

PROJECT REPORT

Unravelling Chicago's Crime Dynamics

Course Name: BUAN 6320 Database Foundations for Business Analytics

Professor: Thiru Pandian

Report submitted by: Group 14

Group Members:

Harikrishna Surabhi (hxs220066)

Jitin Naidu Kondipati (jkn220002)

Kishore Kumar Suresh (kxs220098)

Shrey Sandeep Jain (sxj220064)

Vikram Ratan (vxr220042)

TABLE OF CONTENTS

Introduction	3
Conceptual Design	4
Logical Design	4
Physical Design	5
Data Loading	7
Insights & Visualization	8
Conclusions	18
References	18

INTRODUCTION

Project Description:

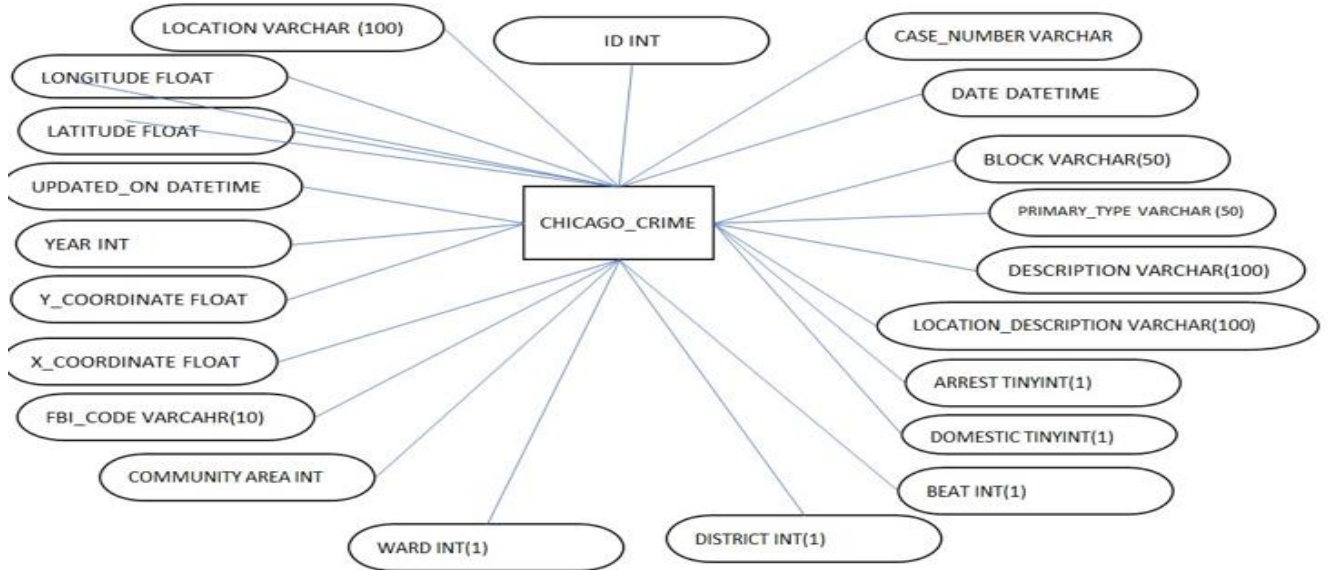
The "Crimes - 2001 to present - Dashboard" dataset is a comprehensive collection of reported crime incidents in the city of Chicago, Illinois, USA, from 2001 to the present. This dataset is maintained by the Chicago Police Department and regularly updated to reflect the most recent information. The data is sourced from the city's CLEAR (Citizen Law Enforcement Analysis and Reporting) system. Utilizing the "Crimes - 2001 to present -Dashboard" dataset, we dive into various aspects of crime, including spatial and temporal characteristics, crime types, and arrest rates. Our dataset contains 1,048,576 rows and 21 columns.

There are multiple applications of our dataset one of the applications is that it can be used by the students wanting to study in Chicago and want to know which parts are safe for them, it will also serve as a valuable resource for stakeholders, policymakers, and law enforcement agencies to make more informed decisions.

Data Dictionary:

Attribute	Data Type
ID	int (NOT NULL)
CASE_NUMBER	varchar(20)
DATE	datetime
BLOCK	varchar(50)
PRIMARY_TYPE	varchar(50)
DESCRIPTION	varchar(100)
LOCATION_DESCRIPTION	varchar(100)
ARREST	boolean
DOMESTIC	boolean
BEAT	int
DISTRICT	int
WARD	int
COMMUNITY_AREA	int
FBI_CODE	varchar(10)
X_COORDINATE	float
Y_COORDINATE	float
YEAR	int
UPDATED_ON	datetime
LATITUDE	float
LONGITUDE	float
LOCATION	varchar(100)

Conceptual Design



Logical Design

chicago_crime	
ID INT	
CASE_NUMBER VARCHAR(20)	
DATE DATETIME	
BLOCK VARCHAR(50)	
PRIMARY_TYPE VARCHAR(50)	
DESCRIPTION VARCHAR(100)	
LOCATION_DESCRIPTION VARCHAR(100)	
ARREST TINYINT(1)	
DOMESTIC TINYINT(1)	
BEAT INT	
DISTRICT INT	
WARD INT	
COMMUNITY_AREA INT	
FBI_CODE VARCHAR(10)	
X_COORDINATE FLOAT	
Y_COORDINATE FLOAT	
YEAR INT	
UPDATED_ON DATETIME	
LATITUDE FLOAT	
LONGITUDE FLOAT	
LOCATION VARCHAR(100)	
Indexes	

Physical Design

-- MySQL Workbench Forward Engineering

SET @OLD_UNIQUE_CHECKS=@@UNIQUE_CHECKS, UNIQUE_CHECKS=0;

SET @OLD_FOREIGN_KEY_CHECKS=@@FOREIGN_KEY_CHECKS, FOREIGN_KEY_CHECKS=0;

SET @OLD_SQL_MODE=@@SQL_MODE,
SQL_MODE='ONLY_FULL_GROUP_BY,STRICT_TRANS_TABLES,NO_ZERO_IN_DATE,NO_ZERO_
DATE,ERROR_FOR_DIVISION_BY_ZERO,NO_ENGINE_SUBSTITUTION';

-- Schema mydb

-- Schema Chicago_Project

-- Schema Chicago_Project

CREATE SCHEMA IF NOT EXISTS `Chicago_Project` DEFAULT CHARACTER SET utf8mb4
COLLATE utf8mb4_0900_ai_ci ;

USE `Chicago_Project` ;

-- Table `Chicago_Project`.`chicago_crime`

CREATE TABLE IF NOT EXISTS `Chicago_Project`.`chicago_crime` (

`ID` INT NOT NULL,

`CASE_NUMBER` VARCHAR(20) NULL DEFAULT NULL,

```
`DATE` DATETIME NULL DEFAULT NULL,  
`BLOCK` VARCHAR(50) NULL DEFAULT NULL,  
`PRIMARY_TYPE` VARCHAR(50) NULL DEFAULT NULL,  
`DESCRIPTION` VARCHAR(100) NULL DEFAULT NULL,  
`LOCATION_DESCRIPTION` VARCHAR(100) NULL DEFAULT NULL,  
`ARREST` TINYINT(1) NULL DEFAULT NULL,  
`DOMESTIC` TINYINT(1) NULL DEFAULT NULL,  
`BEAT` INT NULL DEFAULT NULL,  
`DISTRICT` INT NULL DEFAULT NULL,  
`WARD` INT NULL DEFAULT NULL,  
`COMMUNITY_AREA` INT NULL DEFAULT NULL,  
`FBI_CODE` VARCHAR(10) NULL DEFAULT NULL,  
`X_COORDINATE` FLOAT NULL DEFAULT NULL,  
`Y_COORDINATE` FLOAT NULL DEFAULT NULL,  
`YEAR` INT NULL DEFAULT NULL,  
`UPDATED_ON` DATETIME NULL DEFAULT NULL,  
`LATITUDE` FLOAT NULL DEFAULT NULL,  
`LONGITUDE` FLOAT NULL DEFAULT NULL,  
`LOCATION` VARCHAR(100) NULL DEFAULT NULL,  
PRIMARY KEY (`ID`))
```

ENGINE = InnoDB

DEFAULT CHARACTER SET = utf8mb4

COLLATE = utf8mb4_0900_ai_ci;

SET SQL_MODE=@OLD_SQL_MODE;

SET FOREIGN_KEY_CHECKS=@OLD_FOREIGN_KEY_CHECKS;

SET UNIQUE_CHECKS=@OLD_UNIQUE_CHECKS;

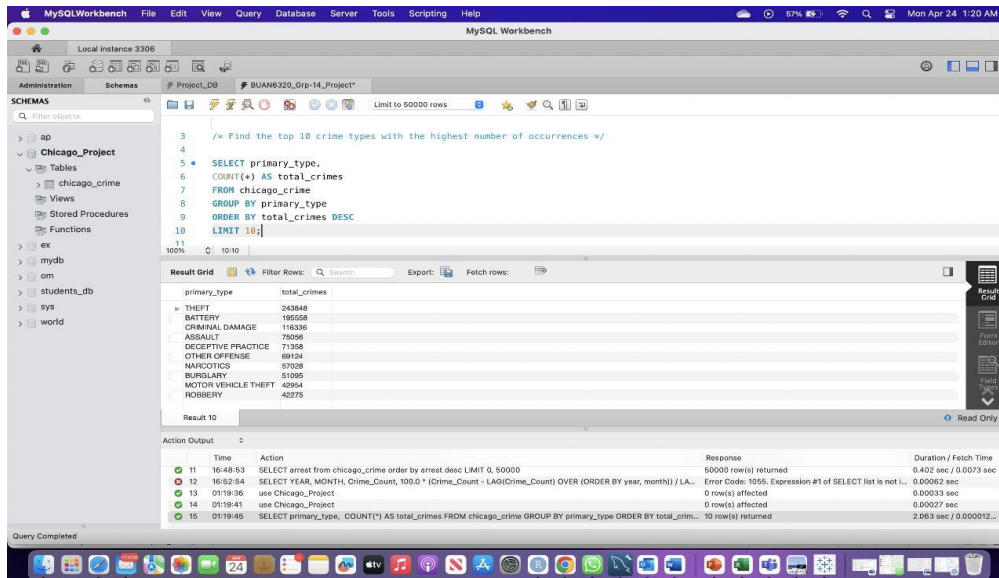
Data Loading

Created tables using commands:

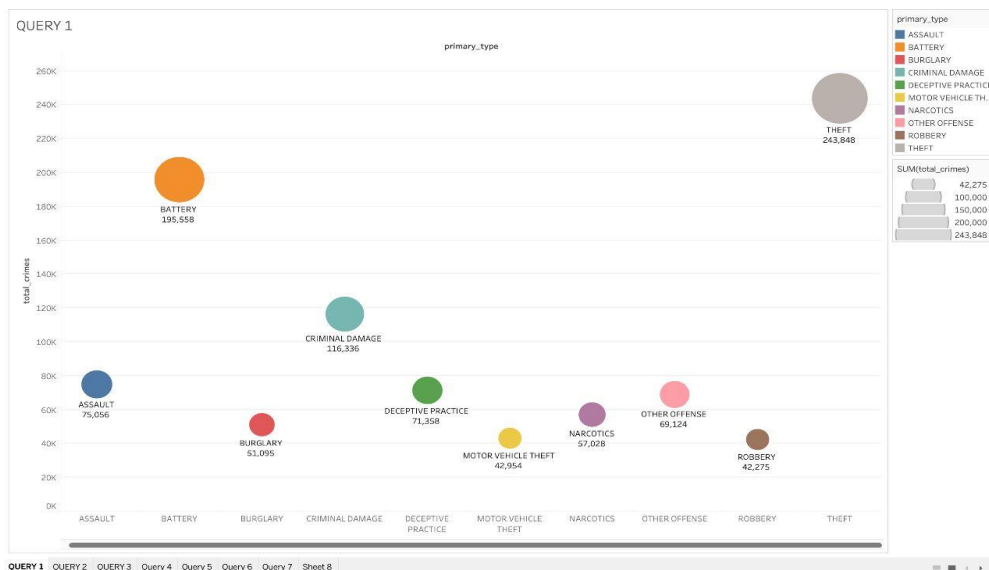
1. Downloaded the dataset from website into C:\ProgramData\MySQL\MySQL Server 8.0\Uploads.
2. Changed the settings and added this line in advanced tab OPT_LOCAL_INFILE=1.
3. Created Database Chicago_Project.
4. CREATE TABLE IF NOT EXISTS `Chicago_Project`.`chicago_crime` (
 `ID` INT NOT NULL,
 `CASE_NUMBER` VARCHAR(20) NULL DEFAULT NULL,
 `DATE` DATETIME NULL DEFAULT NULL,
 `BLOCK` VARCHAR(50) NULL DEFAULT NULL,
 `PRIMARY_TYPE` VARCHAR(50) NULL DEFAULT NULL,
 `DESCRIPTION` VARCHAR(100) NULL DEFAULT NULL,
 `LOCATION_DESCRIPTION` VARCHAR(100) NULL DEFAULT NULL,
 `ARREST` TINYINT(1) NULL DEFAULT NULL,
 `DOMESTIC` TINYINT(1) NULL DEFAULT NULL,
 `BEAT` INT NULL DEFAULT NULL,
 `DISTRICT` INT NULL DEFAULT NULL,
 `WARD` INT NULL DEFAULT NULL,
 `COMMUNITY_AREA` INT NULL DEFAULT NULL,
 `FBI_CODE` VARCHAR(10) NULL DEFAULT NULL,
 `X_COORDINATE` FLOAT NULL DEFAULT NULL,
 `Y_COORDINATE` FLOAT NULL DEFAULT NULL,
 `YEAR` INT NULL DEFAULT NULL,
 `UPDATED_ON` DATETIME NULL DEFAULT NULL,
 `LATITUDE` FLOAT NULL DEFAULT NULL,
 `LONGITUDE` FLOAT NULL DEFAULT NULL,
 `LOCATION` VARCHAR(100) NULL DEFAULT NULL,
 PRIMARY KEY (`ID`))
ENGINE = InnoDB
DEFAULT CHARACTER SET = utf8mb4
COLLATE = utf8mb4_0900_ai_ci;Loaded the data set using the following query.
SET SQL_MODE=@OLD_SQL_MODE;
SET FOREIGN_KEY_CHECKS=@OLD_FOREIGN_KEY_CHECKS;
SET UNIQUE_CHECKS=@OLD_UNIQUE_CHECKS;
5. LOAD DATA LOCAL INFILE ' C:\ProgramData\MySQL\MySQL Server 8.0\Uploads\
Chicago_Project'
 INTO TABLE chicago_crime
 FIELDS TERMINATED BY ',' ENCLOSED BY ''''
 LINES TERMINATED BY '\r\n'
 IGNORE 1 ROWS;

Insights

Insight 1: the top 10 types of crime in Chicago.

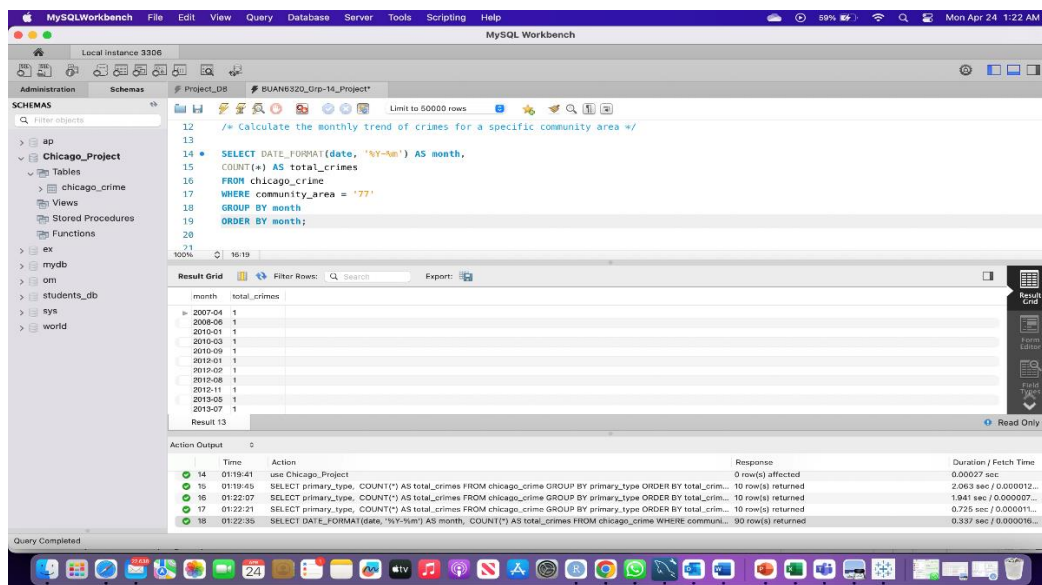


Output:

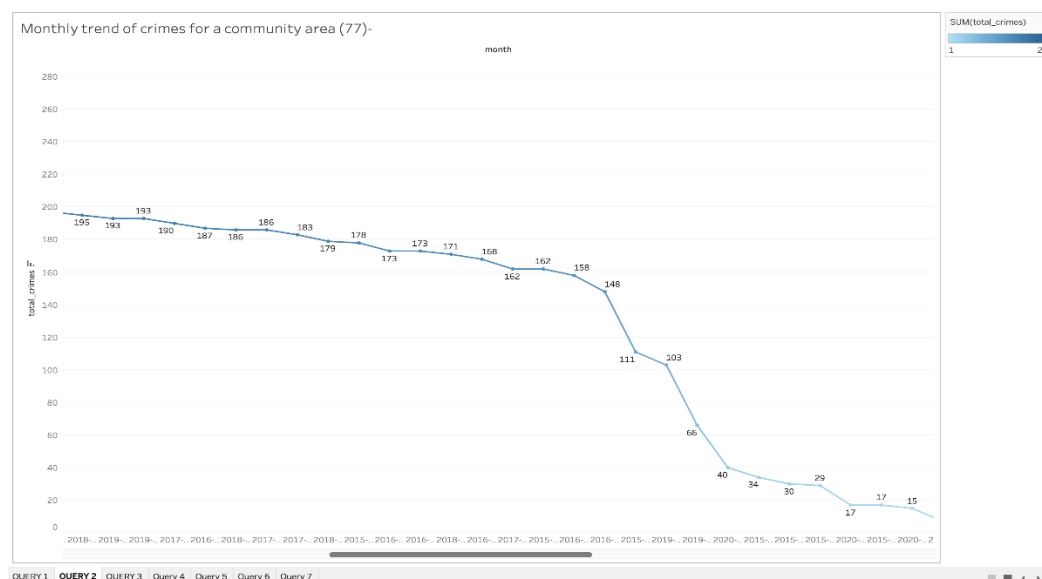


Interpretation: In the above visualization we used a bubble chart to visualize the query and from this graph we can see that theft and battery thefts have the highest crime rate and motorcycle theft is the lowest. High demand for car batteries, low risk of detection and apprehension, and inadequate deterrence measures. and since most of the people use cars in Chicago rather than cars this might have also been a reason for more battery thefts and low motor cycle thefts.

Insight 2: Monthly trend of crimes in a specific community.

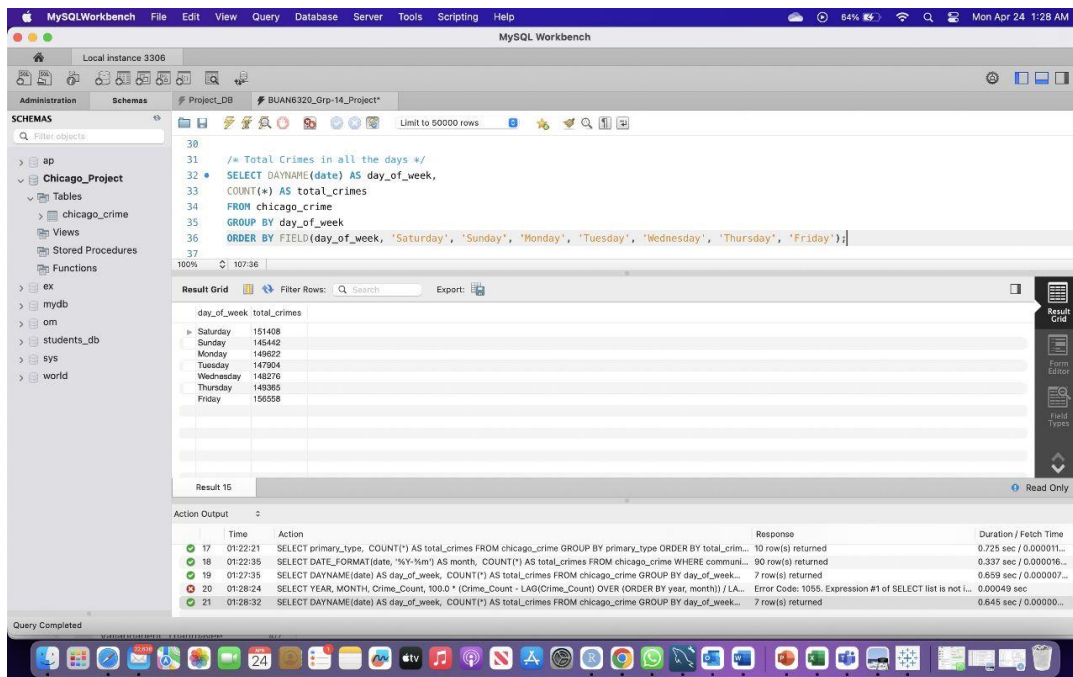


Output:

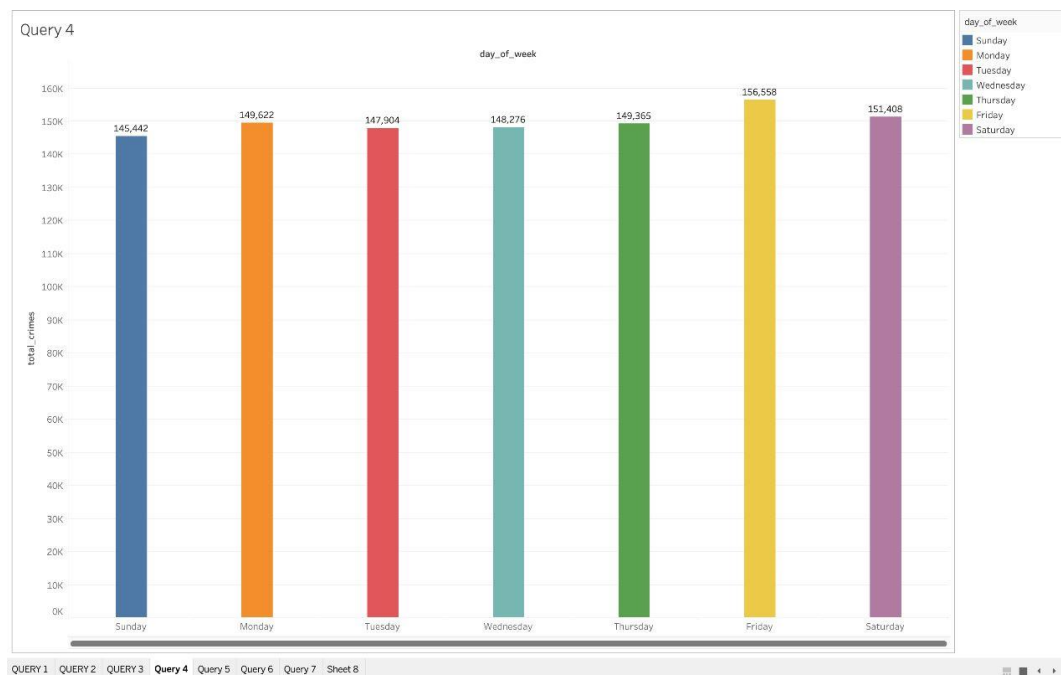


Interpretation: The second insight is regarding the monthly trend of various crimes in a community, from the line graph we can see that though there is a high crime rate in 2015 but as each month passes by the crime rate has followed a declining trend till 2022. It may be due to the increase in public safety. One of the possible reasons for this decline in crime rates could be an increase in public safety measures. This might include a range of initiatives, such as increased police presence and patrols, better community engagement, and increased public awareness of crime prevention strategies. Similarly, if there have been improvements in education or employment opportunities, this could have contributed to a reduction in crime rates.

Insight 3: Total number of crimes per day

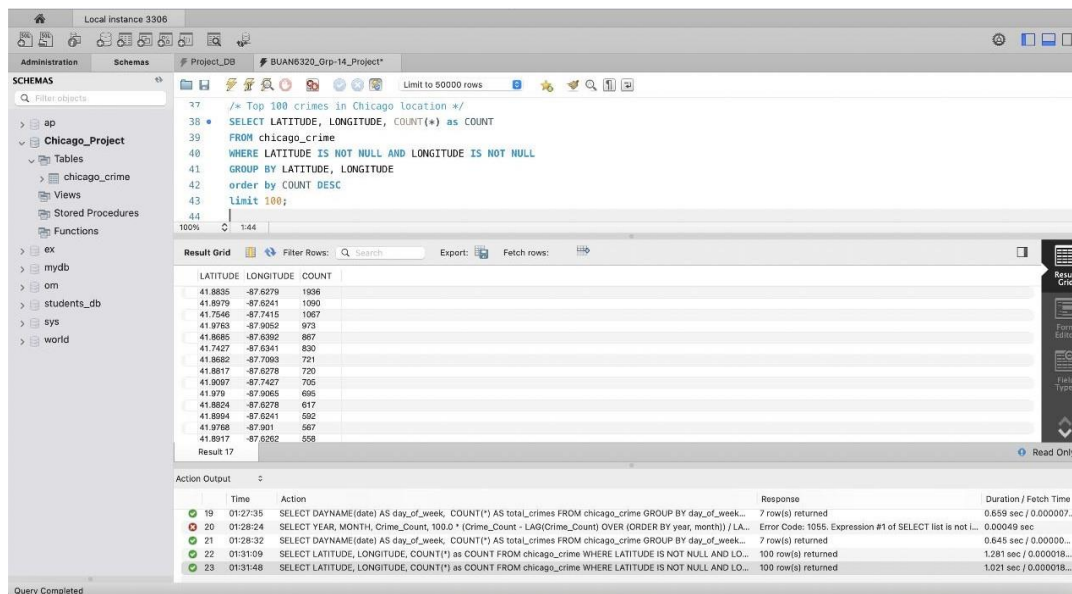


Output:



Interpretation: From the above bar graph we can see that the total number of crimes per day are approximately uniformly distributed with Thursday having the highest crimes.

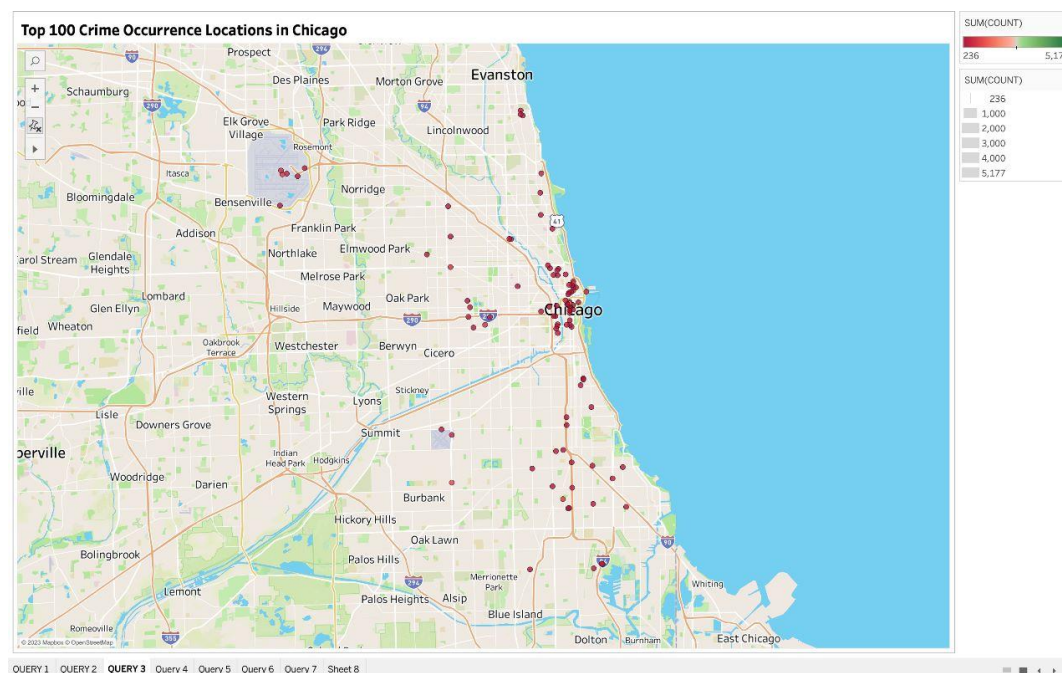
Insight 4: top 100 locations that have the highest crimes



The screenshot shows a database management tool interface. On the left, a 'SCHEMAS' panel lists various databases, with 'Chicago_Project' selected. The main area displays a SQL query: `/* Top 100 crimes in Chicago location */
SELECT LATITUDE, LONGITUDE, COUNT(*) as COUNT
FROM chicago_crime
WHERE LATITUDE IS NOT NULL AND LONGITUDE IS NOT NULL
GROUP BY LATITUDE, LONGITUDE
order by COUNT DESC
Limit 100;` Below the query, a 'Result Grid' shows the top 100 results. The columns are 'LATITUDE', 'LONGITUDE', and 'COUNT'. The first few rows are: (41.8835, -87.6279, 1936), (41.8979, -87.6241, 1090), (41.7546, -87.7415, 1067), (41.9763, -87.9052, 973), (41.8685, -87.5392, 867), (41.7427, -87.6341, 830), (41.8682, -87.7093, 721), (41.8917, -87.6278, 720), (41.8007, -87.7427, 705), (41.979, -87.9055, 695), (41.8824, -87.6278, 617), (41.8994, -87.6241, 592), (41.9768, -87.901, 587), and (41.8917, -87.6262, 558). Below the result grid, an 'Action Output' table shows the execution details of the query.

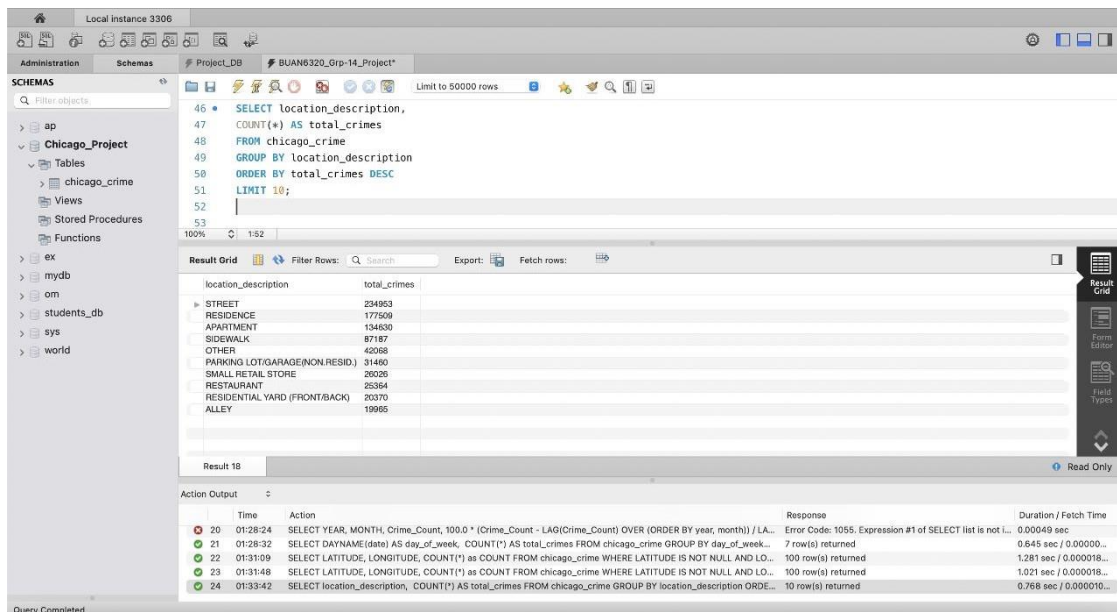
Time	Action	Response	Duration / Fetch Time
19 01:27:35	SELECT DAYNAME(date) AS day_of_week, COUNT(*) AS total_crimes FROM chicago_crime GROUP BY day_of_week...	7 row(s) returned	0.659 sec / 0.000007...
20 01:28:24	SELECT YEAR, MONTH, Crime_Count, 100.0 * (Crime_Count - LAG(Crime_Count) OVER (ORDER BY year, month)) / LA...	Error Code: 1055. Expression #1 of SELECT list is not l...	0.00049 sec
21 01:28:32	SELECT DAYNAME(date) AS day_of_week, COUNT(*) AS total_crimes FROM chicago_crime GROUP BY day_of_week...	7 row(s) returned	0.645 sec / 0.00000...
22 01:31:09	SELECT LATITUDE, LONGITUDE, COUNT(*) as COUNT FROM chicago_crime WHERE LATITUDE IS NOT NULL AND LO...	100 row(s) returned	1.281 sec / 0.000018...
23 01:31:48	SELECT LATITUDE, LONGITUDE, COUNT(*) as COUNT FROM chicago_crime WHERE LATITUDE IS NOT NULL AND LO...	100 row(s) returned	1.021 sec / 0.000018...

Output:

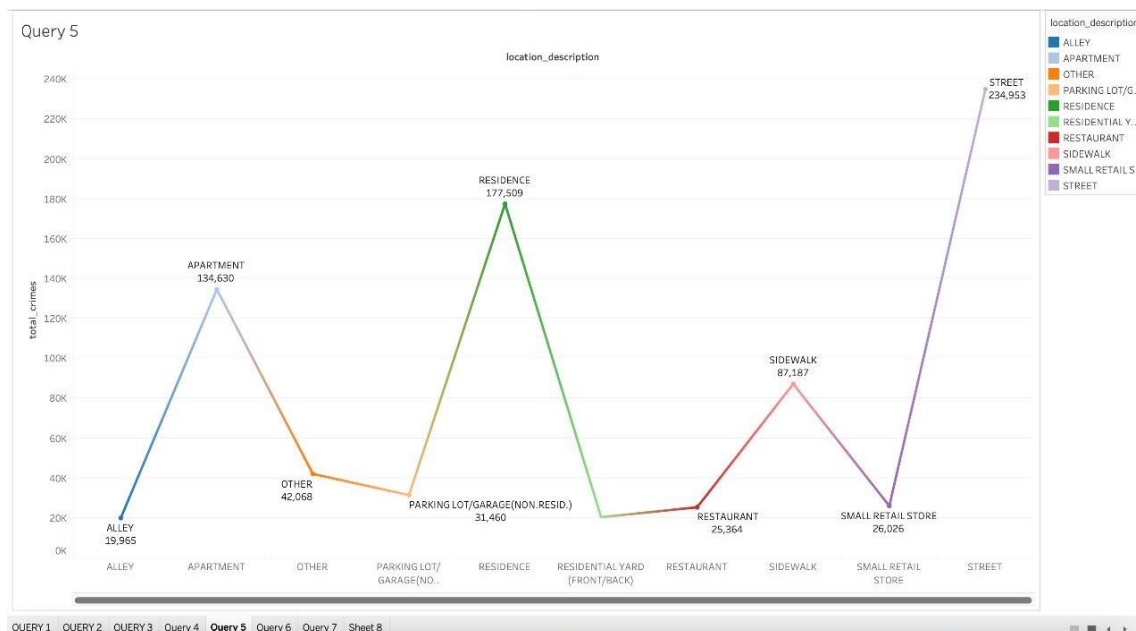


Interpretation: A spatial representation of top 100 locations that have the highest crimes in and around the Chicago city. Here we can see that majority of the crimes occur in Chicago city, we think this is also because the population density in the city is higher as compared to outskirts of the city.

Insight 5: The type of location which has the highest crimes.

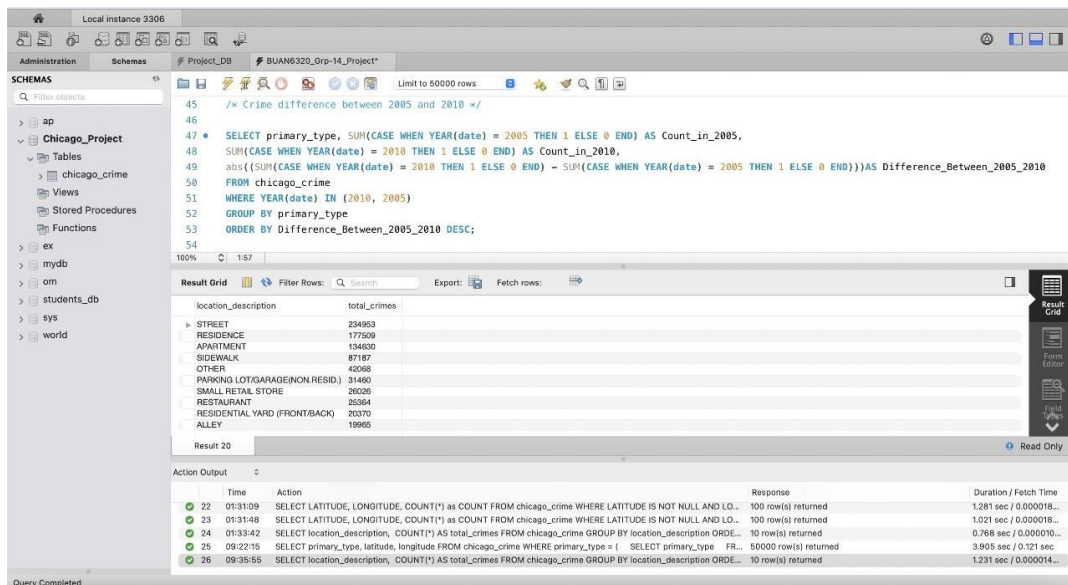


Output:

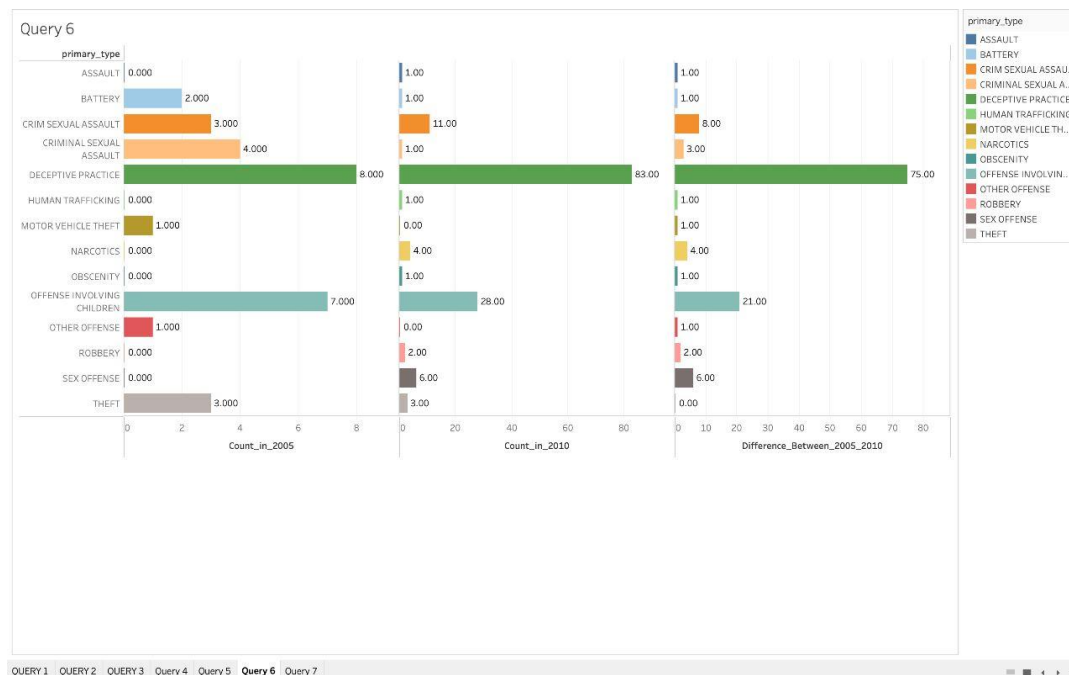


Interpretation: line graph visualizes the type of location which has the highest crimes. We can see that that highest number of crimes occur on street which is sort of expected. One interesting fact about this graph is that on average the number of crimes that occur outdoors vs that occur in private properties is about the same.

Insight 6: Graphical representation of difference in crime types in the years 2005 and 2010.

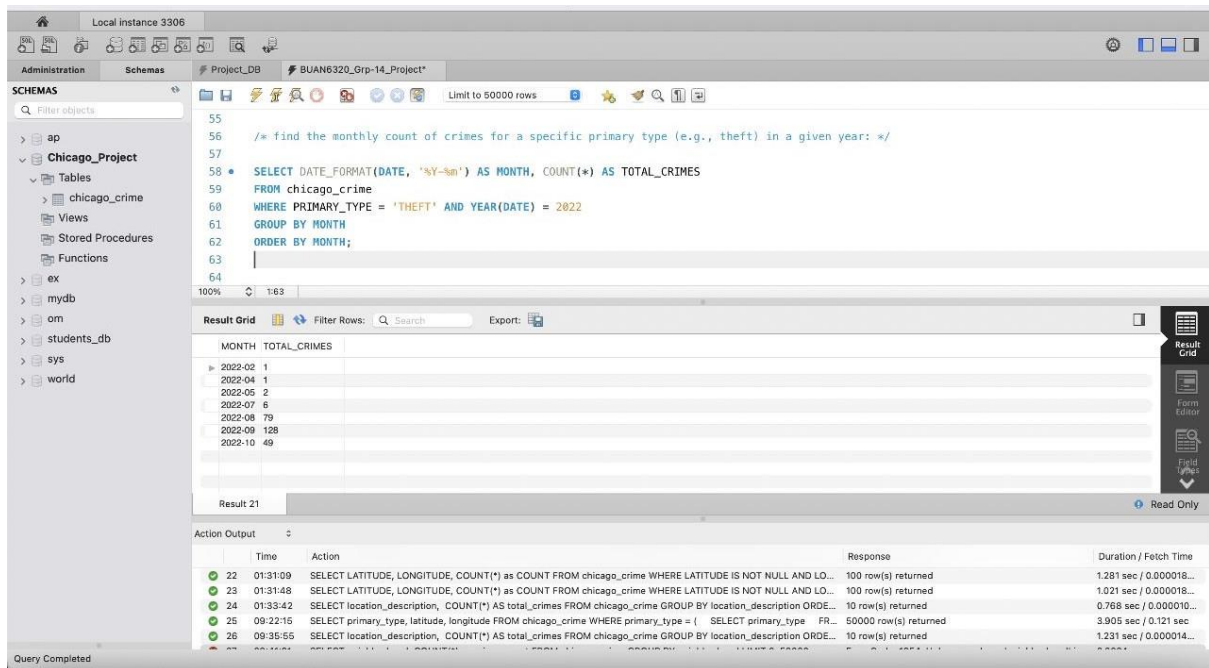


Output:



Interpretation: The above graph represents the difference in crime between year 2005 and year 2015. We can see that all the type of crime have only increase over the decade and the most increment is seen in deceptive practice.

Insight 7: theft rate in the year 2022.



The screenshot shows a database management interface with a SQL query editor and a results grid. The query is designed to find the monthly count of crimes for a specific primary type (e.g., theft) in a given year (2022). The results grid displays the following data:

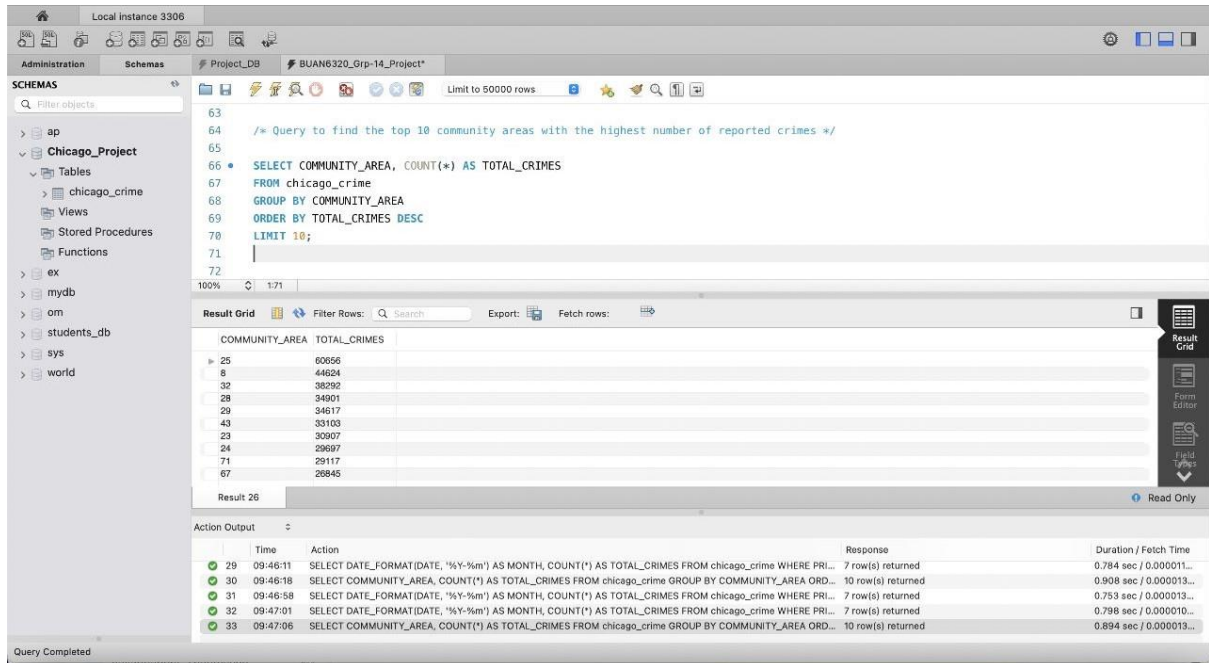
MONTH	TOTAL_CRIMES
2022-02	1
2022-04	1
2022-05	2
2022-07	6
2022-08	79
2022-09	128
2022-10	49

The Action Output section shows the execution details of the query, including the time taken and the response received.

Time	Action	Response	Duration / Fetch Time
22 01:31:09	SELECT LATITUDE, LONGITUDE, COUNT(*) AS COUNT FROM chicago_crime WHERE LATITUDE IS NOT NULL AND LO...	100 row(s) returned	1.281 sec / 0.000018...
23 01:31:48	SELECT LATITUDE, LONGITUDE, COUNT(*) AS COUNT FROM chicago_crime WHERE LATITUDE IS NOT NULL AND LO...	100 row(s) returned	1.021 sec / 0.000018...
24 01:33:42	SELECT location_description, COUNT(*) AS total_crimes FROM chicago_crime GROUP BY location_description ORDE...	10 row(s) returned	0.768 sec / 0.000010...
25 09:22:15	SELECT primary_type, latitude, longitude FROM chicago_crime WHERE primary_type = (SELECT primary_type FR...	50000 row(s) returned	3.905 sec / 0.121 sec
26 09:35:55	SELECT location_description, COUNT(*) AS total_crimes FROM chicago_crime GROUP BY location_description ORDE...	10 row(s) returned	1.231 sec / 0.000014...

Interpretation: In this insight we ran a query counting the total crimes in thousands where we grouped and ordered by month. We can see that in the month of September the rate of crime is highest. One interesting thing we observed here was that in the initial months of the year the crime rate is lower and in the later months of the year the crime rate is higher.

Insight 8: Top 10 communities having the highest number of reported crimes.



The screenshot shows a database management interface with a SQL query editor and a results grid. The query is designed to find the top 10 community areas with the highest number of reported crimes. The results grid displays the following data:

COMMUNITY_AREA	TOTAL_CRIMES
25	60656
8	44624
32	38292
28	34901
29	34617
43	33103
23	30907
24	29697
71	29117
67	26845

Below the results grid, the 'Action Output' section shows a log of database actions and their responses:

Time	Action	Response	Duration / Fetch Time
09:46:11	SELECT DATE_FORMAT(DATE, '%Y-%m') AS MONTH, COUNT(*) AS TOTAL_CRIMES FROM chicago_crime WHERE PRI...	7 row(s) returned	0.784 sec / 0.000011...
09:46:18	SELECT COMMUNITY_AREA, COUNT(*) AS TOTAL_CRIMES FROM chicago_crime GROUP BY COMMUNITY_AREA ORD...	10 row(s) returned	0.908 sec / 0.000013...
09:46:58	SELECT DATE_FORMAT(DATE, '%Y-%m') AS MONTH, COUNT(*) AS TOTAL_CRIMES FROM chicago_crime WHERE PRI...	7 row(s) returned	0.753 sec / 0.000013...
09:47:01	SELECT DATE_FORMAT(DATE, '%Y-%m') AS MONTH, COUNT(*) AS TOTAL_CRIMES FROM chicago_crime WHERE PRI...	7 row(s) returned	0.798 sec / 0.000010...
09:47:06	SELECT COMMUNITY_AREA, COUNT(*) AS TOTAL_CRIMES FROM chicago_crime GROUP BY COMMUNITY_AREA ORD...	10 row(s) returned	0.894 sec / 0.000013...

Interpretation: We have identified and ranked the top 10 communities in our city where crime rates are alarmingly high. By shedding light on these specific areas, we can better allocate resources and develop tailored interventions that address the unique challenges these communities face.

Insight 9: Top 10 most frequently occurring description of crimes with respective count and percentage.

Local instance 3306

Administration Schemas Project_DB BUAN6320_Grp-14_Project*

Limit to 50000 rows

72 /* Query to find the top 10 most frequently occurring
73 descriptions of reported crimes, along with the corresponding count and percentage: */
74
75
76 SELECT DESCRIPTION, COUNT(*) AS TOTAL_CRIMES, (COUNT(*) / (SELECT COUNT(*) FROM chicago_crime)) * 100 AS PERCENTAGE
77 FROM chicago_crime
78 GROUP BY DESCRIPTION
79 ORDER BY TOTAL_CRIMES DESC
80 LIMIT 10;
81

100% 1/75

Result Grid Filter Rows: Search Export: Fetch rows: Read Only

DESCRIPTION	TOTAL_CRIMES	PERCENTAGE
SIMPLE	114109	10.8823
\$500 AND UNDER	96717	9.2237
DOMESTIC BATTERY SIMPLE	95051	9.0648
TO VEHICLE	58670	5.5852
OVER \$500	55592	5.3970
TO PROPERTY	54746	5.2210
FROM BUILDING	39557	3.7725
RETAIL THEFT	38423	3.6643
AUTOMOBILE	36718	3.5017
FORCIBLE ENTRY	30379	2.8972

Result 30

Action Output

	Time	Action	Response	Duration / Fetch Time
37	09:48:49	SELECT WARD, LATITUDE, LONGITUDE, COUNT(*) AS TOTAL_CRIMES FROM chicago_crime GROUP BY WARD ORDER BY TOTAL_CRIMES DESC	Error Code: 1055. Expression #2 of SELECT list is not...	0.00034 sec
38	09:49:19	SELECT WARD, LATITUDE, LONGITUDE, COUNT(*) AS TOTAL_CRIMES FROM chicago_crime ORDER BY TOTAL_CRIMES DESC	Error Code: 1140. In aggregated query without GROUP BY...	0.00034 sec
39	09:49:35	SELECT ARREST, COUNT(*) AS TOTAL_CRIMES FROM chicago_crime WHERE PRIMARY_TYPE = 'ASSAULT' GROUP BY ARREST	1 row(s) returned	0.728 sec / 0.000010...
40	09:50:16	SELECT DOMESTIC, COUNT(*) AS TOTAL_CRIMES FROM chicago_crime WHERE PRIMARY_TYPE = 'THEFT' GROUP BY DOMESTIC	1 row(s) returned	0.875 sec / 0.000010...
41	09:51:04	SELECT DESCRIPTION, COUNT(*) AS TOTAL_CRIMES, (COUNT(*) / (SELECT COUNT(*) FROM chicago_crime)) * 100 AS PERCENTAGE	10 row(s) returned	2.334 sec / 0.000011...

Query Completed

Interpretation: This analysis dives deep into the most common types of crimes occurring in our city, with each crime's respective count and percentage. Armed with this knowledge, we can focus on addressing the root causes of these offenses and developing targeted prevention programs that resonate with the city's most pressing criminal concerns.

Insight 10: Top 10 wards having the highest crimes in the city.

The screenshot displays a database management interface with a SQL query editor and a results grid. The query is as follows:

```
80 LIMIT 10;  
81  
82  
83 • SELECT ward AS WARD_NUMBER, COUNT(*) AS TOTAL_CRIMES  
84 FROM chicago_crime  
85 WHERE YEAR(DATE) = 2022  
86 GROUP BY ward  
87 ORDER BY COUNT(*) DESC  
88 LIMIT 10;  
89
```

The results grid shows the top 10 wards by crime count for the year 2022:

WARD_NUMBER	TOTAL_CRIMES
27	33
42	31
44	27
28	25
3	24
17	22
5	20
9	20
50	18
4	18

The interface also includes a schema tree on the left, a query log at the bottom, and a status bar at the bottom right.

Interpretation: By pinpointing the wards with the highest crime rates, we empower local leaders and law enforcement to collaborate on designing initiatives that specifically cater to the needs of their communities. This data-driven approach enables us to create safer neighbourhoods, one ward at a time.

Conclusion

In conclusion, our comprehensive analysis of the Chicago crime dataset has revealed crucial insights into crime patterns, hotspots, and trends. Our findings, which include the most prevalent crime types, spatial and temporal trends, and affected communities and wards, will serve as a valuable resource for stakeholders, policymakers, and law enforcement agencies. By leveraging this data-driven understanding, we can collectively work towards addressing public safety challenges and fostering a safer environment for all Chicago residents.

References

- "Crimes - 2001 to present - Dashboard" dataset
<https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present-Dashboard/5cd6-ry5gt>