**Milestone 2**

**Group Members & Github Username**

Jingle Wang, [gillianw@seas.upenn.edu](mailto:gillianw@seas.upenn.edu) (gillianwang0325)

Ziqi He, [ziqihe@seas.upenn.edu](mailto:ziqihe@seas.upenn.edu) (HeidiHe2001)

Lingpei Luo, [luol@seas.upenn.edu](mailto:luol@seas.upenn.edu) (JLmm123)

Guangqiuse Hu, [violahu@seas.upenn.edu](mailto:violahu@seas.upenn.edu) (violahu930)

[Crime Data from 2020 to Present - Catalog](https://catalog.data.gov/dataset/crime-data-from-2020-to-present)

**1. Motivation for the idea/description of the problem the application solves.**

The website focused on showcasing crime statistics in different geographic areas or patrol divisions of Los Angeles provides an insightful and comprehensive overview of the city's safety metrics. Through interactive maps, charts, and tables, users can explore crime statistics for different neighborhoods, districts, and communities within LA.

**2. List of features you will definitely implement in the application.**

Map function: The webpage employs color-coded visualizations in map to highlight areas with higher crime rates, allowing users to quickly identify patterns and trends.

Filter function: users can filter the data based on crime types, time periods, and specific geographic regions to obtain more targeted information. Detailed analysis and breakdowns of crime incidents, such as thefts, assaults, burglaries, and more, are presented in a format that is easy to understand.

**3. List of features you might implement in the application, given enough time.**

Search Crime Data Nearest to User: this feature leverages the user's current location or enter address to provide crime data specific to their vicinity. By integrating geolocation services or address input fields, the website can dynamically fetch and display crime statistics, and trends that are closest to the specified address.

Heatmaps and Clustering: Advanced map features like heatmaps and clustering can visually represent the density or intensity of crime incidents in different areas. Heatmaps show areas with higher concentrations of crime, while clustering groups nearby incidents for easier visualization.

**4. List of pages the application will have and a 1-2 sentence description of each page.**

**We expect that the functionality of each page will be meaningfully different from the functionality of the other pages.**

Homepage: The homepage will provide a summary of its features and functionality, along with navigation links to other pages.

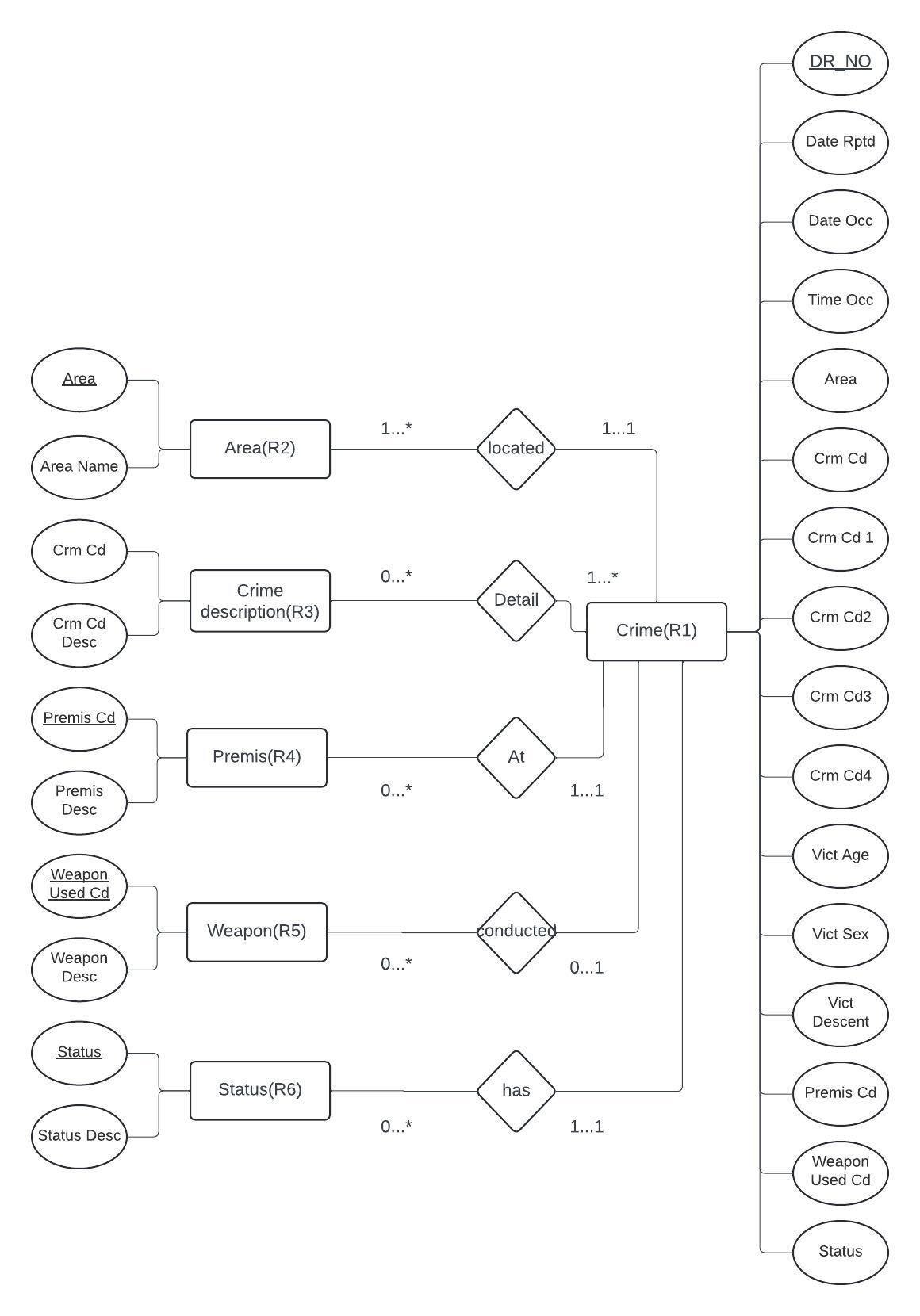
Map Page: The map page allows users to hover with districts that they are interested in, enabling them to search for crime data at specific locations and view geographic data

Search Page: The search page offers users the ability to search and filter through a database of information based on various criteria such as crime code, victim, time period, etc.

Detail Page: The detail page provides comprehensive information about a specific crime offering detailed descriptions, and other relevant data.

About Us Page: The about us page offers insights into the application's background, mission, team members, and any other information about the application.

**5. Relational schema as an ER diagram.**

****

| R1 | DR\_NO, Date Rptd, Date Occ, Time Occ, AREA, Crm Cd, Crm Cd 1, Crm Cd 2, Crm Cd 3, Crm Cd 4, Vict Age, Vict Sex, Vict Descent, Premis Cd, Weapon Used Cd, Status |
| --- | --- |
| R2 | AREA, AREA NAME |
| R3 | Crm Cd, Crm Cd Desc |
| R4 | Premis Cd, Premis Desc |
| R5 | Weapon Used Cd, Weapon Desc |
| R6 | Status, Status Desc |

Attribute Explanations: [Crime Data from 2020 to Present](https://www.opendatanetwork.com/dataset/data.lacity.org/2nrs-mtv8)

**6. SQL DDL for creating the database.**

| CREATE TABLE CrimeData (  DR\_NO int, *-- Unique identifier for each crime report*  Date\_Rptd date, *-- Date when the crime was reported*  Date\_Occ date,  *-- Date when the crime occurred*  Time\_Occ time, *-- Time when the crime occurred*  AREA int NOT NULL, *-- Area code where the crime occurred (1 to 21 inclusive)*  Crm\_Cd int, *-- Primary crime code*  Crm\_Cd1 int, -- *Additional crime code 1*  Crm\_Cd2 int, -- *Additional crime code 2*  Crm\_Cd3 int, -- *Additional crime code 3*  Crm\_Cd4 int, -- *Additional crime code 4*  Vict\_Age int, -- *Age of the victim*  Vict\_Sex char(1), -- *Gender of the victim (M: male, F: female, X: unknown)*  Vict\_Descent varchar(255), -- *Descent or ethnicity of the victim*  Premis\_Cd int, -- *Premises code where the crime occurred*  Weapon\_Used\_Cd int, -- *Weapon code used in the crime*  Status char(2), -- *Status of the crime*  PRIMARY KEY (DR\_NO)  );  CREATE TABLE Areas (  AREA int, *-- Area code*  AREA\_NAME varchar(255),  *-- Name of the area*  PRIMARY KEY (AREA)  );  CREATE TABLE CrimeCodes (  Crm\_Cd int, *-- Crime code*  Crm\_Cd\_Desc varchar(255), *-- Description of the crime code*  PRIMARY KEY (Crm\_Cd)  );  CREATE TABLE Premises (  Premis\_Cd int, *-- Premises code*  Premis\_Desc varchar(255), *-- Description of the premises*  PRIMARY KEY (Premis\_Cd)  );  CREATE TABLE Weapons (  Weapon\_Used\_Cd int,  *-- Weapon code*  Weapon\_Desc VARCHAR(255), *-- Description of the weapon*  PRIMARY KEY (Weapon\_Used\_Cd)  );  CREATE TABLE Status (  Status char(1), *-- Status code*  Status\_Desc VARCHAR(255),  *-- Description of the status*  PRIMARY KEY (Status)  ); |
| --- |

**7. Explanation of how you will clean and pre-process the data.**

* Database Normalization: Our initial step is breaking down the large table into multiple smaller tables to eliminate data redundancy and achieve a more organized and efficient database structure in 3NF form. This separation is based on identifying functional dependencies and grouping related attributes together.
* Data Cleaning:
  + This step includes removing duplicates that check and remove duplicate entries in R2, R3, R4, R5, R6 to avoid redundancy.
  + Additionally, we standardized data formats in R1 to ensure consistent data formats across attributes, such as date formats (Attributes: Date\_Rptd, Date\_Occ), and time formats (Attribute: Time\_Occ).
  + Lastly, remove unnecessary attributes including Mocodes, Location, LON, LAT, etc.

**8. List of technologies you will use.**

* Colab & Python (data cleaning)
* Github
* MySQL database

**9. Description of what each group member will be responsible for.**

Ziqi He and Guangqiuse Hu:

* cleaning and preprocessing the dataset using python
* Analyze functional dependencies and decompose table into 3NF
* get separated tables that can be directly import to SQL database

Jocelyn Luo and Gilian Wang:

* ER diagram & DDL
* frontend page design
* wrap up the milestone report