San Francisco Fire Calls & COVID

Erick Tinoco, Walter Giron, Leslie Vasquez, Xavier Colin

Abstract

- Research project involving tools and methods for San Francisco fire call data analysis.
- Tools utilized: Hadoop, Hive, Excel, Tableau
- Excel and Tableau are used to create visualizations of the data
- Data Analysis between San Francisco fires and COVID

Introduction

- Hadoop and Hive were used to analyze the San Francisco Fire Department calls for service and how COVID was present in this area throughout the years.
- The dataset of 2.2 GB includes all fire units' responses to calls.
- Our belief is COVID cases skyrocketed because of protests, wildfires and domestic activities
- Three major areas affected: Civic Center, Bayview and Dolores Heights
- During a protests, a fire was caused by the protestors which was a sign that many people were in one spot
- All fire calls in the city were reported to the Open Data Portal for the public

Related Work

1. On the DataBricks platform, they used many volunteers to manually input COVID data into there reports. They developed some automation tools to grab information from state websites. Tools used were Python Spark SQL, and DataFrame
2. On the COVID Tracking Project, they worked with many volunteers to input data entry manually in order to have concise data. They developed some automation tools to retrieve data from state websites reporting COVID data.
3. San Francisco's COVID-19 response method: Utilizing established metrics from health departments, the city assessed strategy impact by routinely calculating data gathered through public health surveillance efforts.
4. Despite differences, all three studies aimed for a thorough analysis of COVID-19 data, enhancing understanding and informing versatile public health strategies with unique insights tailored to research challenges.

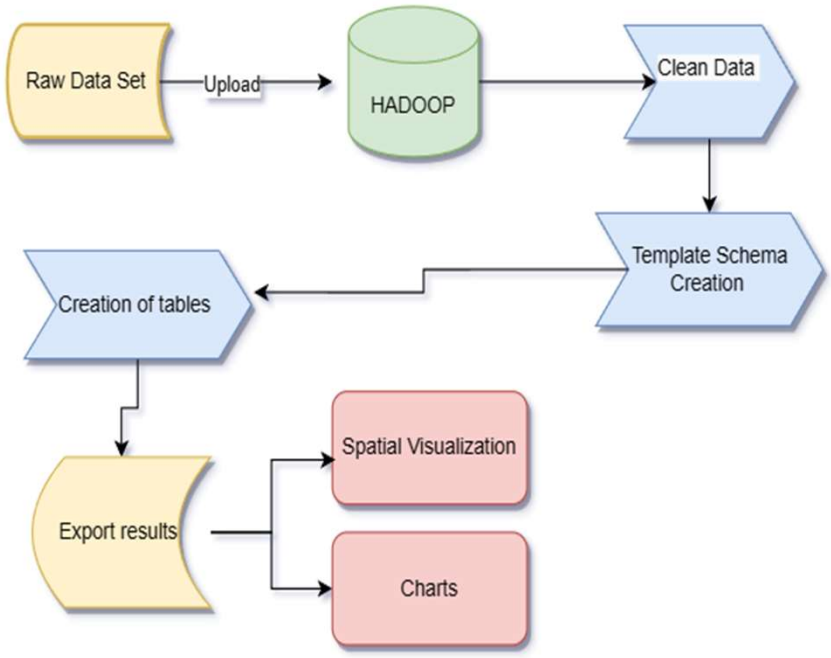
Dataset Specifications

Hardware Specifications	
Nodes	5
CPUs	8
CPU Speed	1995.312 MHZ
Total Memory Size	58 GB

Data Set Specifications	
Data Set	Size (Total 34.467 MB)
Fire_Department_Calls_for_Service_1	30,243 KB
COVID-19 Testing SF	4,224 KB

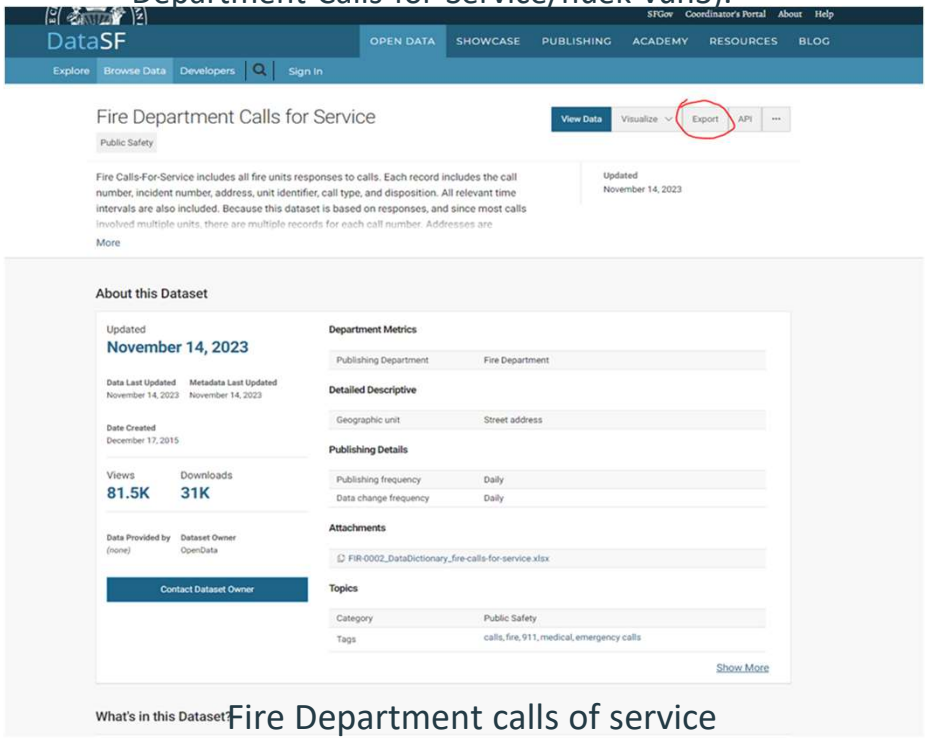
Implementation Flowchart

- Raw Data (San Francisco Gov Site)
- Uploaded using File transfer to Hadoop
- Cleaned in Hadoop (Mad Reduce/Excel)
- Create Database (Hive)
- Tables specified (Hive)
- Downloaded to Explore (Personal PC)
- Analyze (Tableau)

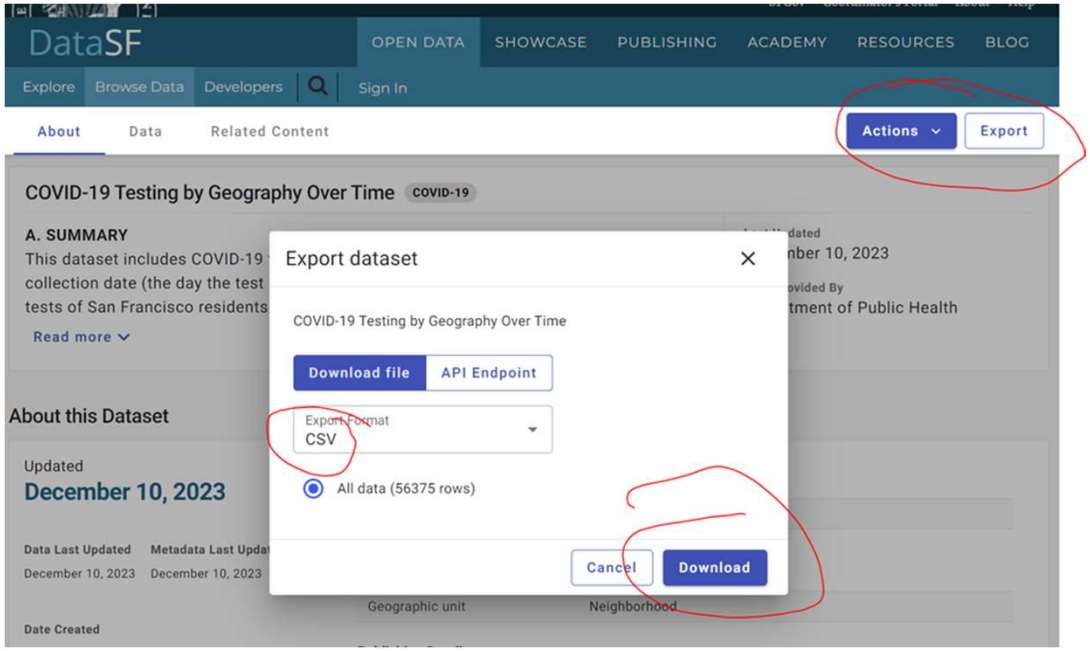


Downloading the Data

- Visit DataSF's official website: [DataSF - Fire Department Calls for Service](https://data.sfgov.org/Public-Safety/Fire-Department-Calls-for-Service/nuek-vuh3).

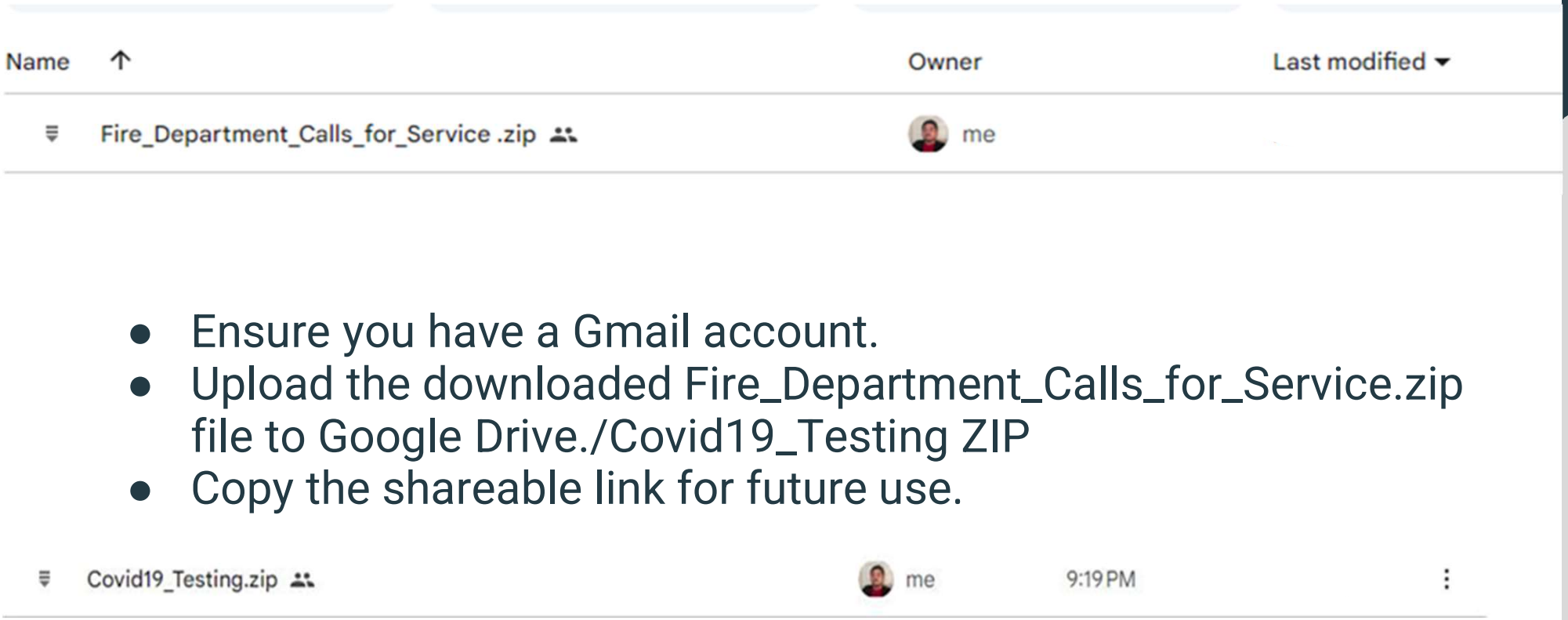


Fire Department calls of service



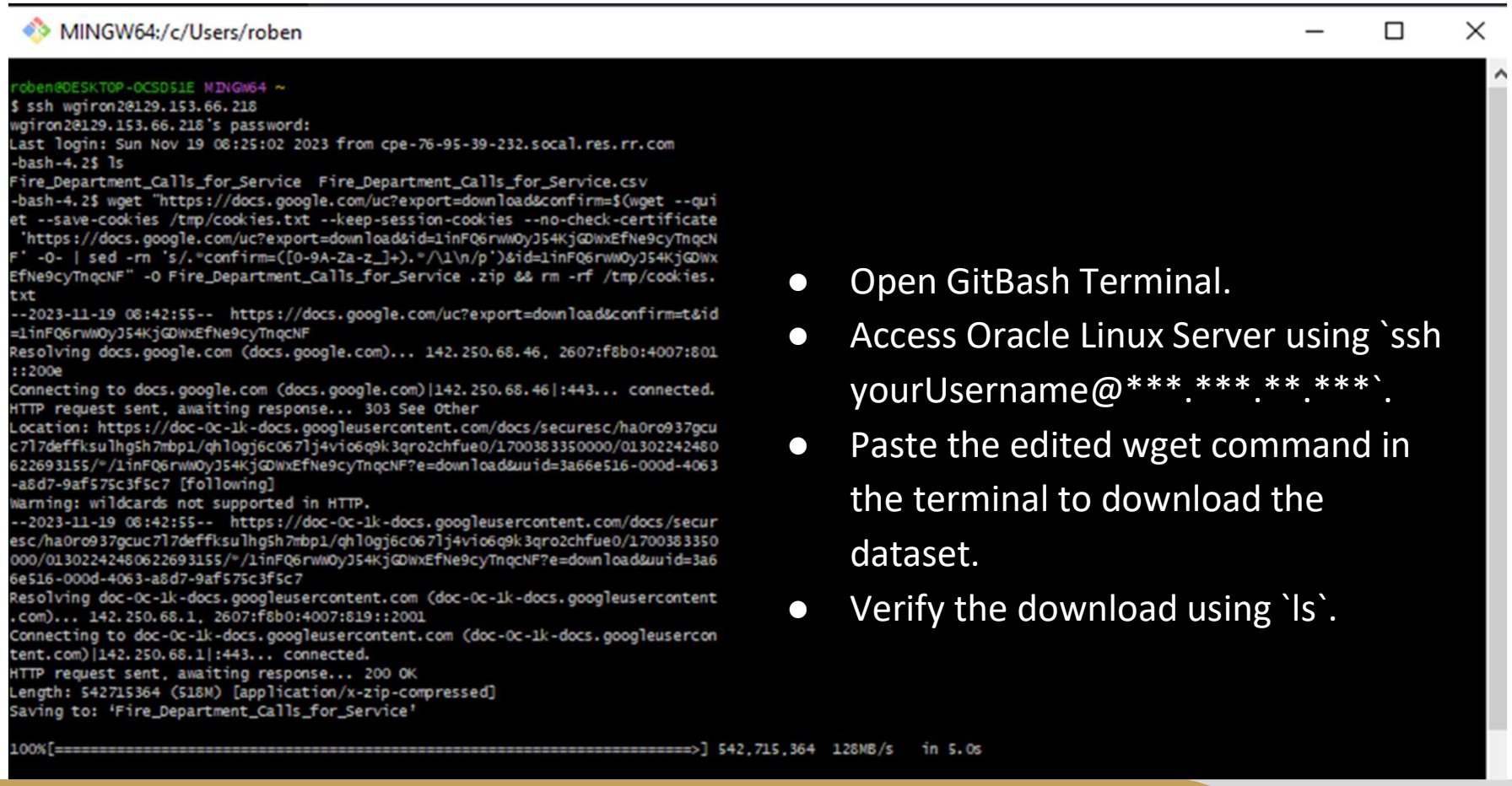
- Click on the desired dataset.
- Download the dataset by selecting the export option and choosing CSV.

Uploading Dataset to Google Drive



- Ensure you have a Gmail account.
- Upload the downloaded Fire_Department_Calls_for_Service.zip file to Google Drive./Covid19_Testing ZIP
- Copy the shareable link for future use.

Extracting File to Oracle Linux Server



- Open GitBash Terminal.
- Access Oracle Linux Server using `ssh yourUsername@***.***.**.***`.
- Paste the edited wget command in the terminal to download the dataset.
- Verify the download using `ls`.

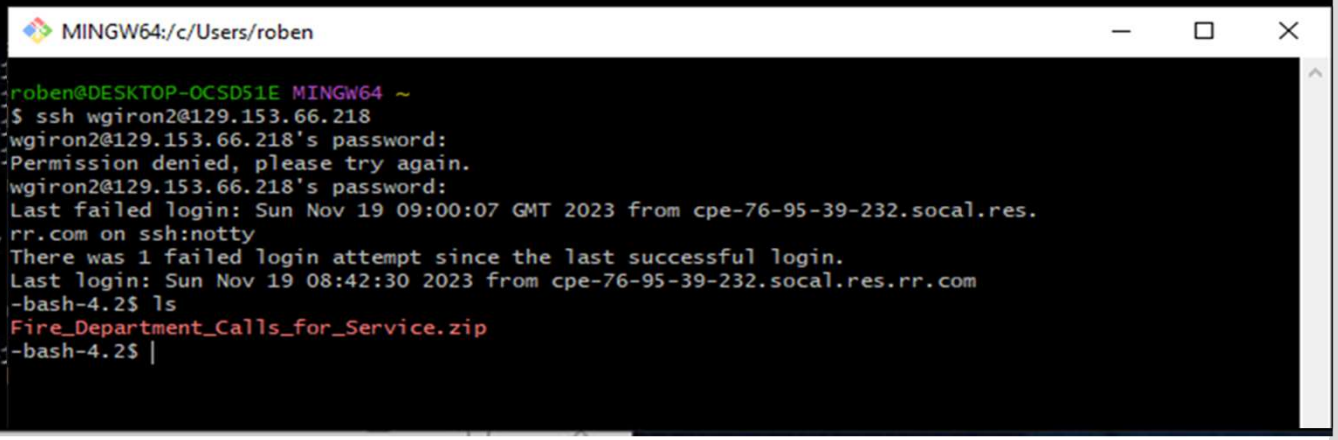
Unzipping Zip FIle to Linux Server

```
100%[=====================================>] 1,599,756   8.14MB/s   in 0.2s

2023-11-30 05:30:12 (8.14 MB/s) - 'Covid19_Testing.zip' saved [1599756/1599756]

rm: cannot remove '/tmp/cookies.txt': Operation not permitted
-bash-4.2$ ls
Covid19_Testing.zip
-bash-4.2$ unzip Covid19_Testing.zip
Archive:  Covid19_Testing.zip
  inflating: Covid19_Testing.csv
-bash-4.2$
-bash-4.2$ du -h Covid19_Testing.csv
8.4M    Covid19_Testing.csv
-bash-4.2$ hdfs dfs -mkdir /user/wgiron2/Covid19_Testing
-bash-4.2$ hdfs dfs -mkdir /user/wgiron2/tmp
mkdir: `/user/wgiron2/tmp': File exists
-bash-4.2$ hdfs dfs -ls
Found 4 items
drwx------   - wgiron2 hdfs          0 2023-
drwxr-xr-x   - wgiron2 hdfs          0 2023-
drwxr-xr-x   - wgiron2 hdfs          0 2023-
drwxr-xr-x   - wgiron2 hdfs          0 2023-


-bash-4.2$ hdfs dfs -put Covid19_Testing.csv
-bash-4.2$
-bash-4.2$ hdfs dfs -ls Covid19_Testing/
Found 1 items
-rw-r--r--   3 wgiron2 hdfs    8739096 2023-
Testing.csv
-bash-4.2$
```

Confirm successful download using `ls`.

Unzip the file using `unzip Fire_Department_Calls_for_Service.zip`.

Unzip the file using ‘Covid19_Testing’



```
8.4M    Covid19_Testing.csv
-bash-4.2$ hdfs dfs -mkdir /user/wgiron2/Covid19_Testing
-bash-4.2$ hdfs dfs -mkdir /user/wgiron2/tmp
mkdir: `/user/wgiron2/tmp': File exists
-bash-4.2$ hdfs dfs -ls
Found 4 items
drwx------   - wgiron2 hdfs          0 2023-11-30 05:05 .Trash
drwxr-xr-x   - wgiron2 hdfs          0 2023-11-21 07:48 .hiveJars
drwxr-xr-x   - wgiron2 hdfs          0 2023-11-30 05:35 Covid19_Testing
drwxr-xr-x   - wgiron2 hdfs          0 2023-11-30 02:34 tmp


-bash-4.2$ hdfs dfs -put Covid19_Testing.csv /user/wgiron2/Covid19_Testing
-bash-4.2$
-bash-4.2$ hdfs dfs -ls Covid19_Testing/
Found 1 items
-rw-r--r--   3 wgiron2 hdfs    8739096 2023-11-30 05:38 Covid19_Testing/Covid19_
Testing.csv
-bash-4.2$
```

Load data set to distribute file system on Hadoop

Create directories in HDFS using `hdfs dfs -mkdir`.

Upload the CSV file to HDFS using `hdfs dfs -put`.

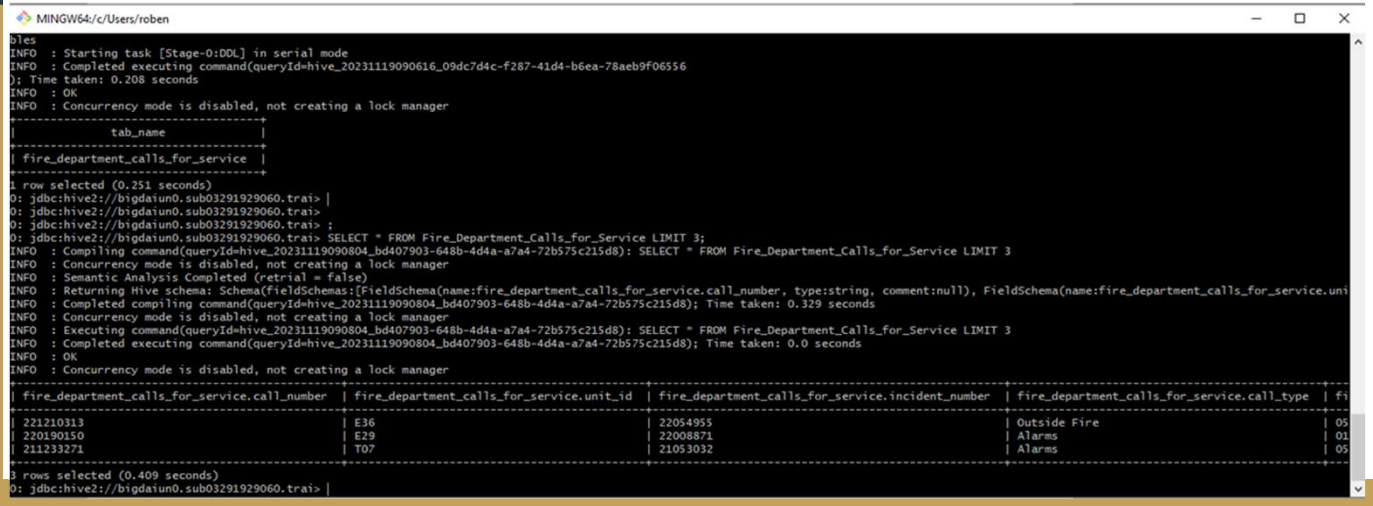
Verify the upload with `hdfs dfs -ls`.

```
-bash-4.2$ hdfs dfs -mkdir /user/wgiron2/tmp
-bash-4.2$ hdfs dfs -ls
Found 3 items
drwx------   - wgiron2 hdfs          0 2023-11-19 06:55 .Trash
drwxr-xr-x   - wgiron2 hdfs          0 2023-11-19 07:39 Fire_Department_Calls_for_Service
drwxr-xr-x   - wgiron2 hdfs          0 2023-11-19 07:39 tmp
-bash-4.2$ hdfs dfs -put Fire_Department_Calls_for_Service.csv /user/wgiron2/Fire_Department_Calls_for_Service
-bash-4.2$ hdfs dfs -ls Fire_Department_Calls_for_Service/
Found 1 items
-rw-r--r--   3 wgiron2 hdfs 2403931416 2023-11-19 07:39 Fire_Department_Calls_for_Service/Fire_Department_Calls_for_Service.csv
-bash-4.2$
```

Establishing Database and Define Tables through Beeline

```
CREATE EXTERNAL TABLE Fire_Department_Calls_for_Service (
Call_Number STRING, Unit_ID STRING, Incident_Number STRING, Call_Type STRING,
Call_Date STRING,Watch_Date STRING,Received_DtTm STRING,Entry_DtTm STRING,Dispatch_DtTm
STRING,
Response_DtTm STRING,On_Scene_DtTm STRING,Transport_DtTm STRING,Hospital_DtTm STRING,
 Call_Final_Disposition STRING,Available_DtTm STRING,Address STRING,City
STRING,Zipcode_of_Incident STRING,Battalion STRING,
Station_Area STRING,Box STRING,Original_Priority STRING,Priority STRING,Final_Priority STRING,
ALS_Unit BOOLEAN,Call_Type_Group STRING,Number_of_Alarms INT,Unit_Type STRING,
Unit_sequence_in_call_dispatch INT,Fire_Prevention_District STRING,
Supervisor_District STRING,Neighborhoods_Analysis_Boundaries STRING,RowID STRING,case_location
STRING)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ","
LOCATION "/user/username/Fire_Department_Calls_for_Service"
TBLPROPERTIES ('skip.header.line.count' = '1');
```

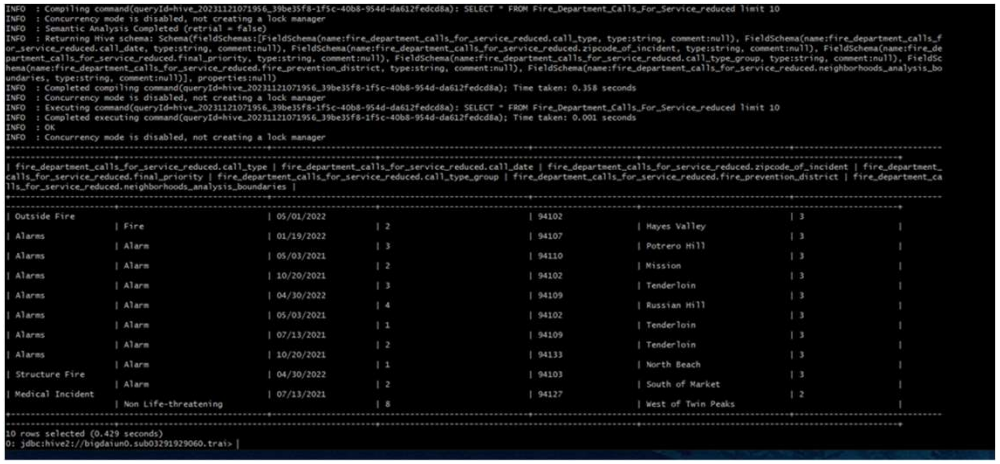
- Initiate Beeline and create a new database.
- Use the database and create an external table to visualize the data.
- Display the first three rows with `SELECT * FROM Fire_Department_Calls_for_Service LIMIT 3`.




```
DROP TABLE IF EXISTS Covid19_Testing;
CREATE EXTERNAL TABLE Covid19_Testing (
   specimen_collection_date STRING,area_type
STRING,
   id STRING,acs_population STRING,new_Test
STRING,
   new_positive_tests STRING,new_negative_tests
STRING,
   new_indeterminate_tests STRING,cumulative_tests
STRING,
   cumulative_positive_tests
STRING,cumulative_negative_tests STRING,
   cumulative_indeterminate_tests STRING,
   cumulative_testing_rate STRING,
   data_as_of STRING,
   data_loaded_at STRING

)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ","
LOCATION "/user/wgiron2/Covid19_Testing"
TBLPROPERTIES ('skip.header.line.count' = '1');
```

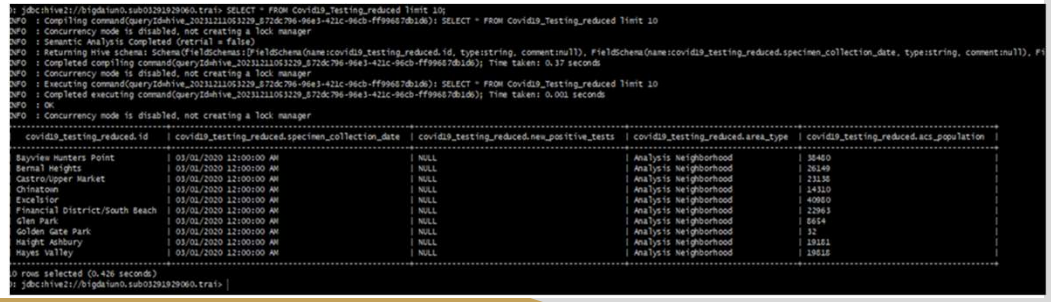
Cleaning Data using MapReduce

CREATE VIEW Fire_Department_Calls_For_Service_reduced AS SELECT Call_Type, Call_Date, Zipcode_of_Incident, Final_Priority, Call_Type_Group, Fire_Prevention_District, Neighborhoods_Analysis_Boundaries FROM Fire_Department_Calls_for_Service;



- Use Beeline to select the database.
- Create a view to select specific columns for analysis.
- Display the first ten rows with `SELECT * FROM Fire_Department_Calls_For_Service_reduced LIMIT 10`.
- Implement MapReduce to optimize and minimize the data.

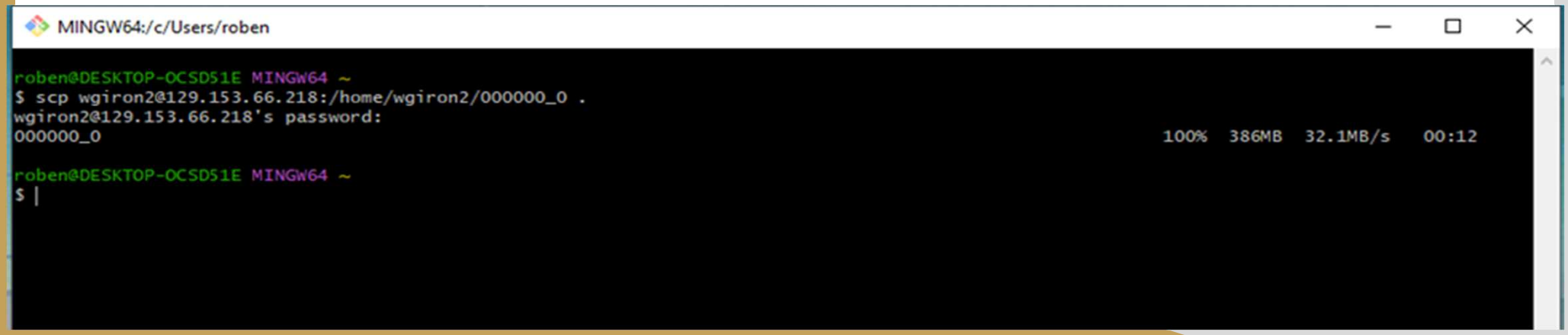
CREATE VIEW Covid19_Testing_reduced AS SELECT id, specimen_collection_date, new_positive_tests, area_type, acs_population FROM Covid19_Testing;



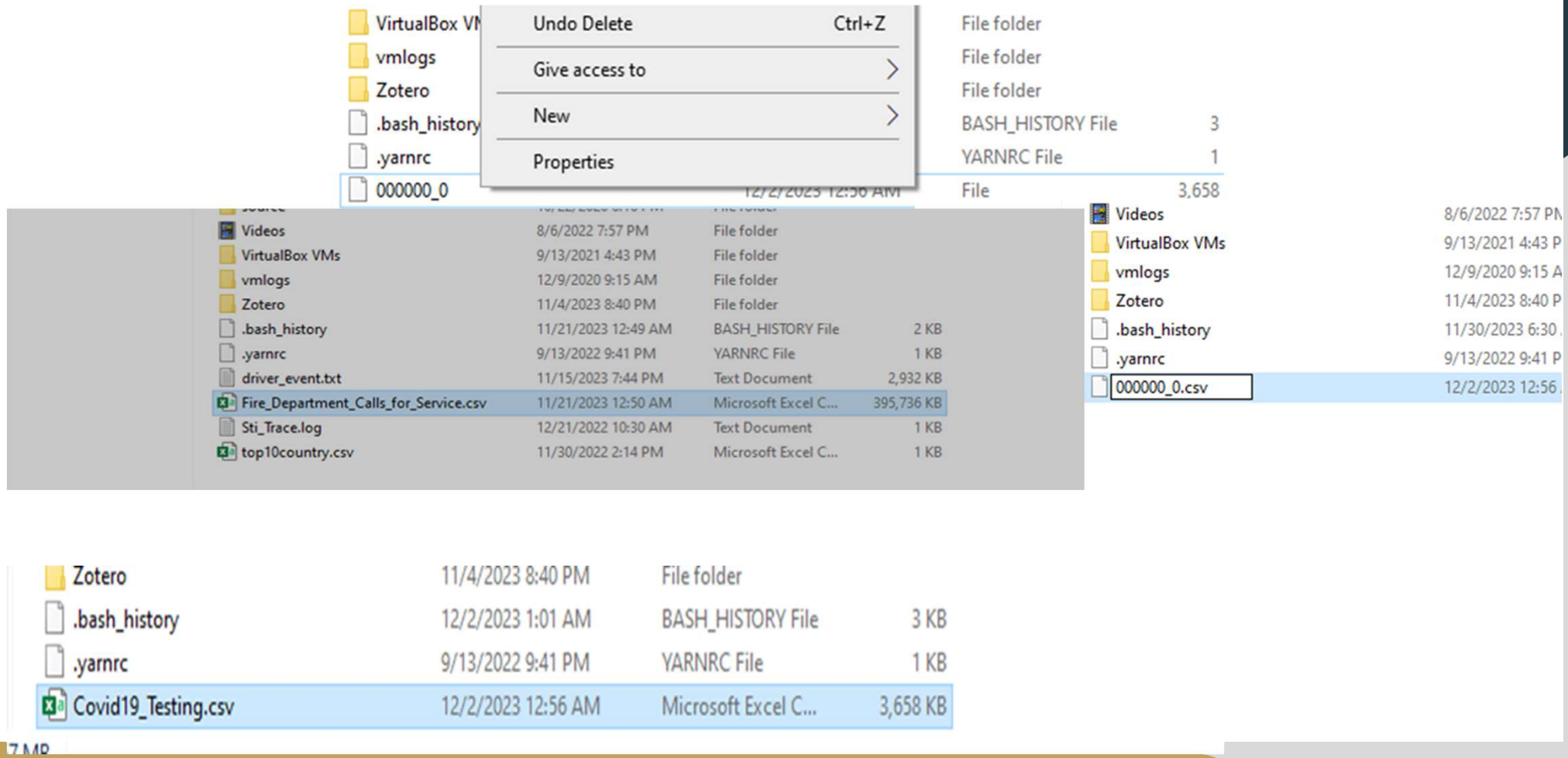
Downloading New Cleaned Data

```
 scp wgiron2@129.153.66.218:/home/wgiron2/000000_0 .
giron2@129.153.66.218's password:
00000_0                                       100% 3658KB    5.6MB/s   00:00

oben@DESKTOP-OCSD51E MINGW64 ~
```

Download the cleaned dataset using `scp`.

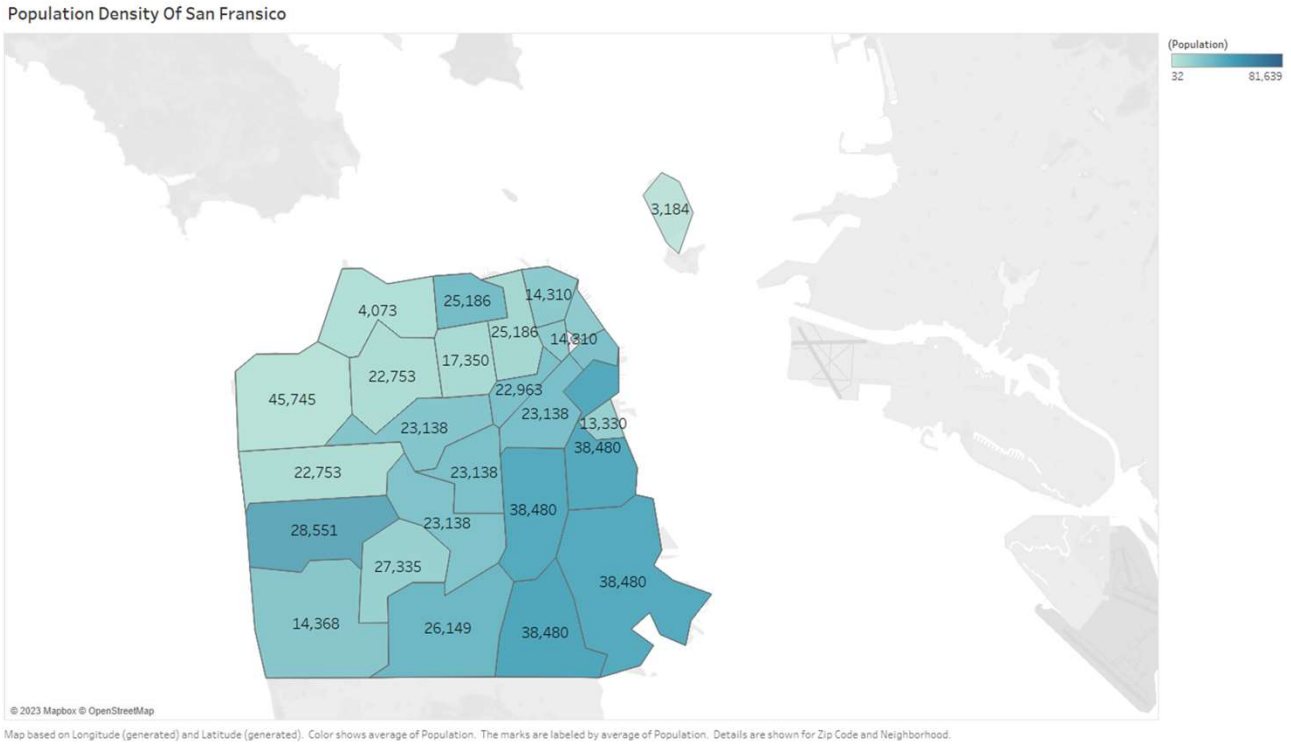


Renaming the files



Analysis and Visualization (Symbol Map)

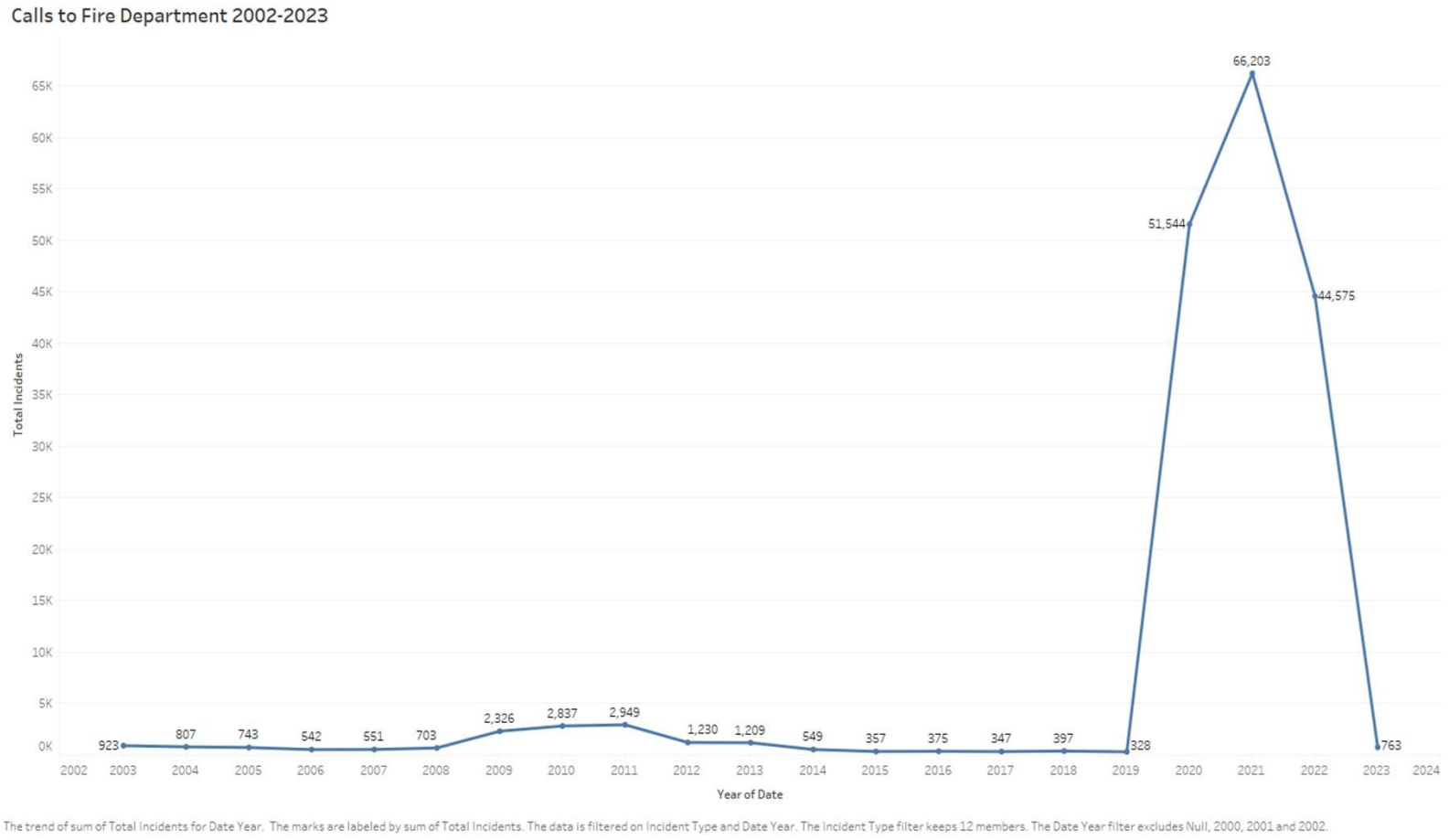
- **The Population of San Francisco as of 2023.**
- **Separated by zip codes**
- **The darker the color the more people**



Map based on Longitude (generated) and Latitude (generated). Color shows average of Population. The marks are labeled by average of Population. Details are shown for Zip Code and Neighborhood.

Analysis and Visualization (Line Chart)

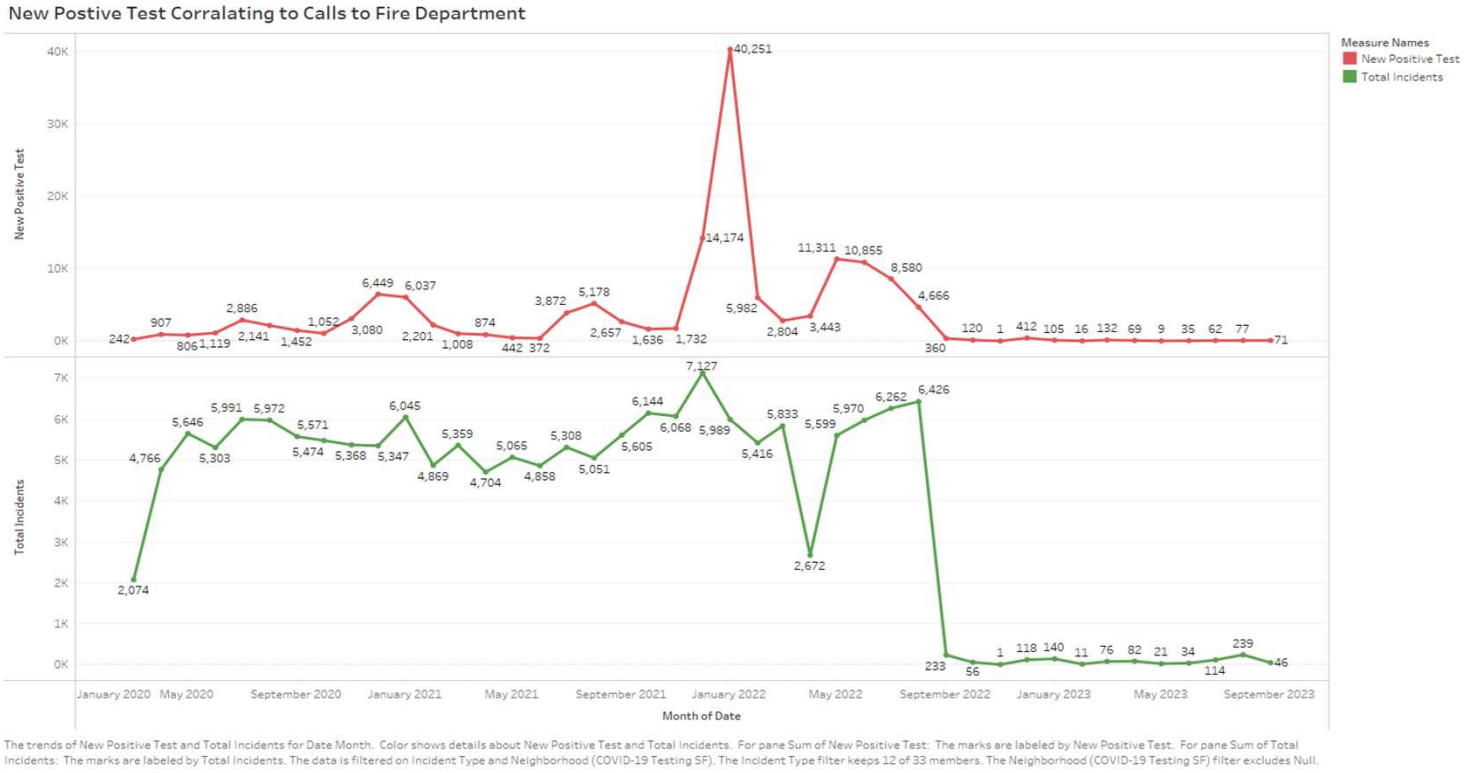
- **Fire related Calls Increased during the Lockdown**
- **Fire related calls were reduced drastically towards the end of 2022.**



Analysis and Visualization (Line Chart)

New Positive Test VS Calls to fire Department 2020-2023

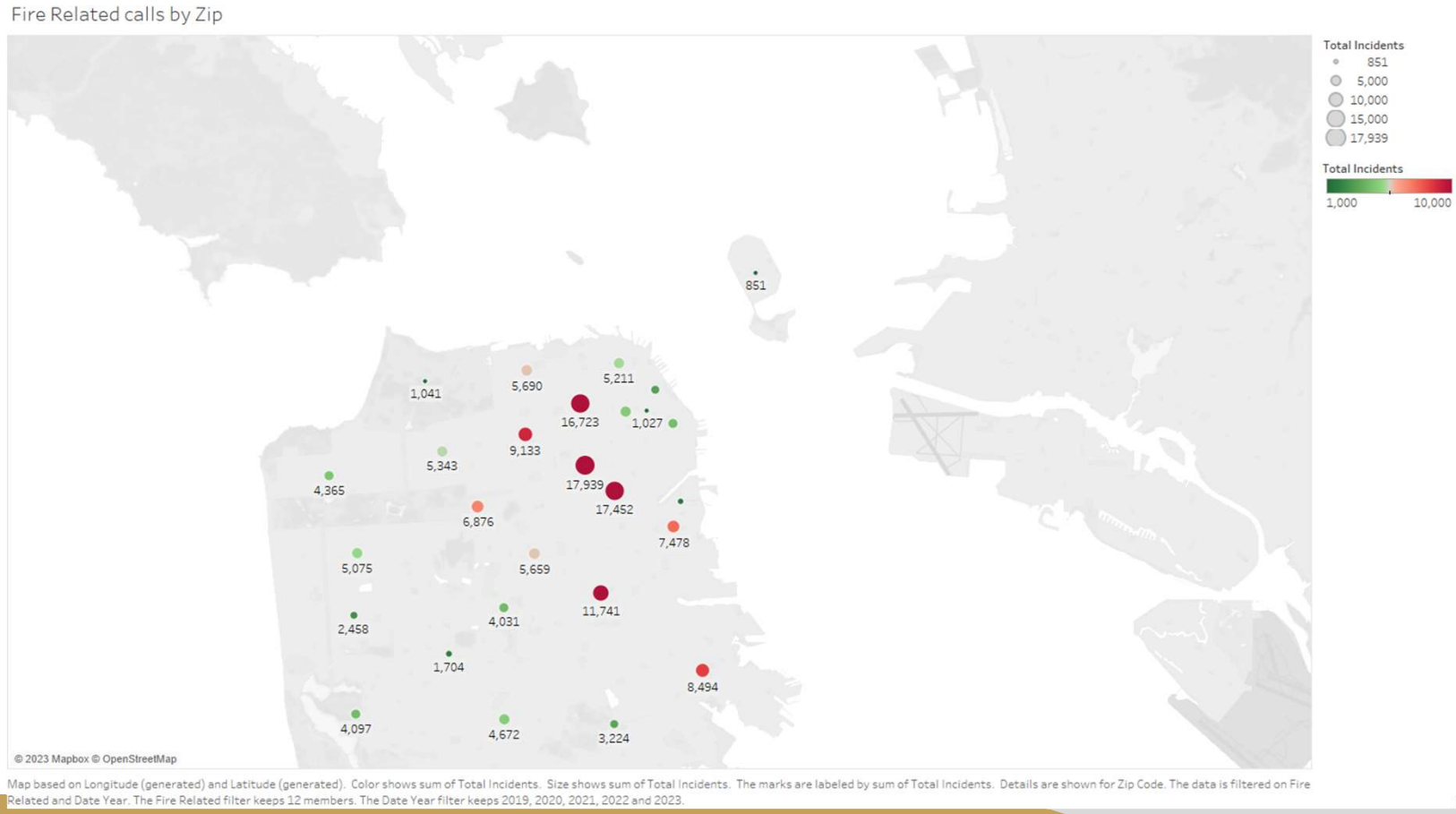
Spike in Fires when Covid Lockdowns began (Mar 2020)



The trends of New Positive Test and Total Incidents for Date Month. Color shows details about New Positive Test and Total Incidents. For pane Sum of New Positive Test: The marks are labeled by New Positive Test. For pane Sum of Total Incidents: The marks are labeled by Total Incidents. The data is filtered on Incident Type and Neighborhood (COVID-19 Testing SF). The Incident Type filter keeps 12 of 33 members. The Neighborhood (COVID-19 Testing SF) filter excludes Null.

Analysis and Visualization (Symbol Map)

This Map Show the Areas most impacted by fires during does covid years

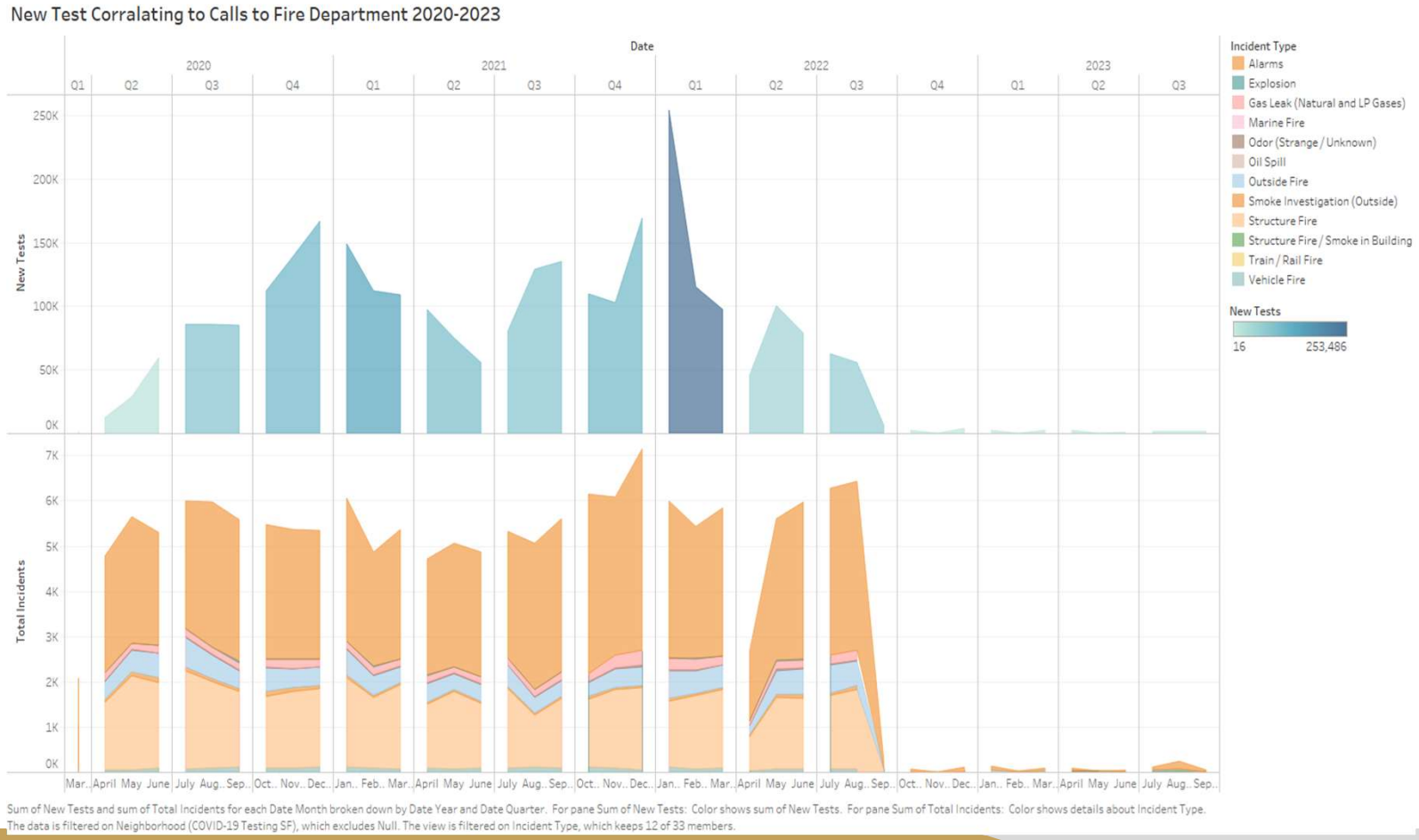


Map based on Longitude (generated) and Latitude (generated). Color shows sum of Total Incidents. Size shows sum of Total Incidents. The marks are labeled by sum of Total Incidents. Details are shown for Zip Code. The data is filtered on Fire Related and Date Year. The Fire Related filter keeps 12 members. The Date Year filter keeps 2019, 2020, 2021, 2022 and 2023.

Analysis and Visualization (Area Chart)

New Covid Test Correlating with Calls to Fire department

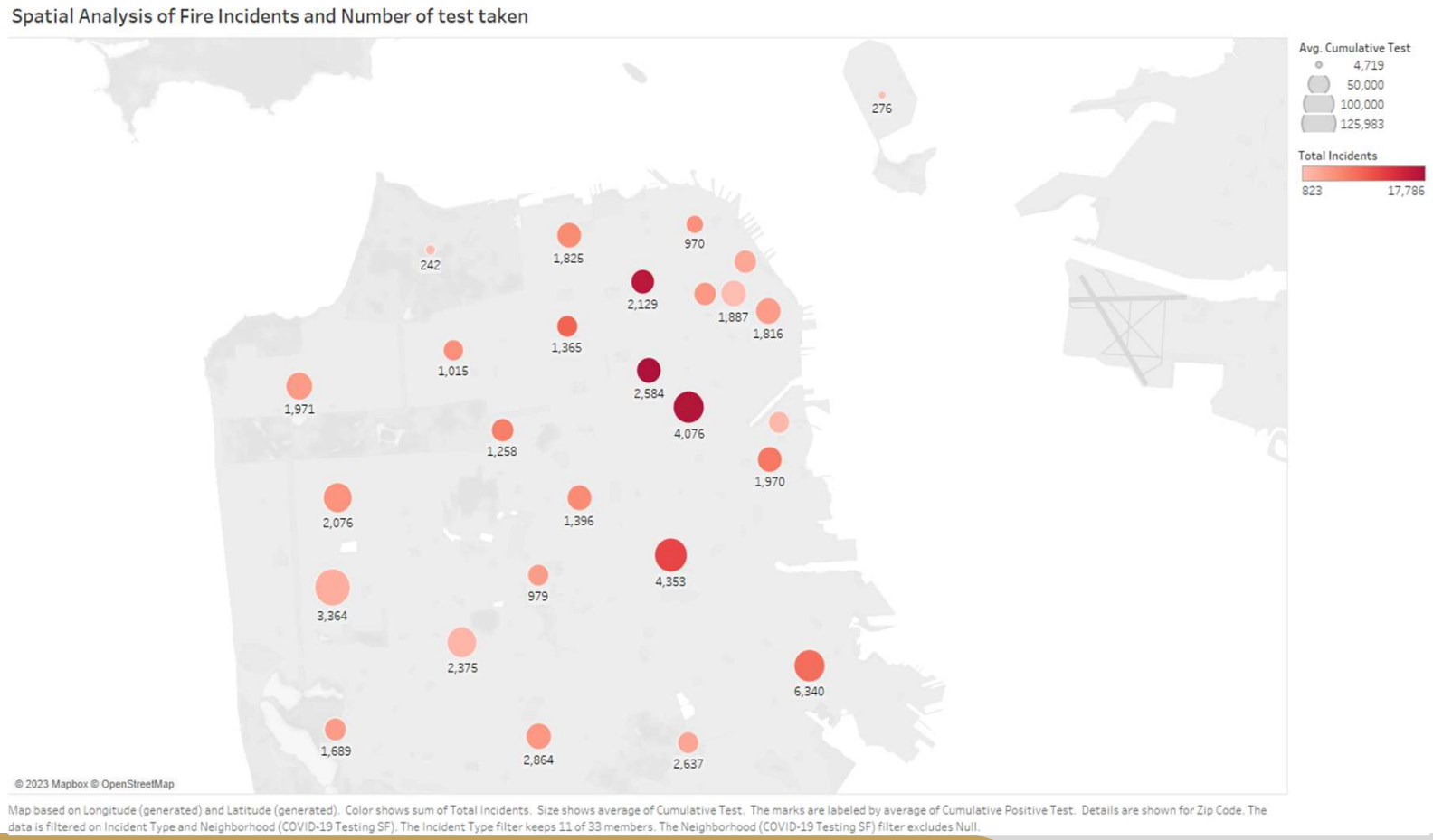
From this chart we can see a pretty decent correlation, between fire calls and new covid test.



Analysis and Visualization (Symbol Map)

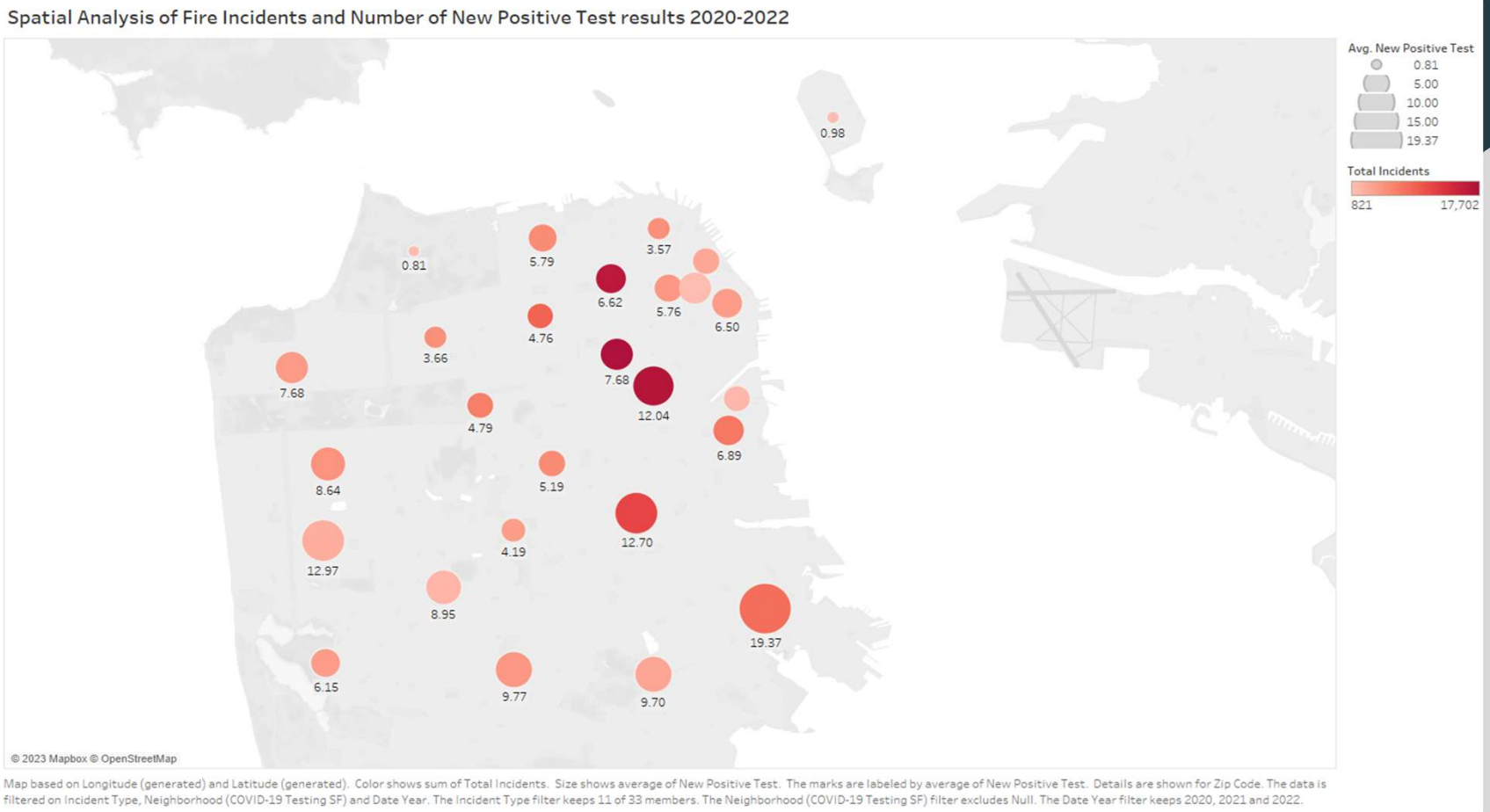
Show Fire Incidents in relation to number covid test taken.

No real Correlation between test taken and amount of people taking test.



Analysis and Visualization (Symbol Map)

We also saw no real correlation between people actually getting sick and where fire calls occurred.



Map based on Longitude (generated) and Latitude (generated). Color shows sum of Total Incidents. Size shows average of New Positive Test. The marks are labeled by average of New Positive Test. Details are shown for Zip Code. The data is filtered on Incident Type, Neighborhood (COVID-19 Testing SF) and Date Year. The Incident Type filter keeps 11 of 33 members. The Neighborhood (COVID-19 Testing SF) filter excludes Null. The Date Year filter keeps 2020, 2021 and 2022.

Summary & Conclusion

- San Francisco Fire Department should create fire awareness programs for the public.

- Whenever there is a protests, the affluent neighborhoods are targeted the most. Retail stores should have fire preventative measures in their stores.

- Residential homes and buildings should have at least one fire extinguisher.

-

Github URLS