



HELLENIC REPUBLIC
**National and Kapodistrian
University of Athens**
—EST. 1837—

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

Ανάπτυξη Βάσης Δεδομένων και Διαδικτυακής Εφαρμογής *LA Crime tracker*

Code repository:

<https://github.com/panagiotagyft/LA-Crime-Tracker-NoSQL.git>



LA Crimes tracker

Συστήματα Βάσεων Δεδομένων
(M149)

Διδάσκοντας: Αλέξης Δελής

Παναγιώτα Γύφτου, 7115112400025

Περιεχόμενα

1. Σχήμα βάσης δεδομένων.....	4
1.1 Σχεδιαστικές επιλογές οργάνωσης βάσης σε συλλογές.....	4
1.2 Ευρετήρια.....	5
1.3 E/R Schema.....	6
2. Κώδικας NO-SQL των <i>Queries</i>	7
2.1 Query1.....	7
2.2 Query2.....	8
2.3 Query3.....	8
2.4 Query4.....	9
2.5 Query5.....	10
2.6 Query6.....	11
2.7 Query7.....	11
2.8 Query8.....	12
2.9 Query9.....	13
2.10 Query10.....	13
3. Web Εφαρμογή.....	14
3.1 Σχεδιάγραμμα Εφαρμογής.....	14
3.2 Τεχνικές Βελτιστοποίησης.....	15
3.2.1 Insert.....	15

3.2.2 Upvoting.....	15
4. Στιγμιότυπα της εφαρμογής.....	16
4.1 Queries.....	16
4.2 Εφαρμογή.....	21
5. Χρήσιμες εντολές.....	23

Σχήμα βάσης δεδομένων

1.1 Σχεδιαστικές επιλογές οργάνωσης βάσης σε πίνακες

Η βάση *NoSQL-LA-CRIME* είναι οργανωμένη σε δύο βασικές **collections**, μία για τα εγκληματολογικά δεδομένα **crime_reports** και μία για τους αστυνομικούς και τις ψήφους τους **upvotes**.

Η συλλογή *crime_reports* περιέχει όλες τις πληροφορίες μιας αναφοράς εγκλήματος. Πεδία με κενές τιμές δεν συμπεριλαμβάνονται· για παράδειγμα, αν κάποια πεδία όπως **όπλα, ηλικία ή φύλο θύματος, mocodes, location** ή **cross_street** δεν έχουν τιμή, διαγράφονται από την αντίστοιχη εγγραφή για εξοικονόμηση χώρου. Επιπλέον, τα μοναδικά *codes* που αποθηκεύονται στη βάση είναι τα *area_id*, *crm_cd* (1,2,3,4) και *premis_cd*. Το *area_id* είναι απαραίτητο, καθώς χρησιμοποιείται για τη δημιουργία του μοναδικού αριθμού αναφοράς *dr_no*, ο οποίος προκύπτει με βάση την ημερομηνία καταχώρησης και το *area_id*. Τα *crime codes* είναι κρίσιμα για τα περισσότερα *queries* και αποθηκεύονται ως *array*. Αν κάποιο από τα *crm_cd* (2,3,4) δεν έχει τιμή, δεν αποθηκεύεται. Το πεδίο *premis_cd* αποθηκεύεται, καθώς σε ορισμένες περιπτώσεις η αντίστοιχη περιγραφή (*premis_desc*) δεν είναι διαθέσιμη. Τέλος, η συλλογή διατηρεί και το συνολικό αριθμό **upvotes** για κάθε αναφορά.

Η συλλογή *upvotes* περιέχει πληροφορίες σχετικά με τις ψήφους (*upvotes*) που έχουν καταχωρήσει οι αστυνομικοί. Συγκεκριμένα, περιλαμβάνει εγγραφές μόνο για τους αστυνομικούς που έχουν πραγματοποιήσει τουλάχιστον μία ψήφο. Κάθε αστυνομικός αναγνωρίζεται από ένα μοναδικό αριθμό σήματος **badge_number** και διαθέτει επιπλέον πεδία όπως **e-mail, όνομα**, καθώς και μια **λίστα με τις αναφορές (DR_NO)** που έχει ψηφίσει.

Η διασύνδεση μεταξύ των συλλογών πραγματοποιείται μέσω του πεδίου *DR_NO*, το οποίο λειτουργεί ως κλειδί σύνδεσης μεταξύ των εγκληματολογικών αναφορών (*crime_reports*) και των ψήφων (*upvotes*). Αυτός ο σχεδιασμός διασφαλίζει αποδοτική αποθήκευση και επιτρέπει ταχεία ανάκτηση δεδομένων μέσω *MongoDB queries*.

1.2 Ευρετήρια

Collection upvotes

Name & Definition	Type	Size	Usage	Properties	Status
> _id_	REGULAR ⓘ	32.8 KB	5 (since Thu Feb 13 2025)	UNIQUE ⓘ	READY
▼ name_1	REGULAR ⓘ	32.8 KB	2 (since Fri Feb 14 2025)		READY
name ↑					

Collection crime_reports

Name & Definition	Type	Size	Usage	Properties	Status
> _id_	REGULAR ⓘ	11.9 MB	10 (since Thu Feb 13 2025)	UNIQUE ⓘ	READY
▼ dr_no_1	REGULAR ⓘ	14.9 MB	1607665 (since Thu Feb 13 2025)		READY
dr_no ↑					
▼ area.area_id_1	REGULAR ⓘ	5.5 MB	0 (since Thu Feb 13 2025)		READY
area.area_id ↑					
▼ date_rptd_1	REGULAR ⓘ	9.4 MB	5 (since Fri Feb 14 2025)		READY
date_rptd ↑					
▼ area.area_name_1	REGULAR ⓘ	5.5 MB	0 (since Fri Feb 14 2025)		READY
area.area_name ↑					
▼ crime_codes.crm_cd_1	REGULAR ⓘ	5.9 MB	1 (since Fri Feb 14 2025)		READY
crime_codes.crm_cd ↑					

Τα ευρετήρια στις συλλογές `upvotes` και `crime_reports` βελτιώνουν σημαντικά την απόδοση των aggregation pipelines, επιταχύνοντας την αναζήτηση, τη συσχέτιση και την ομαδοποίηση δεδομένων. Το index **name_1** στη συλλογή `upvotes` διευκολύνει την αναζήτηση αξιωματικών στα ερωτήματα που σχετίζονται με `upvotes`, ενώ το **dr_no_1** στη συλλογή `crime_reports` επιταχύνει το \$lookup, βελτιώνοντας τη σύνδεση των ψήφων με τις αναφορές εγκλημάτων. Το **crime_codes.crm_cd_1** και το **date_rptd_1** βοηθούν στην αναζήτηση και ομαδοποίηση εγκλημάτων βάσει χρονικού διαστήματος και κωδικών εγκλημάτων, ενώ το **area.area_name_1** και **area.area_id_1** κάνουν πιο αποδοτικό το grouping ανά περιοχή. Αυτά τα indexes μειώνουν σημαντικά τον χρόνο εκτέλεσης queries, ειδικά σε ερωτήματα που περιλαμβάνουν \$match, \$lookup και \$group.

1.3 E/R Schema

crime_reports

- `_id` (ObjectId) - Primary Key
- `dr_no` (Integer) - Unique crime report number
- `date_rptd` (Date) - Date the crime was reported
- `timestamp`
 - `date_occ` (Date) - Date the crime occurred
 - `time_occ` (Time) - Time the crime occurred
- `area`
 - `area_id` (Integer) - Area identifier
 - `area_name` (String) - Name of the area
- `rpt_dist_no` (Integer) - Report district number
- `crime_codes` (Array of `CrimeCode`)
 - `crm_cd` (Integer) - Crime code identifier
 - `crm_cd_desc` (String) - Crime description
- `mocodes` (String) - Modus operandi codes
- `victim`
 - `age` (Integer) - Age of the victim
 - `sex` (String) - Sex of the victim
 - `descent` (String) - Victim's descent
- `premises`
 - `premis_cd` (Integer) - Premises code
 - `premis_desc` (String) - Description of premises
- `weapon` (String) - Weapon used
- `status` (String) - Status of the investigation
- `location`
 - `location` (String) - Address
 - `lat` (Float) - Latitude
 - `lon` (Float) - Longitude
 - `cross_street` (String) - Longitude
- `upvote_count` (Integer) - Number of upvotes received

upvotes

- `_id` (String) - Primary Key (e.g., "officer_1")
- `name` (String) - Officer's name
- `email` (String) - Officer's email
- `badge_number` (Integer) - Unique badge number
- `votes` (Array of `dr_no` references) - List of `CrimeReport.dr_no` that the officer has upvoted

crime_reports ↔ upvotes (Σχέση Πολλά-Προς-Πολλά, M:N)

- Ένας αστυνομικός μπορεί να ψηφίσει (upvote) πολλές αναφορές εγκλημάτων.
- Μία αναφορά εγκλήματος μπορεί να λάβει ψήφους από πολλούς αστυνομικούς.
- Αυτό αναπαρίσταται από τον πίνακα votes στη συλλογή Officer, ο οποίος περιέχει τις τιμές dr_no από τη συλλογή CrimeReport.

Κώδικας NO-SQL των *Queries*

Query 1

Find the total number of reports per “Crm Cd” that occurred within a specified time range and sort them in a descending order.

```
pipeline = [  
  {  
    "$unwind": "$crime_codes"  
  },  
  {  
    "$match": {  
      "timestamp.time_occ": {"$gte": start_time, "$lte": end_time}  
    }  
  },  
  {  
    "$group": {  
      "_id": "$crime_codes.crm_cd",  
      "report_count": {"$sum": 1}  
    }  
  },  
  {  
    "$sort": {"report_count": -1}  
  }  
]
```

Query 2

Find the total number of reports per day for a specific “Crm Cd” and time range.

```
pipeline = [
  {"$match": {
    "crime_codes.crm_cd": crime_code,
    "timestamp.time_occ": {"$gte": start_time, "$lte": end_time}
  }},
  {"$group": {
    "_id": "$timestamp.date_occ",
    "total_reports": {"$sum": 1}
  }},
  {"$sort": {"total_reports": -1}}
]
```

Query 3

Find the three most common crimes committed—regardless of code 1, 2, 3, and 4—per area for a specific day.

```
pipeline = [
  {"$match": { "date_rptd": date }},
  {"$unwind": "$crime_codes"},
  {
    "$group": {
      "_id": {
        "area": "$area.area_name",
        "crime_cd": "$crime_codes.crm_cd"
      },
      "count": { "$sum": 1 }
    }
  },
  {
    "$group": {
      "_id": "$_id.area",
      "crimes": {
        "$push": {
          "crime_cd": "$_id.crime_cd",
          "count": "$count"
        }
      }
    }
  },
  {
    "$project": {
      "area": "$_id",
      "top_crimes": {
        "$slice": [
          {"$sortBy": {"input": "$crimes", "sortBy": {"count": -1}}}, 3
        ]
      }
    }
  }
]
```


Query 4

Find the two least common crimes committed with regards to a given time range.

```
pipeline = [  
  {  
    "$match": {  
      "timestamp.time_occ": {"$gte": start_time, "$lte": end_time}  
    }  
  },  
  {  
    "$unwind": "$crime_codes"  
  },  
  {  
    "$group": {  
      "_id": "$crime_codes.crm_cd",  
      "report_count": {"$sum": 1}  
    }  
  },  
  {  
    "$sort": { "report_count": 1 }  
  },  
  {  
    "$limit": 2  
  }  
]
```

Query 5

Find the types of weapon that have been used for the same crime “Crm Cd” in more than one areas.

```
pipeline = [
  {
    "$addFields": {
      "first_crime_code": { "$arrayElemAt": ["$crime_codes", 0] }
    }
  },
  {
    "$match": {
      "weapon": {"$ne": None}
    }
  },
  {
    "$group": {
      "_id": {
        "crime_cd": "$first_crime_code.crm_cd",
        "weapon": "$weapon"
      },
      "areas": {"$addToSet": "$area.area_name"}
    }
  },
  {
    "$match": {
      "areas.1": {"$exists": True}
    }
  },
  {
    "$group": {
      "_id": "$_id.crime_cd",
      "weapons": {"$addToSet": "$_id.weapon"}
    }
  },
  {
    "$project": {
      "_id": 0,
      "crime_cd": "$_id",
      "weapons": {"$setUnion": ["$weapons", []]}
    }
  },
  {
    "$sort": {"crime_cd": 1}
  }
]
```

Query 6

Find the fifty most upvoted reports for a specific day.

```
pipeline = [  
  {  
    "$match": { "date_rptd": date }  
  },  
  {  
    "$project": {  
      "_id": 0,  
      "dr_no": 1,  
      "upvote_count": 1  
    }  
  },  
  {  
    "$sort": { "upvote_count": -1 }  
  },  
  {  
    "$limit": 50  
  }  
]
```

Query 7

Find the fifty most active police officers, with regard to the total number of upvotes.

```
pipeline = [  
  {  
    "$project": {  
      "name": 1,  
      "badge_number": 1,  
      "total_upvotes": { "$size": "$votes" }  
    }  
  },  
  {  
    "$sort": {  
      "total_upvotes": -1  
    }  
  },  
  {  
    "$limit": 50  
  }  
]
```

Query 8

Find the top fifty police officers, with regard to the total number of areas for which they have upvoted reports.

```
pipeline = [
  { "$unwind": "$votes" },
  {
    "$lookup": {
      "from": "crime_reports",
      "localField": "votes",
      "foreignField": "dr_no",
      "pipeline": [
        { "$project": { "area_id": "$area.area_id", "_id": 0 } }
      ],
      "as": "joined"
    }
  },
  { "$unwind": "$joined" },
  {
    "$group": {
      "_id": {
        "officerName": "$name",
        "badgeNumber": "$badge_number"
      },
      "distinctAreas": { "$addToSet": "$joined.area_id" }
    }
  },
  {
    "$project": {
      "_id": 0,
      "officerName": "$_id.officerName",
      "badgeNumber": "$_id.badgeNumber",
      "totalDistinctAreas": { "$size": "$distinctAreas" }
    }
  },
  { "$sort": { "totalDistinctAreas": -1 } },
  { "$limit": 50 }
]
```

Query 9

Find all reports for which the same e-mail has been used for more than one badge numbers when casting an upvote.

```
pipeline = [
  { "$unwind": "$votes" },
  {
    "$group": {
      "_id": "$email",
      "uniqueBadgeNumbers": { "$addToSet": "$badge_number" },
      "allDrNos": { "$push": "$votes" }
    }
  },
  {
    "$match": { "uniqueBadgeNumbers.1": { "$exists": True } } # If there are at least 2 badge_numbers
  },
  {
    "$project": {
      "_id": 0,
      "email": "$_id",
      "uniqueBadgeNumbers": 1,
      "allDrNos": 1
    }
  }
]
```

Query 10

Find all areas for which a given name has casted a vote for a report involving it.

```
pipeline = [
  { "$match": { "name": officer_name } },
  { "$unwind": "$votes" },
  {
    "$lookup": {
      "from": "crime_reports",
      "localField": "votes",
      "foreignField": "dr_no",
      "as": "crime_info"
    }
  },
  { "$unwind": "$crime_info" },
  { "$group": { "_id": "$crime_info.area.area_name" } }
]
```

Web Εφαρμογή

3.1 Σχεδιάγραμμα Εφαρμογής

1. Αρχική Διαδρομή (/)

- ο Εμφανίζει τη σελίδα **Login**.

2. Διαδρομή /login

- ο Εμφανίζει τη σελίδα **Login**.

3. Διαδρομή /register

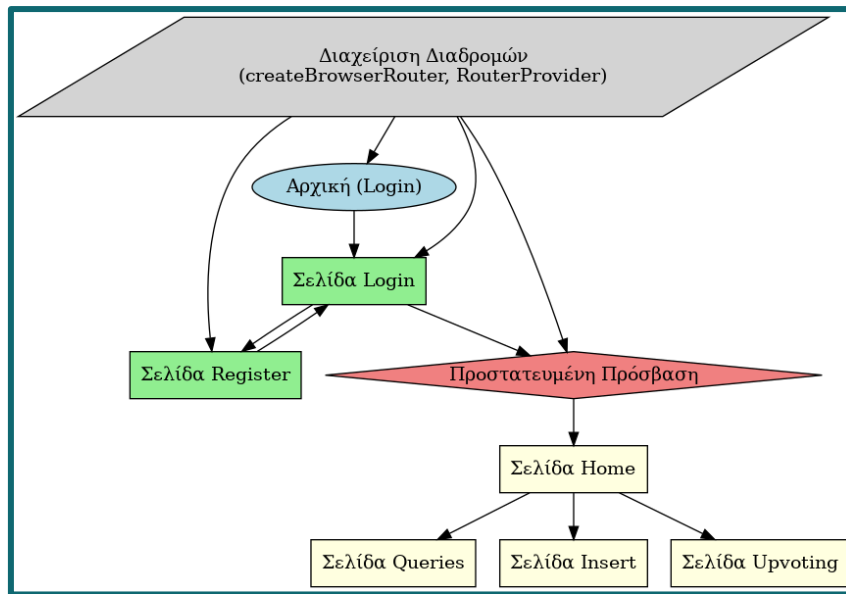
- ο Εμφανίζει τη σελίδα **Register**.

4. Προστατευμένες Διαδρομές (απαιτείται σύνδεση χρήστη μέσω του *ProtectedRoute*):

- ο **/home**: Εμφανίζει τη σελίδα **Home**.
- ο **/queries**: Εμφανίζει τη σελίδα **Queries**.
- ο **/insert**: Εμφανίζει τη σελίδα **Insert**.
- ο **/upvoting**: Εμφανίζει τη σελίδα **Upvoting**.

5. Διαχείριση Διαδρομών

- ο Το `createBrowserRouter` συνδέει κάθε διαδρομή με το αντίστοιχο στοιχείο.
- ο Το `RouterProvider` προσφέρει τη λειτουργικότητα πλοήγησης στην εφαρμογή.



3.2 Τεχνικές Βελτιστοποίησης

3.2.1 Insert

Η δημιουργία του **DR_NO** γίνεται αυτόματα, όταν ο χρήστης επιλέγει τις απαραίτητες παραμέτρους για την δημιουργία του και υλοποιείται από το backend στη μέθοδο `GenerateDRNOView` του `functions_views.py`. Πιο αναλυτικά η παραγωγή του **DR_NO** βασίζεται σε δύο παραμέτρους: το `area_id` (κωδικός περιοχής) και την `date_rptd` (ημερομηνία αναφοράς). Αφού εξάγεται το έτος από την ημερομηνία, εκτελείται ένα ερώτημα στη βάση δεδομένων για να υπολογιστεί ο επόμενος διαθέσιμος αριθμός αναφοράς για την περιοχή και το έτος. Το **DR_NO** σχηματίζεται με τη μορφή `YY + AreaID + NextRecordNumber`, όπου `YY` είναι τα δύο τελευταία ψηφία του έτους, το `AreaID` είναι ο κωδικός περιοχής, και το `NextRecordNumber` ένας πενταψήφιος αριθμός.

3.2.2 Upvoting

Λογω του μεγάλου όγκου δεδομένων, για την βέλτιστη χρήση της μνήμης όταν ο χρήστης θέλει να καταχωρήσει μια ψήφο, για βρει την αναφορά στην οποία θα καταχωρηθεί η ψήφος, το σύστημα αναζητά μερικώς την εγγραφή. Δηλαδή, ο χρήστης πληκτρολογώντας μέρος του αριθμού αναφοράς `dr_no` ζητά από την βάση όχι όλες τις αναφορές αλλά ένα μέρος αυτών, επιτυγχάνοντας την άμεση λήψη της. Η λειτουργία **search** υλοποιείται μέσω της κλάσης `SearchDRNumbersView`, επιτρέποντας την αναζήτηση αριθμών **DR** (`dr_no`) που ταιριάζουν με ένα ερώτημα (`query`) στο πεδίο `crime_reports.dr_no`.

Στιγμιότυπα της εφαρμογής

4.1 Queries

Query 1

1. Find the total number of reports per "Crm Cd" that occurred within a specified time range and sort them in a descending order.

Start Time: 02:18 μμ End Time: 02:19 μμ

Results:

CRIME CODE	REPORT COUNT
440	39
442	34
354	17
343	12
624	10
740	9
310	8
998	8

Reset

Query 2

2. Find the total number of reports per day for a specific "Crm Cd" and time range.

Start Time: 02:46 μμ End Time: 03:00 μμ Crime Code: 250

Results:

DATE	REPORT COUNT
2024-01-12	2
2021-10-24	1
2022-07-01	1
2020-06-02	1
2020-04-04	1
2021-10-11	1

Reset

Query 3

3. Find the three most common crimes committed –regardless of code 1, 2, 3, and 4– per area for a specific day.

Date

01/01/2024

Results:

AREA	THE MOST FREQUENT CRIME
Newton	230, 330, 998
Hollywood	624, 930, 626
N Hollywood	442, 310, 624
Central	440, 330, 998
Van Nuys	310, 440, 930
Wilshire	310, 442, 210
Olympic	626, 310, 330
West Valley	310, 626, 998

Reset

Query 4

4. Find the two least common crimes committed with regards to a given time range.

Start Time

08:20 πμ

End Time

04:20 μμ

Results:

CRIME CODE
904
445

Reset

Query 5

5. Find the types of weapon that have been used for the same crime "Crm Cd" in more than one areas.

Results:

CRIME CODE	WEAPONS
110	<ul style="list-style-type: none">AIR PISTOL/REVOLVER/RIFLE/BB GUNBLUNT INSTRUMENTCLUB/BATDIRK/DAGGERFIREFOLDING KNIFEHAMMERHAND GUNKITCHEN KNIFE
113	<ul style="list-style-type: none">LIQUOR/DRUGS
	<ul style="list-style-type: none">FOLDING KNIFEHAND GUNKITCHEN KNIFE

Reset

Query 6

6. Find the fifty most upvoted reports for a specific day.

Date

01/01/2024

Results:

REPORTS
242004012
241504035
241404120
241004026
242004536
241104041
241504042
240704025

Reset

Query 7

7. Find the fifty most active police officers, with regard to the total number of upvotes.

Results:

NAME	BADGE NUMBER	TOTAL UPVOTES
Anthony White	47344	999
Micheal Johnson	83548	998
Jocelyn Weeks	26275	997
Matthew Garcia	76376	996
Holly Ruiz	42079	996
Robert Nash	54880	994
Jessica Baldwin	78194	992
Kyle Chapman	89015	991

Reset

Query 8

8. Find the top fifty police officers, with regard to the total number of areas for which they have upvoted reports.

Results:

NAME	BADGE NUMBER	TOTAL DISTINCT AREAS
Drew Hunter	45818	21
Heather Singh	20474	21
David Gould	97258	21
Susan Logan	96043	21
Jennifer Wallace	72173	21
Jessica Cannon	96695	21
Angela Olson	70386	21
Michael Hoffman	89068	21

Reset

Query 9

9. Find all reports for which the same e-mail has been used for more than one badge numbers when casting an upvote.

Results:

EMAIL	BADGE NUMBERS	DR_NO LIST
lsmith@example.net	71875, 88054	221305598
		221304048
		201419899
		201015876
		231317616
		201705071
		230508820
		241507789
		201107001
		230116433

Reset

Query 10

10. Find all areas for which a given name has casted a vote for a report involving it.

Specific Office Name

Stacey Love

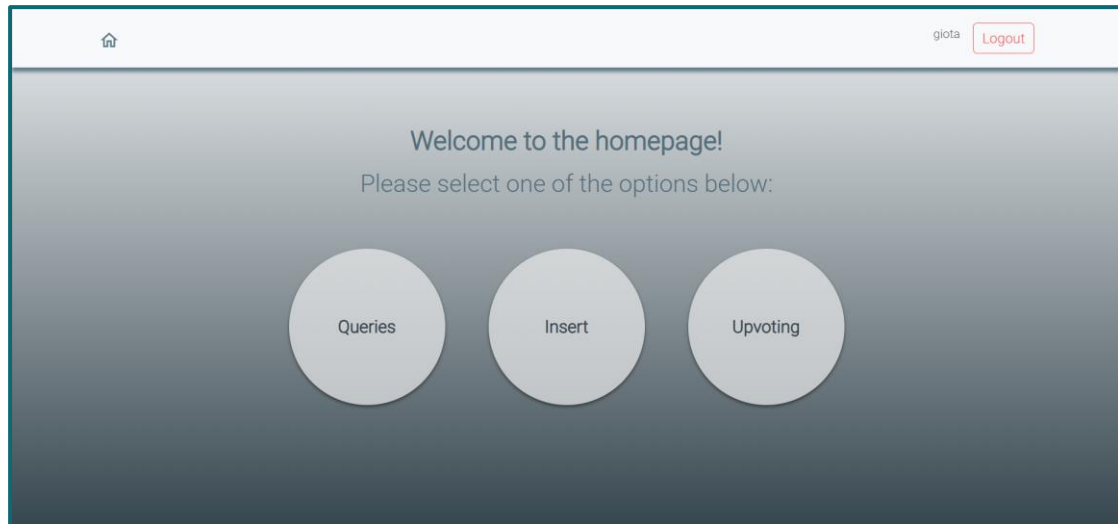
Results:

AREA
Hollywood
N Hollywood
Rampart

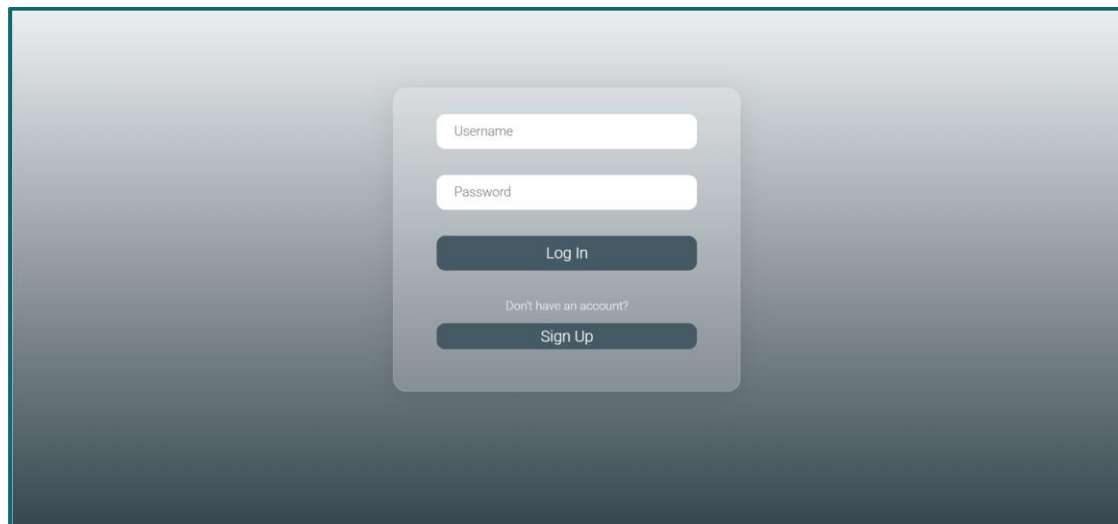
Reset

4.2 Εφαρμογή

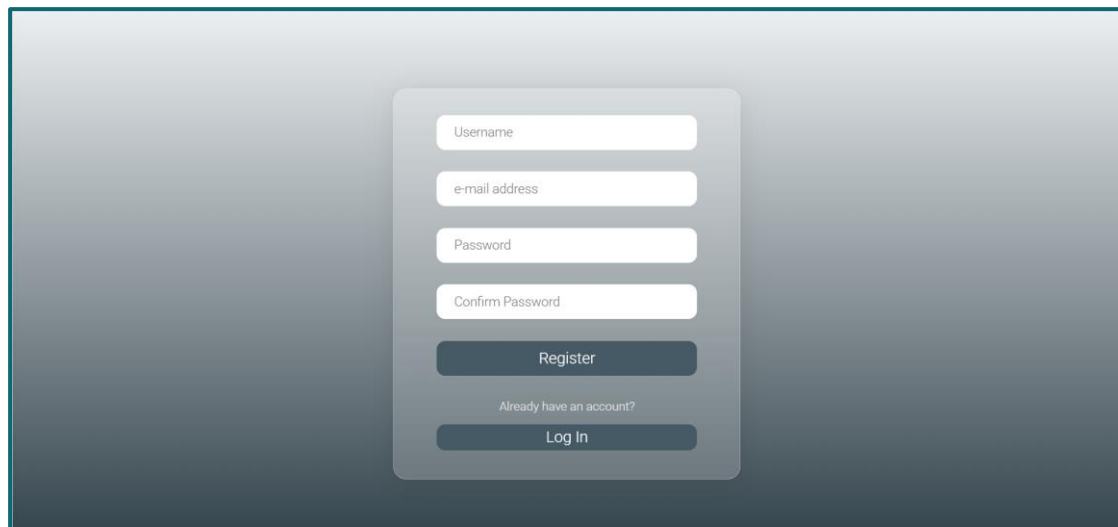
Home Page



Login Page



Register Page



A registration form centered on a dark blue gradient background. The form is enclosed in a light gray rounded rectangle. It contains four text input fields for 'Username', 'e-mail address', 'Password', and 'Confirm Password'. Below these is a dark blue 'Register' button. Underneath the button is a link 'Already have an account?' and a dark blue 'Log In' button.

Username

e-mail address

Password

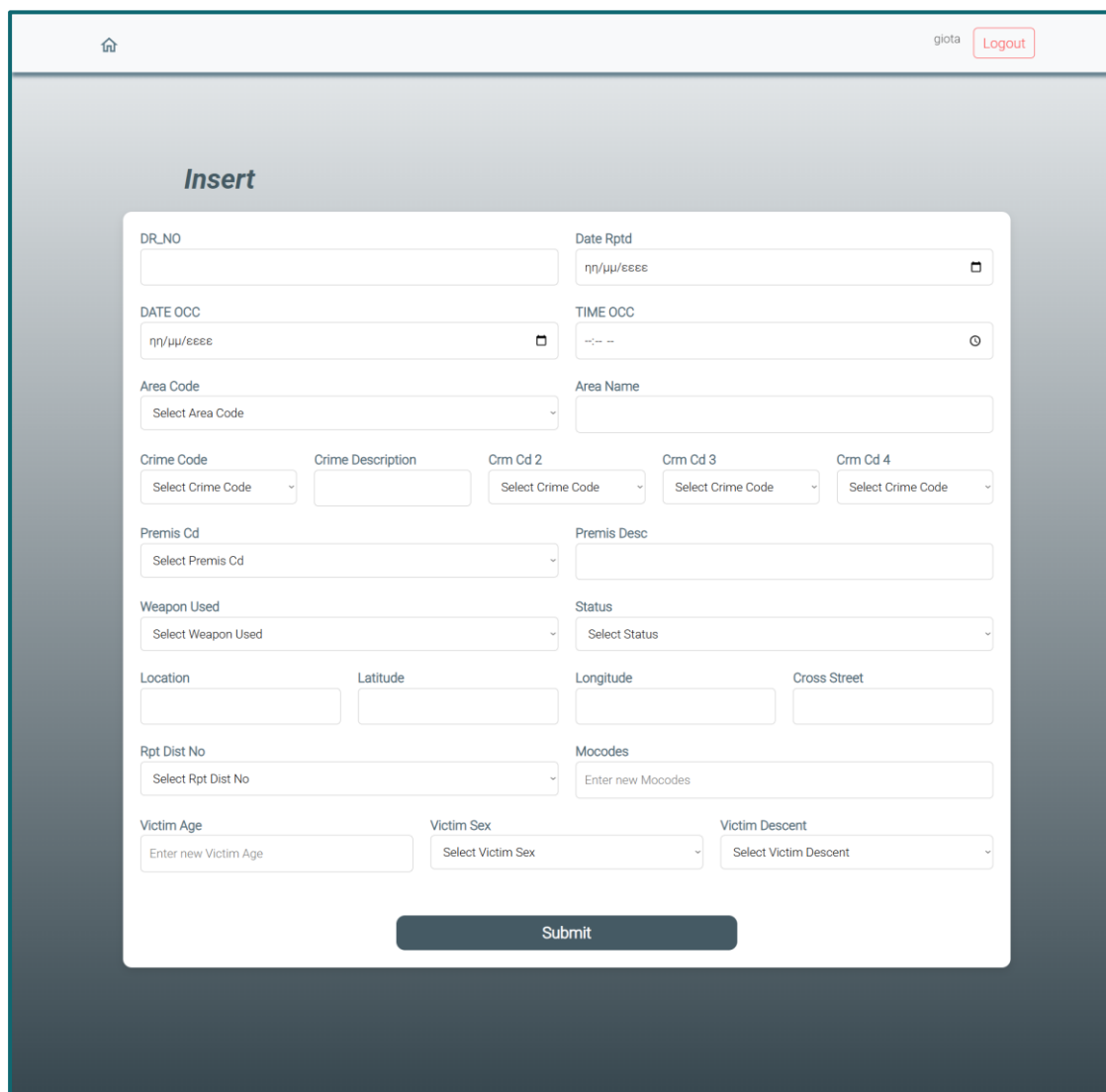
Confirm Password

Register

Already have an account?

Log In

Insert Page



The 'Insert' page is a data entry form within a web application. At the top, there is a header bar with a home icon, the user 'giota', and a 'Logout' button. The main content area has a title 'Insert' and a large form with various input fields. The form includes text inputs for 'DR_NO', 'Date Rptd', 'DATE OCC', 'TIME OCC', 'Area Name', 'Premis Desc', 'Location', 'Latitude', 'Longitude', 'Cross Street', 'Rpt Dist No', 'Mocodes', 'Victim Age', 'Victim Sex', and 'Victim Descent'. It also features several dropdown menus for 'Area Code', 'Crime Code', 'Crm Cd 2', 'Crm Cd 3', 'Crm Cd 4', 'Premis Cd', 'Weapon Used', 'Status', and 'Rpt Dist No'. A 'Crime Description' text input and a 'Victim Sex' dropdown are also present. A dark blue 'Submit' button is at the bottom of the form.

giota Logout

Insert

DR_NO

Date Rptd

DATE OCC

TIME OCC

Area Code

Area Name

Crime Code

Crime Description

Crm Cd 2

Crm Cd 3

Crm Cd 4

Premis Cd

Premis Desc

Weapon Used

Status

Location

Latitude

Longitude

Cross Street

Rpt Dist No

Mocodes

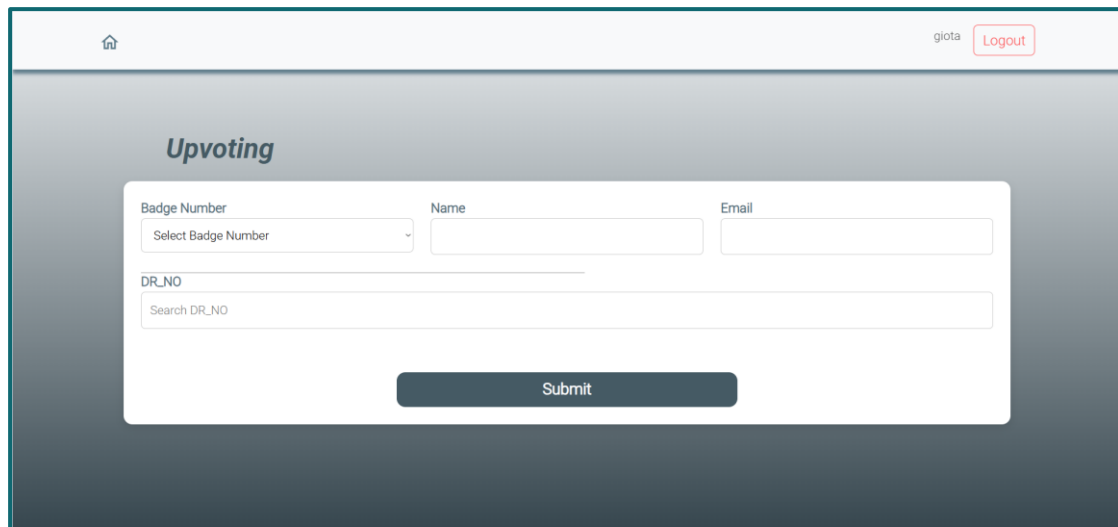
Victim Age

Victim Sex

Victim Descent

Submit

Upvoting Page



The image shows a web application interface for an 'Upvoting' page. At the top, there is a light gray header bar containing a home icon on the left, the text 'giota' in the center, and a red 'Logout' button on the right. Below the header, the main content area has a dark gray background. Centered in this area is a white rectangular form titled 'Upvoting' in a bold, italicized font. The form contains several input fields: a 'Badge Number' dropdown menu with 'Select Badge Number' as the placeholder, a 'Name' text input field, an 'Email' text input field, a 'DR_NO' text input field with 'Search DR_NO' as the placeholder, and a dark gray 'Submit' button at the bottom center.

Χρήσιμες εντολές

React

```
cd my-app  
npm install  
npm start
```

Django

```
cd crime_backend  
python manage.py makemigrations  
python manage.py migrate  
python manage.py runserver
```