

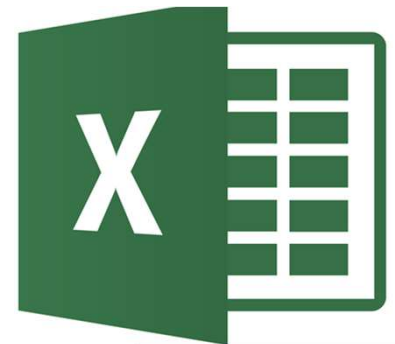


San Francisco Fire Calls & COVID

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.



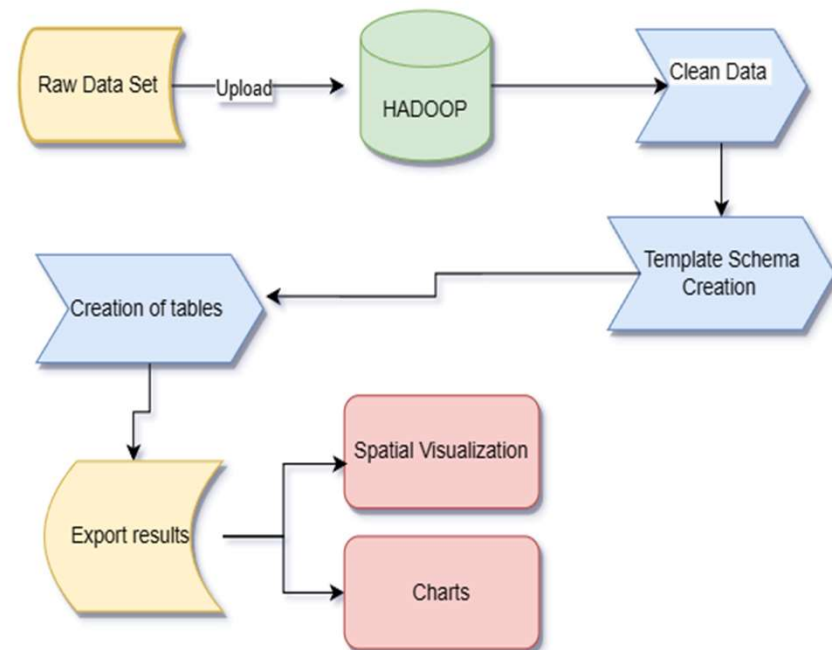
Dataset Specifications

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

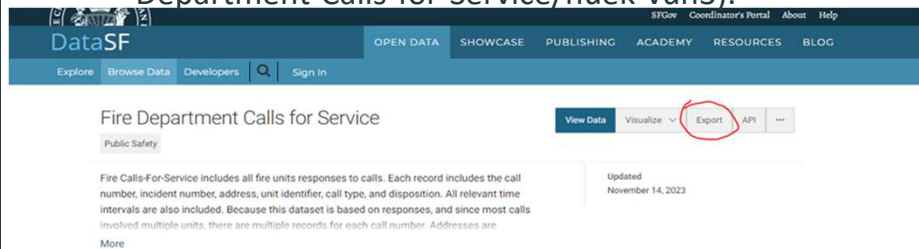
Implementation Flowchart

- Raw Data (San Francisco Gov Site)
- Uploaded using File transfer to Hadoop
- Cleaned in Hadoop (Map 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).



The screenshot shows the DataSF website interface. The top navigation bar includes links for Explore, Browse Data, Developers, and Sign In. The main content area displays the dataset 'Fire Department Calls for Service' under the 'Public Safety' category. A red circle highlights the 'Export' button in the top right corner of the dataset page.

About this Dataset

Updated: **November 14, 2023**

Data Last Updated: November 14, 2023 | Metadata Last Updated: November 14, 2023

Date Created: December 17, 2015

Views: **81.5K** | Downloads: **31K**

Data Provided by: (none) | Dataset Owner: OpenData

[Contact Dataset Owner](#)

Department Metrics

Publishing Department	Fire Department
-----------------------	-----------------

Detailed Descriptive

Geographic unit	Street address
-----------------	----------------

Publishing Details

Publishing frequency	Daily
Data change frequency	Daily

Attachments

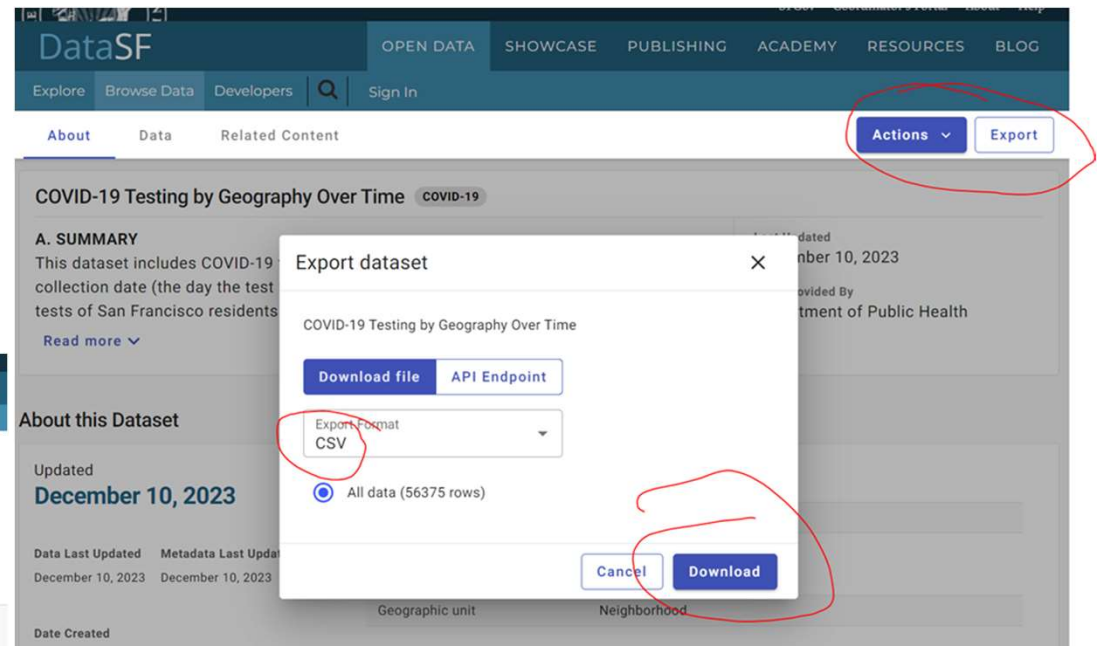
1	FIR-0002_DataDictionary_fire-calls-for-service.xlsx
---	---

Topics

Category	Public Safety
Tags	calls, fire, 911, medical, emergency calls

[Show More](#)

What's in this Dataset? Fire Department calls of service



The screenshot shows the DataSF website interface. The top navigation bar includes links for Explore, Browse Data, Developers, and Sign In. The main content area displays the dataset 'COVID-19 Testing by Geography Over Time' under the 'COVID-19' category. A red circle highlights the 'Export' button in the top right corner of the dataset page.

Export dataset

COVID-19 Testing by Geography Over Time

[Download file](#) [API Endpoint](#)

Export format: **CSV**

☒ All data (56375 rows)

[Cancel](#) [Download](#)

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

Uploading Dataset to Google Drive

Name ↑	Owner	Last modified ▼
☰ Fire_Department_Calls_for_Service .zip 👤	 me	

- 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.

☰ Covid19_Testing.zip 👤	 me	9:19 PM	⋮
-------------------------	--	---------	---

Extracting File to Oracle Linux Server

```
MINGW64:/c/Users/roben
roben@DESKTOP-OCSD51E MINGW64 ~
$ ssh wgiron2@129.153.66.218
wgiron2@129.153.66.218's password:
Last login: Sun Nov 19 08:25:02 2023 from cpe-76-95-39-232.socal.res.rr.com
-bash-4.2$ ls
Fire_Department_Calls_for_Service  Fire_Department_Calls_for_Service.csv
-bash-4.2$ wget "https://docs.google.com/uc?export=download&confirm=$(wget --qui
et --save-cookies /tmp/cookies.txt --keep-session-cookies --no-check-certificate
'https://docs.google.com/uc?export=download&id=1inFQ6rWMOyJ54KjGDWxEfNe9cyTnqCN
F' -O- | sed -rn 's/.*confirm=([0-9A-Za-z_]+).*/\1\n/p')&id=1inFQ6rWMOyJ54KjGDWx
EfNe9cyTnqCNF" -O Fire_Department_Calls_for_Service.zip && rm -rf /tmp/cookies.
txt
--2023-11-19 08:42:55-- https://docs.google.com/uc?export=download&confirm=t&id
=1inFQ6rWMOyJ54KjGDWxEfNe9cyTnqCNF
Resolving docs.google.com (docs.google.com)... 142.250.68.46, 2607:f8b0:4007:801
::200e
Connecting to docs.google.com (docs.google.com)|142.250.68.46|:443... connected.
HTTP request sent, awaiting response... 303 See Other
Location: https://doc-0c-1k-docs.googleusercontent.com/docs/securesc/ha0ro937gcu
c7l7deffksulhg5h7mbp1/ql0gj6c067lj4vio6q9k3qro2chfue0/1700383350000/01302242480
622693155/%1inFQ6rWMOyJ54KjGDWxEfNe9cyTnqCNF?e=download&uuiid=3a66e516-000d-4063
-a8d7-9af575c3f5c7 [following]
Warning: wildcards not supported in HTTP.
--2023-11-19 08:42:55-- https://doc-0c-1k-docs.googleusercontent.com/docs/secu
resc/ha0ro937gcu7l7deffksulhg5h7mbp1/ql0gj6c067lj4vio6q9k3qro2chfue0/1700383350
000/01302242480622693155/%1inFQ6rWMOyJ54KjGDWxEfNe9cyTnqCNF?e=download&uuiid=3a6
6e516-000d-4063-a8d7-9af575c3f5c7
Resolving doc-0c-1k-docs.googleusercontent.com (doc-0c-1k-docs.googleusercontent
.com)... 142.250.68.1, 2607:f8b0:4007:819::2001
Connecting to doc-0c-1k-docs.googleusercontent.com (doc-0c-1k-docs.googleusercon
tent.com)|142.250.68.1|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 542715364 (518M) [application/x-zip-compressed]
Saving to: 'Fire_Department_Calls_for_Service'

100%[=====] 542,715,364 128MB/s in 5.0s
```

- 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`

```
MINGW64:/c/Users/roben
roben@DESKTOP-OCSD51E MINGW64 ~
$ ssh wgiron2@129.153.66.218
wgiron2@129.153.66.218's password:
Permission denied, please try again.
wgiron2@129.153.66.218's password:
Last failed login: Sun Nov 19 09:00:07 GMT 2023 from cpe-76-95-39-232.socal.res.rr.com on ssh:notty
There was 1 failed login attempt since the last successful login.
Last login: Sun Nov 19 08:42:30 2023 from cpe-76-95-39-232.socal.res.rr.com
-bash-4.2$ ls
Fire_Department_Calls_for_Service.zip
-bash-4.2$ |
```

```

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$ |
-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$ |

```

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`.

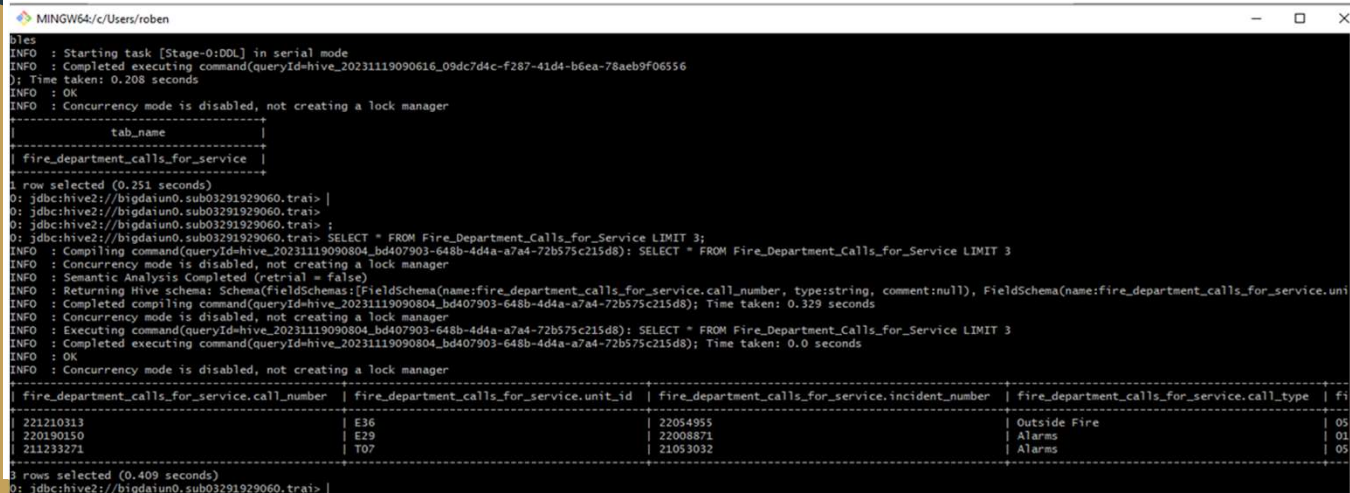
Verify the upload with `hdfs dfs -ls`.

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_DtM STRING, Entry_DtM STRING, Dispatch_DtM  
  STRING,  
  Response_DtM STRING, On_Scene_DtM STRING, Transport_DtM STRING, Hospital_DtM STRING,  
  Call_Final_Disposition STRING, Available_DtM 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');
```



```
MINGW64/c:/Users/roben  
btes  
INFO : Starting task [Stage-0:DDL] in serial mode  
INFO : Completed executing command(queryId=hive_20231119090616_09dc7d4c-f287-41d4-b6ea-78aeb9f06556  
); Time taken: 0.208 seconds  
INFO : OK  
INFO : Concurrency mode is disabled, not creating a lock manager  
-----  
| tab_name |  
-----  
| fire_department_calls_for_service |  
-----  
1 row selected (0.251 seconds)  
0: jdbc:hive2://bigdaiun0.sub03291929060.traib>  
0: jdbc:hive2://bigdaiun0.sub03291929060.traib>  
0: jdbc:hive2://bigdaiun0.sub03291929060.traib>  
0: jdbc:hive2://bigdaiun0.sub03291929060.traib> SELECT * FROM Fire_Department_Calls_for_Service LIMIT 3;  
INFO : Compiling command(queryId=hive_20231119090804_bd407903-648b-4d4a-a7a4-72b575c215d8): SELECT * FROM Fire_Department_Calls_for_Service LIMIT 3  
INFO : Concurrency mode is disabled, not creating a lock manager  
INFO : Semantic Analysis Completed (retrial = false)  
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:fire_department_calls_for_service.call_number, type:string, comment:null), FieldSchema(name:fire_department_calls_for_service.unit_id, type:string, comment:null), FieldSchema(name:fire_department_calls_for_service.incident_number, type:string, comment:null), FieldSchema(name:fire_department_calls_for_service.call_type, type:string, comment:null), FieldSchema(name:fire_department_calls_for_service.call_date, type:string, comment:null), FieldSchema(name:fire_department_calls_for_service.watch_date, type:string, comment:null), FieldSchema(name:fire_department_calls_for_service.received_dt_m, type:string, comment:null), FieldSchema(name:fire_department_calls_for_service.entry_dt_m, type:string, comment:null), FieldSchema(name:fire_department_calls_for_service.dispatch_dt_m, type:string, comment:null), FieldSchema(name:fire_department_calls_for_service.response_dt_m, type:string, comment:null), FieldSchema(name:fire_department_calls_for_service.on_scene_dt_m, type:string, comment:null), FieldSchema(name:fire_department_calls_for_service.transport_dt_m, type:string, comment:null), FieldSchema(name:fire_department_calls_for_service.hospital_dt_m, type:string, comment:null), FieldSchema(name:fire_department_calls_for_service.call_final_disposition, type:string, comment:null), FieldSchema(name:fire_department_calls_for_service.available_dt_m, type:string, comment:null), FieldSchema(name:fire_department_calls_for_service.address, type:string, comment:null), FieldSchema(name:fire_department_calls_for_service.city, type:string, comment:null), FieldSchema(name:fire_department_calls_for_service.zipcode_of_incident, type:string, comment:null), FieldSchema(name:fire_department_calls_for_service.battalion, type:string, comment:null), FieldSchema(name:fire_department_calls_for_service.station_area, type:string, comment:null), FieldSchema(name:fire_department_calls_for_service.box, type:string, comment:null), FieldSchema(name:fire_department_calls_for_service.original_priority, type:string, comment:null), FieldSchema(name:fire_department_calls_for_service.priority, type:string, comment:null), FieldSchema(name:fire_department_calls_for_service.final_priority, type:string, comment:null), FieldSchema(name:fire_department_calls_for_service.als_unit, type:boolean, comment:null), FieldSchema(name:fire_department_calls_for_service.call_type_group, type:string, comment:null), FieldSchema(name:fire_department_calls_for_service.number_of_alarms, type:int, comment:null), FieldSchema(name:fire_department_calls_for_service.unit_type, type:string, comment:null), FieldSchema(name:fire_department_calls_for_service.unit_sequence_in_call_dispatch, type:int, comment:null), FieldSchema(name:fire_department_calls_for_service.fire_prevention_district, type:string, comment:null), FieldSchema(name:fire_department_calls_for_service.supervisor_district, type:string, comment:null), FieldSchema(name:fire_department_calls_for_service.neighborhoods_analysis_boundaries, type:string, comment:null), FieldSchema(name:fire_department_calls_for_service.rowid, type:string, comment:null), FieldSchema(name:fire_department_calls_for_service.case_location, type:string, comment:null)], partitionKeys:[], name:fire_department_calls_for_service, comment:null)  
INFO : Completed compiling command(queryId=hive_20231119090804_bd407903-648b-4d4a-a7a4-72b575c215d8); Time taken: 0.329 seconds  
INFO : Concurrency mode is disabled, not creating a lock manager  
INFO : Executing command(queryId=hive_20231119090804_bd407903-648b-4d4a-a7a4-72b575c215d8): SELECT * FROM Fire_Department_Calls_for_Service LIMIT 3  
INFO : Completed executing command(queryId=hive_20231119090804_bd407903-648b-4d4a-a7a4-72b575c215d8); Time taken: 0.0 seconds  
INFO : OK  
INFO : Concurrency mode is disabled, not creating a lock manager  
-----  
| Fire_department_calls_for_service.call_number | Fire_department_calls_for_service.unit_id | Fire_department_calls_for_service.incident_number | Fire_department_calls_for_service.call_type | fi  
-----  
| 221210313 | E36 | 22054955 | Outside Fire | 05  
| 220190150 | E29 | 22008871 | Alarms | 01  
| 211233271 | T07 | 21053032 | Alarms | 05  
-----  
3 rows selected (0.409 seconds)  
0: jdbc:hive2://bigdaiun0.sub03291929060.traib>
```


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.

```
INFO : Compiling command(queryId=hive_20231121071956_39b35f8-1f5c-4088-954d-da612fedcd8a): SELECT * FROM Fire_Department_Calls_For_Service_reduced limit 10
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Semantic Analysis Completed (retry=0)
INFO : Returning hive schema: Schema(fieldsSchemas:[fieldSchema(name=fire_department_calls_for_service_reduced.call_type, type=string, comment=null), fieldSchema(name=fire_department_calls_for_service_reduced.call_date, type=string, comment=null), fieldSchema(name=fire_department_calls_for_service_reduced.zipcode_of_incident, type=string, comment=null), fieldSchema(name=fire_department_calls_for_service_reduced.final_priority, type=string, comment=null), fieldSchema(name=fire_department_calls_for_service_reduced.call_type_group, type=string, comment=null), fieldSchema(name=fire_department_calls_for_service_reduced.fire_prevention_district, type=string, comment=null), fieldSchema(name=fire_department_calls_for_service_reduced.neighborhoods_analysis_boundaries, type=string, comment=null)], properties=null)
INFO : Completed compiling command(queryId=hive_20231121071956_39b35f8-1f5c-4088-954d-da612fedcd8a); Time taken: 0.358 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20231121071956_39b35f8-1f5c-4088-954d-da612fedcd8a): SELECT * FROM Fire_Department_Calls_For_Service_reduced limit 10
INFO : Completed executing command(queryId=hive_20231121071956_39b35f8-1f5c-4088-954d-da612fedcd8a); Time taken: 0.001 seconds
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager

+-----+-----+-----+-----+-----+-----+
| fire_department_calls_for_service_reduced.call_type | fire_department_calls_for_service_reduced.call_date | fire_department_calls_for_service_reduced.zipcode_of_incident | fire_department_calls_for_service_reduced.final_priority | fire_department_calls_for_service_reduced.call_type_group | fire_department_calls_for_service_reduced.fire_prevention_district | fire_department_calls_for_service_reduced.neighborhoods_analysis_boundaries |
+-----+-----+-----+-----+-----+-----+
| Outside Fire | 05/01/2022 | 94102 | 3 |  |  |  | |
| Alarms | Fire | 01/19/2022 | 2 | 94107 | Hayes Valley | 3 |  |
| Alarms | Alarm | 06/03/2021 | 3 | 94110 | Potrero Hill | 3 |  |
| Alarms | Alarm | 10/20/2021 | 2 | 94102 | Mission | 3 |  |
| Alarms | Alarm | 04/30/2022 | 3 | 94109 | Tenderloin | 3 |  |
| Alarms | Alarm | 05/03/2021 | 4 | 94102 | Russian Hill | 3 |  |
| Alarms | Alarm | 07/13/2021 | 1 | 94109 | Tenderloin | 3 |  |
| Alarms | Alarm | 10/20/2021 | 2 | 94133 | Tenderloin | 3 |  |
| Structure Fire | Alarm | 04/30/2022 | 1 | 94103 | North Beach | 3 |  |
| Medical Incident | Non Life-threatening | 07/13/2021 | 2 | 94127 | South of Market | 2 |  |
|  |  |  |  |  | West of Twin Peaks | 2 |  |
+-----+-----+-----+-----+-----+-----+
10 rows selected (0.429 seconds)
jdbcTemplate://bigdataumb-sub03291929060-train>
```

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

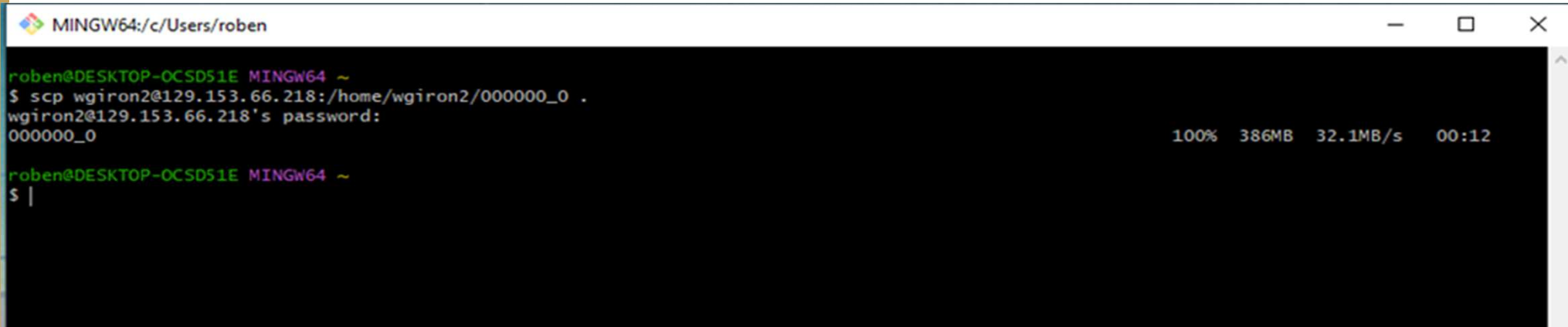
```
jdbcTemplate://bigdataumb-sub03291929060-train> SELECT * FROM Covid19_Testing_reduced limit 10
INFO : Compiling command(queryId=hive_20231121071956_39b35f8-1f5c-4088-954d-da612fedcd8a): SELECT * FROM Covid19_Testing_reduced limit 10
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Semantic Analysis Completed (retry=0)
INFO : Returning hive schema: Schema(fieldsSchemas:[fieldSchema(name=covid19_testing_reduced.id, type=string, comment=null), fieldSchema(name=covid19_testing_reduced.specimen_collection_date, type=string, comment=null), fieldSchema(name=covid19_testing_reduced.new_positive_tests, type=string, comment=null), fieldSchema(name=covid19_testing_reduced.area_type, type=string, comment=null), fieldSchema(name=covid19_testing_reduced.acs_population, type=string, comment=null)], properties=null)
INFO : Completed compiling command(queryId=hive_20231121071956_39b35f8-1f5c-4088-954d-da612fedcd8a); Time taken: 0.37 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20231121071956_39b35f8-1f5c-4088-954d-da612fedcd8a): SELECT * FROM Covid19_Testing_reduced limit 10
INFO : Completed executing command(queryId=hive_20231121071956_39b35f8-1f5c-4088-954d-da612fedcd8a); Time taken: 0.001 seconds
INFO : OK
INFO : Concurrency mode is disabled, not creating a lock manager

+-----+-----+-----+-----+-----+
| covid19_testing_reduced.id | covid19_testing_reduced.specimen_collection_date | covid19_testing_reduced.new_positive_tests | covid19_testing_reduced.area_type | covid19_testing_reduced.acs_population |
+-----+-----+-----+-----+-----+
| 35460 | 03/01/2020 12:00:00 AM | NULL | Analysis neighborhood | 35460 |
| 25149 | 03/01/2020 12:00:00 AM | NULL | Analysis neighborhood | 25149 |
| 23136 | 03/01/2020 12:00:00 AM | NULL | Analysis neighborhood | 23136 |
| 14110 | 03/01/2020 12:00:00 AM | NULL | Analysis neighborhood | 14110 |
| 40860 | 03/01/2020 12:00:00 AM | NULL | Analysis neighborhood | 40860 |
| 22961 | 03/01/2020 12:00:00 AM | NULL | Analysis neighborhood | 22961 |
| 8614 | 03/01/2020 12:00:00 AM | NULL | Analysis neighborhood | 8614 |
| 12 | 03/01/2020 12:00:00 AM | NULL | Analysis neighborhood | 12 |
| 18181 | 03/01/2020 12:00:00 AM | NULL | Analysis neighborhood | 18181 |
| 18115 | 03/01/2020 12:00:00 AM | NULL | Analysis neighborhood | 18115 |
+-----+-----+-----+-----+-----+
10 rows selected (0.426 seconds)
jdbcTemplate://bigdataumb-sub03291929060-train>
```

Downloading New Cleaned Data

```
scp wgiron2@129.153.66.218:/home/wgiron2/000000_0 .  
wgiron2@129.153.66.218's password:  
000000_0 100% 3658KB 5.6MB/s 00:00  
roben@DESKTOP-OCSD51E MINGW64 ~  
$ |
```

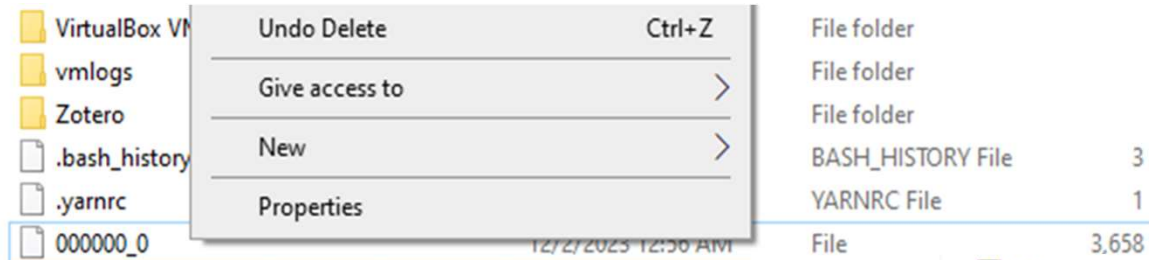
Download the cleaned dataset using `scp`.



The screenshot shows a terminal window titled "MINGW64: c:/Users/roben". The user "roben" is at the prompt "roben@DESKTOP-OCSD51E MINGW64 ~". They enter the command "scp wgiron2@129.153.66.218:/home/wgiron2/000000_0 .". The terminal shows the password prompt "wgiron2@129.153.66.218's password:" and then the file transfer progress: "000000_0 100% 386MB 32.1MB/s 00:12". The prompt returns to "roben@DESKTOP-OCSD51E MINGW64 ~" and the user enters a dollar sign "\$" followed by a vertical bar "|".

```
MINGW64: c:/Users/roben  
roben@DESKTOP-OCSD51E MINGW64 ~  
$ scp wgiron2@129.153.66.218:/home/wgiron2/000000_0 .  
wgiron2@129.153.66.218's password:  
000000_0 100% 386MB 32.1MB/s 00:12  
roben@DESKTOP-OCSD51E MINGW64 ~  
$ |
```

Renaming the files



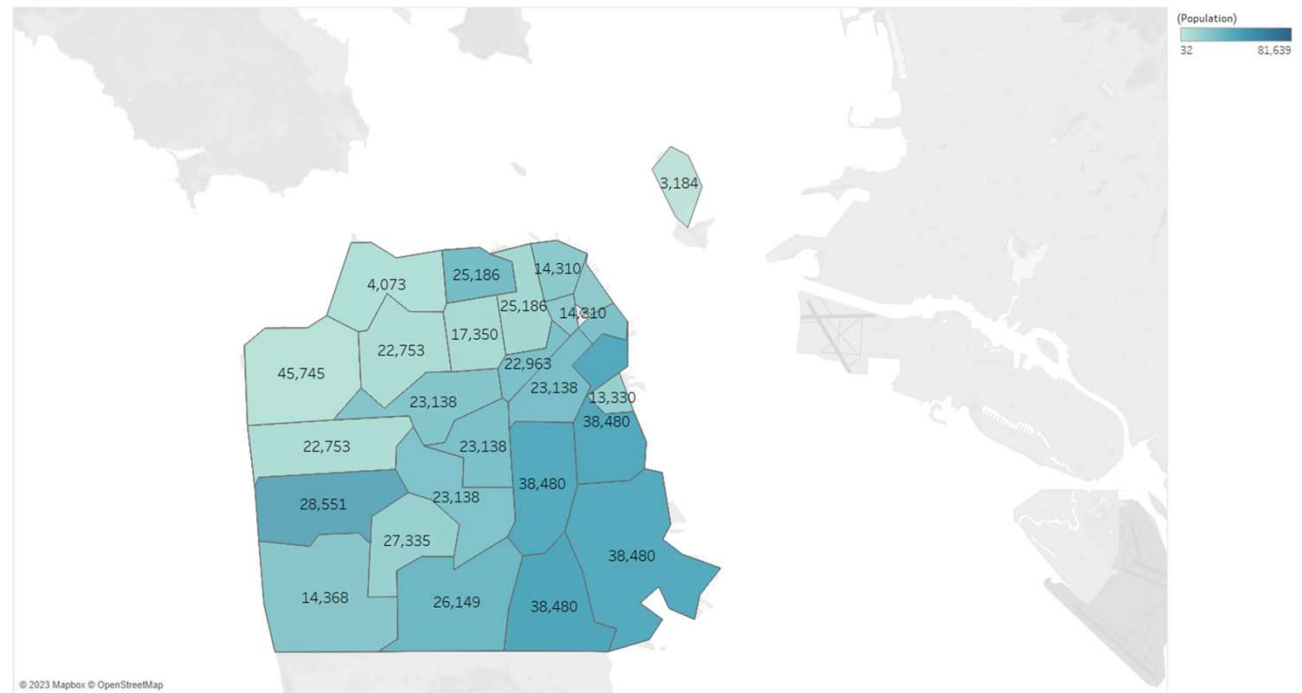
Videos	8/6/2022 7:57 PM	File folder		Videos	8/6/2022 7:57 PM
VirtualBox VMs	9/13/2021 4:43 PM	File folder		VirtualBox VMs	9/13/2021 4:43 PM
vmlogs	12/9/2020 9:15 AM	File folder		vmlogs	12/9/2020 9:15 AM
Zotero	11/4/2023 8:40 PM	File folder		Zotero	11/4/2023 8:40 PM
.bash_history	11/21/2023 12:49 AM	BASH_HISTORY File	2 KB	.bash_history	11/30/2023 6:30 PM
.yarnrc	9/13/2022 9:41 PM	YARNRC File	1 KB	.yarnrc	9/13/2022 9:41 PM
driver_event.txt	11/15/2023 7:44 PM	Text Document	2,932 KB	000000_0.csv	12/2/2023 12:56 PM
Fire_Department_Calls_for_Service.csv	11/21/2023 12:50 AM	Microsoft Excel C...	395,736 KB		
Sti_Trace.log	12/21/2022 10:30 AM	Text Document	1 KB		
top10country.csv	11/30/2022 2:14 PM	Microsoft Excel C...	1 KB		

Zotero	11/4/2023 8:40 PM	File folder	
.bash_history	12/2/2023 1:01 AM	BASH_HISTORY File	3 KB
.yarnrc	9/13/2022 9:41 PM	YARNRC File	1 KB
Covid19_Testing.csv	12/2/2023 12:56 AM	Microsoft Excel C...	3,658 KB

Analysis and Visualization (Symbol Map)

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

Population Density Of San Fransico

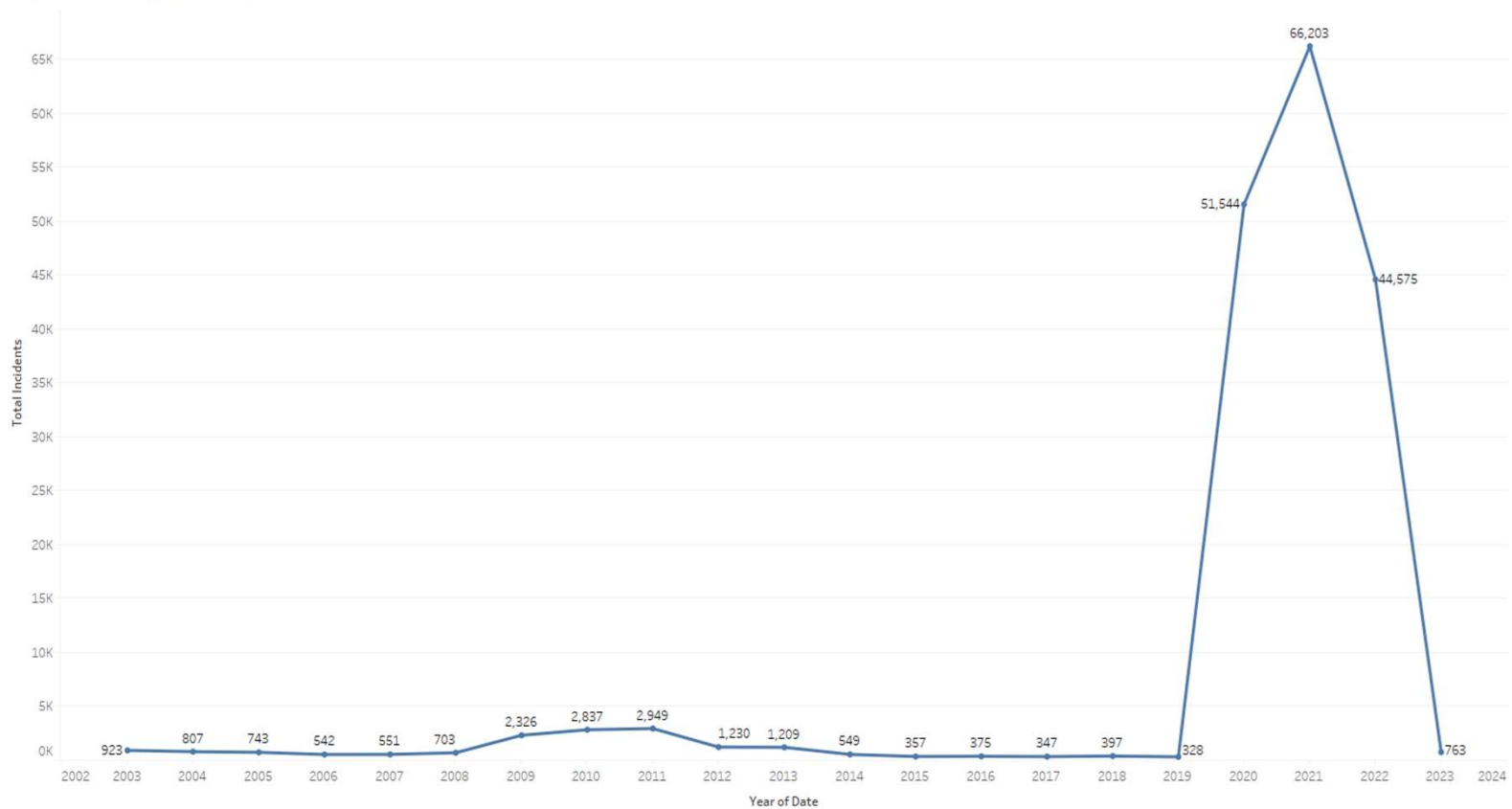


Map based on Longitude (generated) and Latitude (generated). Color shows average of Population. The marks are labeled by average of Population. Details are shown for Zip Code and Neighborhood.

Analysis and Visualization (Line Chart)

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

Calls to Fire Department 2002-2023



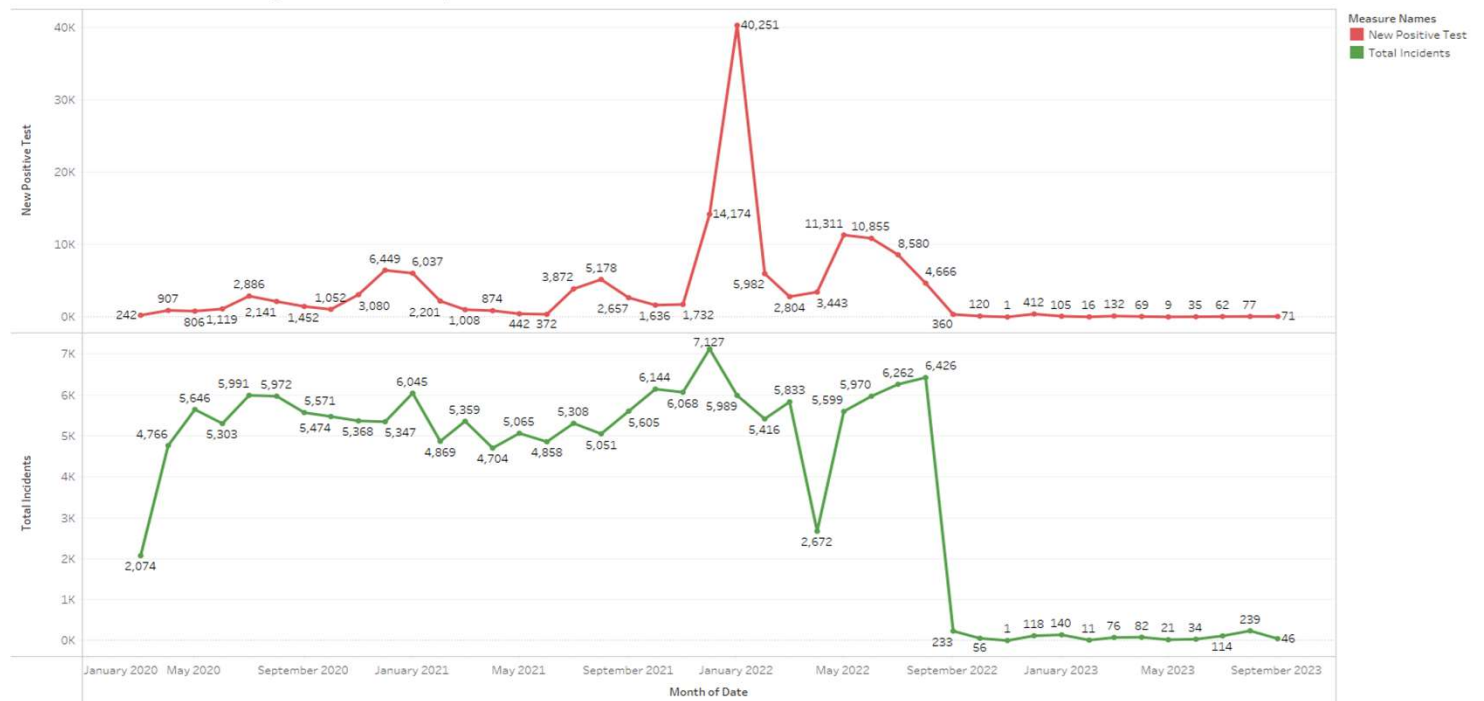
The trend of sum of Total Incidents for Date Year. The marks are labeled by sum of Total Incidents. The data is filtered on Incident Type and Date Year. The Incident Type filter keeps 12 members. The Date Year filter excludes Null, 2000, 2001 and 2002.

Analysis and Visualization (Line Chart)

**New Positive Test VS
Calls to fire
Department 2020-
2023**

**Spike in Fires when
Covid Lockdowns
began (Mar 2020)**

New Postive Test Corralating to Calls to Fire Department

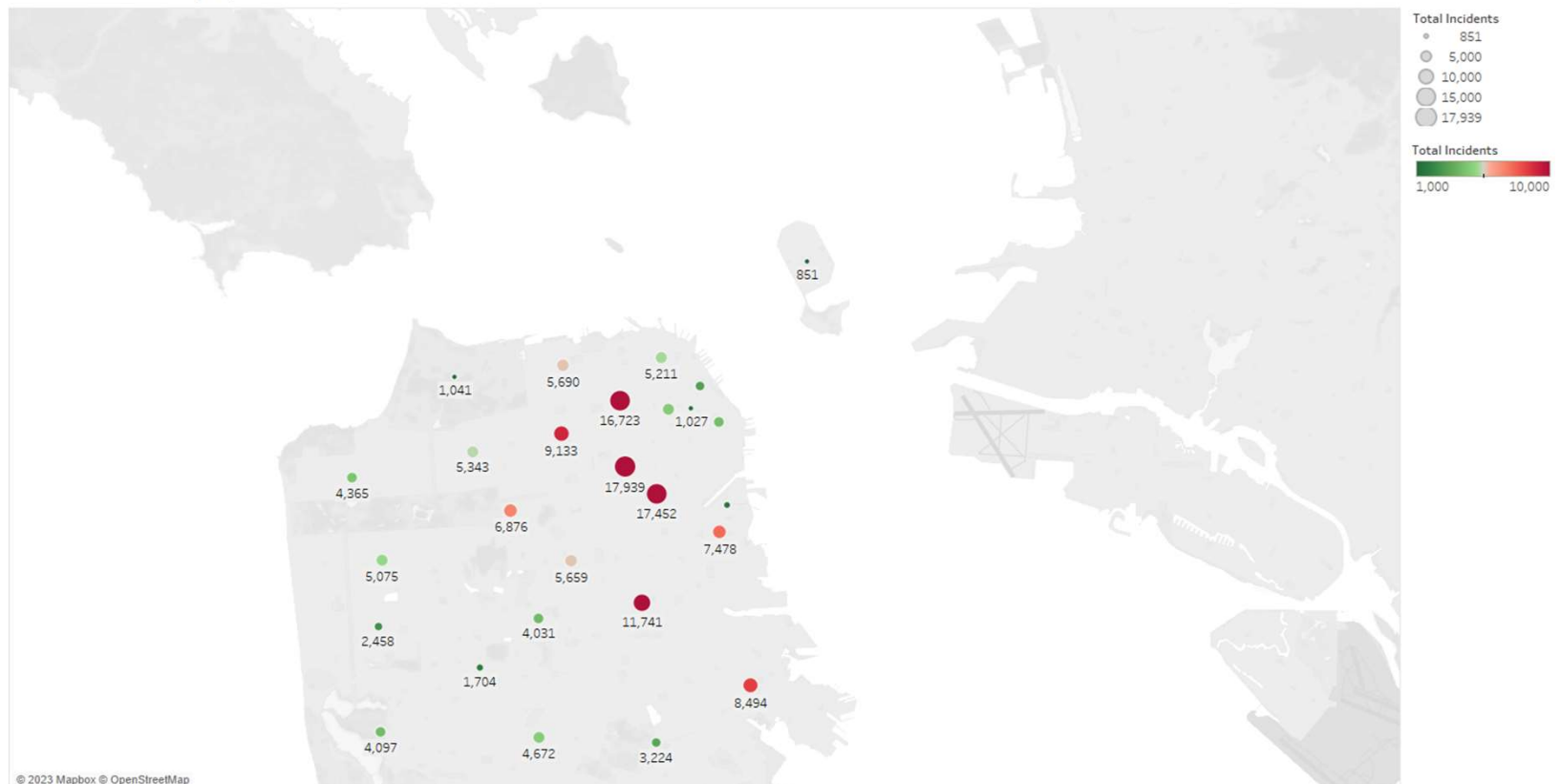


The trends of New Positive Test and Total Incidents for Date Month. Color shows details about New Positive Test and Total Incidents. For pane Sum of New Positive Test: The marks are labeled by New Positive Test. For pane Sum of Total Incidents: The marks are labeled by Total Incidents. The data is filtered 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

Fire Related calls by Zip



© 2023 Mapbox © OpenStreetMap

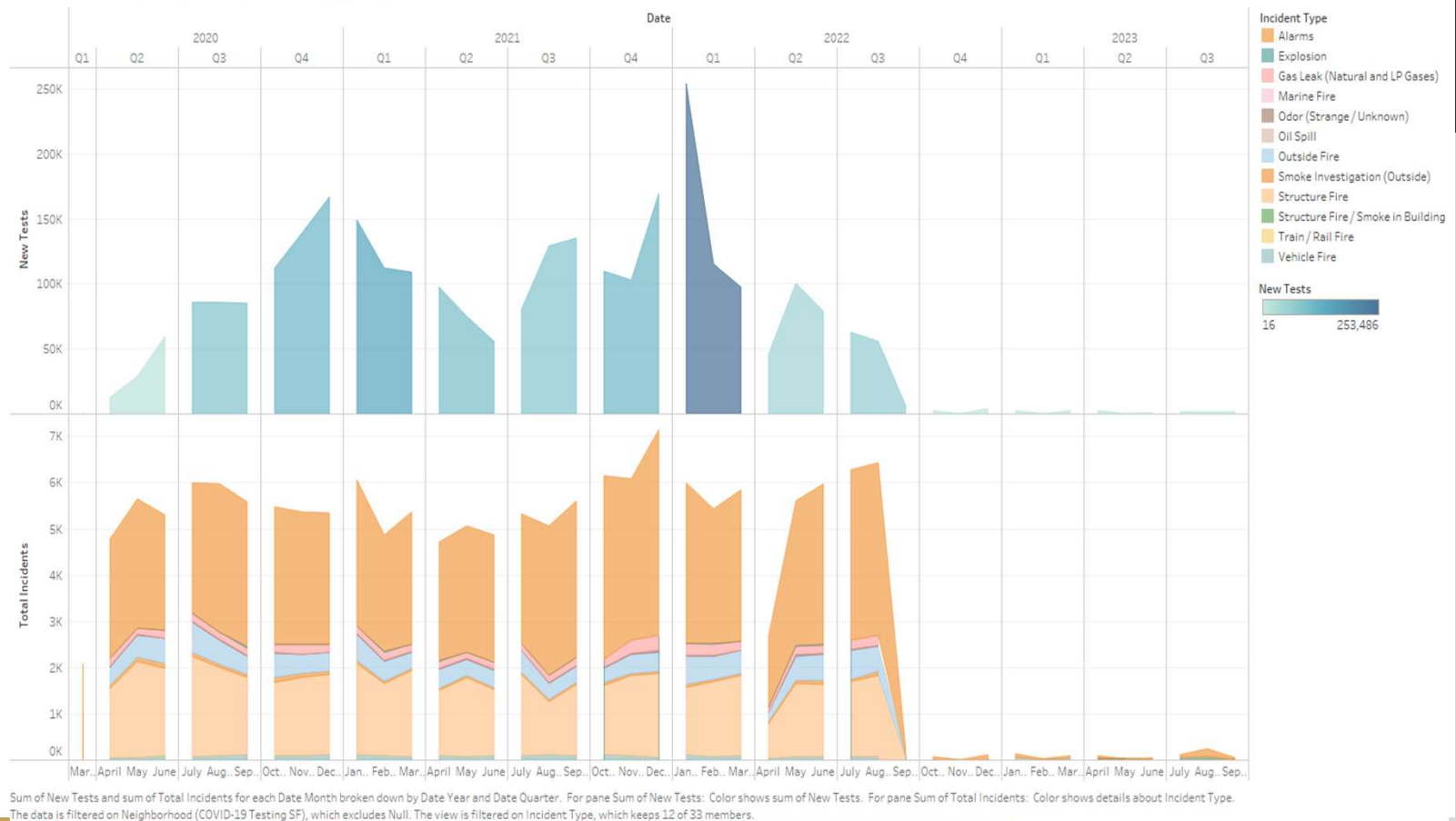
Map based on Longitude (generated) and Latitude (generated). Color shows sum of Total Incidents. Size shows sum of Total Incidents. The marks are labeled by sum of Total Incidents. Details are shown for Zip Code. The data is filtered on Fire Related and Date Year. The Fire Related filter keeps 12 members.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

New Test Corralating to Calls to Fire Department 2020-2023

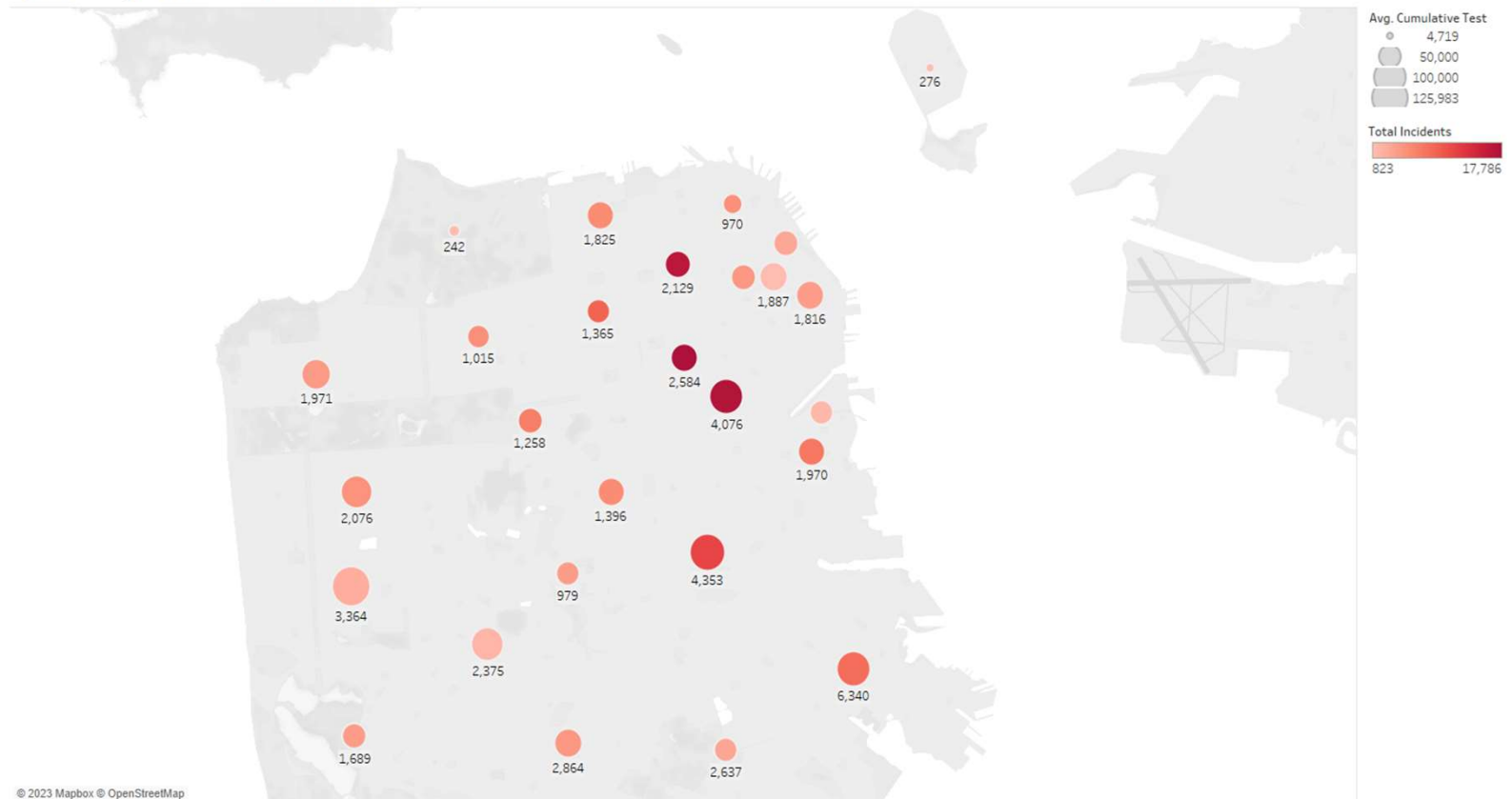


Analysis and Visualization (Symbol Map)

Spatial Analysis of Fire Incidents and Number of test taken

Show Fire Incidents in relation to number covid test taken.

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

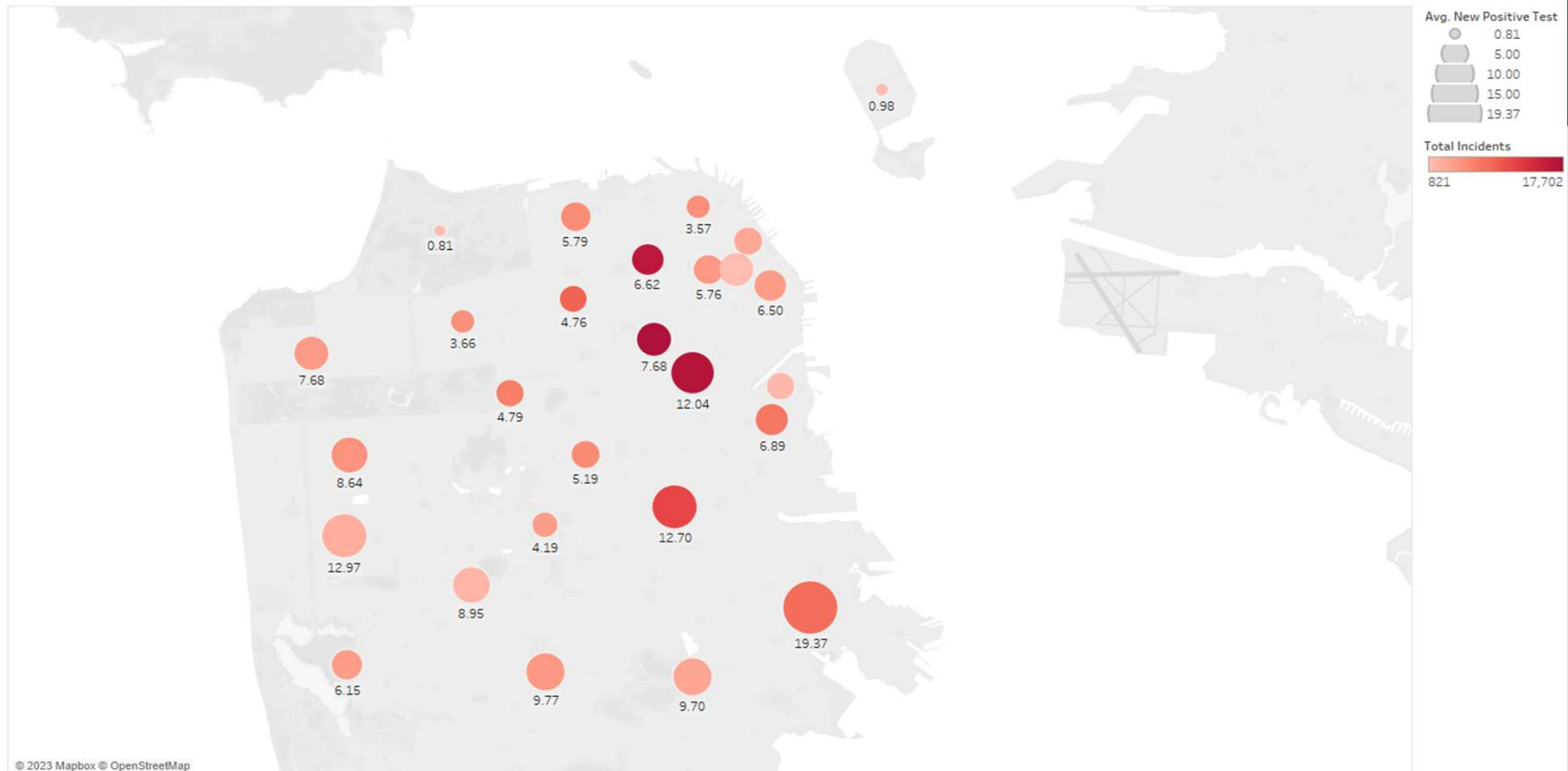


Map based on Longitude (generated) and Latitude (generated). Color shows sum of Total Incidents. Size shows average of Cumulative Test. The marks are labeled by average of Cumulative Positive Test. Details are shown for Zip Code. The data is filtered on Incident Type and Neighborhood (COVID-19 Testing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

Analysis and Visualization (Symbol Map)

Spatial Analysis of Fire Incidents and Number of New Positive Test results 2020-2022

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