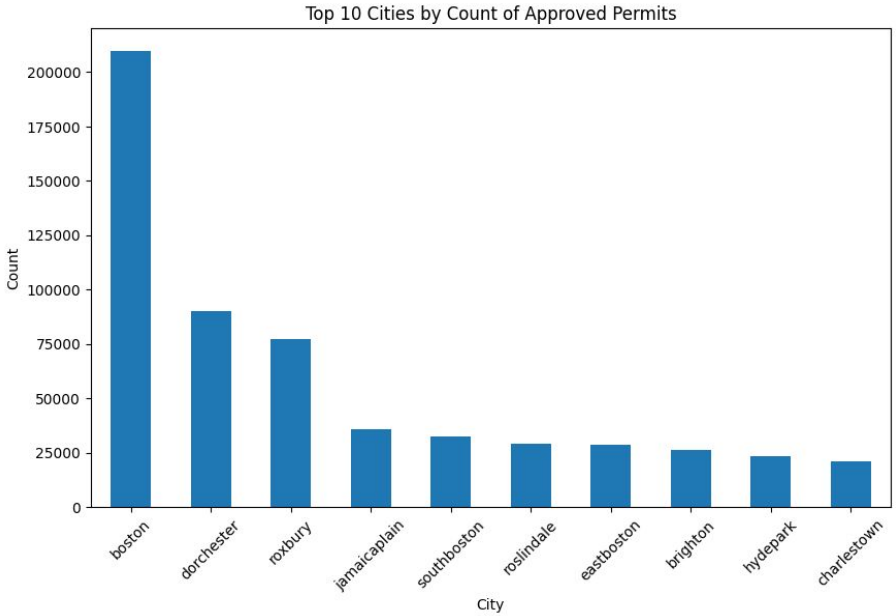
Prediction Features

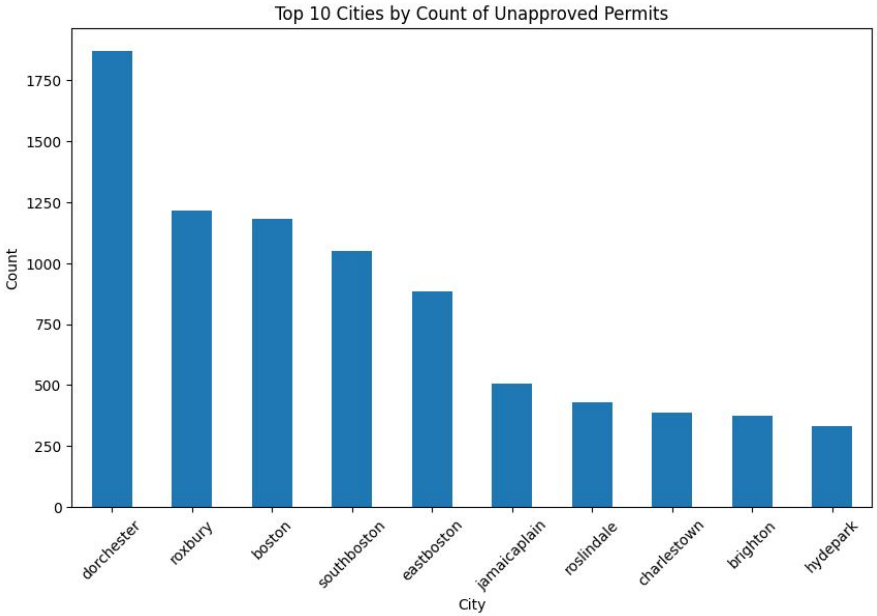
Words used in Approved

Words used in Rejected



Prediction Features

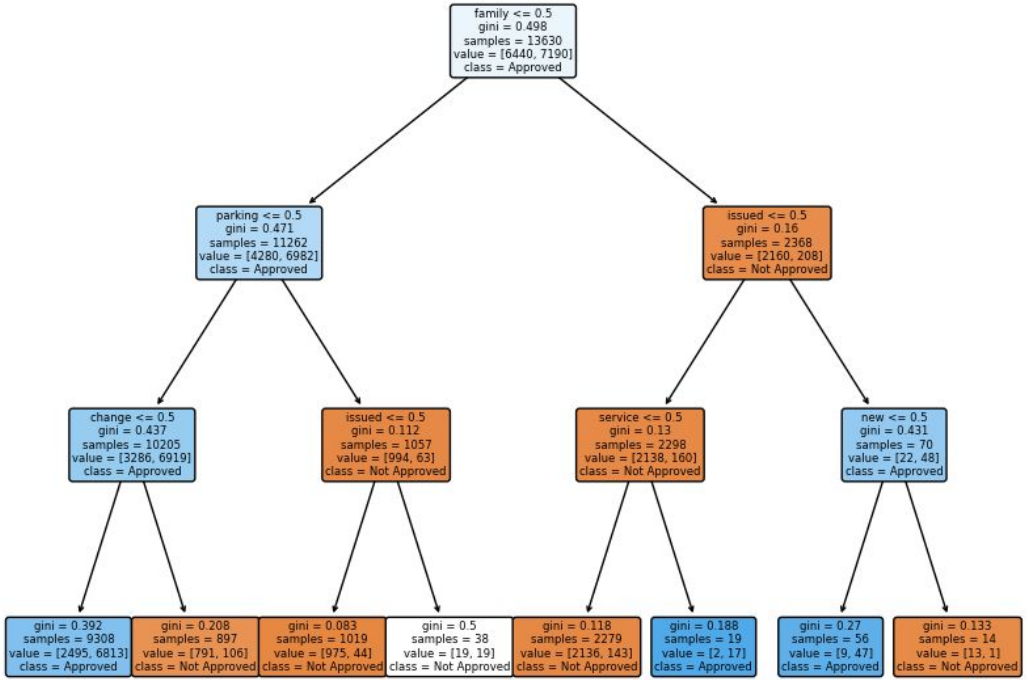




Prediction Features

Accuracy: 0.79

	precision	recall	f1-score	support
False	0.92	0.61	0.73	1598
True	0.73	0.95	0.83	1810
accuracy			0.79	3408
macro avg	0.83	0.78	0.78	3408
weighted avg	0.82	0.79	0.78	3408



Takeaways

- Assumption: Location, work type, valuation, and size are key influencers in permit decisions.
- Correlation between COVID cases/deaths and project counts
- Increase on Covid death leads a drop on permitted projects and increase in denied projects
- Type of family, development type and zoning information is important to predicting the decisions
-

Next Steps

- Finish analysis for extension project and complete final report.

Individual Contribution

David Kim: Extension Project, Dataset Cleaning, Prediction Wordcloud, Analysis

Zhihuan Hao: Merging abp and zba with covid-19; Then did Covid-19 Correlation Analysis with line charts and correlation matrix and linear regression.

Lukas: Merged abp and zba data, performed nlp on text features and generated decision trees to find new features. Drafted the final report.

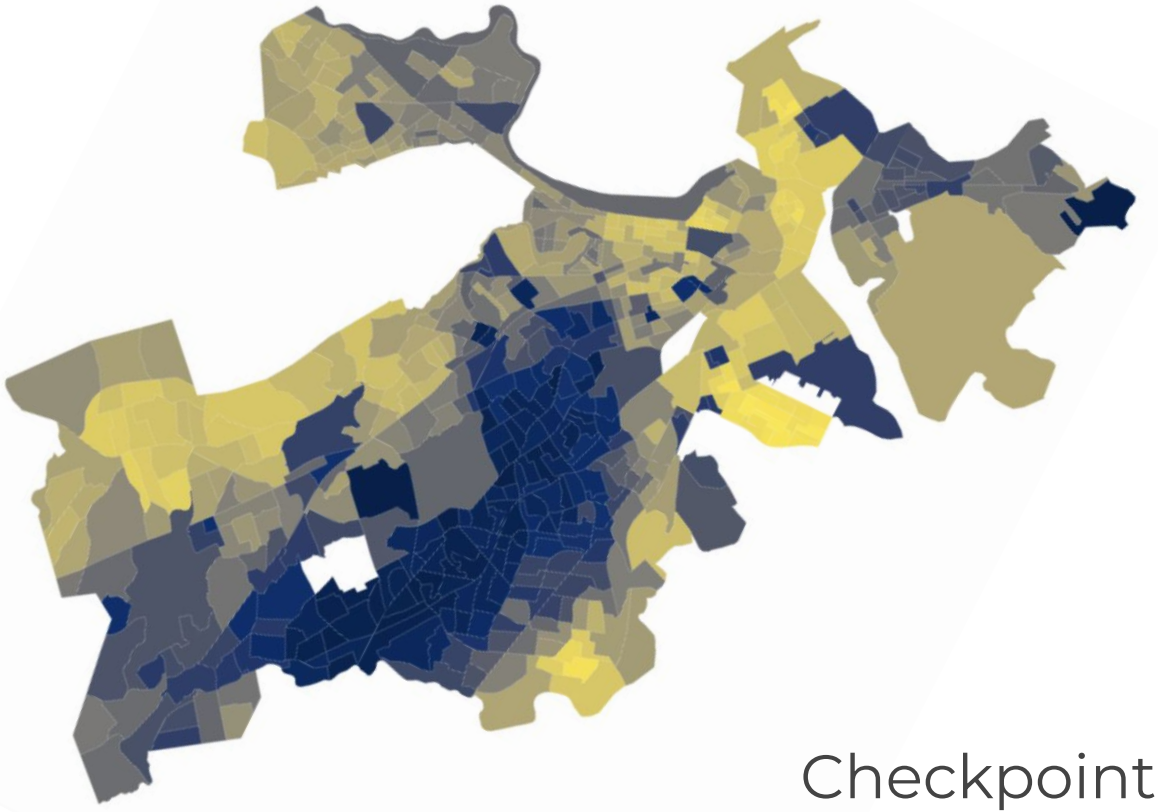
Efim: Presentation, COVID Correlation Analysis, Q-Q plots, Regression analysis

Jackson:

CITY *of* BOSTON
PERMITTING

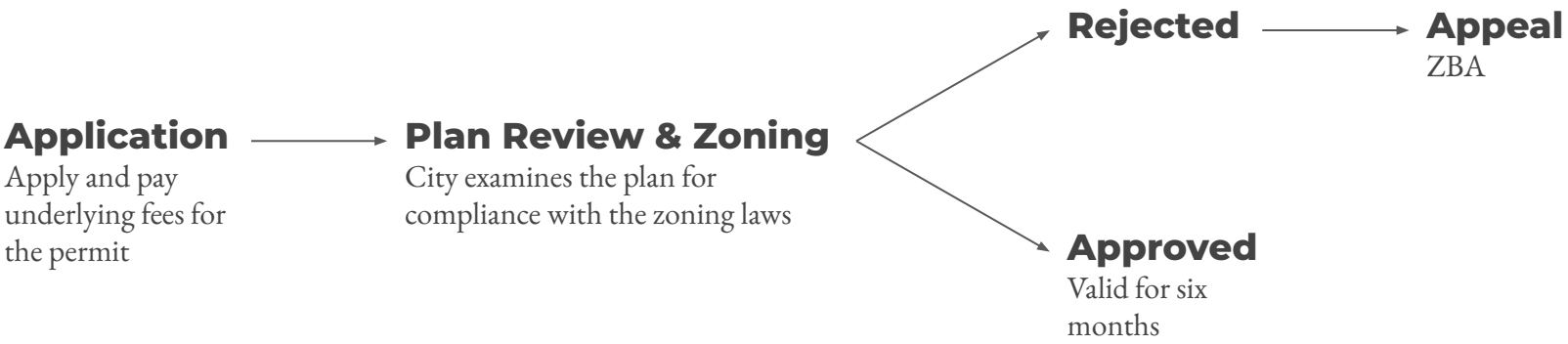
CS 506 - Team D

David E. Kim (Team Lead)
Lukas Werk
Richard Hao
Efim Sokolov
Jackson Fisk



Checkpoint A

Boston Permitting Process



Datasets

Approved Permits

Approved Building Permits

Article 80 Permits

Approved Permits subject to Article 80

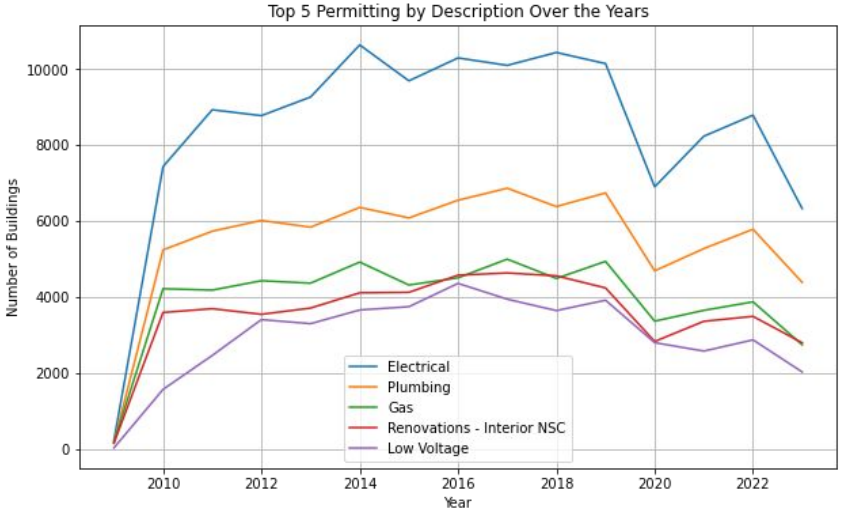
Zoning Board of Appeal

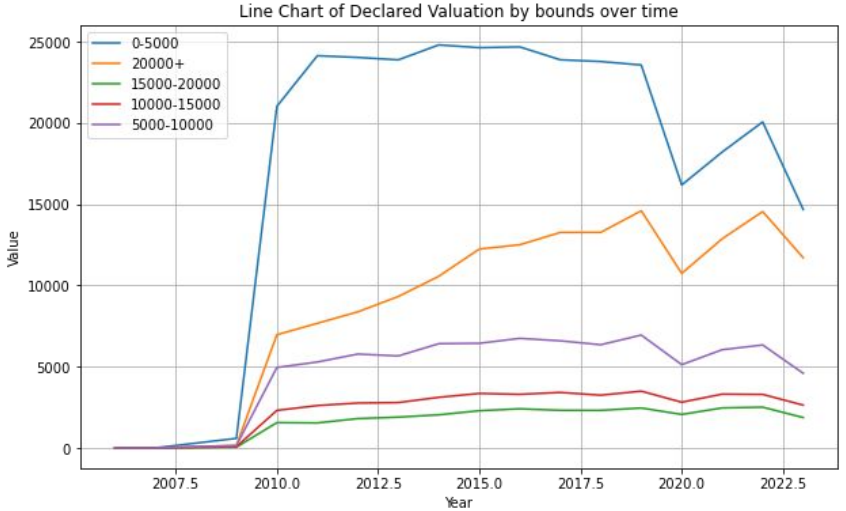
ZBA decisions

Census Data

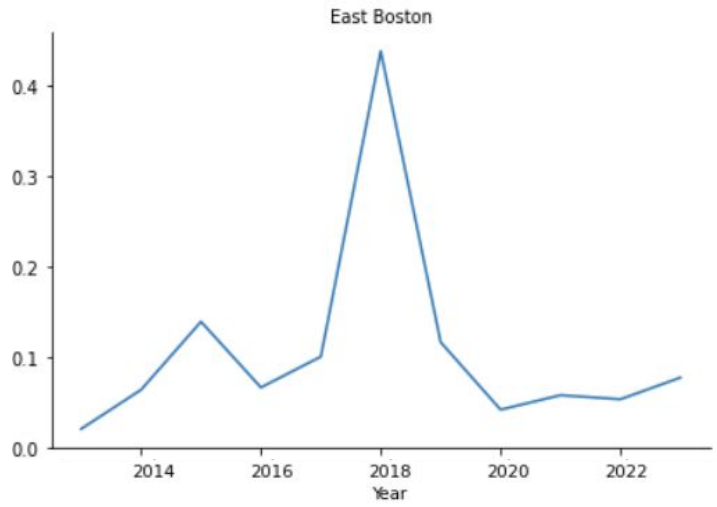
For additional analysis

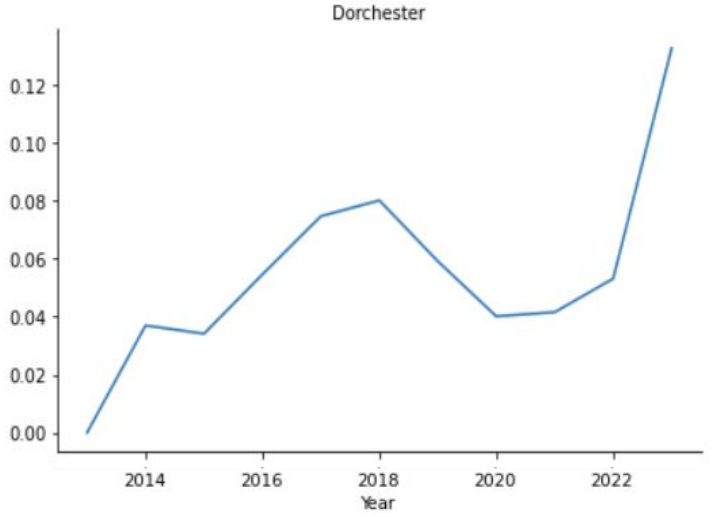
Approved Permits by Type and Valuation



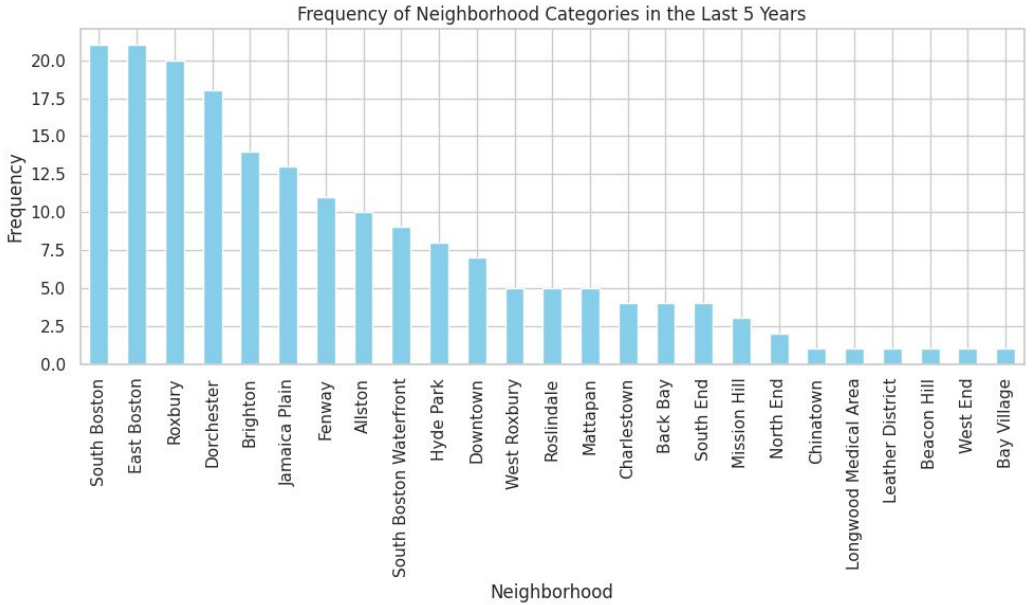


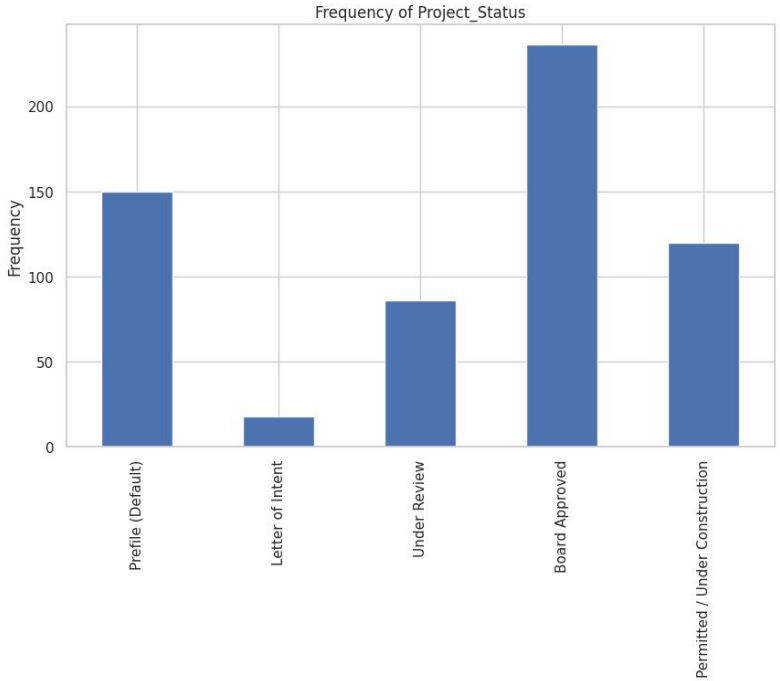
Appeal Distributions by Neighborhood



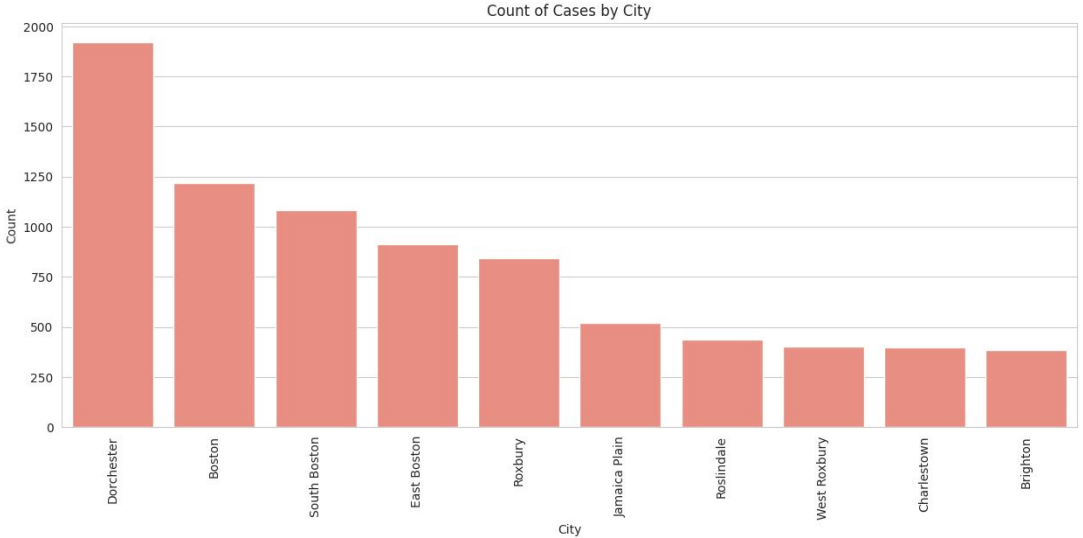


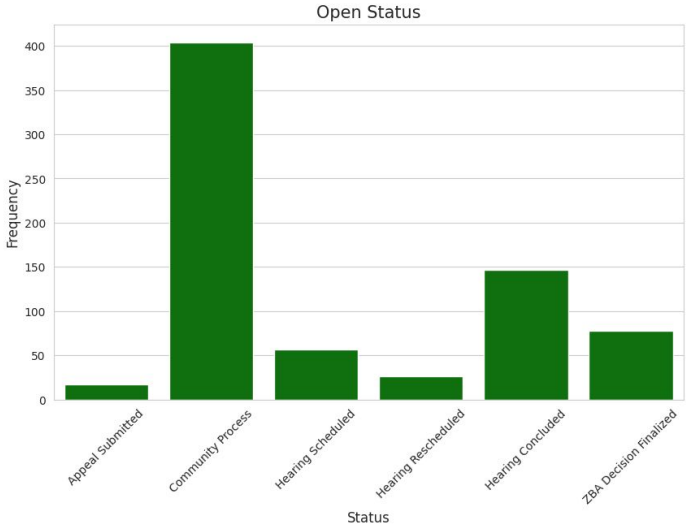
Categories and Status Article 80



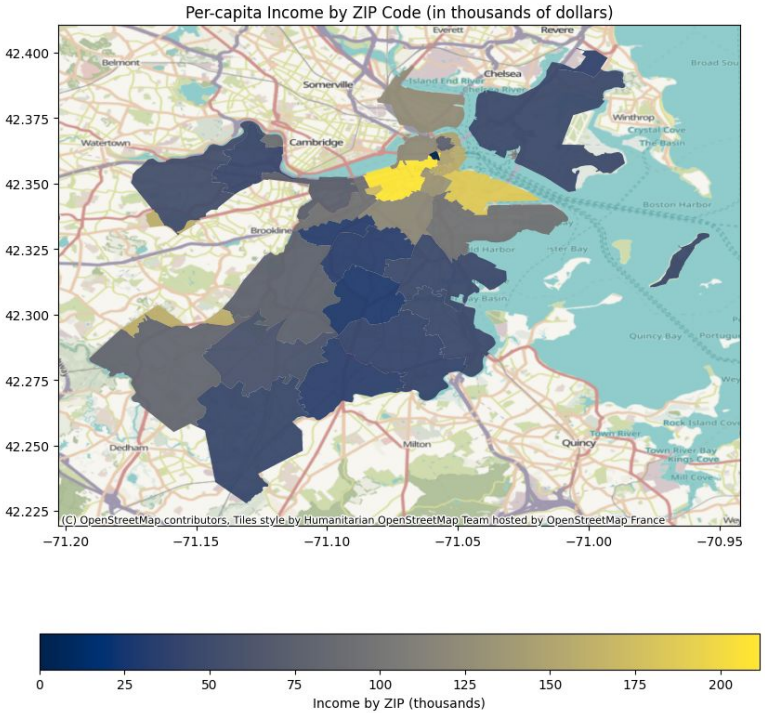


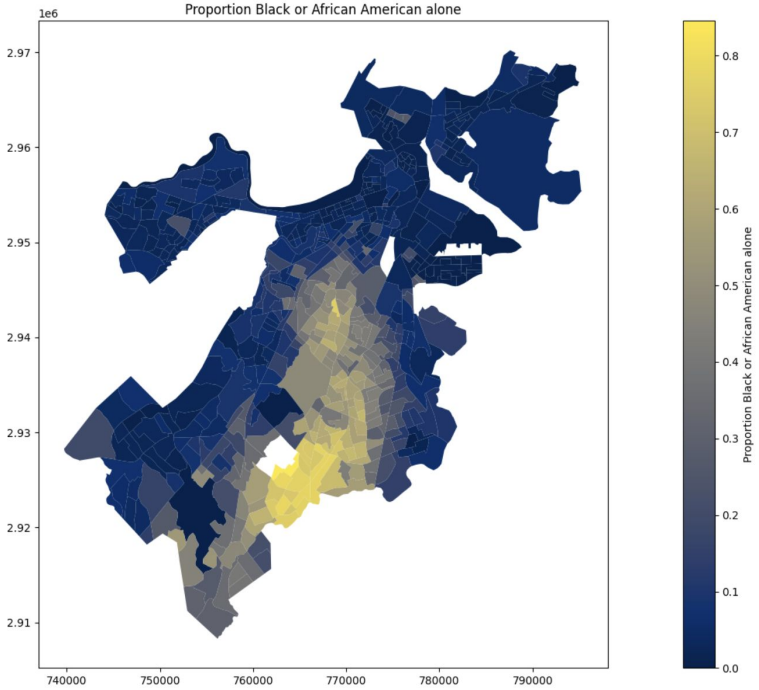
Appealed Permits Counts and Status



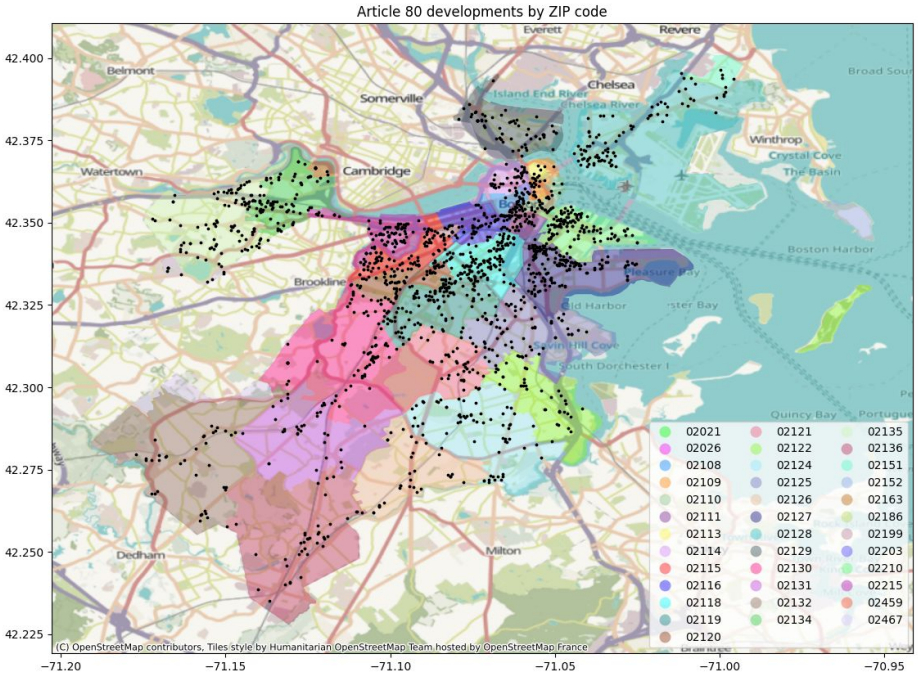


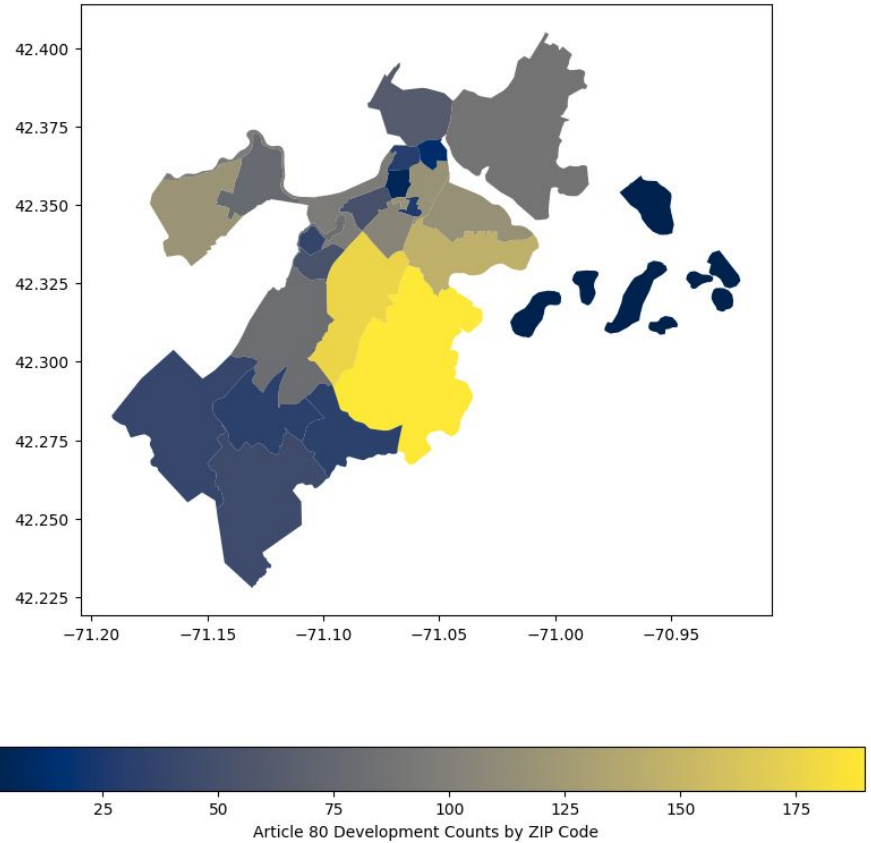
Income and Demographic Distributions



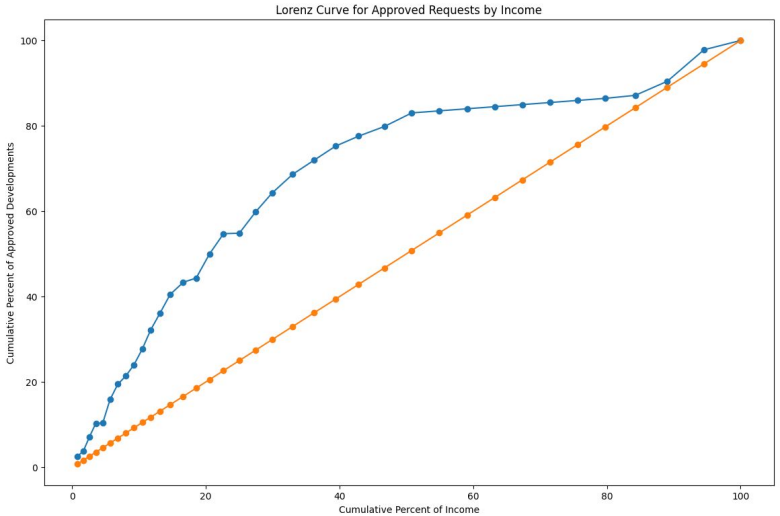


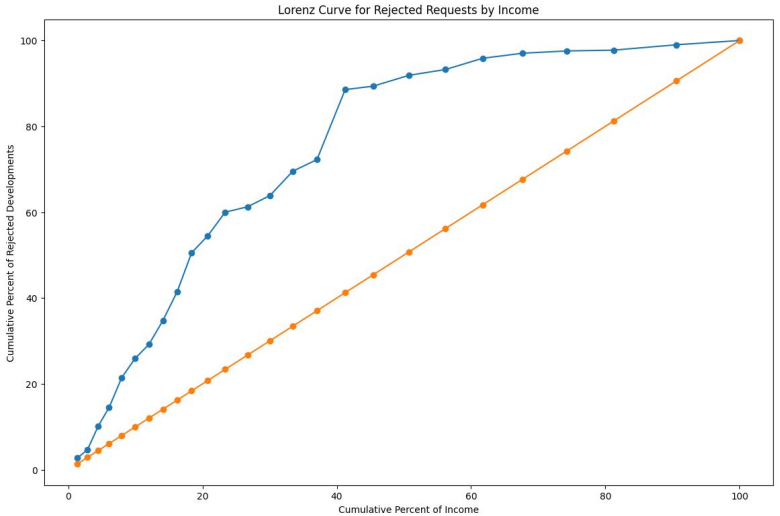
Development Distribution





Income Lorenz Curves and Gini Coefficients

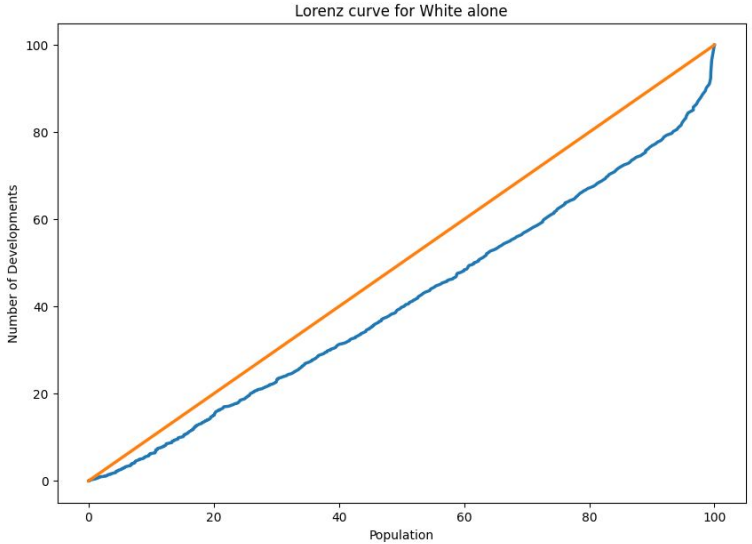


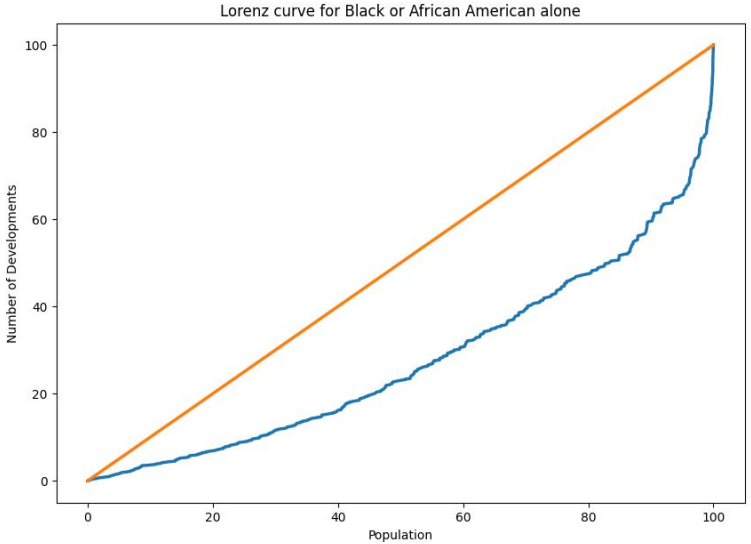


Gini Coefficients

- Income-Population: 0.48 (US is 0.47)
- Income-Development for approved: -0.395
- Income-Development for rejected: -0.512

Demographic Lorenz Curves and Gini Coefficients





Gini Coefficients by census demographic group:

- Total: 0.303
- White alone: 0.181
- Black or African American alone: 0.446
- Asian alone: 0.335
- Hispanic or Latino: 0.409
- American Indian and Alaska Native alone: 0.409

Goal

Analyze both the approved and rejected permits, including those assessed under Article 80, to **uncover insights linked to broader societal, political, and environmental concerns**.

Meetings

Fridays	9 AM
Sundays	9:30 PM

Division of Labor

D.K.	Project Lead and ZBA
R.H.	Approved Permits
E.S.	Census & Geospatial
J.F.	Article 80 Permits
L.W.	Data Prep. & Cleaning

Current Progress

- Initial data cleaning
- Initial visualizations
- Preliminary Analyses
- Census data merging
- Geospatial visualizations
- Project base questions answered
- Exploratory Analyses
- Extension project started

Takeaways

- Additional data and merging techniques are needed
- Assumptions such as the impact of COVID-19 will need to be backed up by additional analysis
- One or two more please

Next Steps

- Continue finding environmental and political patterns and selectively investigating further
- Improve our results by exploring different data subsets and cleaning optimizations
- Work toward telling a story with our data and shaping our final report
- Begin the initial stages of our extension project