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

Project Description:

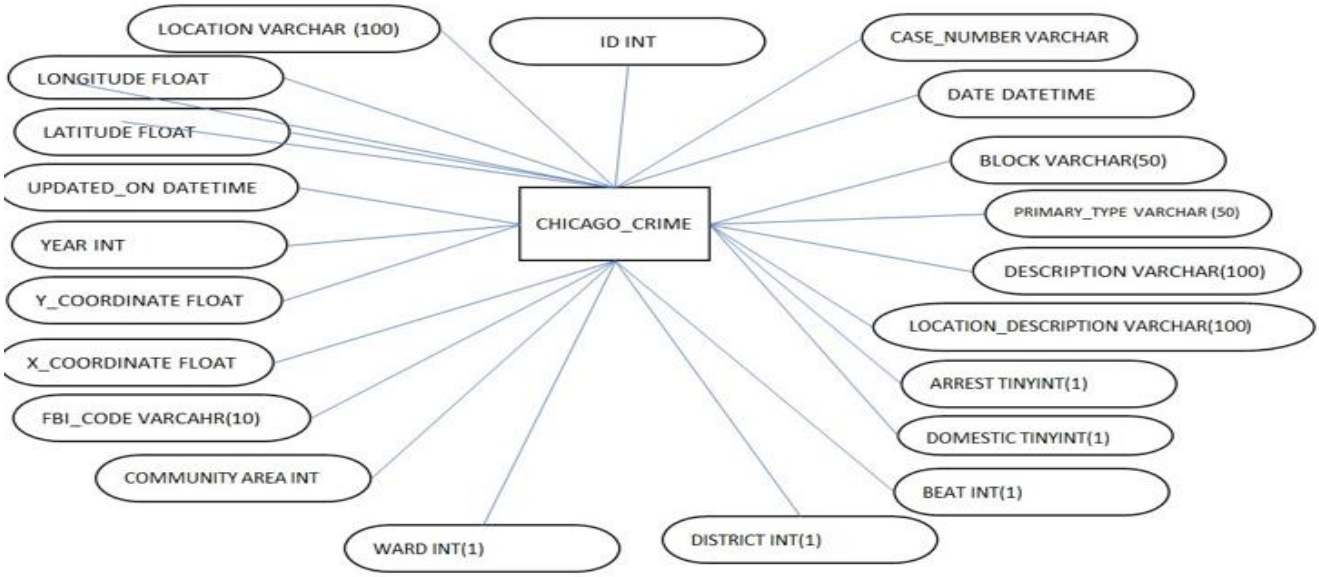
The "Crimes - 2001 to present - Dashboard" dataset is a comprehensive collection of reported crime incidents in the city of Chicago, Illinois, USA, from 2001 to the present. This dataset is maintained by the Chicago Police Department and regularly updated to reflect the most recent information. The data 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.

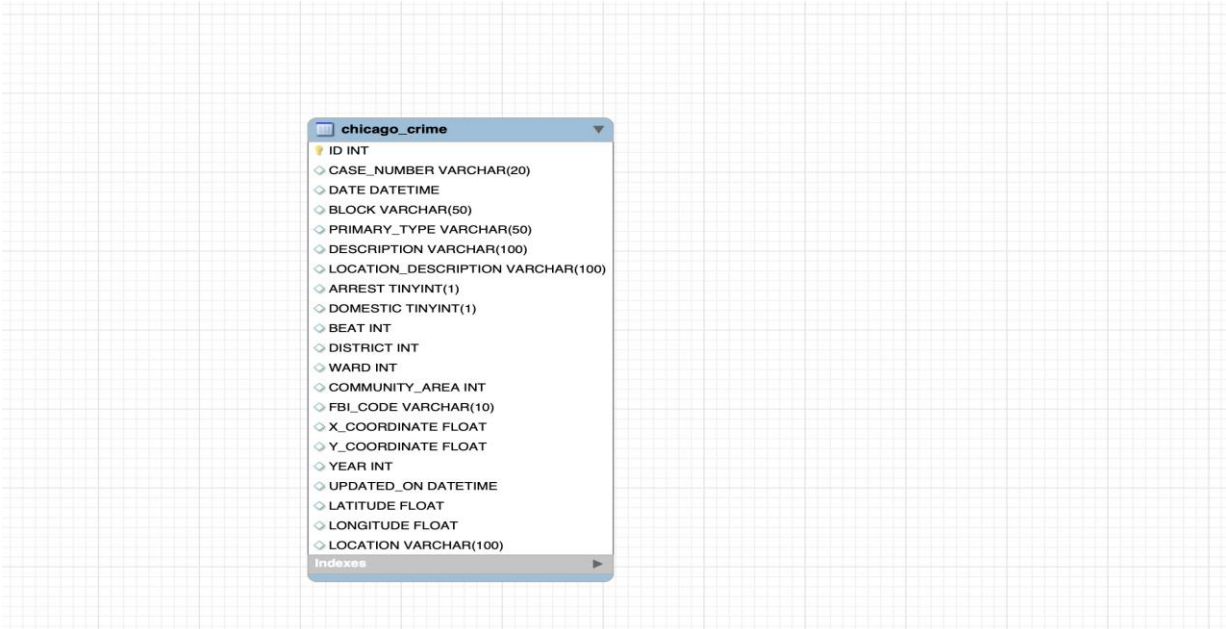
Data Dictionary:

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



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

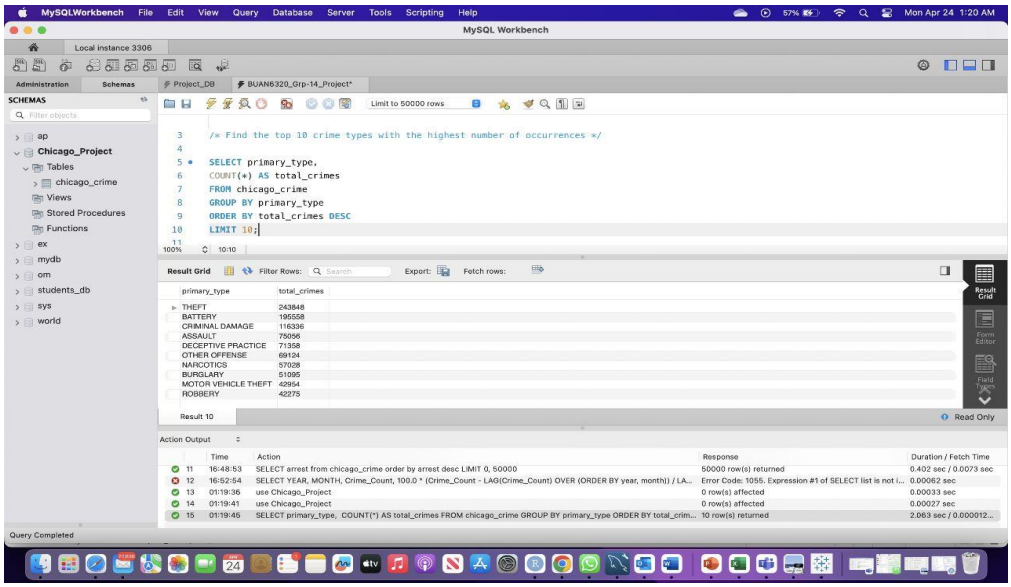
Created tables using commands:

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;

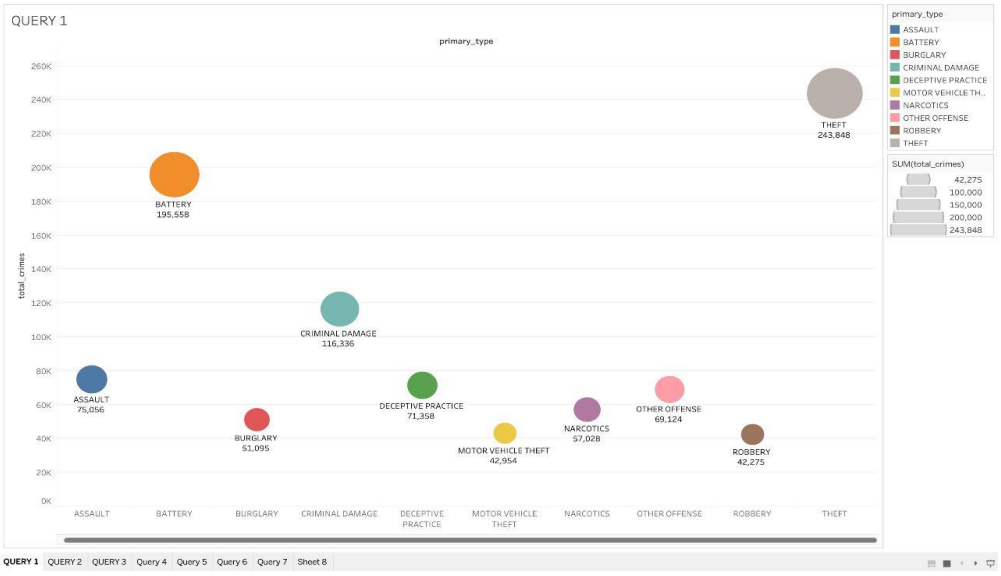
5. LOAD DATA LOCAL INFILE ' C:\ProgramData\MySQL\MySQL Server 8.0\Uploads\ Chicago_Project'
   INTO TABLE chicago_crime
   FIELDS TERMINATED BY ',' ENCLOSED BY '"'
   LINES TERMINATED BY '\r\n'
   IGNORE 1 ROWS;

Insights

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

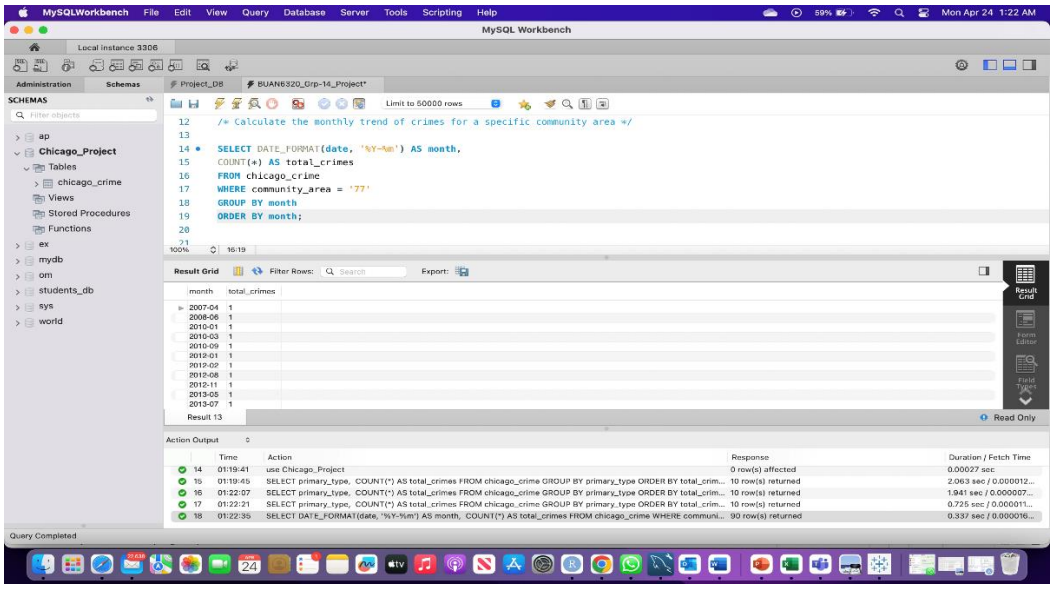


Output:

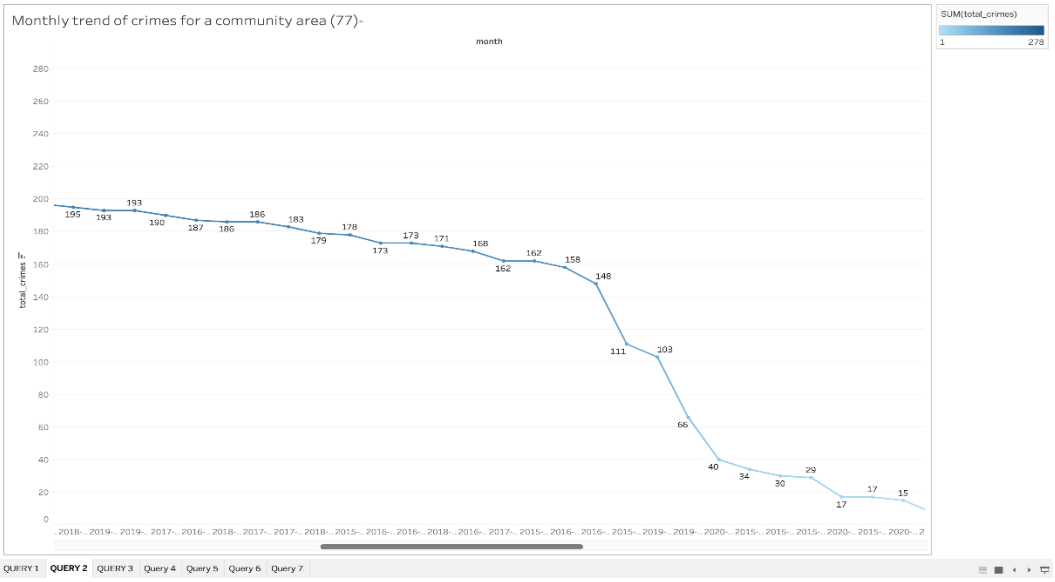


Interpretation: In the above visualization we used a bubble chart to visualize the query and from this graph we can see that theft and battery thefts have the highest crime rate and motorcycle theft is the lowest. High demand for car batteries, low risk of detection and apprehension, and inadequate deterrence 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.

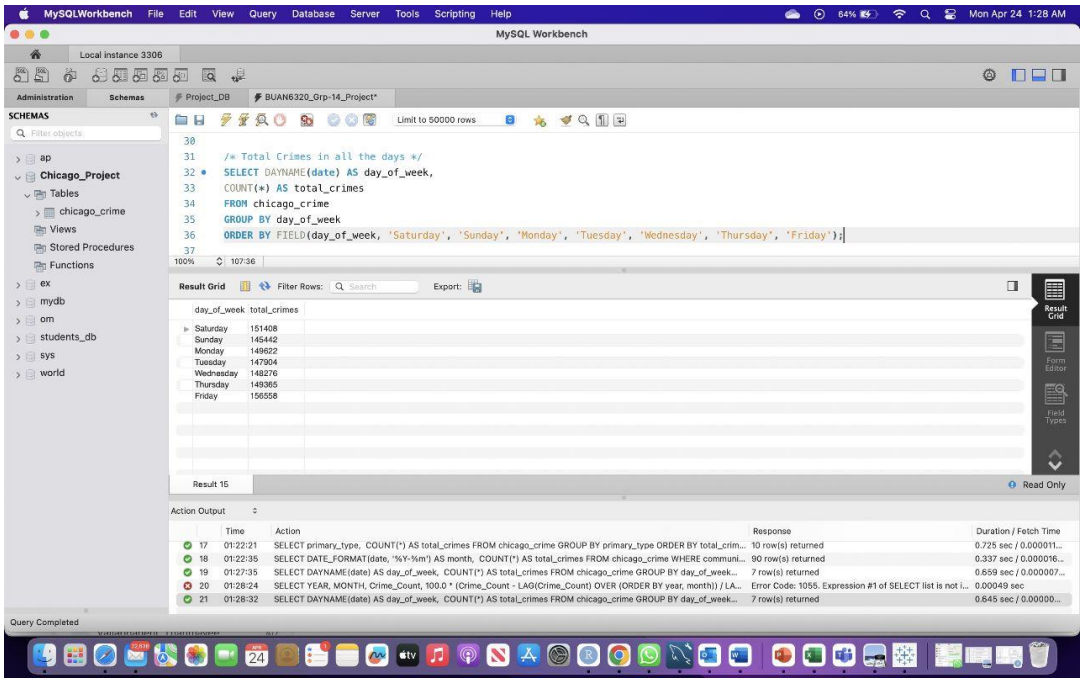


Output:

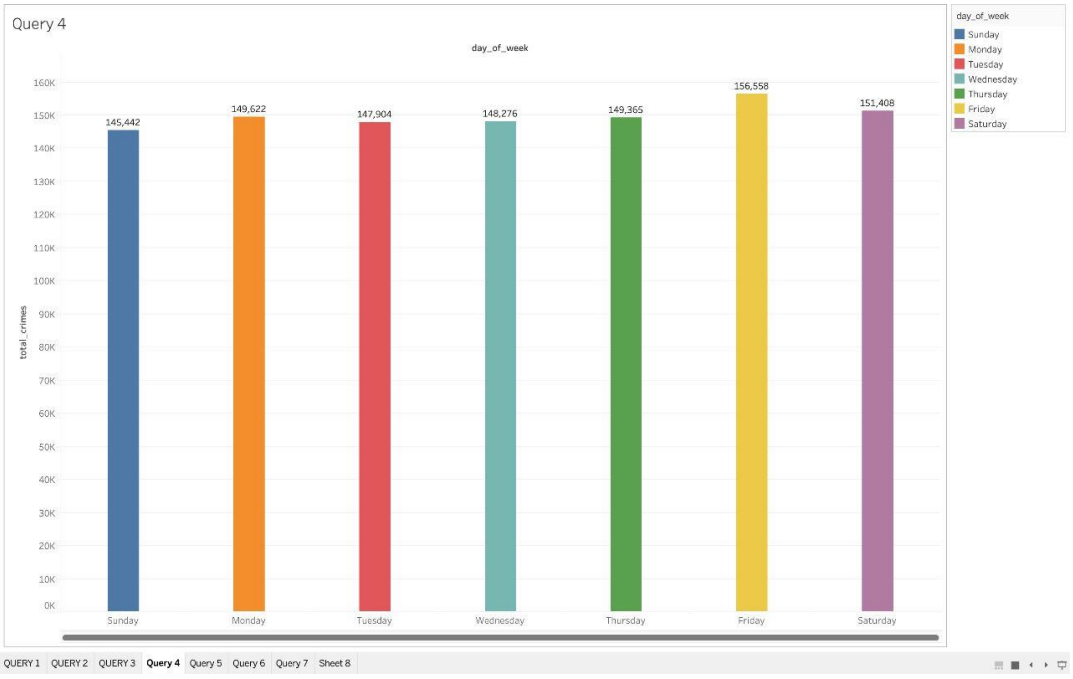


Interpretation: The second insight is regarding the monthly trend of various crimes in a community, from the line graph we can see that though there is a high crime rate in 2015 but as each month passes by the crime rate has followed a declining trend till 2022. It may be due to the increase in public 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

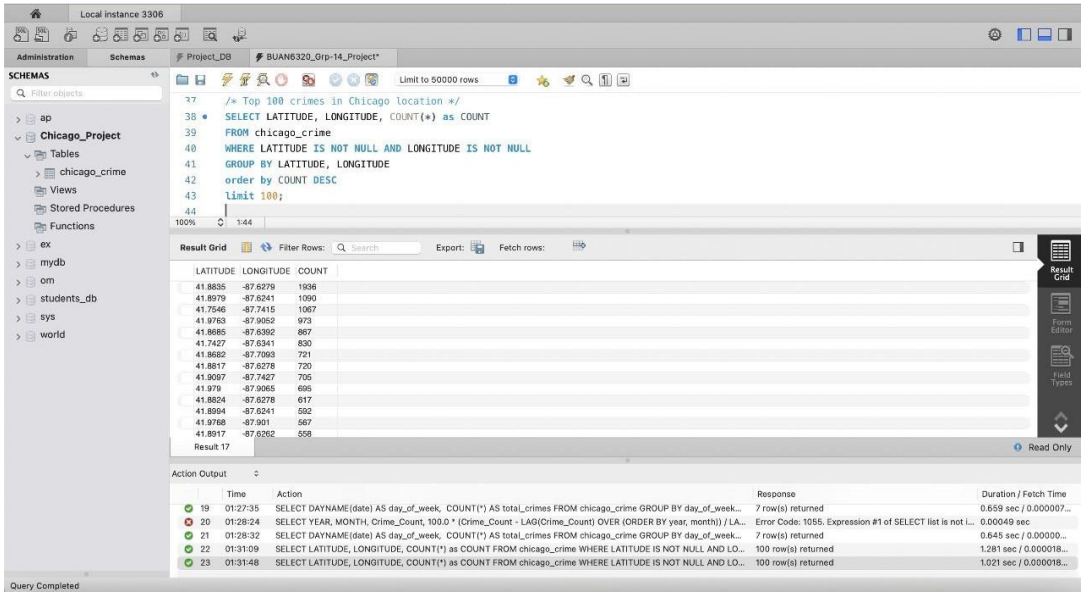


Output:

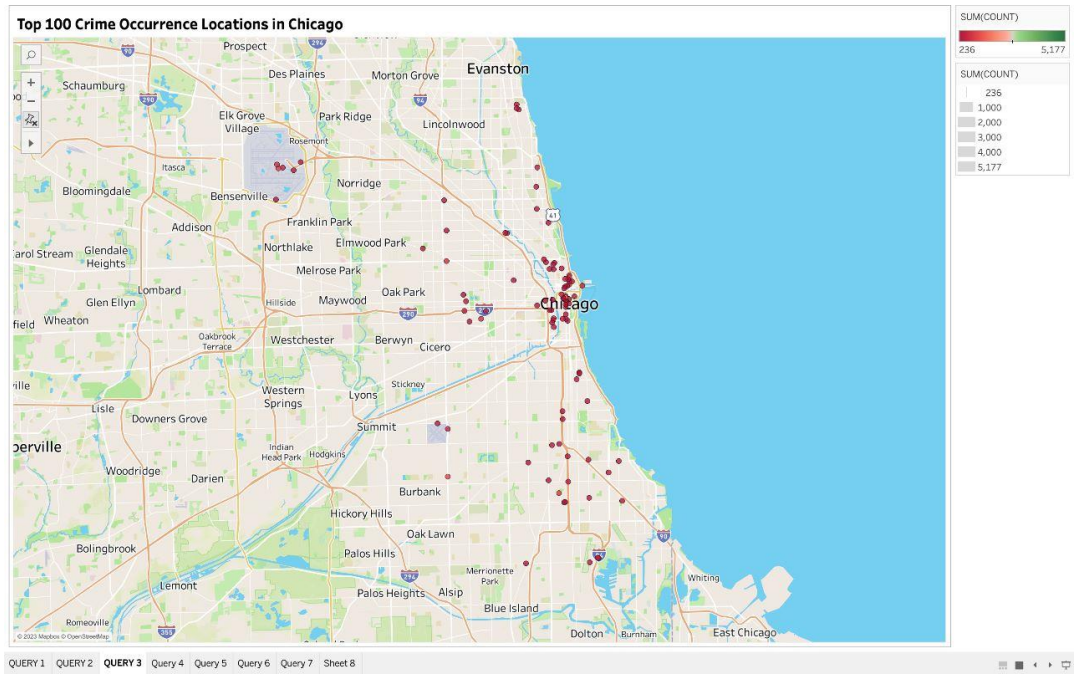


Interpretation: From the above bar graph we can see that the total number of crimes per day are approximately uniformly distributed with Thursday having the highest crimes.

Insight 4: top 100 locations that have the highest crimes

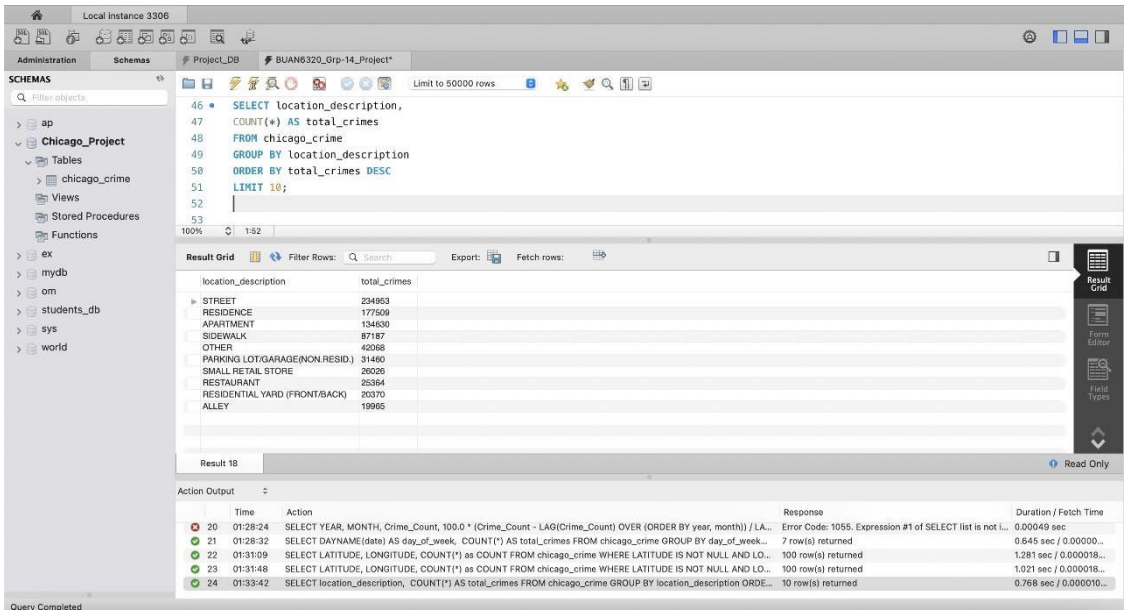


Output:

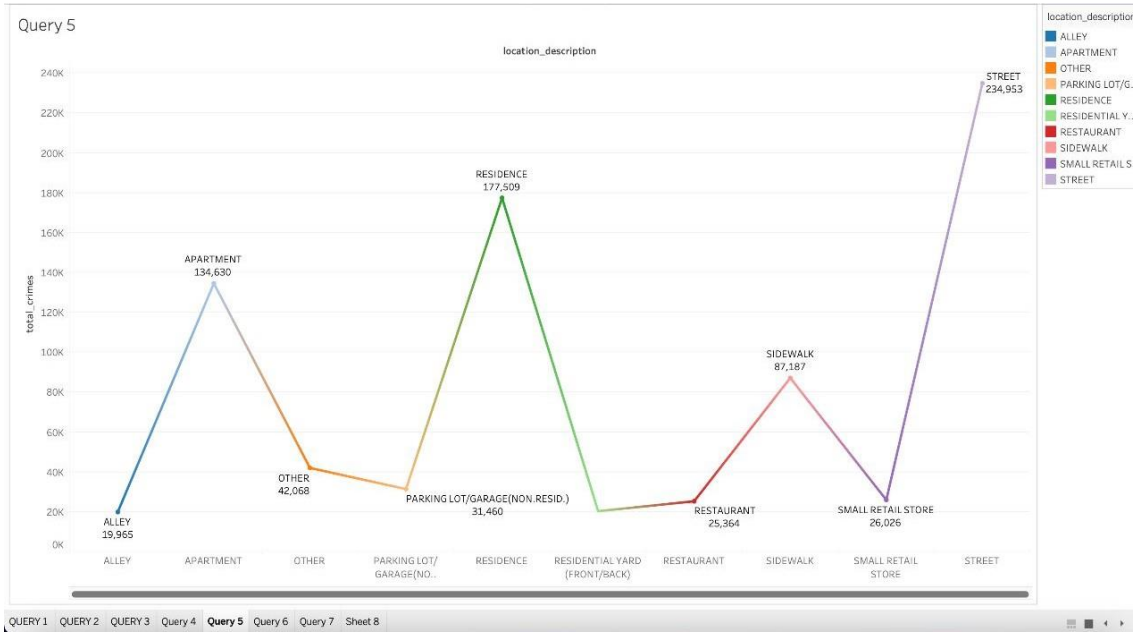


Interpretation: A spatial representation of top 100 locations that have the highest crimes in and around the Chicago city. Here we can see that majority of the crimes occur in Chicago city, we think this is also because the population density in the city is higher as compared to outskirts of the city.

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

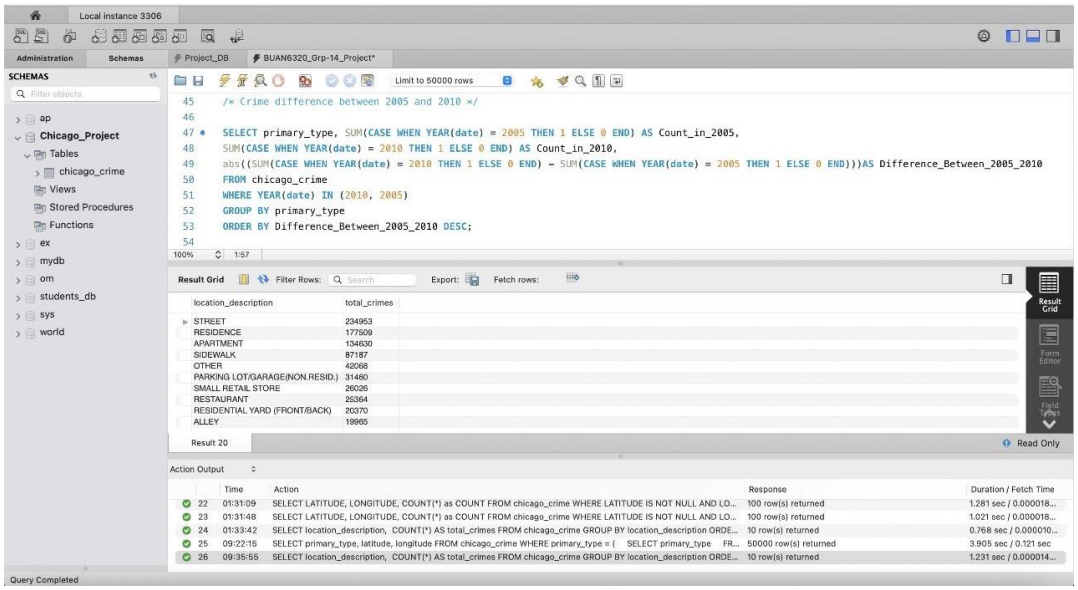


Output:

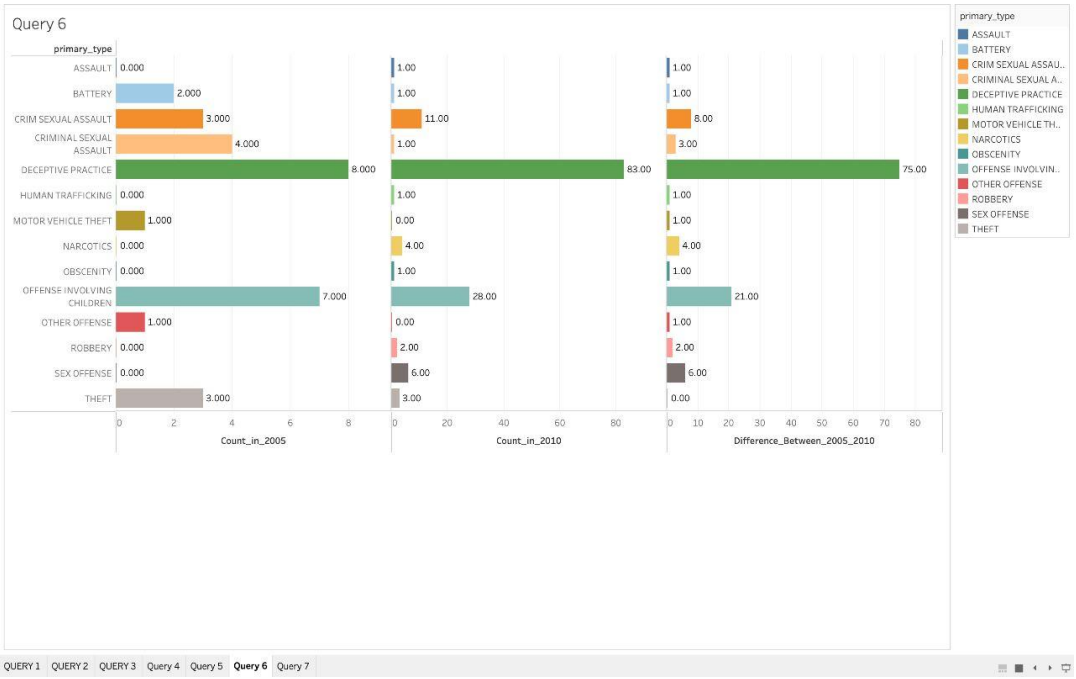


Interpretation: line graph visualizes the type of location which has the highest crimes. We can see that that highest number of crimes occur on street which is sort of expected. One interesting fact about this graph is that on average the number of crimes that occur outdoors vs that occur in private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

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

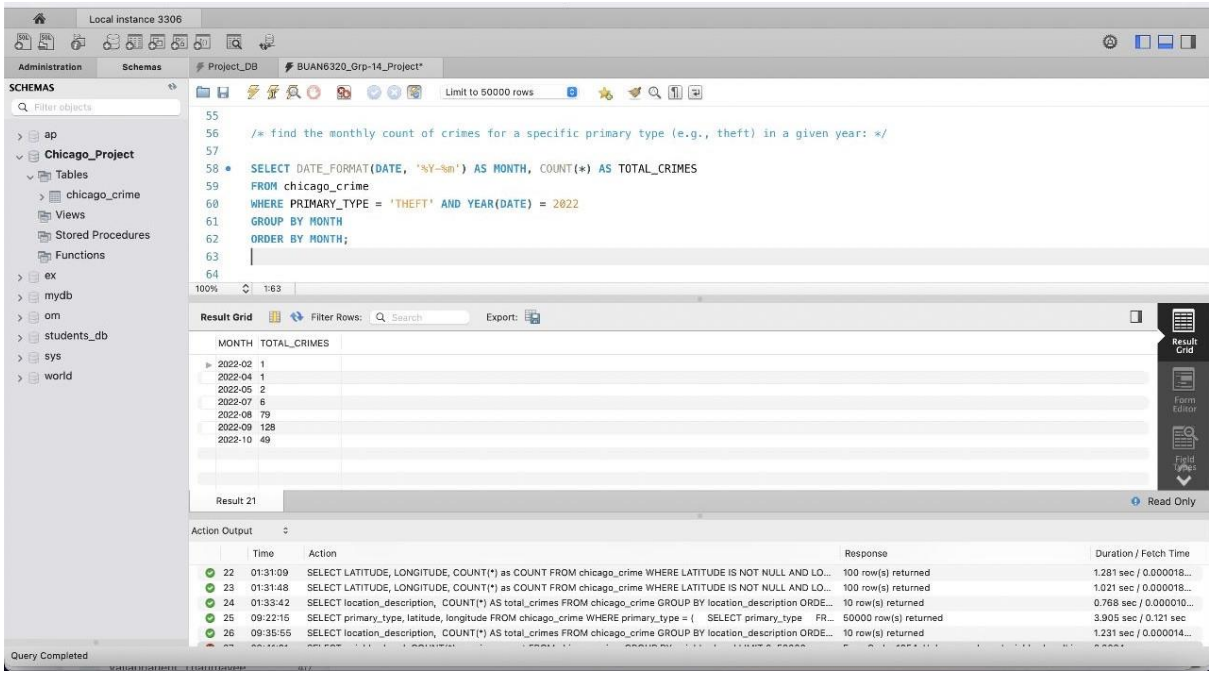


Output:



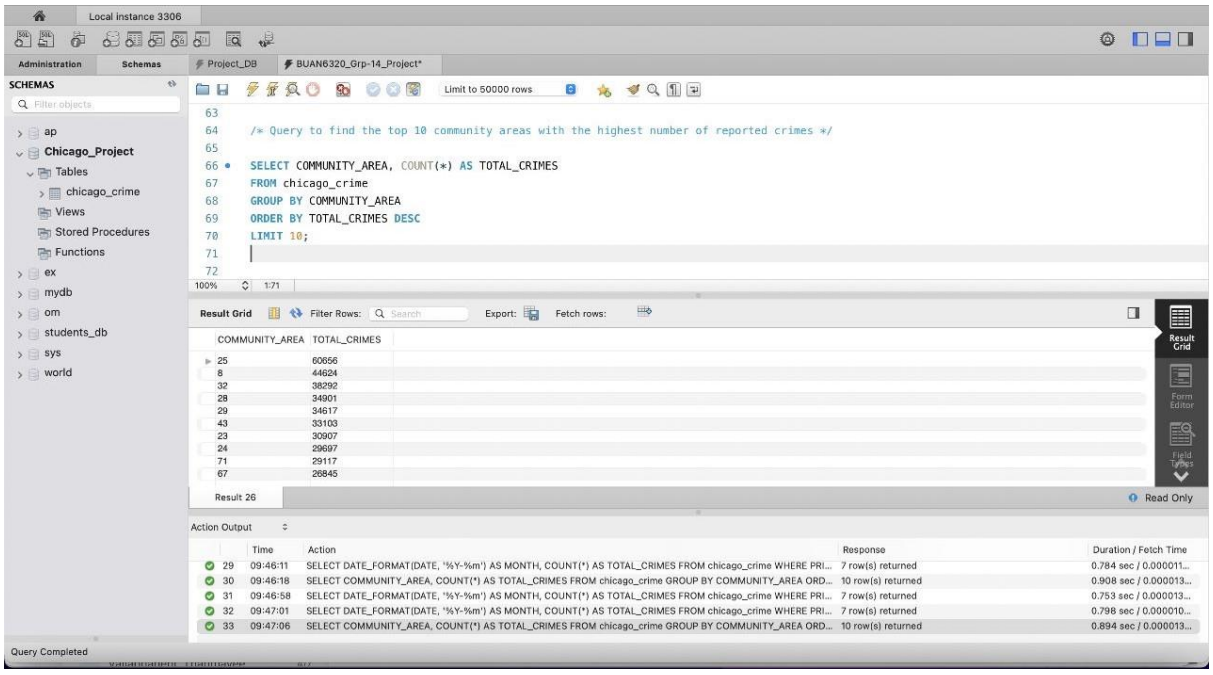
Interpretation: The above graph represents the difference in crime between year 2005 and year 2015. We can see that all the type of crime have only increase over the decade and the most increment is seen in deceptive practice.

Insight 7: theft rate in the year 2022.



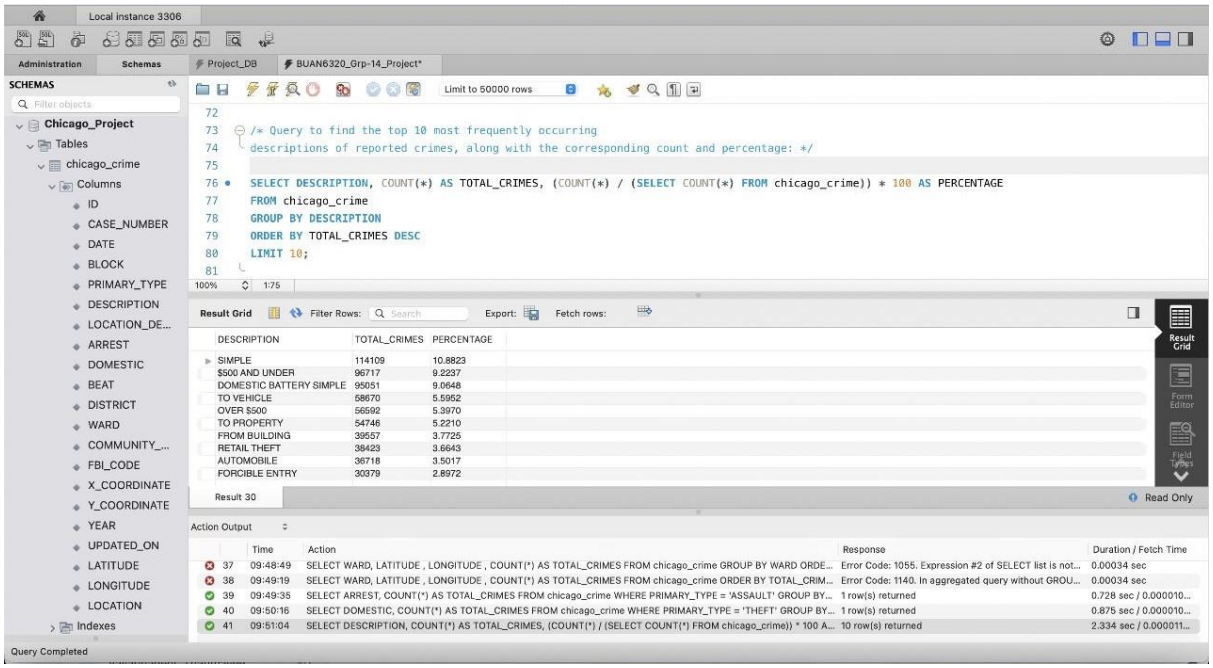
Interpretation: In this insight we ran a query counting the total crimes in thousands where we grouped and ordered by month. We can see that in the month of September the rate of crime is highest. One interesting thing we observed here was that in the initial months of the year the crime rate is lower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



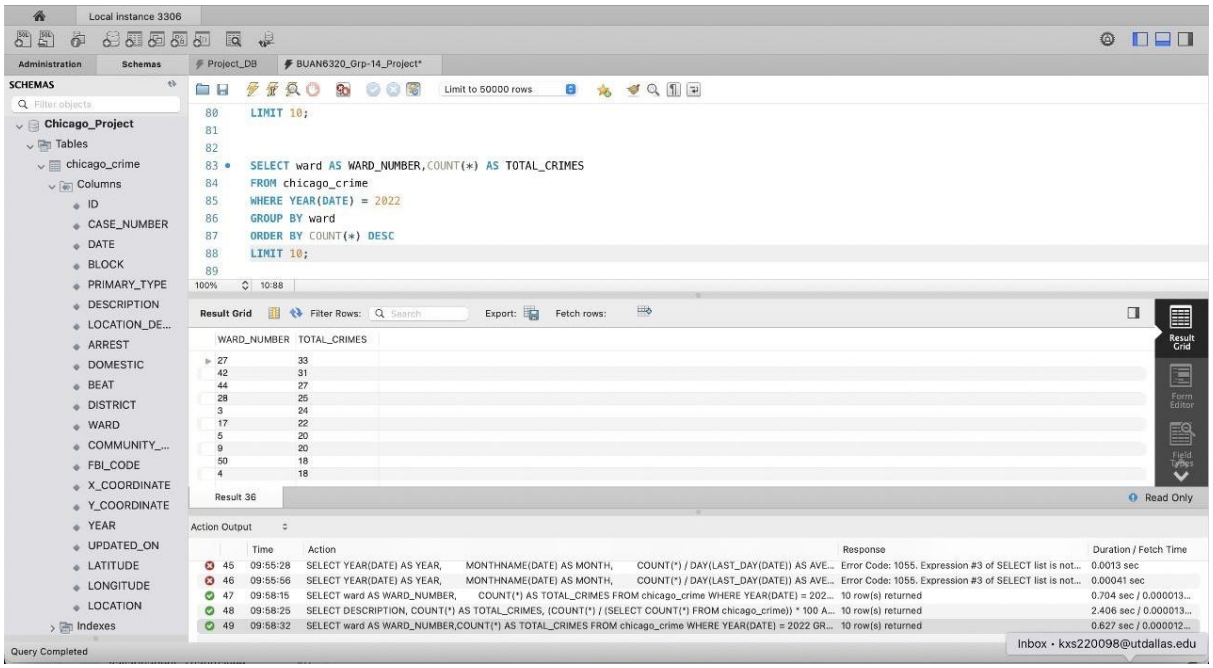
Interpretation: We have identified and ranked the top 10 communities in our city where crime rates are alarmingly high. By shedding light on these specific areas, we can better allocate resources and develop tailored interventions that address the unique challenges these communities face.

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



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.



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