NYPD Calls for Service User Search Engine

Presented by 5400 Team 1:

Jasmine(Keyi) Jiang, Jiajun Ma, Heyou Pan, Haonan Yao, Tianren Xie

Project Landscape

Crime is a major concern for many people, and it can have a significant impact on people quality of life. The New York City Police Department (NYPD) maintains a comprehensive database of crime incidents reported across the city, which can provide valuable insights into crime rates and patterns in different neighborhoods.

The focus of this project is to analyze crime rates and NYPD dispatching efficiency in various areas of New York City using data provided in the NYPD database. Specifically, we will examine the incidence of different types of crime in each neighborhood and visualize the data to help users search for the relative safety of different areas. The insights gained from this analysis might be of interest to a range of stakeholders.

Calls for Service to NYPD's 911 system
&
NYPD Precincts

- Call takers and dispatchers use the NYPD 911 system (ICAD) to communicate with callers and the NYPD, and each record in the dataset represents an entry into the system, generated by both members of the public and self-initiated entries by NYPD Members of Service. The data can be utilized to address issues being responded to by the NYPD.

- NYPD Precincts records the office contact information (i.e., full name, ID, phone number and address) of each precincts in different districts of New York.

Dataset	Rows	Columns
Calls for Service	**7.4M**	**19**
NYPD Precincts	**77**	**4**

Data source:

1. *NYPD Calls for Service https://data.cityofnewyork.us/Public-Safety/NYPD-Calls-for-Service-Year-to-Date-/n2zq-pubd (Last updated in February 1, 2023)*
2. *NYPD Precincts https://www.nyc.gov/site/nypd/bureaus/patrol/precincts-landing.page*

Calls for Service Dataset after processing at a glimpse:

Data Preprocessing Steps:

- **Check Null** values
- **Create timestamp** from date & time columns
- **Compute efficiency** (overall_efficiency, responding_efficiency, police_dispatch_efficiency, and solve_case_efficiency) to facilitate evaluations
- **Factor** borough and patrol columns

Columns added:

- ***INCIDENT_TS***
- ***OVERALL_EFF***
- ***RESPOND_EFF***
- ***DISPATCH_EFF***
- ***SOLVE_EFF***

Date & time to display

Geographic locations to visualize

Timestamp

Column Name	Column Description
OBJECTID	Row identifier for each call
CAD_EVNT_ID	Unique identifier generated by the the ICAD 911 system
CREATE_DATE	Date of call
INCIDENT_DATE	Date of incident
INCIDENT_TIME	Time of incident
NYPD_PCT_CD	NYPD precinct call is in
BORO_NM	Borough call is in
PATRL_BORO_NM	NYPD patrol Borough call is in
GEO_CD_X	The X-Coordinate of the midblock of the street segment where the violation was issued
GEO_CD_Y	The Y-Coordinate of the midblock of the street segment where the violation was issued
RADIO_CODE	NYPD code used to inform NYPD member of service the nature of the call
TYP_DESC	Description based on RADIO_CODE
CIP_JOBS	Flag indicating if the call relates to a Crime In Progress (CIP)
ADD_TS	Timestamp of when the call was added to the system
DISP_TS	Timestamp of when the call was dispatched to a responding unit
ARRIVD_TS	Timestamp of when the responding unit arrived on the scene
CLOSNG_TS	Timestamp of when the call was marked closed
Latitude	The Latitude of the midblock of the street segment where the violation was issued
Longitude	The Longitude of the midblock of the street segment where the violation was issued

Business Use Case

Potential Residents

- Potential buyers or renters of properties in New York City may use this tool to help them make informed decisions about where to live based on their personal safety concerns.

City planners and Policymakers

- City planners and policymakers may use this tool to identify areas that require additional resources or interventions to improve public safety.

Law enforcement agencies

- Law enforcement agencies can use this project to track crime trends and patterns, and allocate resources accordingly.

Public Safety Guide

- The project provides New York City residents, visitors and businesses with a concise public safety resource that clarifies neighborhood safety and crime trends. In addition, the tool assists users in locating the appropriate NYPD precinct in a specific area, providing contact information for support or crime reports.

This project is designed to help you make informed decisions and drive positive change.

Selected Technologies and Technical Viabilities

R

Use R for data preprocessing: factor column transformation, computed NYPD efficiency data based on timestamp info and created new feature columns.

Python

Data preprocessing: check null values, clean dataset and merge dataframe for further analysis.

PostgreSQL

1. Store the information & columns directly relevant to all NYPDs in PostgreSQL RDBMS.

2. Leverage a small size of data tables to avoid low speed, high cost and memory space.

3. Perform query operations quickly for data manipulation.

MongoDB

1. Store all relevant information of the NYPD service calls dataset, prepared to provide detailed information of a certain case/observation.

2. Store each piece of entry as nodes to make retrieval more efficient and intuitive.

Python - Flask

Develop interactive search engine web page for users/clients to directly retrieve desired information.

Implementation of Design choices

1. Use a common design pattern:
    a. **Jupyter Notebook** - data preprocessing and connecting to Postgresql and MongoDB database.
    b. **Postgresql** - NYPD information - ensures data structure consistency, data security, easy maintenance, multi-person access, and prevents data redundancy.
    c. **MongoDB** - large-scale event files - opens to redundant data and low requirements in hardware and software, low costs for huge storage.
    d. Insert each dataset into the database, create variables to store input value from user search engine and set as filters in the query language.

2. Create a flow for the user:
    a. Give user instructions on the **Flask web page** about how to search for specific information with necessary input key words.
    b. Track down NYPD borough and patrol information to find all NYPDs around the area.
    c. Search specific event details based on the NYPD precinct, date, type of event, etc.
    d. Offer various options for conditional queries including incident time, NYPD efficiency, location, etc., and is able to provide more information if needed.

3.

ETL Pipeline

PostgreSQL

Create: CREATE TABLE NYPD_PCT (
BORO_NM VARCHAR(25),
PATRL_BORO_NM VARCHAR(255),
NYPD_PCT_CD INT NOT NULL,
NYPD_FULL_NAME VARCHAR(255),
PHONE_NUMBER VARCHAR(15),
ADDRESS VARCHAR(255),
Primary Key(NYPD_PCT_CD))

Insert: INSERT INTO NYPD_PCT
(BORO_NM, PATRL_BORO_NM,
NYPD_PCT_CD, NYPD_FULL_NAME,
PHONE_NUMBER, ADDRESS)
VALUES (%s,%s,%s,%s,%s,%s)

Query: Select * from NYPD_PCT where BORO_NM LIKE '%QUEENS%'
and PATRL_BORO_NM LIKE '%NORTH%'

MongoDB

Construct: client = MongoClient('localhost',27017)
db = client.apan5400
collection = db.calls

Insert: collection.insert_many(newsfeeds)

Query: pipeline = [
{ "$match":
{ "NYPD_PCT_CD": nypd_var,
"CREATE_DATE": date_time,
"TYP_DESC":{"$regex":type_incident,
"$options" :'i'}}
}
]

Data Governance Policies & Project Cost

1. **Data Governance policies**
   a. Data licensing: public domain – retrieve data from NYC open data
   b. Data availability: easily accessible and shareable for users
   c. Data quality:

      - consistency: using R to clean the data and unite the time format

      -free-of-error: keep the null or na objects to make sure the data is correct and reliable

2. **Data scalability and Cost:**
   a. Scalability: Local storage is enough for our current implementation, but we would prefer more resources if storing the complete data in RDBMS
   b. Storage cost: 1.69GB(original calls dataset) + 2.11GB(cleaned dataset) + 4KB(NYPD info dataset) on disk
   c. Compute cost: Localhost Port

Performance Evaluation Criteria

Time Usage - **%%time**

1. Time to ingest data from local disk to jupyter notebook:

   CPU time: 38.1s Wall time: 47.8s

2. Time to insert data into Postgresql database:

   CPU time: 20.9ms Wall time: 148ms

3. Time to use SQL query to retrieve NYPD data from RDBMS:

   CPU time: 3.91ms Wall time: 23.3ms

4. Time to insert complete data into MongoDB:

   CPU time: 9min 48s Wall time: 28min 39s

5. Time to use MongoDB query to retrieve event detailed data from NoSQL database:

   CPU time: 74.5ms Wall time: 20.5s

MongoDB Space Usage

```
apan5400> show dbs
admin      40.00 KiB
apan5400    1.18 GiB
config     72.00 KiB
local      72.00 KiB
apan5400> show collections
calls
```

Initial UI Web Page

Enter some text to search:

The following fields stand for: Borough, Patrol Borough(North or South), NYPD Precinct ID, Call Created date, and key word of incident type.

You can input the first two fields to look for the NYPD Precincts in searched areas.

Then you can input the last three fields to look for specific incident details.

Please clear before input new, you must input either the first field or the third field.

Borough

Patrol Borough

NYPD Precinct ID

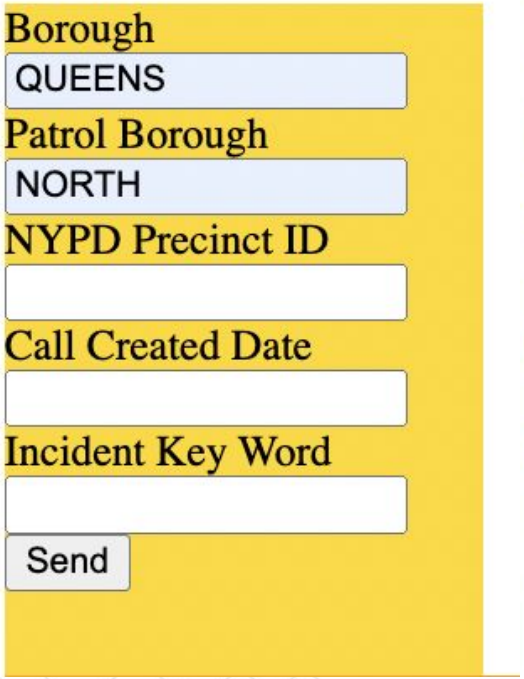
Call Created Date

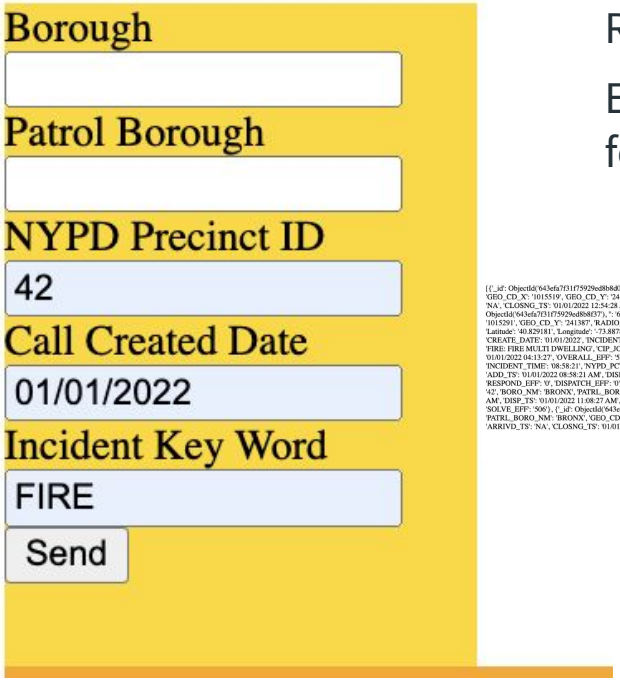
Incident Key Word

Send

Jasmine(Keyi) Jiang, Jiajun Ma, Heyou Pan, Haonan Yao, Tianren Xie

Interactive Search Engine





Returns string lists of results:

Eg: please refer to the html for clear result.

{'_id': ObjectId('643efa7f31f75929ed8b8d00'), '': '83', 'CAD_EVNT_ID': '82284594', 'CREATE_DATE': '01/01/2022', 'INCIDENT_DATE': '01/01/2022', 'INCIDENT_TIME': '00:06:56', 'NYPD_PCT_CD': '42', 'BORO_NM': 'BRONX', 'PATRL_BORO_NM': 'BRONX', 'GEO_CD_X': '1015519', 'GEO_CD_Y': '241579', 'RADIO_CODE': '105Z', 'TYP_DESC': 'INVESTIGATE/POSSIBLE CRIME: SHOTS FIRED/OUTSIDE', 'CIP_JOBS': 'Critical', 'ADD_TS': '01/01/2022 12:06:56 AM', 'DISP_TS': '01/01/2022 12:07:41 AM', 'ARRIVD_TS': 'NA', 'CLOSNG_TS': '01/01/2022 12:54:28 AM', 'Latitude': '40.829707', 'Longitude': '-73.887011', 'INCIDENT_TS': '01/01/2022 00:06:56', 'OVERALL_EFF': '2852', 'RESPOND_EFF': '0', 'DISPATCH_EFF': '45', 'SOLVE_EFF': '2807'}, {'_id': ObjectId('643efa7f31f75929ed8b8f37'), '': '650', 'CAD_EVNT_ID': '82285334', 'CREATE_DATE': '01/01/2022', 'INCIDENT_DATE': '01/01/2022', 'INCIDENT_TIME': '00:42:39', 'NYPD_PCT_CD': '42', 'BORO_NM': 'BRONX', 'PATRL_BORO_NM': 'BRONX', 'GEO_CD_X': '1015291', 'GEO_CD_Y': '241387', 'RADIO_CODE': '59W', 'TYP_DESC': 'FIRE: OUTSIDE RUBBISH', 'CIP_JOBS': 'Non CIP', 'ADD_TS': '01/01/2022 12:42:39 AM', 'DISP_TS': '01/01/2022 12:52:22 AM', 'ARRIVD_TS': 'NA', 'CLOSNG_TS': '01/01/2022 12:52:26 AM', 'Latitude': '40.829181', 'Longitude': '-73.887836', 'INCIDENT_TS': '01/01/2022 00:42:39', 'OVERALL_EFF': '587', 'RESPOND_EFF': '0', 'DISPATCH_EFF': '583', 'SOLVE_EFF': '4'}, {'_id': ObjectId('643efa8031f75929ed8b99a4'), '': '3319', 'CAD_EVNT_ID': '82288833', 'CREATE_DATE': '01/01/2022', 'INCIDENT_DATE': '01/01/2022', 'INCIDENT_TIME': '04:13:27', 'NYPD_PCT_CD': '42', 'BORO_NM': 'BRONX', 'PATRL_BORO_NM': 'BRONX', 'GEO_CD_X': '1008796', 'GEO_CD_Y': '239165', 'RADIO_CODE': '59R4', 'TYP_DESC': 'FIRE: FIRE MULTI DWELLING', 'CIP_JOBS': 'Non CIP', 'ADD_TS': '01/01/2022 04:13:27 AM', 'DISP_TS': '01/01/2022 05:39:36 AM', 'ARRIVD_TS': 'NA', 'CLOSNG_TS': '01/01/2022 05:44:40 AM', 'Latitude': '40.823102', 'Longitude': '-73.911314', 'INCIDENT_TS': '01/01/2022 04:13:27', 'OVERALL_EFF': '5473', 'RESPOND_EFF': '0', 'DISPATCH_EFF': '5169', 'SOLVE_EFF': '3043'}, {'_id': ObjectId('643efa8031f75929ed8ba316'), '': '5737', 'CAD_EVNT_ID': '82291949', 'CREATE_DATE': '01/01/2022', 'INCIDENT_DATE': '01/01/2022', 'INCIDENT_TIME': '08:58:21', 'NYPD_PCT_CD': '42', 'BORO_NM': 'BRONX', 'PATRL_BORO_NM': 'BRONX', 'GEO_CD_X': '1008141', 'GEO_CD_Y': '239675', 'RADIO_CODE': '59R4', 'TYP_DESC': 'FIRE: FIRE MULTI DWELLING', 'CIP_JOBS': 'Non CIP', 'ADD_TS': '01/01/2022 08:58:21 AM', 'DISP_TS': '01/01/2022 08:58:21 AM', 'ARRIVD_TS': 'NA', 'CLOSNG_TS': '01/01/2022 09:05:10 AM', 'Latitude': '40.824504', 'Longitude': '-73.913678', 'INCIDENT_TS': '01/01/2022 08:58:21', 'OVERALL_EFF': '409', 'RESPOND_EFF': '0', 'DISPATCH_EFF': '0', 'SOLVE_EFF': '409'}, {'_id': ObjectId('643efa8031f75929ed8ba8e8'), '': '7227', 'CAD_EVNT_ID': '82293898', 'CREATE_DATE': '01/01/2022', 'INCIDENT_DATE': '01/01/2022', 'INCIDENT_TIME': '11:07:33', 'NYPD_PCT_CD': '42', 'BORO_NM': 'BRONX', 'PATRL_BORO_NM': 'BRONX', 'GEO_CD_X': '1012275', 'GEO_CD_Y': '243254', 'RADIO_CODE': '105J', 'TYP_DESC': 'INVESTIGATE/POSSIBLE CRIME: SHOTS FIRED/INSIDE', 'CIP_JOBS': 'Critical', 'ADD_TS': '01/01/2022 11:07:33 AM', 'DISP_TS': '01/01/2022 11:08:27 AM', 'ARRIVD_TS': 'NA', 'CLOSNG_TS': '01/01/2022 11:16:53 AM', 'Latitude': '40.834315', 'Longitude': '-73.898727', 'INCIDENT_TS': '01/01/2022 11:07:33', 'OVERALL_EFF': '560', 'RESPOND_EFF': '0', 'DISPATCH_EFF': '54', 'SOLVE_EFF': '506'}, {'_id': ObjectId('643efa8031f75929ed8bb82e'), '': '11137', 'CAD_EVNT_ID': '82299641', 'CREATE_DATE': '01/01/2022', 'INCIDENT_DATE': '01/01/2022', 'INCIDENT_TIME': '16:47:18', 'NYPD_PCT_CD': '42', 'BORO_NM': 'BRONX', 'PATRL_BORO_NM': 'BRONX', 'GEO_CD_X': '1014850', 'GEO_CD_Y': '242038', 'RADIO_CODE': '59R3', 'TYP_DESC': 'FIRE: RESIDENCE PRIVATE HOUSE', 'CIP_JOBS': 'Non CIP', 'ADD_TS': '01/01/2022 04:47:18 PM', 'DISP_TS': '01/01/2022 04:48:18 PM', 'ARRIVD_TS': 'NA', 'CLOSNG_TS': '01/01/2022 07:56:01 PM', 'Latitude': '40.830969', 'Longitude': '-73.889427', 'INCIDENT_TS': '01/01/2022 16:47:18', 'OVERALL_EFF': '11323', 'RESPOND_EFF': '0', 'DISPATCH_EFF': '60', 'SOLVE_EFF': '11263'}]

Conclusions & Future Recommendations

√Completed

1. Use RDBMS to store NYPD's confidential information for its high-degree, multi-level security.
2. Use DBMS to store complete information for quick search at low cost.
3. Allow users to retrieve information via Flask web page.

×To Improve

1. Construct a better organized and functional relational database schema.
2. Reduce time cost in NoSQL database by leveraging key-value patterns.
3. Ignoring data storage costs, we could store all data in RDBMS to calculate NYPD efficiency and other measurements.
4. Improve UI design on user interface.

Team Member Contributions

	Keyi J	Heyou P	Jiajun M	Haonan Y	Tianren X
Background & Use Case				√	
Data Source Specification					√
Design choices & Techs	√	√			
Data Governance		√	√		
Performance Evaluation	√				
Conclusion & Recommendations	√				

Thank you for listening