IgniTech: Fighting Fire With Data

Anoushka Gupta
New York University
ag8733@nyu.edu

Arunima Mitra
New York University
am13018@nyu.edu

Divyanshi Parashar
New York University
dp3635@nyu.edu

Abstract

Despite advancements in fire safety technology and emergency response systems, government organizations continue to face challenges in effectively preventing and managing fire emergencies. In this project, we explore the dynamics of fire accidents and investigate the causal relationship between incidents and various socio-demographic and environmental factors, including income levels, population density and weather conditions. We also evaluate the effectiveness of fire inspections and access to essential firefighting resources such as fire-hydrants.This paper offers valuable insights into the complex dynamics of fire incidents and suggest areas where targeted intervention is required.

I. Introduction

On the morning of 9th January 2022, a massive fire in a 19-story residential building in the Bronx claimed the lives of seventeen individuals including eight children. Investigation revealed that the tragic incident resulted from a "malfunctioning electric space heater"[1]. The National Fire Protection Association (NFPA) reported over 1.3 million fires in the US in 2019, resulting in an estimated $3,700$ civilian deaths and \$14.8 billion in direct property damage. Fire accidents continue to be one of the primary hazards to public safety and property, threatening communities and compromising security.

Our project is an initiative aiming to understand and explore the intricacies of fire dynamics and the diverse socio-demographic and environmental factors that impact the frequency of such incidents in the city of New York. The significance of our analysis lies in its potential to provide actionable insights for government agencies, urban planners, and emergency responders. By identifying vulnerable populations and high-risk areas, stakeholders can prioritize resources and implement tailored strategies to enhance fire safety and emergency preparedness. Furthermore, this research contributes to the ongoing discourse on effective disaster management practices, emphasizing the importance of evidence-based approaches in mitigating the impact of fire emergencies.

II. Related Works

As urban development and planning evolve and improve, driven by the ever-growing needs of modern societies, researchers globally are working to understand the factors that influence fire incidents. With cities expanding and populations densifying, the complexities of urban environments present unique challenges in fire prevention and management. [2] Zhang,Yao et al used fire incident data and satellite imagery to explore urban fire dynamics and its relationship with urban growth in Nanjing, China. They concluded that in cities there was an increase in fires with increase in population and urban land. [3] U.S Fire Administration's National Fire Estimation Paper suggested that population is one of the most important predictive variables in estimating fire incidents. They also suggest that type of fuel used in an area in industry and homes may also contribute heavily to fire incidents. The fire frequency analysis study carried out by Xu, Zhisheng et al [4] indicates that a majority of fires in urban populations are caused during winters due to consumption of high-power electrical appliances. Their study reveals electric fires cause 47.81% of fires in urban areas. Hosain, Smirnov et al. [5] studied the risk factors of residential fires in urban and rural Ohio. They surmise that in urban areas population density and vacant housing influence fire incidents while in rural areas economic activity greatly influences fire incidents.

III. Datasets

A. Fire Incident Data [6]

Provided by the Fire Department of New York (FDNY), the fire incident dataset captures detailed information about incidents handled by the FDNY and includes fire, medical and non-medical emergencies The raw unfiltered dataset has 10,275,092 rows, spanning from 2005 to 2023. With 3.1 GB size, this dataset is our primary dataset. It was last updated on January 22, 2024. 1 shows the raw data sample.

B. Fire Hydrant Data [7]

The Fire Hydrants dataset is provided by the Department of Environmental Protection(DEP). It captures the location of fire hydrants across New York in 2023 and is updated annually. This dataset, with a size of 12.1 MB, has the latitude, longitude, borough, id and community board of each hydrant. 2 shows the raw data sample.

STARFIRE_INCIDENT_ID	INCIDENT_DATETIME	ALARM_BOX_BOR...	ALARM_BOX_NUMBER	ALARM_BOX_LOCATION	INCIDENT_BOROUGH	ZIPCODE	POLICEPRECINCT	CITYCOUNCILDISTRI
2100402630110024	01/04/2021 12:36:00 AM	MANHATTAN	263	ESSEX ST & BROOME ST	MANHATTAN	10002	7	
2100428210120019	01/04/2021 12:37:00 AM	BRONX	2821	CASTLE HILL & QUIMBY AVE	BRONX	10473	43	
2100434290140027	01/04/2021 12:38:00 AM	BROOKLYN	3429	S/S OF AVENUE V OPP W 13	BROOKLYN	11223	60	
2100416730210026	01/04/2021 12:38:00 AM	MANHATTAN	1673	EDWARD M. MORGAN PL &	MANHATTAN	10032	33	
2100425570120020	01/04/2021 12:40:00 AM	BRONX	2557	E167 ST & STEBBINS/REV P	BRONX	10459	42	
2100411070140028	01/04/2021 12:41:00 AM	BROOKLYN	1107	4 AVE & 68 ST	BROOKLYN	11220	68	
2100429540120021	01/04/2021 12:41:00 AM	BRONX	2954	BATHGATE AVE & 176 ST	BRONX	10457	48	

1 to 50 of 10,275,092

Fig. 1: Fire Incident Dataset Snippet

the_geom	BORO	UNITID	POINT_X	POINT_Y	CB	LATITUDE	LONGITUDE
POINT (-73.79456804377382 40.77221777714884)	4	H425919a	1041150.586	220683.273	407	40.7722168	-73.79457092
POINT (-73.91289250895464 40.64434661766513)	3	H325449	1008423.396	174041.245	318	40.64434814	-73.9128952
POINT (-73.95303997851815 40.72505714515936)	3	H307276	997266.21899999	203437.906	301	40.72505569	-73.95304108
POINT (-73.99463256503688 40.69398892775767)	3	H301843	985738.421	192115.373	302	40.6939888	-73.99462891
POINT (-73.93569187481359 40.73528872722269)	4	H439410	1002071.97	207168.6435	402	40.73529053	-73.93569183
POINT (-73.91147293739287 40.63402697951728)	3	H328476	1008821.14399999	170281.91799999	318	40.63402557	-73.91147614
POINT (-73.94400116043501 40.6773924260893)	3	H315750	999782.669	186073.70399999	308	40.67739105	-73.94400024

Fig. 2: Fire Hydrant Dataset Snippet

C. Fire Inspection Data [8]

The Fire Inspection dataset (size 72.3 MB) captures historical information on inspections conducted by the Bureau of Fire Prevention, focusing on permits and inspection details categorized by account holders from 2014 to 2019. The data is provided by the Fire Department of New York (FDNY) and was last updated in March 2019.

ACCT_ID	ALPHA	ACCT_NUM	OWNER_NAME	LAST_VISIT_DT	LAST_FULL_INSP_DT	LAST_INSP_STAT	PREM_ADDR	BIN	COMMUNITY	COUNCIL	BBL	LAT	LONG	POST	BOROUGH	Number	Street	Cer	NTA
426	AA	34156844	NYC ECONOMIC DI	12/23/2014	12/23/2014	NOT APPROVAL(W	10 SOUTH ST	1000	101	1	1000	40.7	-74.01	1000	MN	10	SOUTH S	9	Batte
427	A	34230839	MARK & SON META	09/02/2014	09/02/2014	APPROVAL	10 SOUTH ST	1000	101	1	1000	40.7	-74.01	1000	MN	10	SOUTH S	9	Batte
428	AA	34318352	NYC ECONOMIC DI	12/11/2014	12/11/2014	NOT APPROVAL(W	10 SOUTH ST	1000	101	1	1000	40.7	-74.01	1000	MN	10	SOUTH S	9	Batte
429	A	34334987	DIAMOND TECH PF	12/29/2014	12/29/2014	APPROVAL	10 SOUTH ST	1000	101	1	1000	40.7	-74.01	1000	MN	10	SOUTH S	9	Batte
430	AA	35111616	NYC ECONOMIC DI	04/27/2015	04/27/2015	NOT APPROVAL(W	10 SOUTH ST	1000	101	1	1000	40.7	-74.01	1000	MN	10	SOUTH S	9	Batte
431	A	35169192	LEADDOG MARKET	06/24/2015	06/24/2015	APPROVAL	10 SOUTH ST	1000	101	1	1000	40.7	-74.01	1000	MN	10	SOUTH S	9	Batte

Fig. 3: Fire Inspection Dataset Snippet

D. Census & Income Data [9]

The dataset is compiled by the United States Census Bureau and the American Housing Survey. It consists of three major categories-

- Mean Income Data - It has 216 rows and 218 columns.
- Median Income Data - It has 216 rows and 110 columns
- Age Demographics - It has 216 rows and 122 columns

Each category spans 10 files, corresponding to the years 2011 through 2020. This comprehensive dataset offers insights into various aspects of NYC zip codes, including mean and median income breakdowns, population demographics by age, gender, and race. The size of the dataset is 8MB.

E. Decennial Population & Housing

The dataset, compiled by the United States Census Bureau, comprises two distinct input files. One file contains data on zip codes along with corresponding average decennial population estimates, while the other file includes zip codes alongside average housing unit information. Each file contains 220 rows and 5 columns. This is a concise dataset with an approximate size of 2 MB.

F. Weather Data [10]

The weather dataset sourced from the Open-mateo website offers detailed hourly weather data spanning the years 2016 to 2022. This dataset has $60,000$ rows and 10 columns and has an approximate size of 3 MB. It describes the temperature, precipitation, rain, cloud coverage, wind speed and wind direction.

IV. Data Cleaning & Transformation

A. Fire Incident Data

During the data profiling phase, we assessed the completeness of data by running a map-reduce job on the raw dataset to count the number of entries that did not conform to the table schema. The job also calculated the count of rows with empty or null fields. The implementation is captured as part of InvalidEntres.java.
While assessing the data quality, we observed several minute inconsistencies with regards to the alarm_box_location field where the same location used different street identifiers. Many rows used the word "STREET", others wrote "ST". The same inconsistencies occurred with "BROADWAY", "BDWY", "AVENUE", "AV", "AVE", "TURNPIKE", "TPKE", etc and other street identifiers.

We take findings from the profiling phase to apply them into the data cleaning phase. The implementation can be found in the DataCleaning.java file.

- Since this dataset also captures non-fire incidents which were handled by FDNY, we added appropriate filtering mechanisms in place to only consider fire incidents.
- To ensure data completeness, we remove rows which do not conform to the specified schema and have null values for the necessary columns.
- We perform further cleaning by dropping columns which we did not require for our analysis.
- We add an additional timestamp column that is consistent with the weather dataset and is rounded to the nearest hour. This proved useful while performing the inner join of the two datasets.
- We also modified the format for the two timestamp fields (incident-datetime and incident-closetime). This modification is necessary to store the field with timestamp type in the Hive tables.
- The location column was modified to follow a standard mapping for different street identifiers.

To profile the clean dataset, I ran two map-reduce jobs and performed some summarizations. The implementation is captured as part of WordCounter.jar and NumericalSummarizations.jar.

B. Hydrant Data

In our study, we initially streamlined the dataset by retaining only the relevant attributes: geographical coordinates (latitude and longitude) and borough identifiers.

- The borough identifiers were initially encoded as numerical codes, which we subsequently transformed into their respective borough names.
- The dataset, however, lacked ZIP codes, presenting only the geographical coordinates for reference. Our preliminary strategy to append ZIP codes based on the geographical coordinates of the fire hydrants involved spatial mapping. However, given the average distance of 400–500 meters between hydrants, this method required precise ZIP code allocation. To address this, we evaluated each set of geographical coordinates against multipolygon representations of the ZIP code areas. These multipolygons were constructed from various polygons and points, and the point-in-polygon tests for each hydrant proved time-consuming. We initially optimized this process by creating bounding boxes for each multipolygon and assessing whether the hydrant's coordinates fell within the bounding box before testing against the multipolygon. Despite this optimization, the process remained inefficient.

Subsequently, we utilized the Google Maps Geocoding API to reverse geocode the hydrants' coordinates and extract the corresponding ZIP codes from the returned addresses. This approach proved more efficient and accurate for our purposes, and we successfully incorporated the ZIP codes into the dataset as part of our MapReduce job. During this phase, we also standardized the borough names to align with other datasets and removed the geographical coordinates column, "the_geom," to further streamline the data.

C. Fire Inspection Data

The fire inspection dataset initially comprised 20 columns and 401,000 rows. To focus on relevant data, we eliminated columns such as "community board," "latitude," "longitude," and "owner name" that were not pertinent to our research. We selected columns including "zip code," "borough," "inspection date" (for extracting the year for year-wise analysis), and "inspection status." Rows lacking data in these columns were removed.

- Additionally, the "borough" column contained abbreviations, which we expanded to the full borough names for consistency with other datasets.
- The date format also differed from the "fire incident" dataset, so we standardized it to the format: yyyy-MM-dd'T'HH:mm.

- Some rows in the txt file were not consistent with the column names after conversion from csv to txt, so the columns could not be fetched using just the index. The date and borough column were fetched after checking the length of data in the column, and status was fetched after string matching for all categories.

D. Census & Income Data

The yearly census and income data posed significant challenges due to their extensive nature, comprising 30 files across three categories of data. Each category contained a large number of columns, necessitating a thorough understanding of the data structure to extract meaningful columns. Moreover, the data presented additional complexities, as numerous null values were inconsistently labeled as "X," "-", or "/," indicating no observations likely due to small sample sizes or missing data. The structure of the files varied between the years 2011 to 2016 and those beyond 2016. This discrepancy required the adoption of different parsing logic and the selection of different columns during Map-Reduce.

E. Decennial Population & Housing

In the decennial data, redundancy was identified among columns such as geographical area, Zip Code Tabulation Area (ZCTA), and ZIP code, all essentially representing the same geographical information. To streamline the dataset, only the ZIP code column was retained, while the redundant columns were dropped. Null values were replaced with zeros. The original dataset lacked a borough column, which is crucial for contextualizing the geographical location of ZIP codes within New York City. To address this gap, a borough column was added during the reduce phase in the map-reduce task.

F. Weather Data

The weather data was cleaned to remove rows with null values. Each record in the dataset was checked to ensure that the record follows a consistent date-time format.

V. Design Diagram

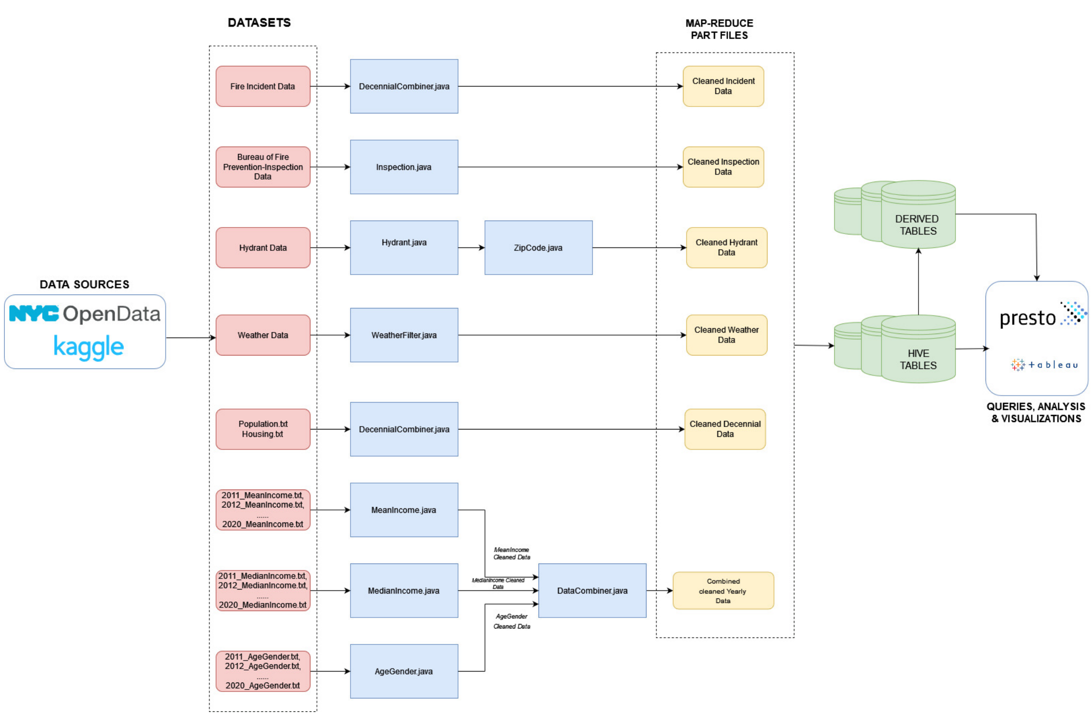


Fig. 4: Design Diagram

- We implemented a series of MapReduce jobs to clean and transform various datasets, including Fire Incident, Fire Inspection, Weather and Hydrant data.
- Subsequently, the cleaned Hydrant dataset underwent another MapReduce job to incorporate zip codes, leveraging the Google Maps Platform API.
- The Decennial Data, comprising two raw input files, necessitated a MapReduce job with MultipleInputs.

- Additionally, the Census and Income data, distributed across multiple files within each category, underwent initial cleaning through MapReduce jobs tailored for each category. Following this, a subsequent MapReduce job aggregated the cleaned dataset from the three categories, consolidating yearly census data.
- The resulting part files from the MapReduce jobs were ingested into Hive tables. To facilitate more complex queries, derived tables were created within Hive.
- For querying, Presto served as the engine, offering high-performance processing. Tableau, integrated with a JDBC connector, was then employed to generate insightful visualizations based on the queried data.

VI. Individual Analysis

A. Fire Incident Data

Our analysis for fire incidents were done across four categories, results of which are presented below.

1) Borough-wise analysis: Looking at the borough-wise distribution 5, we can clearly see that the frequency of incidents is maximum in Brooklyn, followed by Manhattan and Queens. Staten Island has the least frequencies reported. We can note a positive trend depicted in 6, illustrating the year-on-year distribution of fire incidents. Incidents rose from 2011 to 2016 before declining. However, considering the entire graph, it's evident that the overall number of incidents decreased from 2005 to 2023. The pattern is observed consistently for all boroughs.

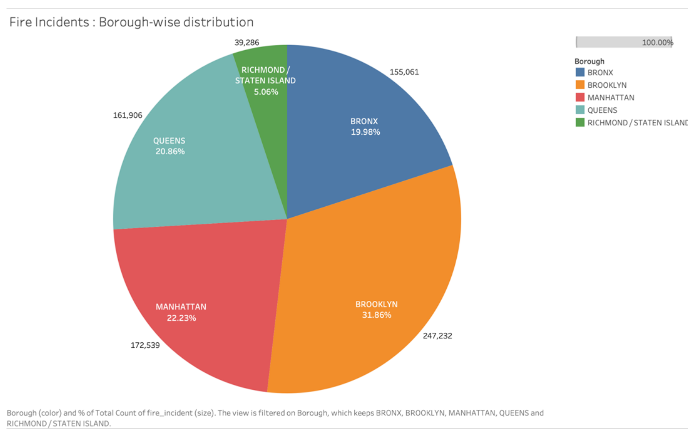


Fig. 5: Borough-wise distribution of fire incidents

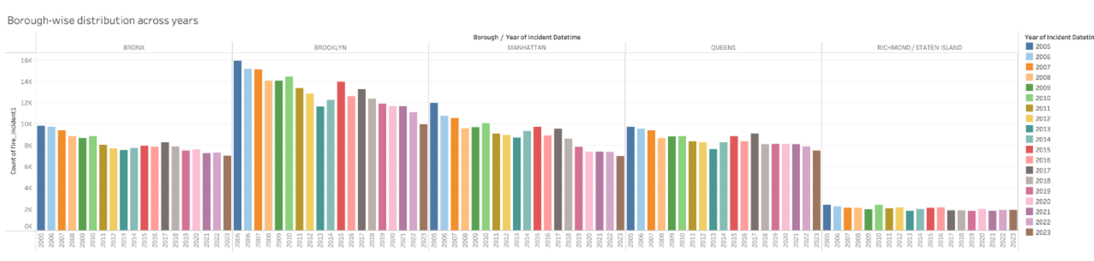


Fig. 6: Borough-wise distribution across years

2) Response Time analysis: Response time, as defined by the dataset is the elapsed time in seconds between the incident_datetime (time when the incident was reported) and the first_onscene_datetime (time when the first unit arrived on scene). We examined this average response time (in seconds) across both years and boroughs to gain insight into the efficiency of the NYPD's response to fire incidents. Contrary to our expectations, the average response time has actually increased over the years, as illustrated 24. This trend remains consistent across all boroughs. It underscores the urgent need for additional personnel and resources to facilitate quicker incident response.

3) Incident Class Group analysis: The dataset comprises two fire-related incident class groups. 8 illustrates see the two class groups and the various class of fires inside each group. Structural fires pertain to fires occurring in residential and commercial indoor structures, while non-structural fires encompass outdoor fire incidents. 9 illustrates that the city experiences a notably higher number of structural fire incidents compared to non-structural fires. Further analysis of the yearly frequency indicates a more rapid decline in non-structural fires than in structural fires. This highlights the necessity for increased inspections of indoor dwellings to address this discrepancy.

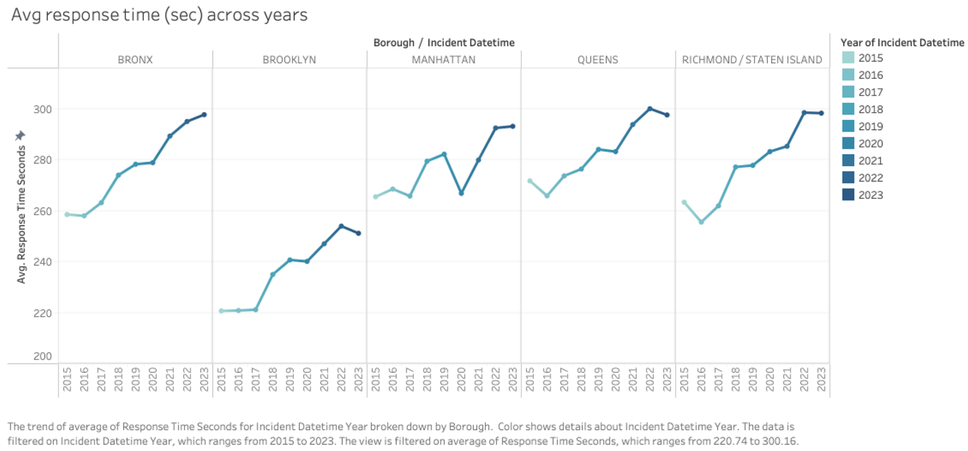


Fig. 7: Average response time (in seconds) across years in different boroughs

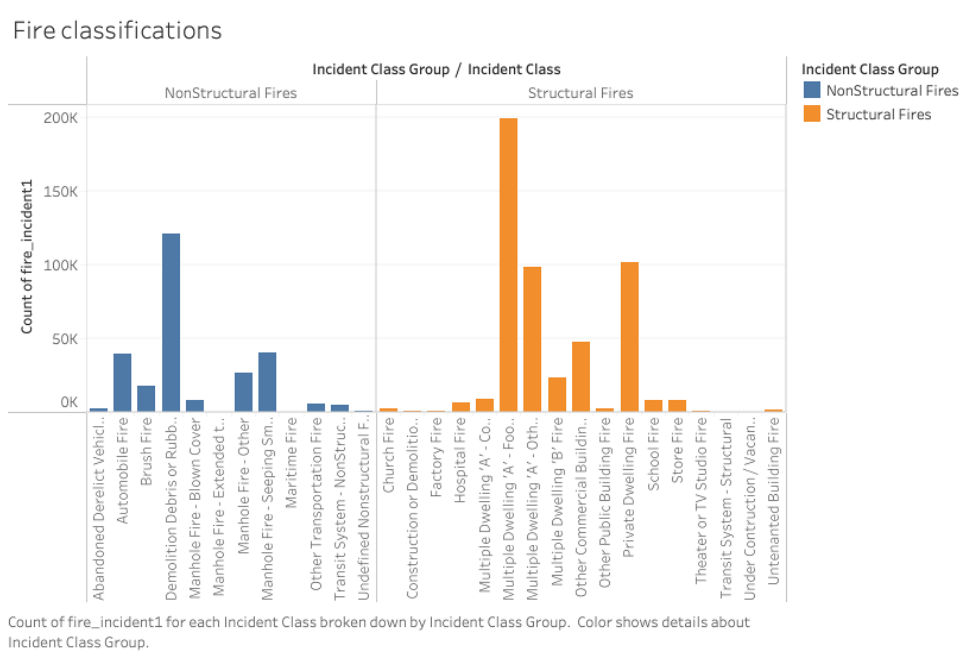


Fig. 8: Different classifications under Structural Fires and Non Structural Fires

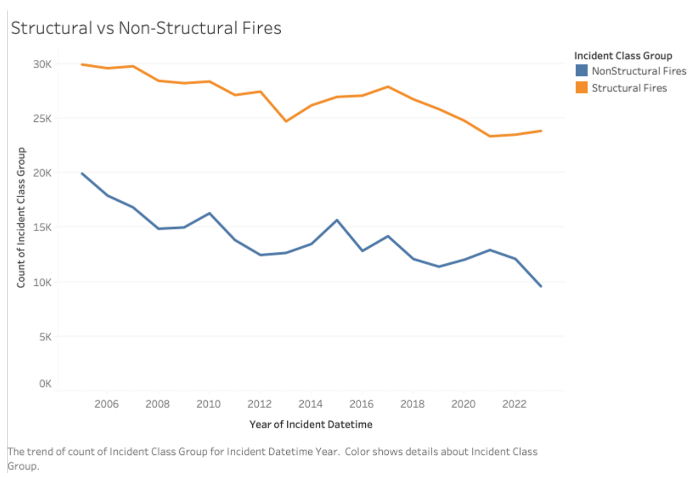
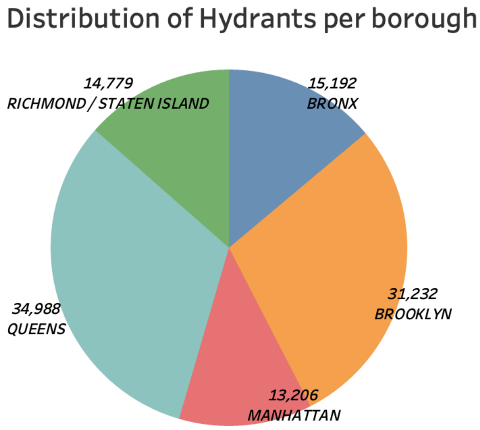
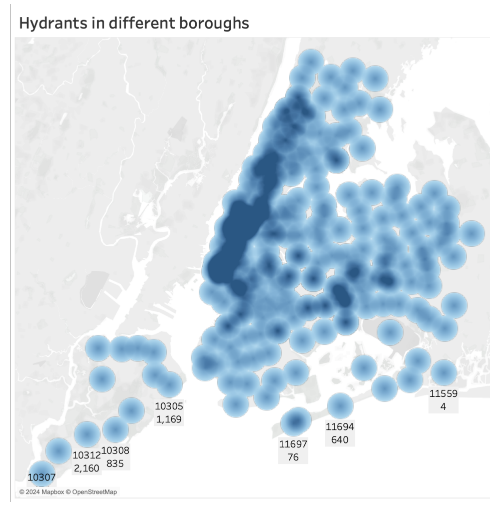


Fig. 9: Structural and Non-structural fires across years

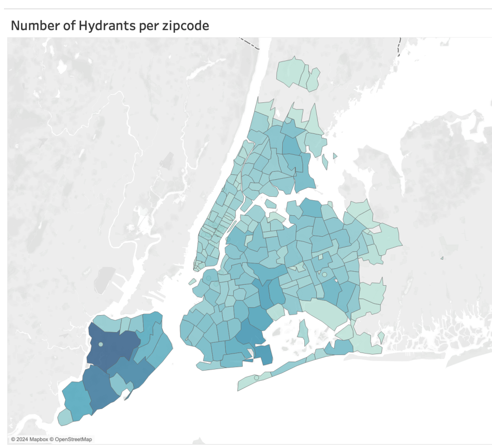


(a) Distribution of hydrants across all boroughs

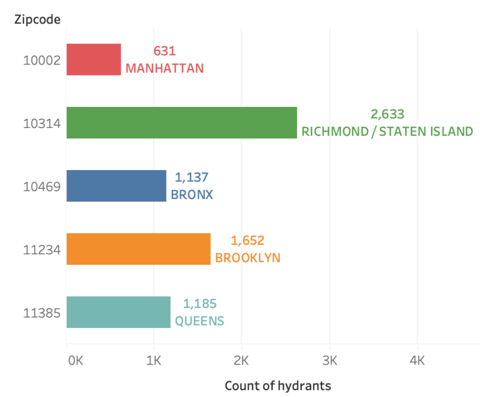


(b) Density of hydrants in New York

Fig. 10: Hydrant distribution analysis



(a) Density of hydrants per ZIP code



(b) The ZIP code having the highest hydrant count of each borough

Fig. 11: Hydrant-ZIP code analysis

B. Hydrant Data

1) Borough-Wise: The initial analysis involved comparing the number of hydrants and density across all five boroughs and ZIP codes. New York has a total of $109,397$ hydrants. The results show that Queens has the highest number of hydrants $(34,998)$ followed by Brooklyn $(31,232)$ and the density of hydrants is the most in Manhattan.

2) Zip-code wise: We also visualized the distribution of hydrants across all ZIP codes. From figure 2(c) and 2(d), we can see that ZIP codes in Staten Island have the most number of hydrants compared to other ZIP codes though Manhattan has the highest density of hydrants, as visible in Figure 2(b). This is because of the difference in area-wise distribution of hydrants since Staten Island ZIP codes are much larger in area as compared to those in Manhattan. Fire hydrant density was evaluated across boroughs, with a focus on areas with high fire incidents. The analysis revealed that Brooklyn, which had the most fire incidents, population and housing units, had lower hydrant density, suggesting a need for increased hydrant installations. This will be covered in detail in the Joint Analysis section.

C. Fire Inspection Data

1) Borough-Wise: Fire inspection data was analyzed to assess total inspections carried out in all five boroughs for every year. Manhattan has the most number of inspections $(145,768)$, followed by Brooklyn $(87,968)$.

2) Status-wise for each year: The analysis also revealed patterns in inspection numbers and outcomes in chronological order. There were 3 categories- Units that were approved, not approved due to violation, not approved (reason unknown).

During the years 2014-2016, the 'approval' results were as equal to the 'not approval' results, but from 2017, the 'approval' results started drastically increasing.

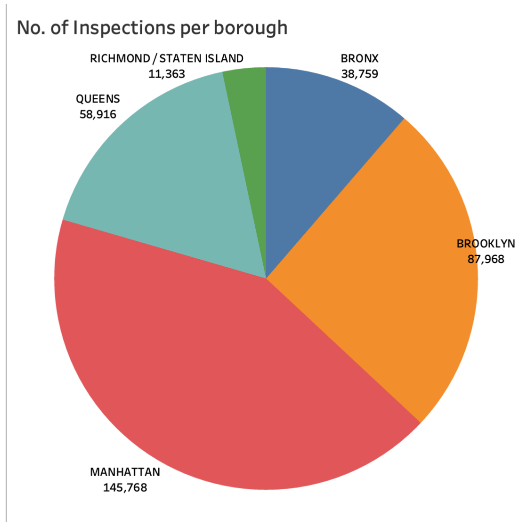


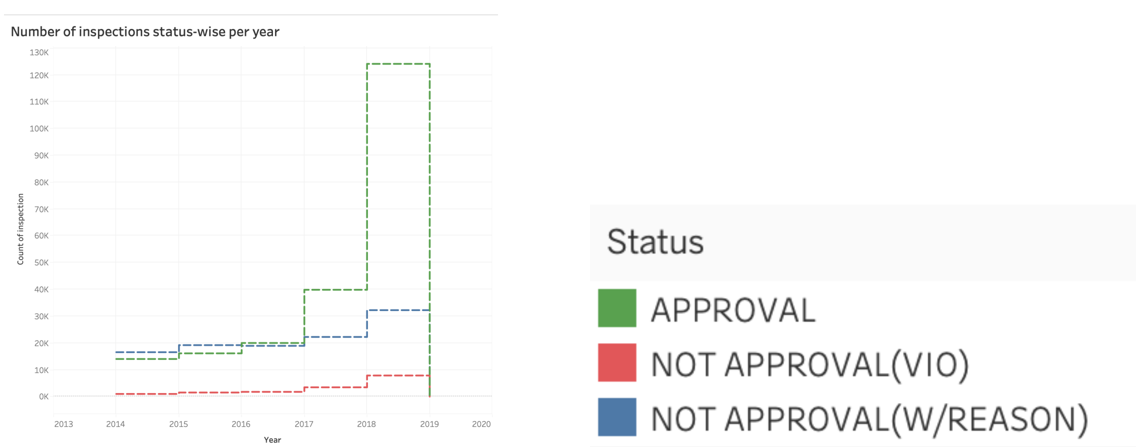
Fig. 12: Distribution of inspections across all five boroughs of New York

3) Borough wise for each year: Year 2018 witnessed a huge increase in inspections due to multiple reasons-

- The boom in construction across New York City likely led to a rise in the number of inspections to ensure safety compliance on a growing number of job sites.
- Several new laws and reforms were implemented in 2018 which increased the scope and requirements for safety inspections. This included enhanced training requirements for workers and stricter penalties for non-compliance, which necessitated more inspections to ensure adherence to these new standards.
- The rollout of the Department of Building (DOB) NOW online portal, which expanded to include more kinds of inspection requests, made it easier to schedule and manage inspections.

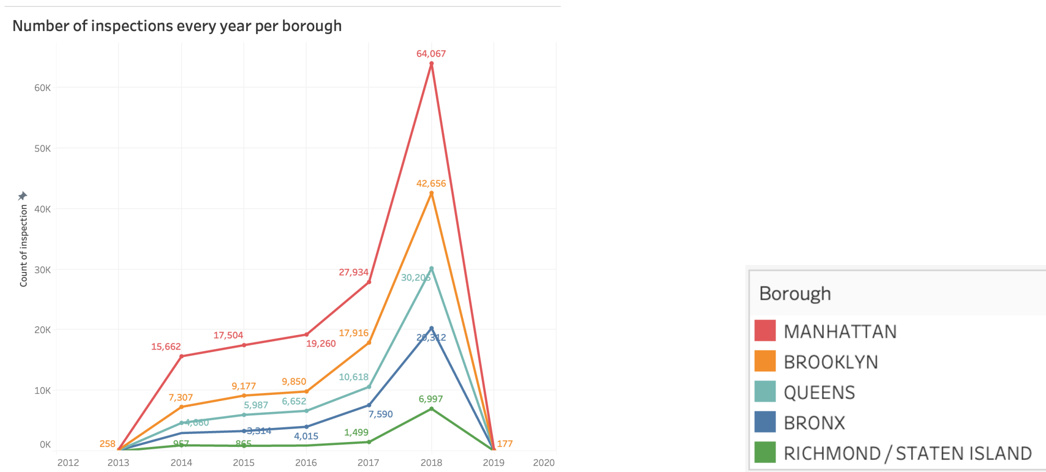
D. Census & Income Data

1) Population Analysis: Analyzing the population in various boroughs over the years showed a slow steady increase in population for the first 7 years followed by a small decline in the total population. The total number of children and seniors in NYC also slowly increases over the years.The decline 15 in the total population in 2020 can be attributed to the onset of the COVID-19 global pandemic.



(a) (b)

Fig. 13: shows the yearwise inspection count for each status category (Approval, Not Approval(VIO) and Not Approval(W/REASON)



(a) (b)

Fig. 14: Yearwise inspection count for each borough from 2012 to 2019

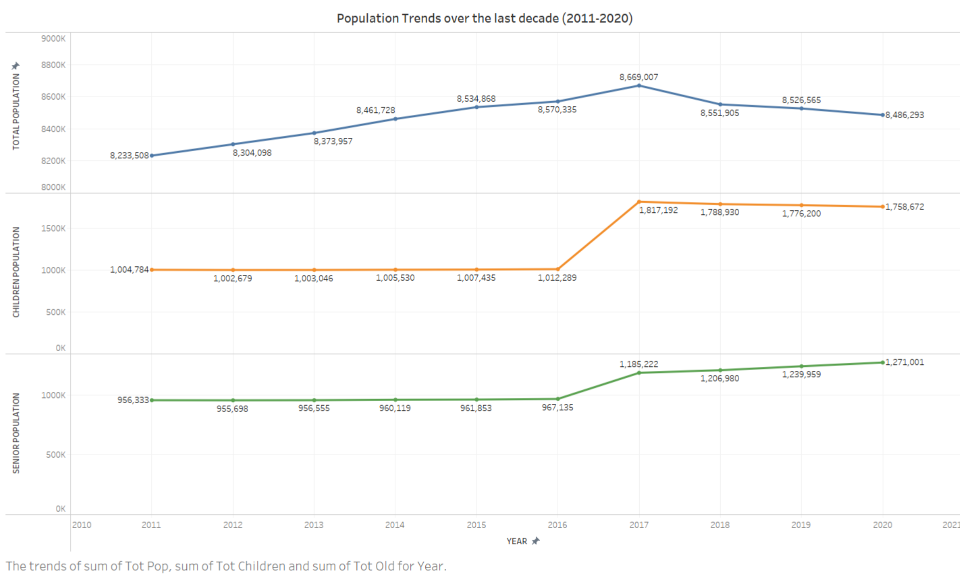


Fig. 15: Population trends for 2011-2020

2) Mean Income of Borough: The figure illustrates the mean income of the entire population and population by races in a borough.We focus primarily on White, Asian, and Black demographics due to their substantial representation for meaningful analysis. Manhattan emerges with the highest mean income among all boroughs. Notably, across all boroughs, the mean income for White individuals surpasses that of Asian and Black individuals. However, an exception arises in Queens 16 where the mean income of Black individuals exceeds that of Asian individuals.

3) Per Capita Mean Income and Median Income across Boroughs: Manhattan has the highest per-capita mean income of $\$119,124$. However, we see that the median income of a person living in Manhattan is $\$56,749$ which is half the mean income. This stark contrast implies the presence of a small yet immensely affluent group within Manhattan, exerting significant influence on the borough's mean income. Consequently, this also highlights the significant financial inequality among Manhattan's population, reflective of broader socioeconomic disparities within the borough.

E. Weather Data

Weather data analysis examined the influence of temperature, precipitation, and wind speed on fire incidents. The analysis revealed that low temperatures were associated with higher fire incidents

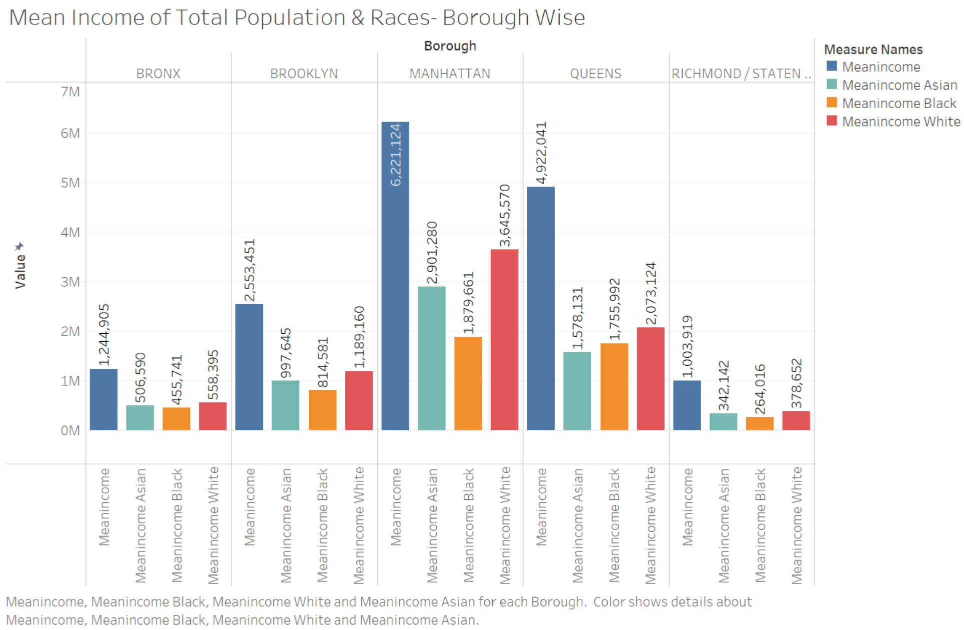


Fig. 16: Mean Income of Total Population & Races- Borough Wise

F. Decennial Population & Housing

The decennial data provides insights into the population and housing unit distribution in the various boroughs of New York City. This is an average estimate for the last decade (2011-2020). NYC has a total population of $8,992,164$ and $3,661,181$ housing units.18 The first pie chart illustrates the population distribution across different boroughs, highlighting the relative proportions of residents in each area. Brooklyn has the highest decennial population followed by Queens and Manhattan. The second pie chart depicts the distribution of housing units among the various boroughs. Brooklyn, Queens and Manhattan have the highest number of housing units. The ratio of the total population to housing units is lowest for Manhattan and is comparable for the rest of the boroughs.

VII. Joint Analysis

The analysis conducted in Section 6 gives us insights about our individual datasets, highlighting intriguing patterns observed across boroughs and over time. However, as our primary project objective is to determine the variables influencing fire incidents in NYC, it is important that we delve deeper into conducting joint analyses of datasets. By analyzing multiple datasets together, we can uncover correlations and patterns that may not be apparent when examining each dataset individually.

We use the Karl-Pearson's correlation coefficient (r) to quantify the linear relationship between two variables. We use this coefficient to gauge the impact of different datasets with respect to the fire incident dataset.

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

The Pearson correlation coefficient r ranges from -1 to 1, where 1 describes a perfect positive linear relationship and -1 describes a perfect negative relationship.

A. Weather vs Fire Incident

We capture three important metrics in the weather data and compute the correlation coefficient of each metric with the fire incident dataset. We follow this process to understand the impact of each metric on incidents. The three metrics we capture from the weather data are temperature, wind-speed, and precipitation.

1) Temperature vs Fire incidents: We get a negative correlation coefficient of -0.3079 19 suggesting us that more fires occur during the colder months in NYC. This aligns with our initial hypothesis that winters in New York require constant use of heating appliances in the whole city. Xu, Zhisheng et al . [8] mentioned how electrical and heating appliances are heavily used for heating, boiling, drying and indoor cooking since most people stay indoor. This causes more malfunctions and hence, can likely cause more fire incidents.

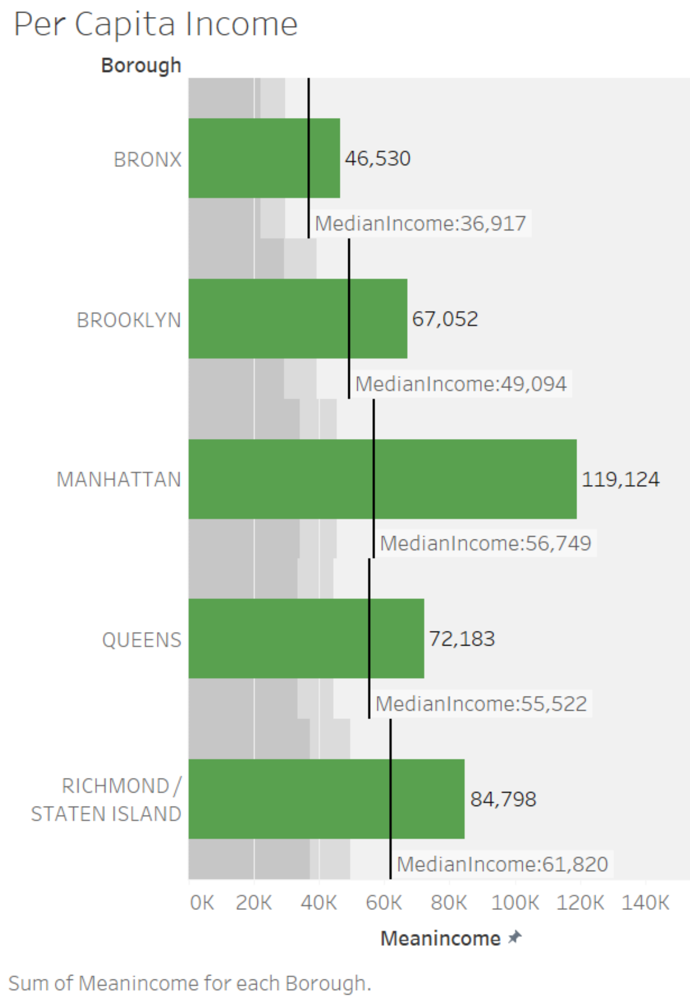


Fig. 17: Borough Wise Per capita Mean Income and Median Income

2) Precipitation vs Fire incidents: We observe that on days with moderate and heavy showers, the number of incidents decrease significantly. This can also be seen from the negative correlation of $-0.408\,19$ we get by combining precipitation and fire incident dataset.

3) Wind Speed vs Fire incidents: There is a strong negative correlation between windspeed and fire incidents in NYC. We observe that the number of incidents (ratio of incident count to wind speed frequency count) significantly reduce when the wind speed exceeds 20km/hr and decreases even further when wind speed exceed 30km/hr. This supports our initial hypothesis that strong winds can help extinguish small fires and inhibit spread of fire. We may note that this hypothesis is only for NYC where a majority of the fires are structural fires. In cities which see higher counts of forest fires, we may need to account for wind direction to make an informed decision.

B. Census & Income vs Fire Incident

From the census and income data we consider three important metrics: population, mean income and housing units data and study the effect of each of these factors on fire incidents by computing the Pearson correlation coefficient.

1) Population vs Fire Incidents 21: The Pearson Correlation coefficient for population and fire incidents is $0.90\,19$. This significant correlation suggests a strong positive relationship between population size and frequency of fire incidents in NYC. As the population in NYC increases ,the number of fire incidents also increases. Firstly, a larger population often necessitates more extensive infrastructure and transportation networks. With greater urban density comes increased proximity between structures, amplifying the potential for fire spread in the event of an ignition. Higher population density can also strain public resources and services, potentially impacting fire prevention and emergency response capabilities.

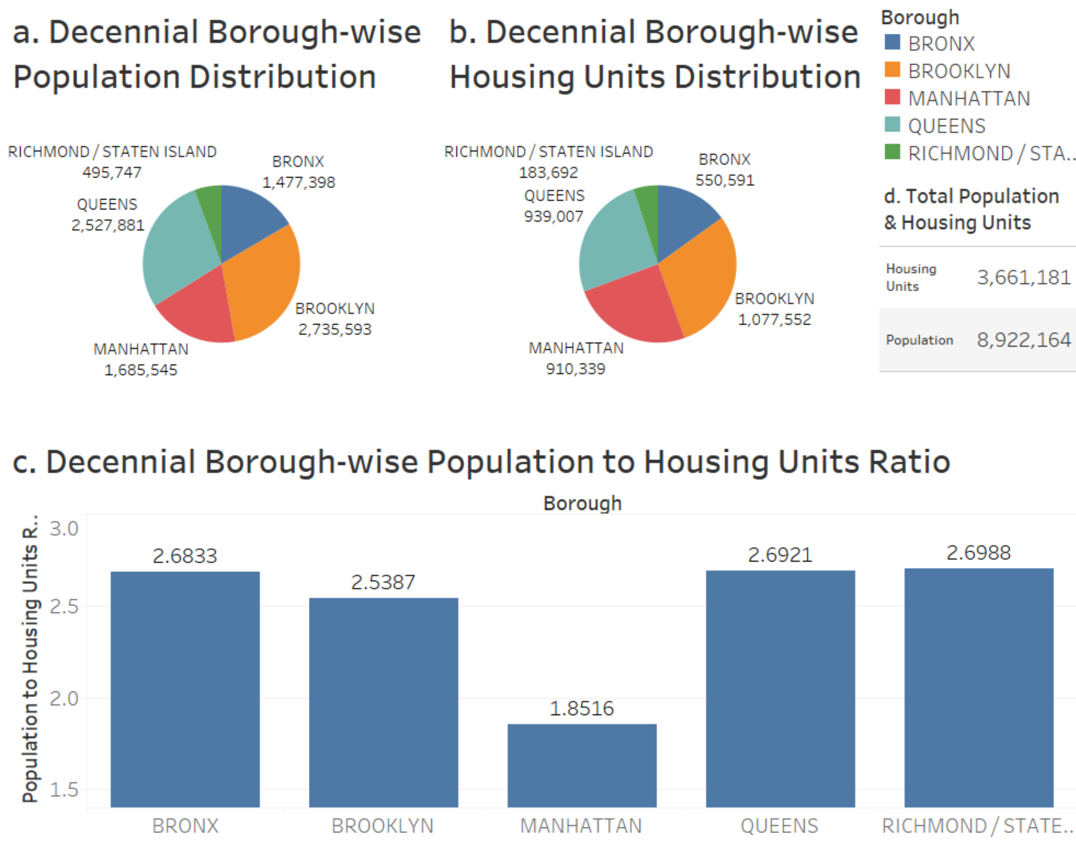


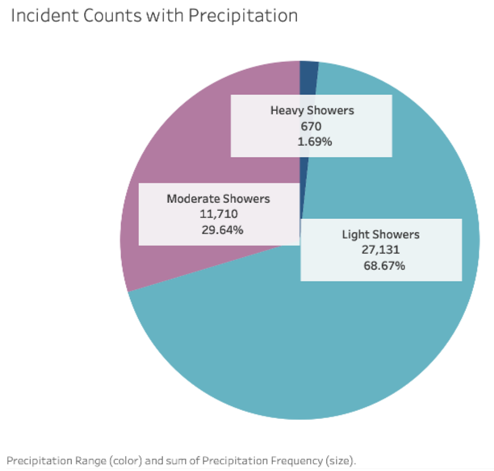
Fig. 18: Borough-wise Decennial Population distribution,Housing Units Distribution & Population to Housing Units Ratio

Census data	POPULATION	**FIRE INCIDENTS**	**0.902**
	HOUSING UNITS		**0.927**
	MEAN INCOME		**-0.131**
Inspection data	FIRE INSPECTIONS		**0.547**
Weather data	TEMPERATURE		**-0.3079**
	PRECIPITATION		**-0.408**
	WINDSPEED		**-0.7409**

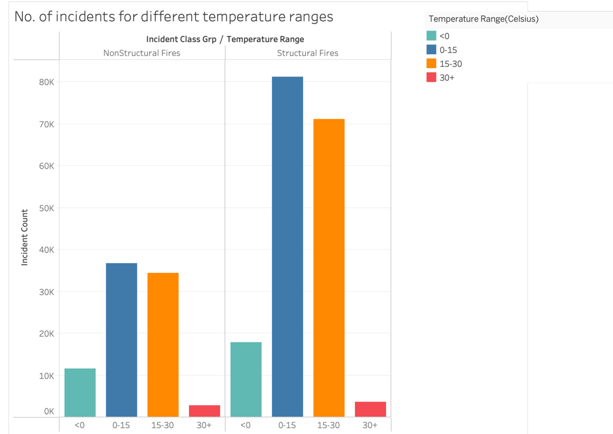
Fig. 19: Pearson Correlation Coefficient of various factors and Fire Incidents

2) Housing Units vs Fire Incidents: The Pearson Correlation coefficient for population and fire incidents is 0.92 19. This suggests an almost positive linear relationship between the two variables. The prevalence of multifamily residential buildings or increased housing units can exacerbate fire risk due to shared walls and ventilation systems. Increase in housing units also leads to increase in use of equipment like stoves which are a fire hazard.

3) Mean Income vs Fire Incidents: The correlation coefficient was -0.131 19. This negative correlation indicates that as the mean income of individuals living in New York City decreases, their exposure to fire incidents tends to increase, albeit to a limited extent. While the correlation is relatively weak, it highlights the potential relationship between socioeconomic factors and fire risk. Lower-income individuals and communities may face greater challenges in maintaining fire-safe living conditions, such as access to fire prevention resources, and education on fire safety practices. Additionally, socioeconomic disparities may influence the prevalence of fire-related hazards, such as faulty electrical wiring or substandard housing conditions, which can contribute to an increased risk of fire incidents.



(a) Number of incidents for different precipitation ranges



(b) Number of incidents for different temperature ranges

Fig. 20: Analyzing key weather metrics and fire incident data

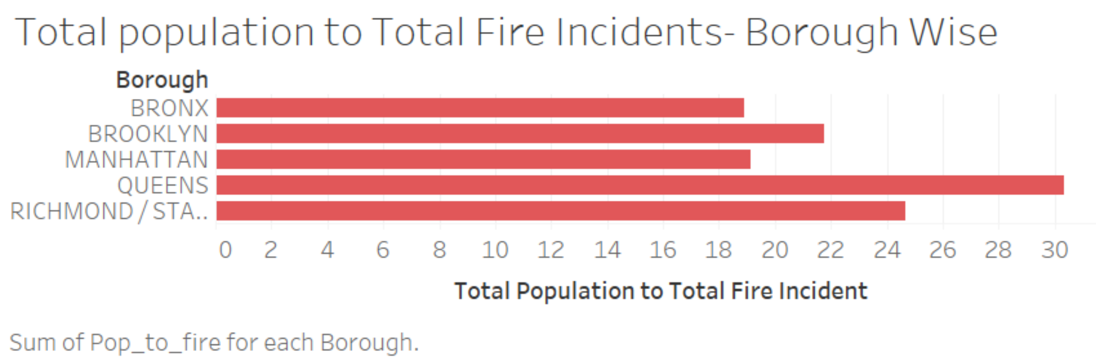


Fig. 21: Total Population to Total Fire Incidents-Borough Wise

C. Hydrants vs Fire Incident

From the borough-wise fire incident distribution, we observed that Brooklyn had the highest number of fire incidents. However the density of hydrants is not the highest in this borough. On the basis of this analysis, it seems that a good idea would be to increase hydrant density to ensure a decrease in fire incidents, given Brooklyn also has the highest population and housing units.

But investigating further, we calculated the ratio of incidents to hydrants for each borough. Even though the density of hydrants in Manhattan is more, Manhattan still needs more hydrants followed by Bronx and Brooklyn.

D. Inspection VS Fire Incident

We observed a correlation factor of 0.547 19, which was unexpected initially. The incidents should decrease as the inspections increase. But looking at it form another way , we can also say that the Government kept increasing the fire inspections because of increasing fire incidents every year. Despite this, the incidents did not decline. But in 2018 when the inspections were increased by a huge amount, the fire incidents actually started declining. We visualized this in Figure ().

Next we visualized the year wise inspections and incidents for each borough. Staten Island showed the least decline because the increase in inspection for this borough was less in comparison to other boroughs. We can also see that Manhattan saw the highest increment in inspections from 2017 to 2018, which resulted in a significant decrease in incidents as well. If the number of inspections in Brooklyn are also increased, the number of incidents will also keep declining. Queens witnessed the highest decrease in fire incidents as the inspections were increased.

Since the inspection data is historical and was last updated in March 2019, the results cannot be compared for this year. However, seeing the trends, we can assume that if the inspections keep increasing, incidents will also keep decreasing given the DOB's increased range of inspection requests and the launch of a new online inspection request portal in 2018. In conclusion, to decrease the number of fire-incidents the areas where interventions is required are:

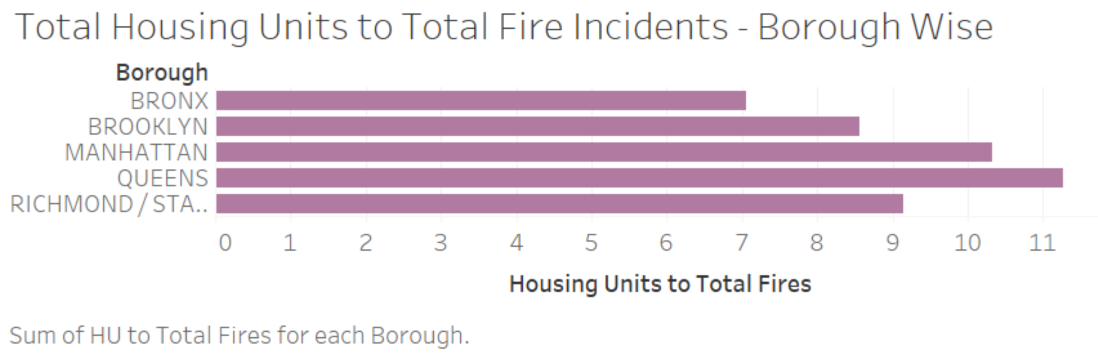


Fig. 22: Total Housing Units to Total Fire Incidents-Borough Wise

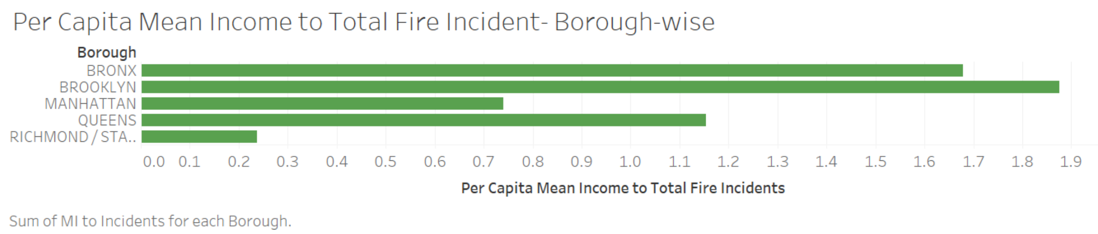


Fig. 23: Per Capita Mean Income to Total Fire Incidents-Borough Wise

- Inspections: Increasing the number of inspections will help in decreasing fire incidents across all boroughs, especially Richmond/Staten Island and Bronx.
- Hydrants: Initiative should be taken in Increasing the hydrant density across all boroughs, especially Manhattan, Bronx and Brooklyn.

VIII. Challenges

We encountered several obstacles while working on our project. Handling a total of six diverse datasets was a challenge, particularly while dealing with the wide-ranging time periods covered by the datasets. We had to ensure a careful alignment for getting an accurate analysis. Running the map-reduce jobs was significantly complex for our census data. Having a huge number of columns(230) and scattered data in around thirty different files took a large amount of time to run and designing the entire process took a lot of effort.

Additionally, obtaining precise NYC zip codes from the hydrants' latitude and longitude posed a significant technical challenge that was resolved through reverse geocoding API. Employing a reputable existing API ensured the accuracy of the postal codes extracted from the RESTful responses. Since all the project members were new to Tableau, the team had a steep learning curve for visualizing the data. Applying very complex queries was a technical challenge we faced in the beginning of the project.

IX. Conclusion

Our project undertook a comprehensive examination of fire incidents in New York City, seeking to understand their dynamics and the socio-demographic and environmental factors influencing their occurrence. Our analysis leveraged multiple datasets, including fire incidents, hydrants, inspections, census, and weather data, which were meticulously cleaned and analyzed to provide actionable insights. The analyses provided insights into how these factors interact and influence fire risks. The findings indicate that population size and housing units have a strong positive correlation with the frequency of fire incidents.

Additionally, our analysis showed that higher temperatures, higher precipitation, and higher wind speeds were associated with reduced fire incidents, underscoring the impact of environmental factors on fire risks. Socioeconomic disparities are evident, as lower-income communities tend to experience a higher frequency of fire incidents, highlighting the need for targeted interventions. The importance of fire inspections was also highlighted, with increased inspections correlating with a subsequent decline in fire incidents.

Despite the challenges faced during the project, including the handling of diverse datasets, and the alignment of time periods, our research demonstrates the value of data-driven approaches in understanding and managing fire incidents. Our research contributes to the ongoing discourse on effective disaster management practices, emphasizing the importance of evidence-based approaches in enhancing public safety and resilience against fire emergencies.

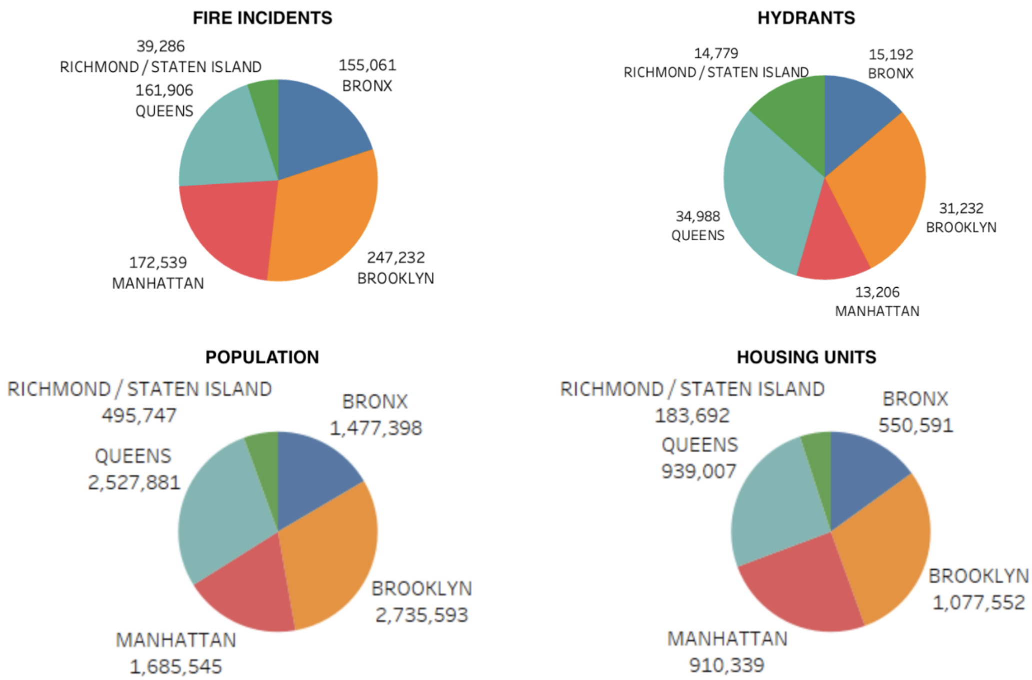


Fig. 24: Borough-Wise Distribution of Fire Incidents, Hydrants, Population & Housing Units

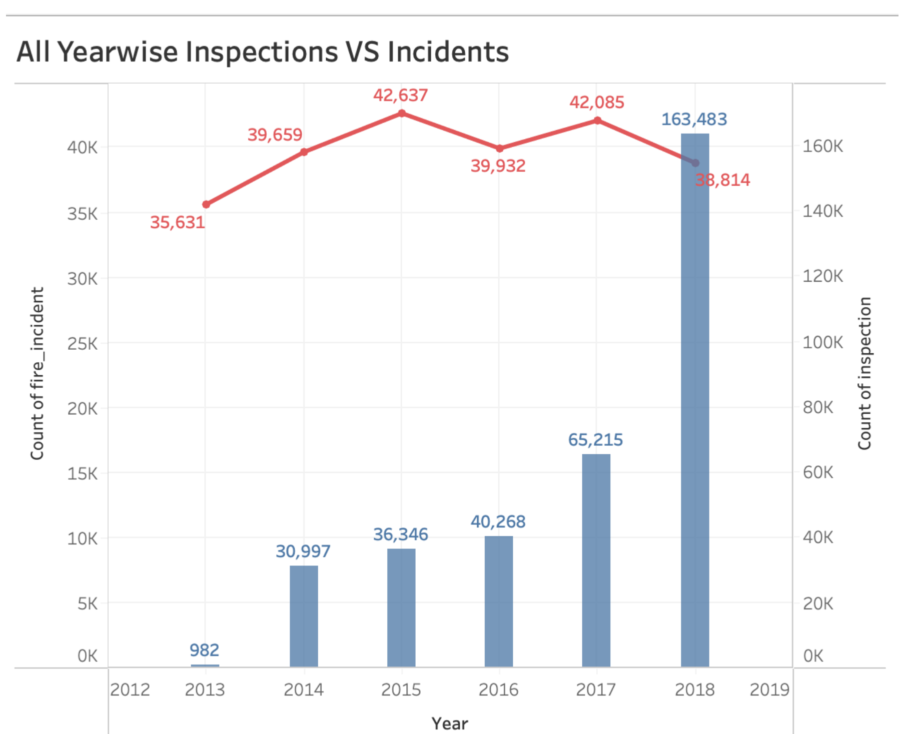


Fig. 25: Year-wise Fire Inspections Vs Fire Incidents

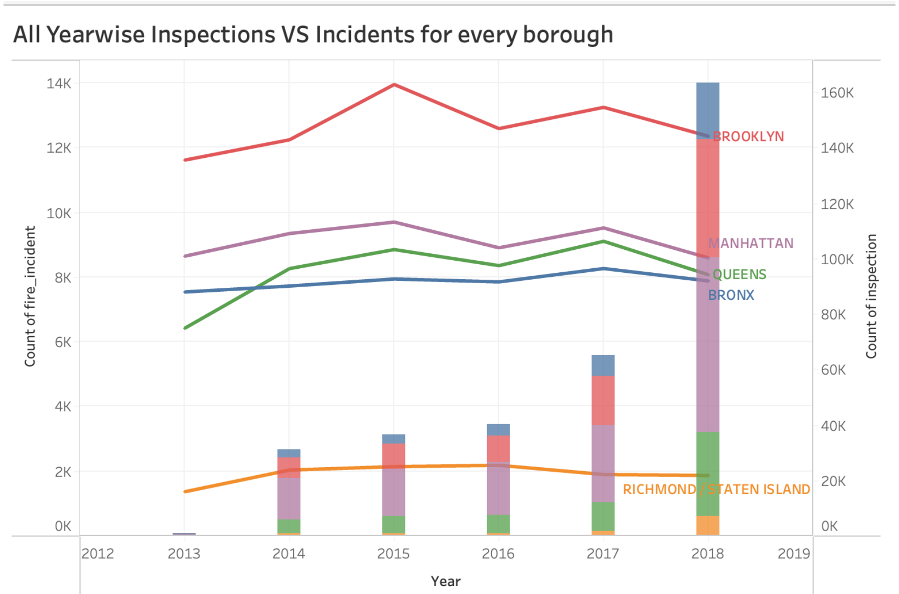


Fig. 26: Year-wise & Borough-wise Fire Inspections Vs Fire Incidents

X. Future Scope

Our research has practical implications for the FDNY and city government, offering valuable insights that can drive improvements in fire safety measures and emergency response strategies. The findings of this research can inform better resource allocation and hydrant placements, enabling the FDNY to deploy personnel and equipments more effectively to areas with higher fire incident rates. Moreover, the insights garnered from our research can guide the development of targeted policies aimed at enhancing fire safety in New York City. Moving forward, there is a need for continued exploration of causal relationships between fire incidents and various influencing factors such as building codes, community education programs etc. Developing targeted interventions and resource allocation strategies based on empirical evidence will be crucial for mitigating fire risks and enhancing emergency response capabilities in NYC.

References

[1] New York Times Article https://www.nytimes.com/live/2022/01/10/nyregion/bronx-fire-nyc
[2] Zhang, X.; Yao, J.; Sila-Nowicka, K.; Jin, Y. Urban Fire Dynamics and Its Association with Urban Growth: Evidence from Nanjing, China. ISPRS Int. J. Geo-Inf. 2020, 9, 218.
[3] National Fire Estimation using NFIRS Data
[4] Xu, Zhisheng, Dingli Liu, and Long Yan. ”Temperature-based fire frequency analysis using machine learning: A case of Changsha, China.” Climate Risk Management 31 (2021): 100276.
[5] Hossain, M.R. and Smirnov, O., 2023. Analyzing the risk factors of residential fires in urban and rural census tracts of Ohio using panel data analysis. Applied geography, 151, p.102863.
[6] Fire Department of New York City (FDNY). NYC open data. Fire Incident Dispatch Data — NYC Open Data. https://data.cityofnewyork.us/Public-Safety/Fire-Incident-Dispatch-Data/8m42-w767
[7] NYCDEP citywide hydrants. NYC Open Data. (n.d.). https://data.cityofnewyork.us/Environment/NYCDEP-Citywide-Hydrants/6pui-xhxz
[8] Fire Inspection Data https://data.cityofnewyork.us/Public-Safety/Bureau-of-Fire-Prevention-Inspections-Historical-/ssq6-fkht/about_data
[9] Census Data https://data.census.gov/profile/ZCTA5_10001?g=860XX00US10001
[10] Weather Data https://data.cityofnewyork.us/Environment/NYCDEP-Citywide-Hydrants/6pui-xhxz