FSPAN

Austin Crime Report

By: Fariha Syed

What to expect...

1. **Business Objective**
2. **Team Intro / Strategy**
3. **Dataset Attributes & Data Loading**
4. **Model Design Evolution**
5. **Visualizations & Insights**

FSPAN

Business Objective

New tech-startup seeking to relocate their **HQ** to **Austin, TX**

Investors requested a market analysis to identify a **safe**, **low-risk** area for HQ

Strategic team enabled based on the **business requirements**

Launch initiative and execute the strategy to provide **final recommendation**

FSPAN

Our Strategy

Overall: data-centric based on trusted and reliable public information

Goal: identify Top 5 “safe” zip_codes and relevant statistics

Data: 2003-2021 Austin Crime Report - accurate and reliable information

Parameters: $20k budget, multi-skilled team, 30-day turnaround

FSPAN

Dataset Attributes

Dataset includes **crime rates**, **trends**, and **patterns** various zip codes throughout Austin

The data file consists of **2,348,000 records** and **27 columns**

Total number of missing values: **591,629**. That's about **25.2%** of all the data

Crimes range from Theft, Burglary to Family Violence.

```
> names(Crime_Reports)
 [1] "Incident Number"              "Highest Offense Description" "Highest Offense Code"
 [4] "Family Violence"              "Occurred Date Time"          "Occurred Date"
 [7] "Occurred Time"                "Report Date Time"            "Report Date"
[10] "Report Time"                  "Location Type"               "Address"
[13] "Zip Code"                     "Council District"            "APD Sector"
[16] "APD District"                 "PRA"                         "Census Tract"
[19] "Clearance Status"             "Clearance Date"              "UCR Category"
[22] "Category Description"         "X-coordinate"                "Y-coordinate"
[25] "Latitude"                     "Longitude"                   "Location"
```

FSPAN

Data Loading Concepts / SW

Tableau | Power BI

MySQL | MySQL Workbench

```
CREATE TABLE `crime_report` (
  `incident_no` varchar(45) DEFAULT NULL,
  `highest_offense_description` varchar(45) DEFAULT NULL,
  `highest_offense_code` int DEFAULT NULL,
  `family_violence` enum('Y','N') DEFAULT NULL,
  `occurred_datetime` varchar(20) DEFAULT NULL,
  `occurred_date` varchar(10) DEFAULT NULL,
  `occurred_time` varchar(5) DEFAULT NULL,
  `report_datetime` varchar(20) DEFAULT NULL,
  `report_date` varchar(10) DEFAULT NULL,
  `report_time` varchar(5) DEFAULT NULL,
  `location_type` varchar(45) DEFAULT NULL,
  `address` varchar(45) DEFAULT NULL,
  `zip_code` varchar(5) DEFAULT NULL,
  `council_district` varchar(2) DEFAULT NULL,
  `apd_sector` varchar(2) DEFAULT NULL,
  `apd_district` varchar(2) DEFAULT NULL,
  `report_id` int NOT NULL,
  `census_tract` decimal(5,2)  DEFAULT NULL,
  `clearance_status` enum('C','N') DEFAULT NULL,
  `clearance_date` varchar(10) DEFAULT NULL,
  `UCR_category` varchar(3) DEFAULT NULL,
  `category_description` varchar(45) DEFAULT NULL,
  `x_coordinate` int DEFAULT NULL,
  `y_coordinate` int DEFAULT NULL,
  `latitude` decimal(10,8) DEFAULT NULL,
  `longitude` decimal(10,8) DEFAULT NULL,
  `location` varchar(45) DEFAULT NULL
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci;
```

```
UPDATE crime_report
 SET report_date = STR_TO_DATE(report_date, '%m/%d/%Y') WHERE report_date LIKE '__/__/___';
 UPDATE crime_report
 SET occurred_date = STR_TO_DATE(occurred_date, '%m/%d/%Y') WHERE occurred_date LIKE '__/__/___';
 UPDATE crime_report
 SET clearance_date = STR_TO_DATE(clearance_date, '%m/%d/%Y') WHERE clearance_date LIKE '__/__/___';

UPDATE crime_report
 SET report_time = LPAD(report_time, 4, '0')
 WHERE LENGTH(report_time) < 4;
 UPDATE crime_report
 SET report_time = DATE_FORMAT(STR_TO_DATE(report_time, '%H%i'), '%H:%i');

UPDATE crime_report
 SET occurred_time = LPAD(occurred_time, 4, '0')
 WHERE LENGTH(occurred_time) < 4;
 UPDATE crime_report
 SET occurred_time = DATE_FORMAT(STR_TO_DATE(occurred_time, '%H%i'), '%H:%i');
```

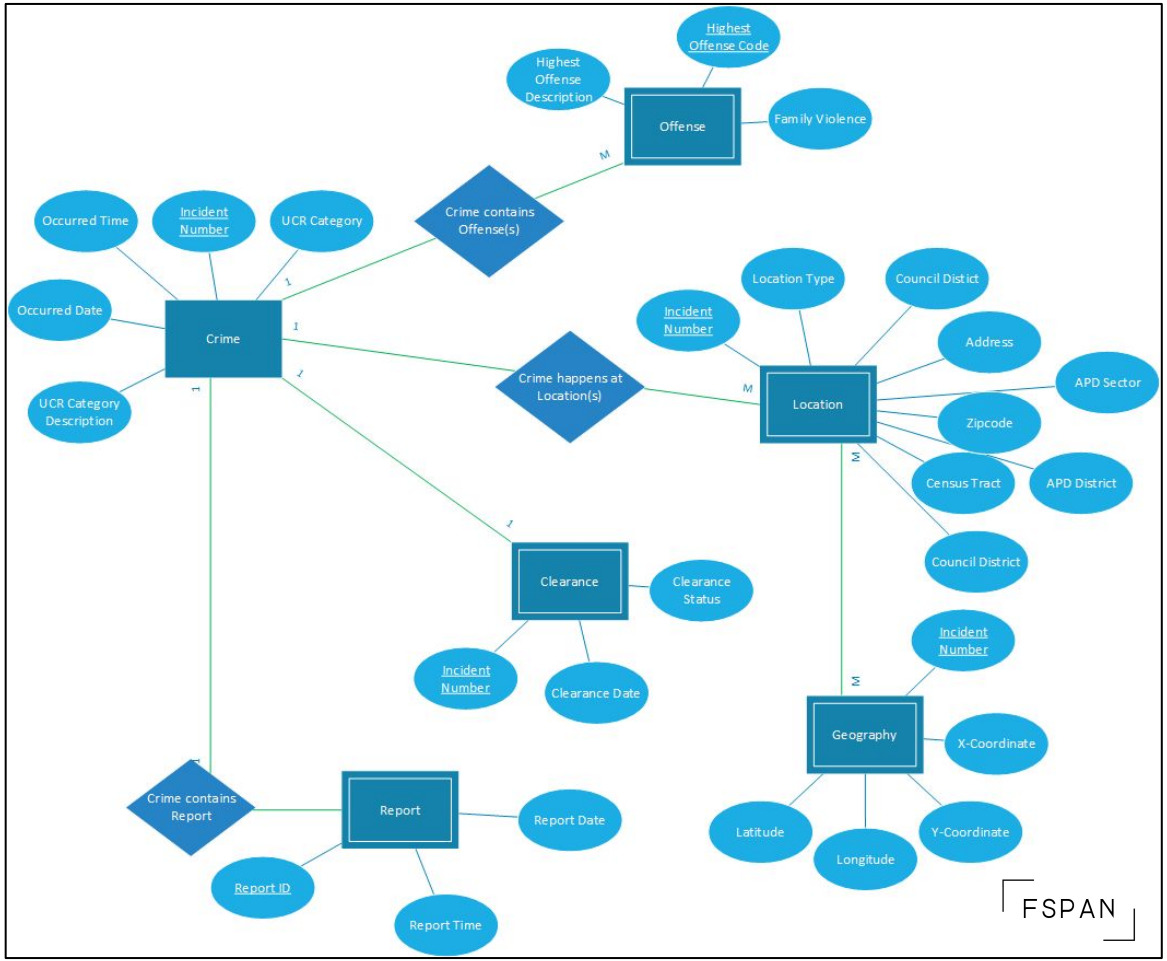
FSPAN

Chen's Model

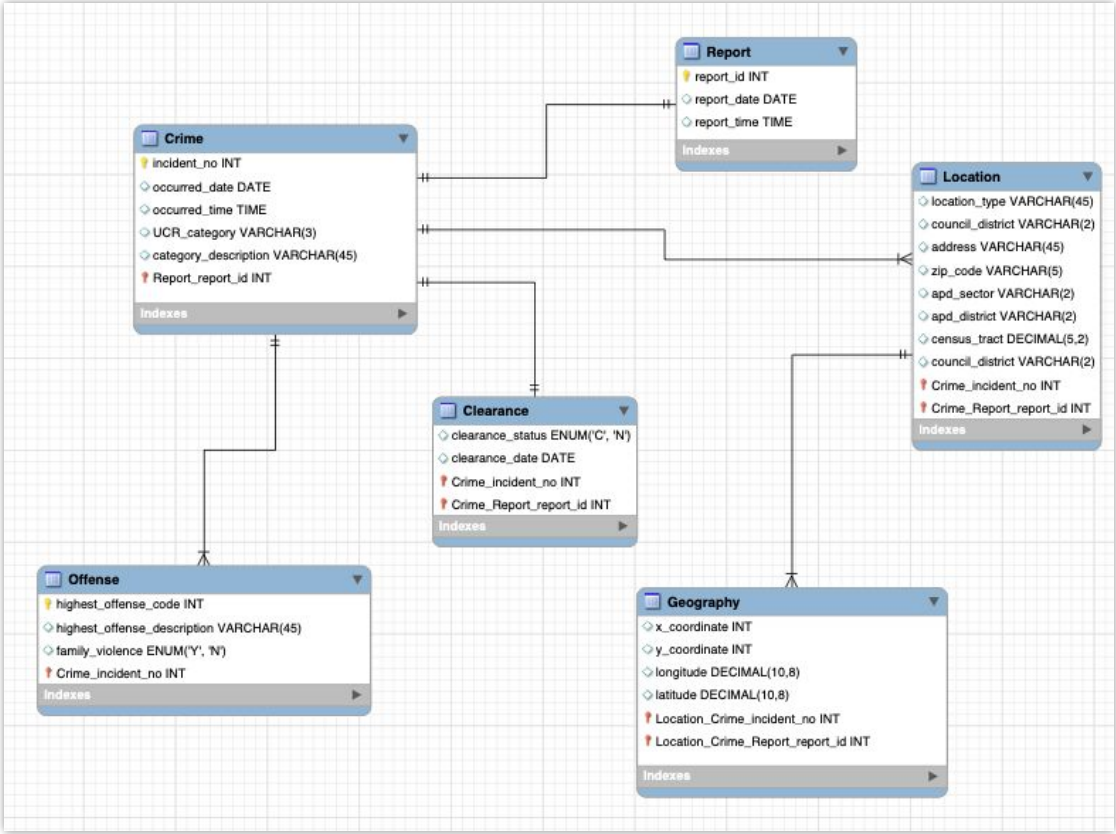
A crime can have many offenses, but one offense can be associated with one crime **(1 to M)**

A type of crime can happen at many locations, but one specific location can have one type of crime **(1 to M)**

A crime can contain multiple reports, and a report can have multiple types of crimes **(1 to M)**



Physical Design



FSPAN

Insight 1: Has crime decreased or increased over the years?

SQL Query: queried the number of incidents by occurrence date (year)

Data Visualization: plotted a line-graph chart by total incidents by month and year

Results: Austin reached a peak in 2009 although data visualization shows that in a ten year period after 2009, crime **decreased overall at a steady rate**

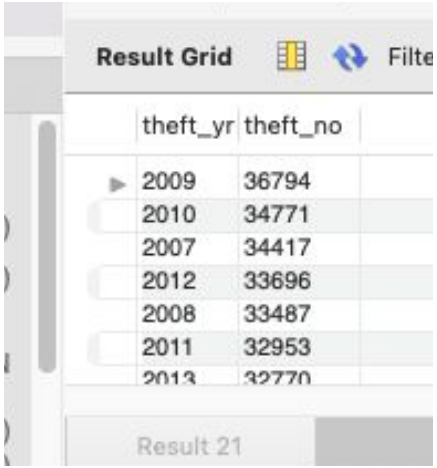
Recommendation: based on our analysis, decreasing crime rates would indicate that **Austin to be an attractive location for HQ**

FSPAN

SQL Used

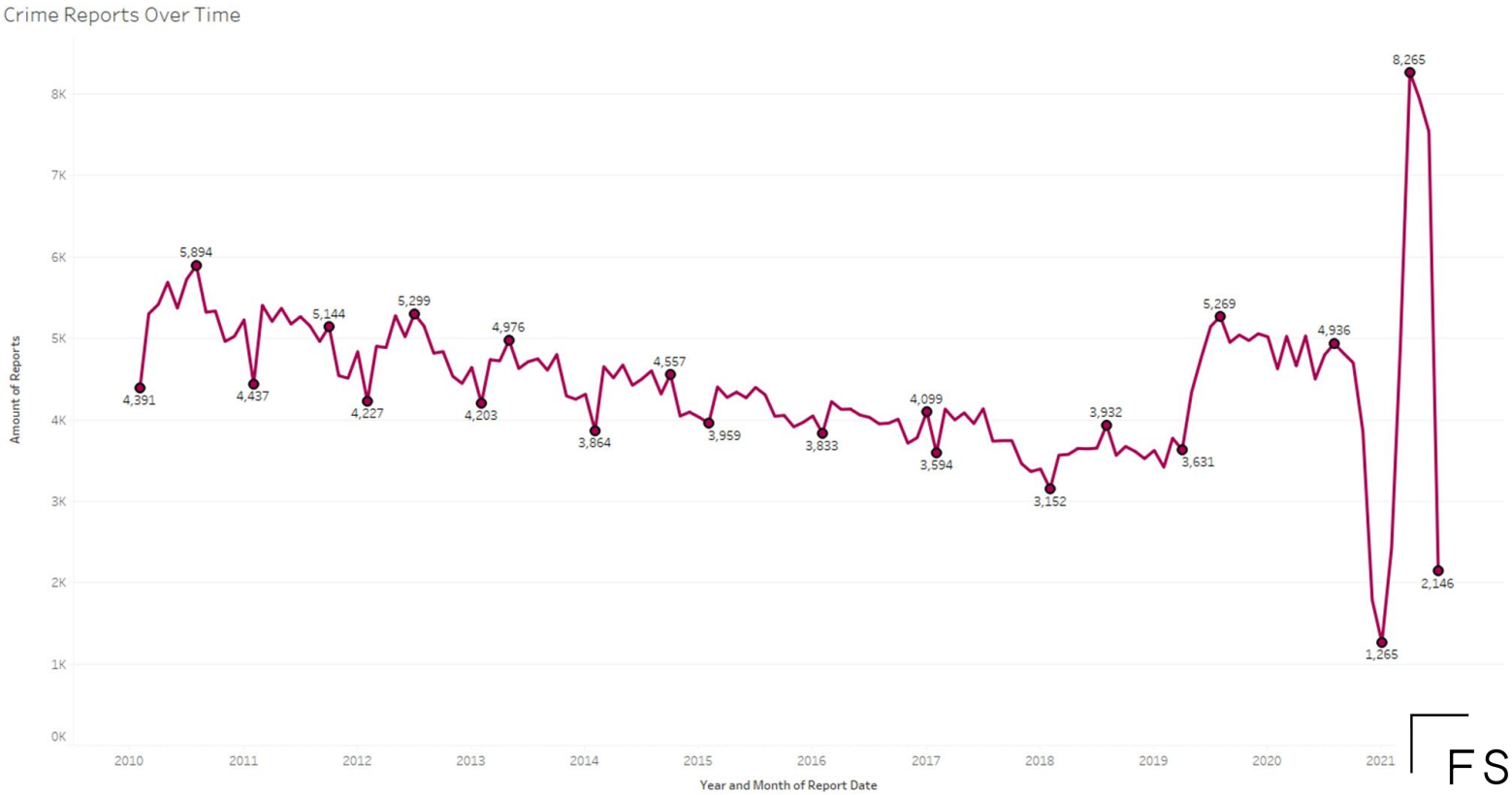
Which years had the highest reported incidents for theft/burglary? 2009

```
SELECT YEAR(occurred_date) AS theft_yr, COUNT(*) AS
theft_no
FROM crime_report
WHERE category_description = 'Theft'
GROUP BY theft_yr, category_description
ORDER BY theft_no DESC
LIMIT 10;
```



FSPAN

Visualization 1



The trend of count of Incident Number for Report Date Month. The view is filtered on Report Date Month, which ranges from January 2010 to July 2021.

FSPAN

Insight 2: Does clearance status affect Crime rates?

SQL Query: queried the number of incidents by occurrence date and grouped by clearance status

Data Visualization: plotted a line-graph chart of clearance closure rates by year

Results: the clearance status is a **key indicator** of whether a case has been closed or remains open

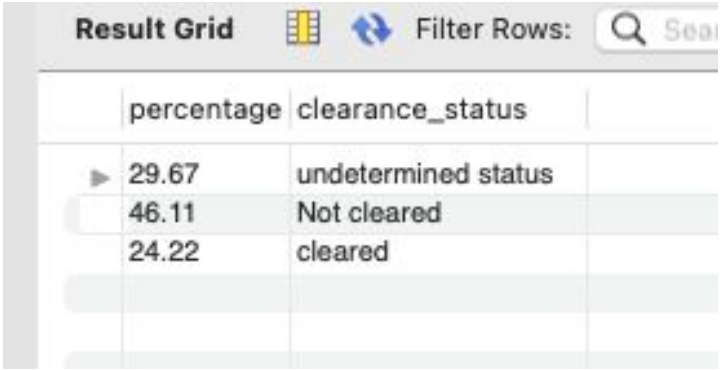
Recommendation: over the past two decades, the clearance rate percentage has steadily declined, indicating a **confirming decrease in reported crimes**

FSPAN

SQL Used

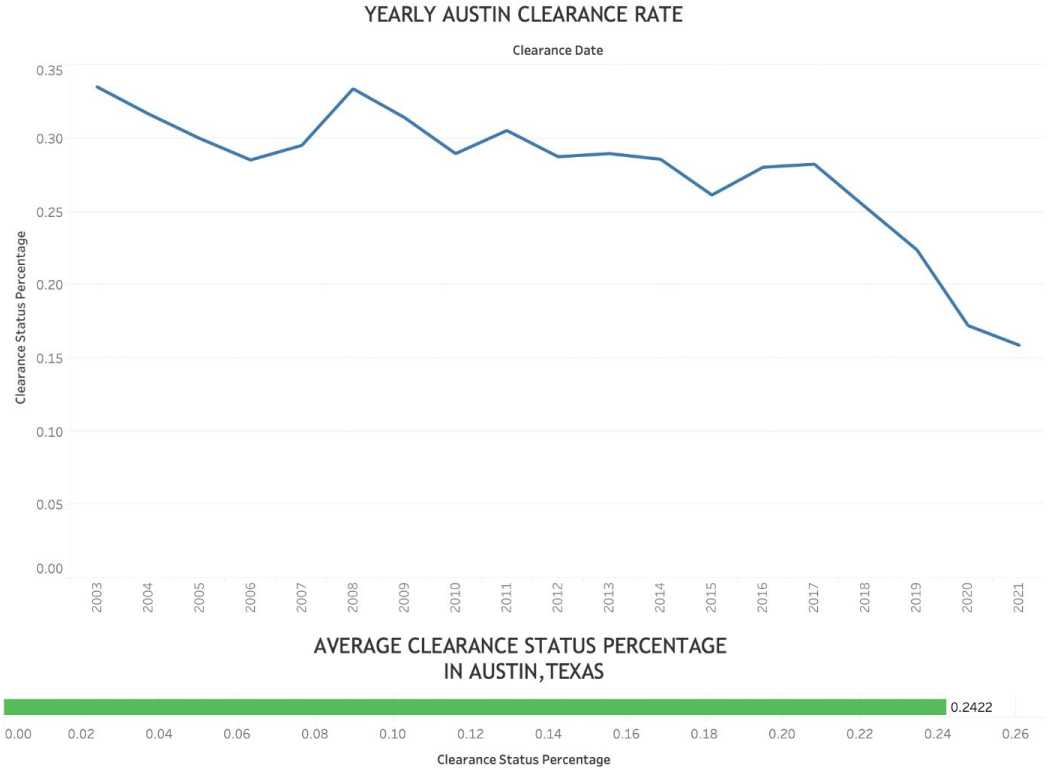
How many incident reports are considered “cleared,” or in other words, the number of cases that have been solved, percentage-wise? 24.22% are solved

```
SELECT
  ROUND(COUNT(*) * 100 / SUM(COUNT(*)) OVER (), 2) AS percentage,
  CASE WHEN clearance_status = 'C' THEN 'cleared' WHEN clearance_status = 'N' THEN 'Not cleared' ELSE 'undetermined status' END AS clearance_status
FROM crime_report
WHERE clearance_status IN ('C', 'N', '')
GROUP BY clearance_status;
```



FSPAN

Visualization 2



FSPAN

Insight 3: What were the effects of COVID-19 on crime?

SQL Query: queried the number of incidents by occurrence date and filtered for only dates between March 2020 and September 2021

Data Visualization: plotted a line-graph chart by total incidents by month and year

Results: number of incidents reported remained relatively stable at approximately **8k-9k per month**

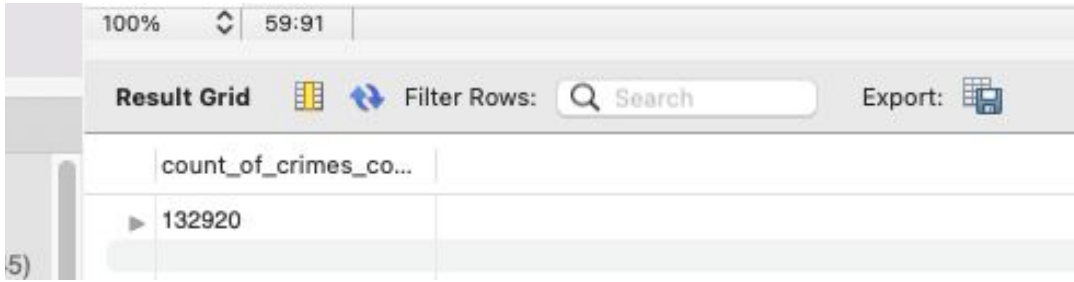
Recommendation: we observed a decline in the number of incidents reported, indicating a potential impact of COVID on crime rates - **disregard COVID data**

FSPAN

SQL Used

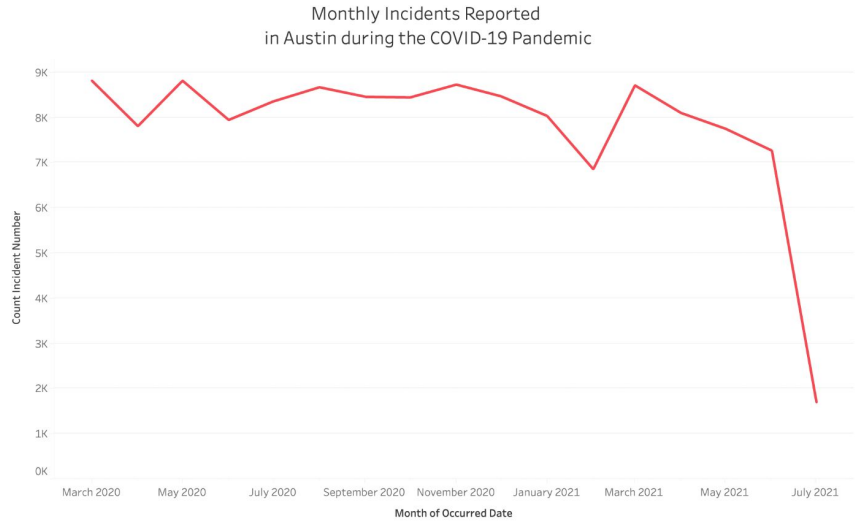
How many incident reports were made during the months of the COVID pandemic? 132,920

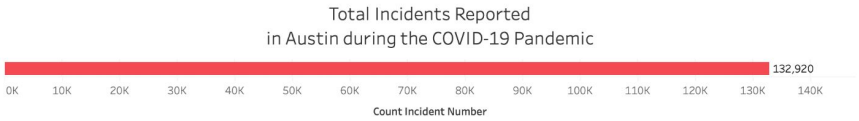
```
SELECT COUNT(*) AS count_of_crimes_covid
FROM crime_report
WHERE  occurred_date BETWEEN '2020-03-01'
AND '2021-09-01';
```

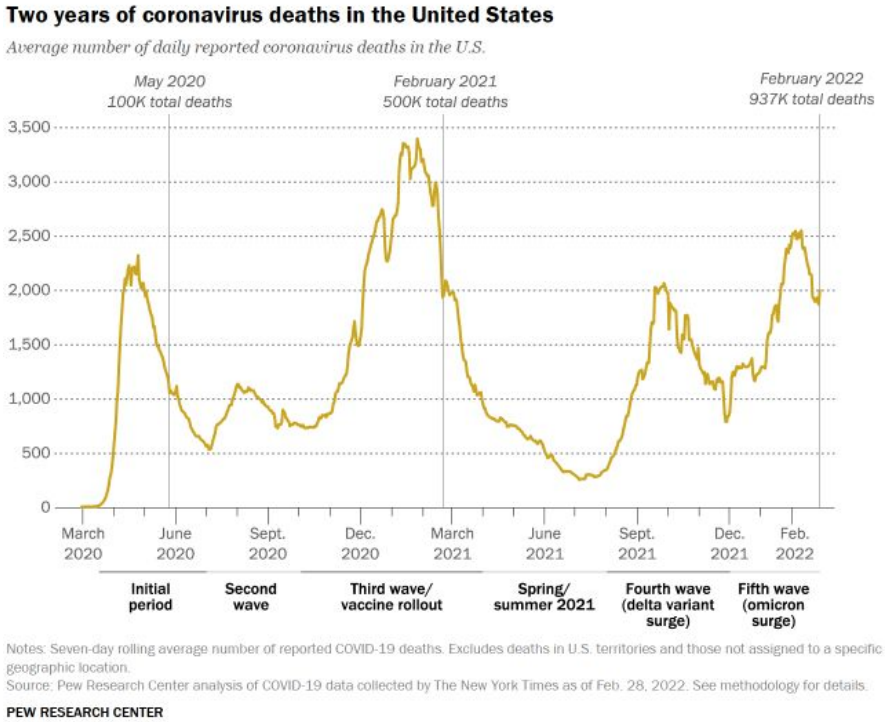


FSPAN

Visualization 3







FSPAN

Insight 4: Where do most crimes occur in Austin?

SQL Query: total incidents by latitude and longitude for Austin area **zip_codes**

Data Visualization: plotted a heat-map by total incidents by **zip_code**

Results: plotted data points into heat map to optically illustrate the results

Recommendations:

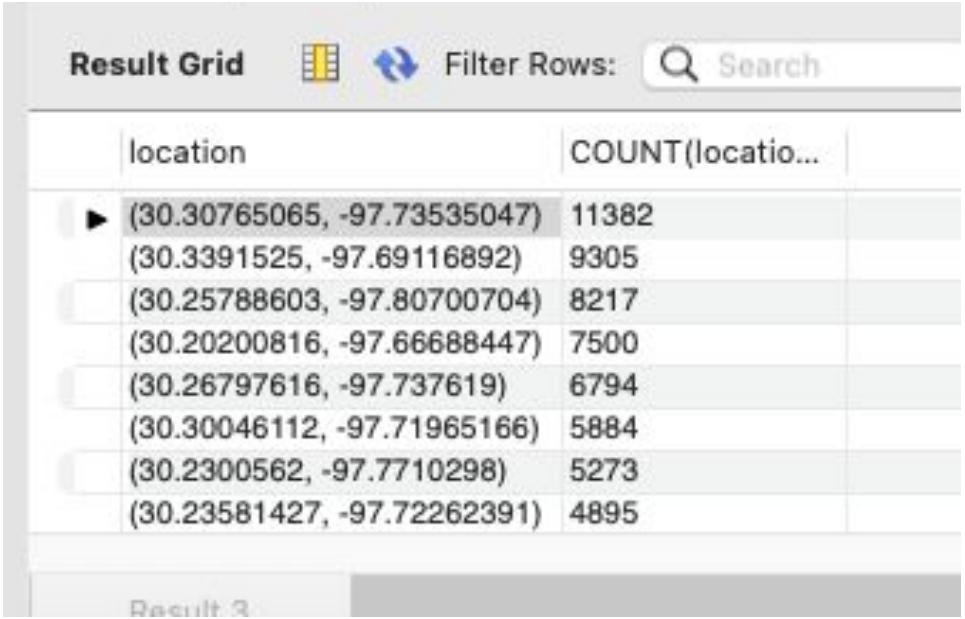
- **Central Austin** and **Southern Austin** have the most occurrences of crime
- **Downtown Austin** has lower crime incidents but would require additional security budget
- **Northern Austin** has the least amount of occurrences of crime

FSPAN

SQL Used

What is the most common location for crimes committed in Austin, Texas?
'(30.30765065, -97.73535047)'

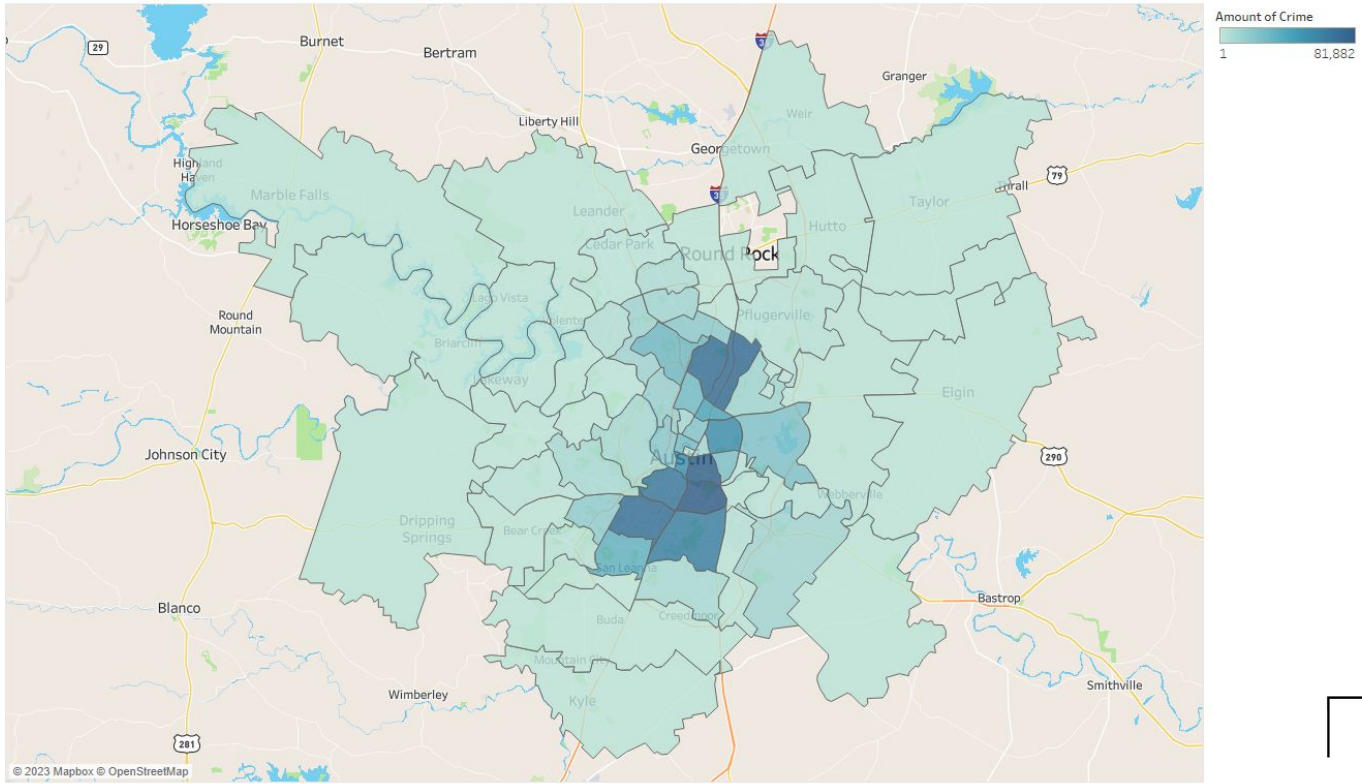
```
SELECT location, COUNT(location)
FROM crime_report
GROUP BY location
ORDER BY COUNT(location) DESC
LIMIT 10;
```



FSPAN

Visualization 4

Location of Crimes by Zipcode, latitude, longitude



Map based on Longitude (generated) and Latitude (generated). Color shows count of Latitude. Details are shown for Zip Code. The view is filtered on Latitude (generated) and Longitude (generated). The Latitude (generated) filter keeps non-Null values only. The Longitude (generated) filter keeps non-Null values only.

FSPAN

Insight 5: Family Violence effects on crime

SQL Query: queried the number of incidents by zip_code and occurrence date and filtered for family violence

Data Visualization: plotted a google heat-map by total incidents per zip_code with Family Violence

Results: based on the data, it appears that **Northern Austin** and far **South Austin** would be **"safer"** locations

Recommendation: we recommend that employees look at **Round Rock area for living**.

FSPAN

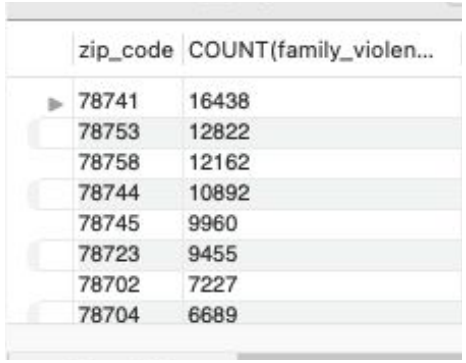
SQL Used

What is the most most incidents by zipcode for family violence? 78741

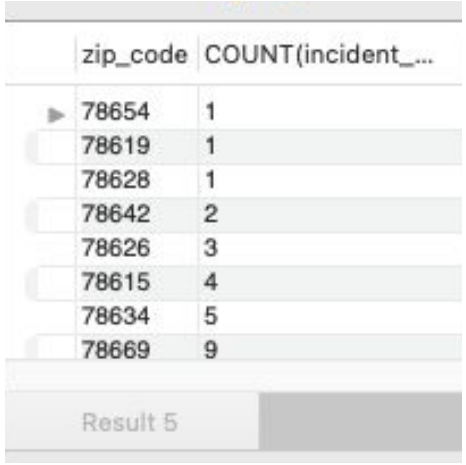
```
SELECT zip_code, COUNT(family_violence)
FROM crime_report
WHERE family_violence = "Y"
GROUP BY zip_code
ORDER by COUNT(family_violence) desc
LIMIT 10;
```

Least incidents by zip code for family violence?

```
SELECT zip_code, COUNT(family_violence)
FROM crime_report
WHERE family_violence = "Y"
GROUP BY zip_code
ORDER by COUNT(family_violence) asc
LIMIT 10;
```



zip_code	COUNT(family_violen...
78741	16438
78753	12822
78758	12162
78744	10892
78745	9960
78723	9455
78702	7227
78704	6689

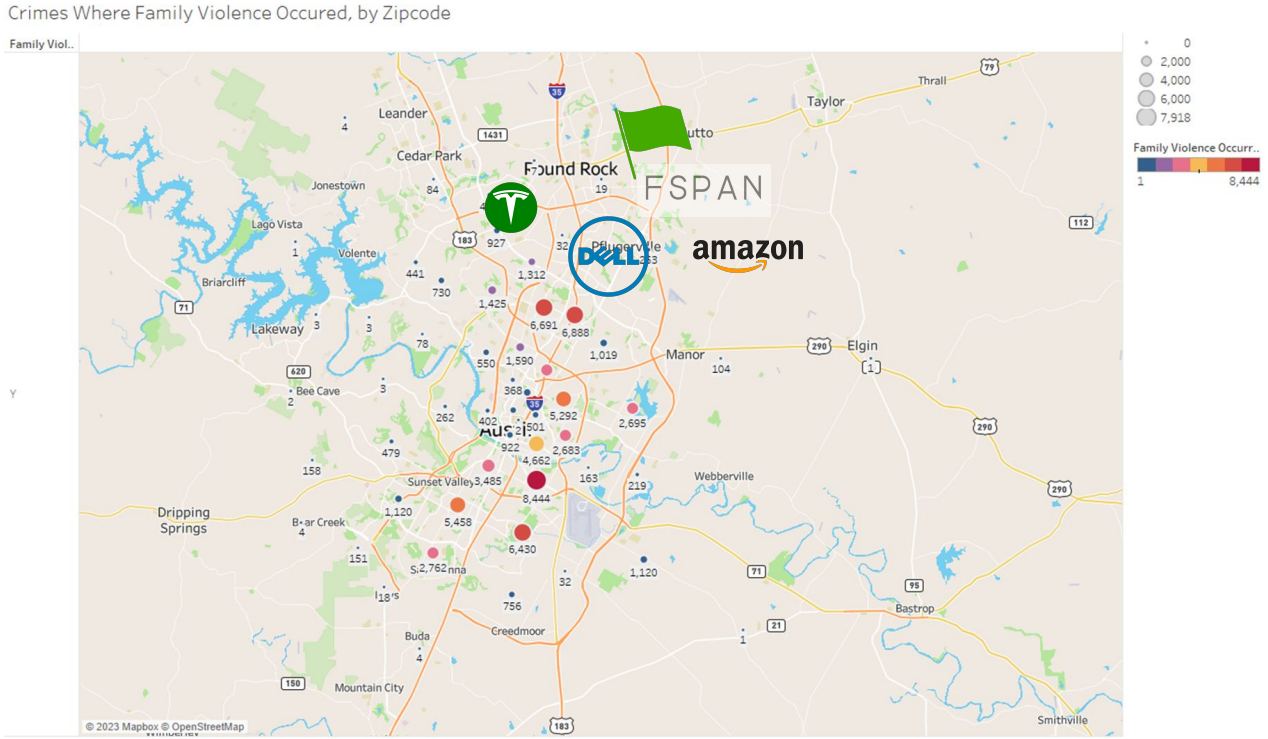


zip_code	COUNT(incident_...
78654	1
78619	1
78628	1
78642	2
78626	3
78615	4
78634	5
78669	9

Result 5

FSPAN

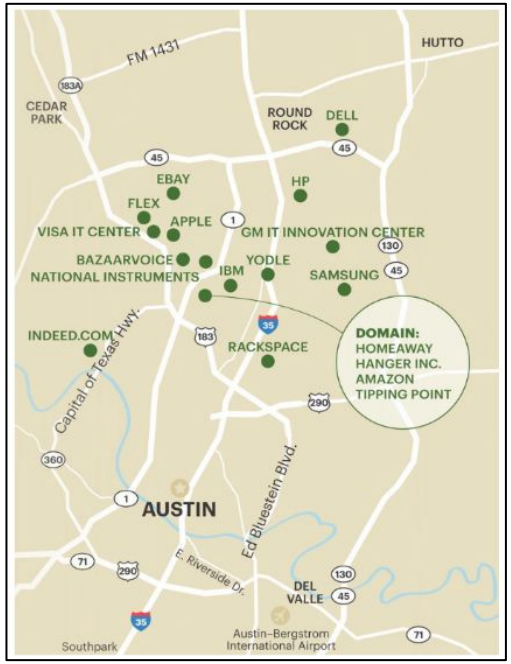
Visualization 5



Map based on Longitude (generated) and Latitude (generated) broken down by Family Violence. Color shows count of Family Violence. Size shows count of Latitude. Details are shown for Zip Code. The view is filtered on Family Violence, which keeps Y.

FSPAN

Recommendations for FSPAN



- We recommend FSPAN move their HQ to Northern Austin.
- Several other companies such as Tesla and Dell have chosen that location as well, indicating that it is a safer area for corporate offices.
- Surrounding cities include Round Rock and Pflugerville.
- The pattern for theft is quite low, especially since the trend for remote work is on the rise.

Data Source

Austin Crime Report 2003-7/2021 | Kaggle - https://www.kaggle.com/datasets/sdallman/austin-crime-report-200372021

Visualization 3 | https://www.pewresearch.org/politics/2022/03/03/the-changing-political-geography-of-covid-19-over-the-last-two-years/

https://www.bizjournals.com/austin/print-edition/2016/02/26/go-big-go-north-north-austin-is-the-place-to-be.html

FSPAN