



Juan Carlos Ceballos

Final Project



Data Source



Austin, Texas, USA Police Department

- Real data from crimes committed from 2003 to 2023
- Data shape **2,462,417 rows × 27 columns**
- Updated Weekly
- Public information

Source: <https://catalog.data.gov/dataset/crime-reports-bf2b7>

Data Cleaning



In [6]:	1 df																																																																																																																																																																							
Out[6]:																																																																																																																																																																								
<table border="1"> <thead> <tr> <th>Incident Number</th><th>Highest Offense Description</th><th>Highest Offense Code</th><th>Family Violence</th><th>Occurred Date Time</th><th>Occurred Date</th><th>Occurred Time</th><th>Report Date Time</th><th>Report Date</th><th>Report Time</th><th>...</th><th>Census Tract</th><th>Clearance Status</th><th>Clearance Date</th></tr> </thead> <tbody> <tr> <td>0</td><td>2006471156</td><td>FAMILY DISTURBANCE</td><td>3400</td><td>N 02/16/2006 02:25:00 PM</td><td>02/16/2006</td><td>1425.0</td><td>02/16/2006 02:25:00 PM</td><td>02/16/2006</td><td>1425.0</td><td>...</td><td>NaN</td><td>NaN</td><td>NaN</td></tr> <tr> <td>1</td><td>20045044338</td><td>TAMPERING WITH ID NUMBER</td><td>2719</td><td>N 09/14/2004 03:32:00 PM</td><td>09/14/2004</td><td>1532.0</td><td>09/14/2004 03:32:00 PM</td><td>09/14/2004</td><td>1532.0</td><td>...</td><td>NaN</td><td>N</td><td>09/14/2004</td></tr> <tr> <td>2</td><td>2006960811</td><td>FAMILY DISTURBANCE</td><td>3400</td><td>N 04/06/2006 10:29:00 AM</td><td>04/06/2006</td><td>1029.0</td><td>04/06/2006 10:29:00 AM</td><td>04/06/2006</td><td>1029.0</td><td>...</td><td>NaN</td><td>N</td><td>05/01/2000</td></tr> <tr> <td>3</td><td>2013851154</td><td>SEXUAL ASSAULT OF CHILD/OBJECT</td><td>1707</td><td>Y 01/01/2009 12:01:00 AM</td><td>01/01/2009</td><td>1.0</td><td>03/26/2013 04:56:00 PM</td><td>03/26/2013</td><td>1656.0</td><td>...</td><td>4.00</td><td>C</td><td>04/11/2011</td></tr> <tr> <td>4</td><td>20161800084</td><td>RAPE OF A CHILD</td><td>204</td><td>Y 06/28/2016 01:05:00 AM</td><td>06/28/2016</td><td>105.0</td><td>06/28/2016 01:05:00 AM</td><td>06/28/2016</td><td>105.0</td><td>...</td><td>24.41</td><td>C</td><td>07/01/2011</td></tr> <tr> <td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td></tr> <tr> <td>2462412</td><td>20215001631</td><td>DEBIT CARD ABUSE</td><td>1108</td><td>N 09/24/2020 12:00:00 PM</td><td>09/24/2020</td><td>1200.0</td><td>01/13/2021 10:27:00 AM</td><td>01/13/2021</td><td>1027.0</td><td>...</td><td>24.47</td><td>N</td><td>01/14/2022</td></tr> <tr> <td>2462413</td><td>20105013578</td><td>BURGLARY OF VEHICLE</td><td>601</td><td>N 03/19/2010 07:00:00 PM</td><td>03/19/2010</td><td>1900.0</td><td>03/19/2010 11:16:00 PM</td><td>03/19/2010</td><td>2316.0</td><td>...</td><td>19.11</td><td>N</td><td>03/25/2011</td></tr> <tr> <td>2462414</td><td>20085066651</td><td>PROTECTIVE ORDER</td><td>3829</td><td>N 12/16/2008 03:56:00 PM</td><td>12/16/2008</td><td>1556.0</td><td>12/16/2008 03:56:00 PM</td><td>12/16/2008</td><td>1556.0</td><td>...</td><td>13.00</td><td>NaN</td><td>NaN</td></tr> <tr> <td>2462415</td><td>20101570072</td><td>VIOL CITY ORDINANCE - OTHER</td><td>3299</td><td>N 06/06/2010 12:31:00 AM</td><td>06/06/2010</td><td>31.0</td><td>06/06/2010 12:31:00 AM</td><td>06/06/2010</td><td>31.0</td><td>...</td><td>431.00</td><td>C</td><td>08/07/2011</td></tr> <tr> <td>2462416</td><td>20102371234</td><td>FORGERY AND PASSING</td><td>1000</td><td>N 08/25/2010 04:05:00 PM</td><td>08/25/2010</td><td>1605.0</td><td>08/25/2010 04:05:00 PM</td><td>08/25/2010</td><td>1605.0</td><td>...</td><td>13.00</td><td>N</td><td>08/26/2011</td></tr> </tbody> </table>	Incident Number	Highest Offense Description	Highest Offense Code	Family Violence	Occurred Date Time	Occurred Date	Occurred Time	Report Date Time	Report Date	Report Time	...	Census Tract	Clearance Status	Clearance Date	0	2006471156	FAMILY DISTURBANCE	3400	N 02/16/2006 02:25:00 PM	02/16/2006	1425.0	02/16/2006 02:25:00 PM	02/16/2006	1425.0	...	NaN	NaN	NaN	1	20045044338	TAMPERING WITH ID NUMBER	2719	N 09/14/2004 03:32:00 PM	09/14/2004	1532.0	09/14/2004 03:32:00 PM	09/14/2004	1532.0	...	NaN	N	09/14/2004	2	2006960811	FAMILY DISTURBANCE	3400	N 04/06/2006 10:29:00 AM	04/06/2006	1029.0	04/06/2006 10:29:00 AM	04/06/2006	1029.0	...	NaN	N	05/01/2000	3	2013851154	SEXUAL ASSAULT OF CHILD/OBJECT	1707	Y 01/01/2009 12:01:00 AM	01/01/2009	1.0	03/26/2013 04:56:00 PM	03/26/2013	1656.0	...	4.00	C	04/11/2011	4	20161800084	RAPE OF A CHILD	204	Y 06/28/2016 01:05:00 AM	06/28/2016	105.0	06/28/2016 01:05:00 AM	06/28/2016	105.0	...	24.41	C	07/01/2011	2462412	20215001631	DEBIT CARD ABUSE	1108	N 09/24/2020 12:00:00 PM	09/24/2020	1200.0	01/13/2021 10:27:00 AM	01/13/2021	1027.0	...	24.47	N	01/14/2022	2462413	20105013578	BURGLARY OF VEHICLE	601	N 03/19/2010 07:00:00 PM	03/19/2010	1900.0	03/19/2010 11:16:00 PM	03/19/2010	2316.0	...	19.11	N	03/25/2011	2462414	20085066651	PROTECTIVE ORDER	3829	N 12/16/2008 03:56:00 PM	12/16/2008	1556.0	12/16/2008 03:56:00 PM	12/16/2008	1556.0	...	13.00	NaN	NaN	2462415	20101570072	VIOL CITY ORDINANCE - OTHER	3299	N 06/06/2010 12:31:00 AM	06/06/2010	31.0	06/06/2010 12:31:00 AM	06/06/2010	31.0	...	431.00	C	08/07/2011	2462416	20102371234	FORGERY AND PASSING	1000	N 08/25/2010 04:05:00 PM	08/25/2010	1605.0	08/25/2010 04:05:00 PM	08/25/2010	1605.0	...	13.00	N	08/26/2011
Incident Number	Highest Offense Description	Highest Offense Code	Family Violence	Occurred Date Time	Occurred Date	Occurred Time	Report Date Time	Report Date	Report Time	...	Census Tract	Clearance Status	Clearance Date																																																																																																																																																											
0	2006471156	FAMILY DISTURBANCE	3400	N 02/16/2006 02:25:00 PM	02/16/2006	1425.0	02/16/2006 02:25:00 PM	02/16/2006	1425.0	...	NaN	NaN	NaN																																																																																																																																																											
1	20045044338	TAMPERING WITH ID NUMBER	2719	N 09/14/2004 03:32:00 PM	09/14/2004	1532.0	09/14/2004 03:32:00 PM	09/14/2004	1532.0	...	NaN	N	09/14/2004																																																																																																																																																											
2	2006960811	FAMILY DISTURBANCE	3400	N 04/06/2006 10:29:00 AM	04/06/2006	1029.0	04/06/2006 10:29:00 AM	04/06/2006	1029.0	...	NaN	N	05/01/2000																																																																																																																																																											
3	2013851154	SEXUAL ASSAULT OF CHILD/OBJECT	1707	Y 01/01/2009 12:01:00 AM	01/01/2009	1.0	03/26/2013 04:56:00 PM	03/26/2013	1656.0	...	4.00	C	04/11/2011																																																																																																																																																											
4	20161800084	RAPE OF A CHILD	204	Y 06/28/2016 01:05:00 AM	06/28/2016	105.0	06/28/2016 01:05:00 AM	06/28/2016	105.0	...	24.41	C	07/01/2011																																																																																																																																																											
...																																																																																																																																																											
2462412	20215001631	DEBIT CARD ABUSE	1108	N 09/24/2020 12:00:00 PM	09/24/2020	1200.0	01/13/2021 10:27:00 AM	01/13/2021	1027.0	...	24.47	N	01/14/2022																																																																																																																																																											
2462413	20105013578	BURGLARY OF VEHICLE	601	N 03/19/2010 07:00:00 PM	03/19/2010	1900.0	03/19/2010 11:16:00 PM	03/19/2010	2316.0	...	19.11	N	03/25/2011																																																																																																																																																											
2462414	20085066651	PROTECTIVE ORDER	3829	N 12/16/2008 03:56:00 PM	12/16/2008	1556.0	12/16/2008 03:56:00 PM	12/16/2008	1556.0	...	13.00	NaN	NaN																																																																																																																																																											
2462415	20101570072	VIOL CITY ORDINANCE - OTHER	3299	N 06/06/2010 12:31:00 AM	06/06/2010	31.0	06/06/2010 12:31:00 AM	06/06/2010	31.0	...	431.00	C	08/07/2011																																																																																																																																																											
2462416	20102371234	FORGERY AND PASSING	1000	N 08/25/2010 04:05:00 PM	08/25/2010	1605.0	08/25/2010 04:05:00 PM	08/25/2010	1605.0	...	13.00	N	08/26/2011																																																																																																																																																											
AM	08/25/2010 04:05:00 PM	08/25/2010	1605.0	...	13.00	N	08/26/2010	NaN	NaN	3105754.0	10060029.0	30.239131	-97.770694	(30.23913102, -97.77069399)																																																																																																																																																										
AM	08/25/2010 04:05:00 PM	08/25/2010	1605.0	...	13.00	N	08/26/2010	NaN	NaN	310528.0	10104030.0	30.358489	-97.688952	(30.35848873, -97.68895198)																																																																																																																																																										
AM	08/25/2010 04:05:00 PM	08/25/2010	1605.0	...	13.00	N	08/26/2010	NaN	NaN	3107741.0	10061626.0	30.243395	-97.764287	(30.24339493, -97.76428662)																																																																																																																																																										

Criteria for filtering data:

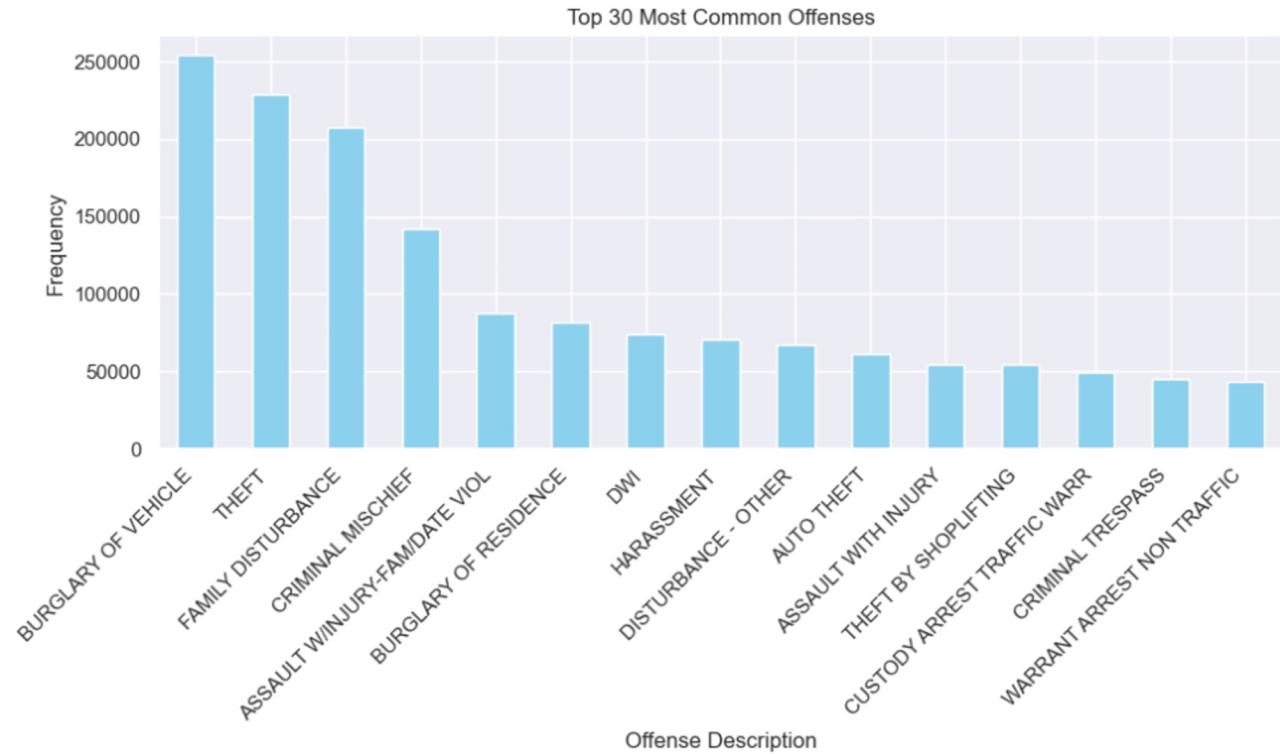
- 2019 to 2023 Analysis
- Filtered data size **27,865 rows x 9 columns**

```
In [60]: 1 # Reduced df:
2 sampled_df = reduced_df.sample(frac=0.10, random_state=1)
```

Performing some EDA

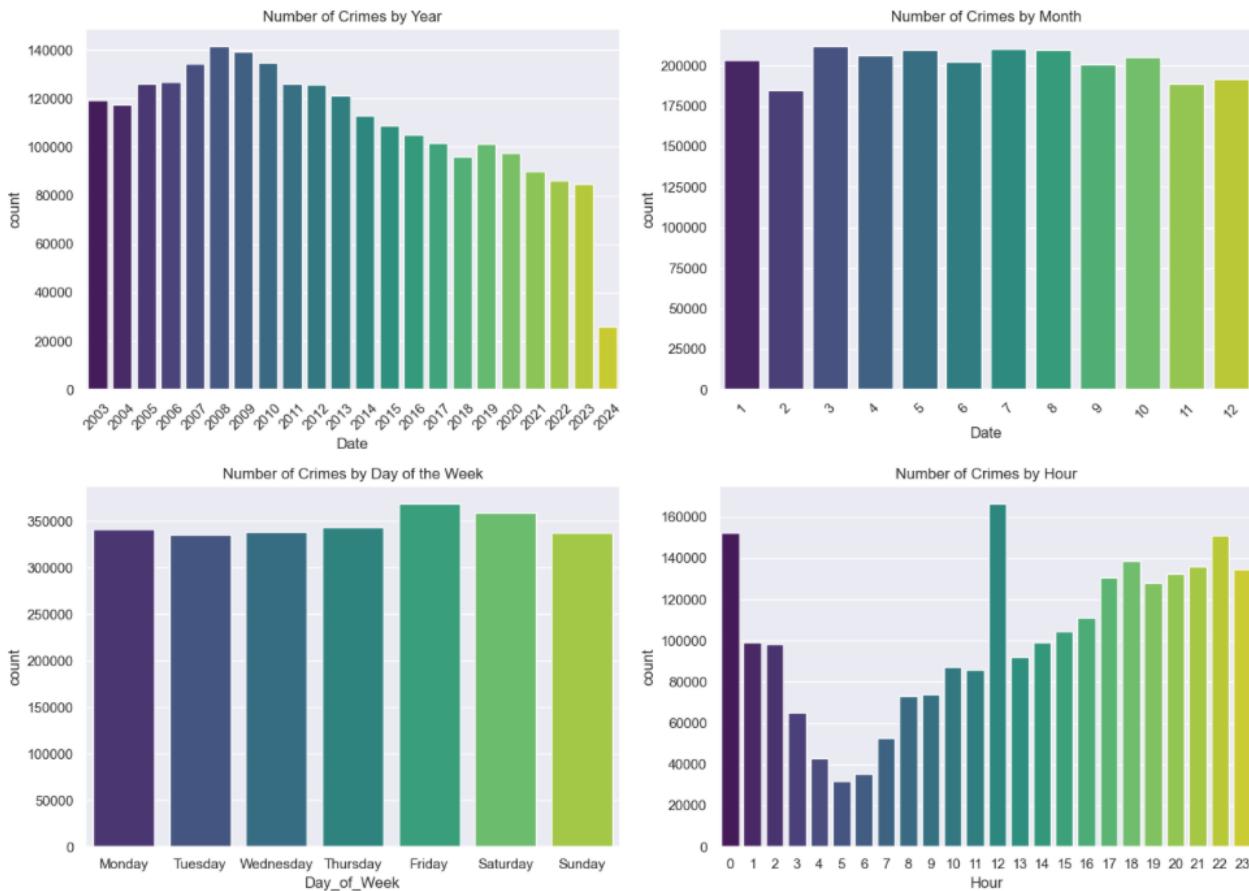
In [24]:

```
1 # Plotting
2 plt.figure(figsize=(10, 6))
3 offense_counts.plot(kind='bar', color='skyblue')
4 plt.title('Top 30 Most Common Offenses')
5 plt.xlabel('Offense Description')
6 plt.ylabel('Frequency')
7 plt.xticks(rotation=45, ha='right')
8 plt.tight_layout()
9 plt.show()
```

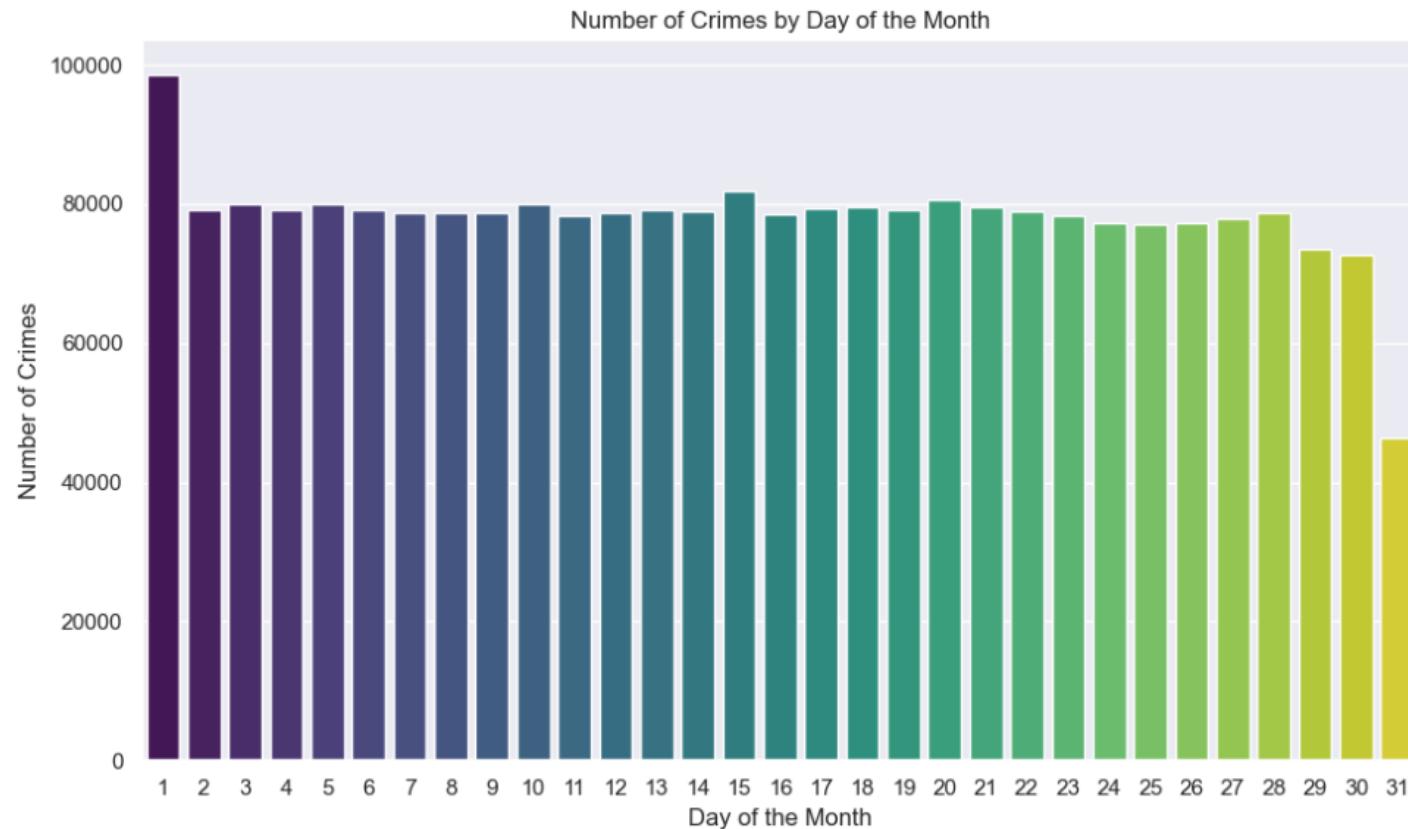


Performing some EDA

```
8 # Plot 1: Number of crimes by year
9 plt.subplot(2, 2, 1)
10 sns.countplot(data=df, x=df['Date'].dt.year, palette='viridis')
11 plt.title('Number of Crimes by Year')
12 plt.xticks(rotation=45)
13
14 # Plot 2: Number of crimes by month
15 plt.subplot(2, 2, 2)
16 sns.countplot(data=df, x=df['Date'].dt.month, palette='viridis')
17 plt.title('Number of Crimes by Month')
18 plt.xticks(rotation=45)
19
20 # Plot 3: Number of crimes by day of the week
21 plt.subplot(2, 2, 3)
22 sns.countplot(data=df, x=df['Day_of_Week'], order=['Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday', 'Saturday', 'Sunday'],
23 plt.title('Number of Crimes by Day of the Week')
24
25 # Plot 4: Number of crimes by hour
26 plt.subplot(2, 2, 4)
27 sns.countplot(data=df, x=df['Hour'], palette='viridis')
28 plt.title('Number of Crimes by Hour')
29
30 plt.tight_layout()
31 plt.show()
32
```



Performing some EDA



```
: 1 # Plotting the number of crimes over different time periods
2 plt.figure(figsize=(10, 6))
3
4 # Plot: Number of crimes by day of the month
5 sns.countplot(data=df, x=df['Date'].dt.day, palette='viridis')
6 plt.title('Number of Crimes by Day of the Month')
7 plt.xlabel('Day of the Month')
8 plt.ylabel('Number of Crimes')
9
10 plt.tight_layout()
11 plt.show()
```

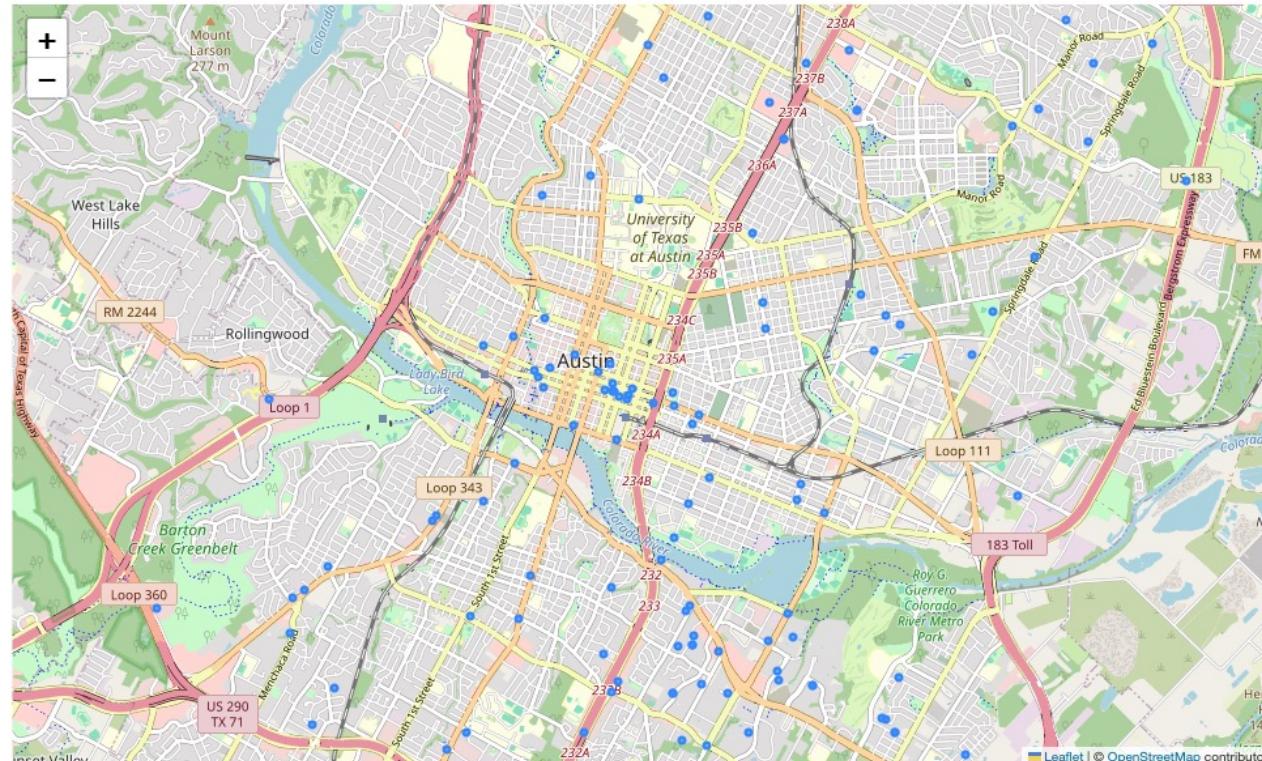
Trying to map the crime locations

Using a map of Austin Texas instead, I used the folium library

In [33]:

```
1 # Since my sample has more than 2 million crimes i am using just a sample to visualize
2
3 sample_df = df.sample(frac=0.0001)
4
5 # Trying to map the data in a map using folium
6
7 # Creating a map centered around Austin, Texas
8 austin_map = folium.Map(location=[30.2672, -97.7431], zoom_start=11)
9
10 # Adding markers for each crime location
11 for index, row in sample_df.iterrows():
12     folium.CircleMarker([row['Latitude'], row['Longitude']], popup=row['Highest Offense Description'],
13                         radius=2
14                         ).add_to(austin_map)
15
16 # Display the map
17 austin_map
```

Out[33]:



Trying to map the crime locations



Safari Archivo Edición Visualización Historial Marcadores Desarrollo Ventana Ayuda

localhost

Desktop/7. Programming /0. Bootcamp/8. Final Project / alternative_project - Jupyter Notebook Crime Reports - Comma Separated Values File - Catalog

1 de 12 coincidencias Empieza con folium Logout

jupyter alternative_project Last Checkpoint: hace 7 horas (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel) O

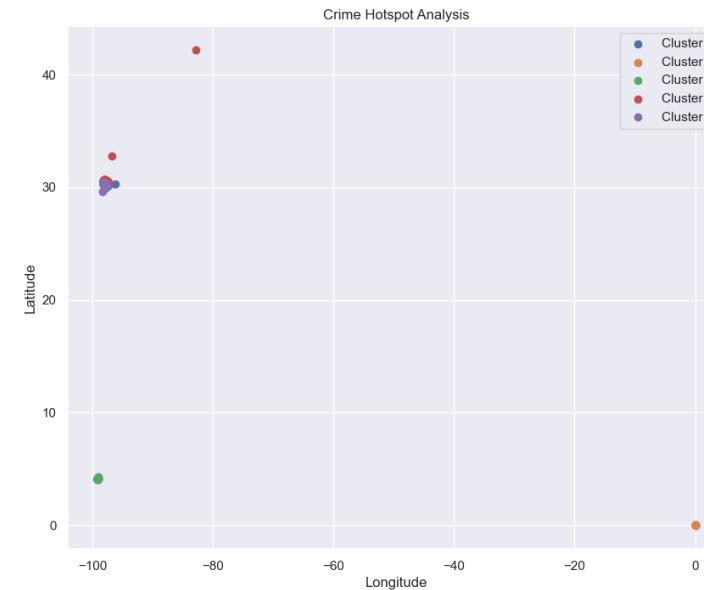
17 austin_map

out [33]:

A map of Austin, Texas, showing crime locations marked with blue dots. The map includes major roads like Loop 1, Loop 111, and the Colorado River, as well as landmarks such as the University of Texas at Austin and Zilker Park. A hand cursor is visible on the map.

Leaflet | © OpenStreetMap contributors

Trying to cluster crimes with K-Means



Safari Archivo Edición Visualización Historial Marcadores Desarrollo Ventana Ayuda localhost

Desktop/7. Programming /0. Bootcamp/8. Final Project / alternative_project - Jupyter Notebook Crime Reports - Comma Separated Values File - Catalog 1 de 12 coincidencias Empieza con folium Dom may. 26 07:40

jupyter alternative_project Last Checkpoint: hace 7 horas (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel) Logout

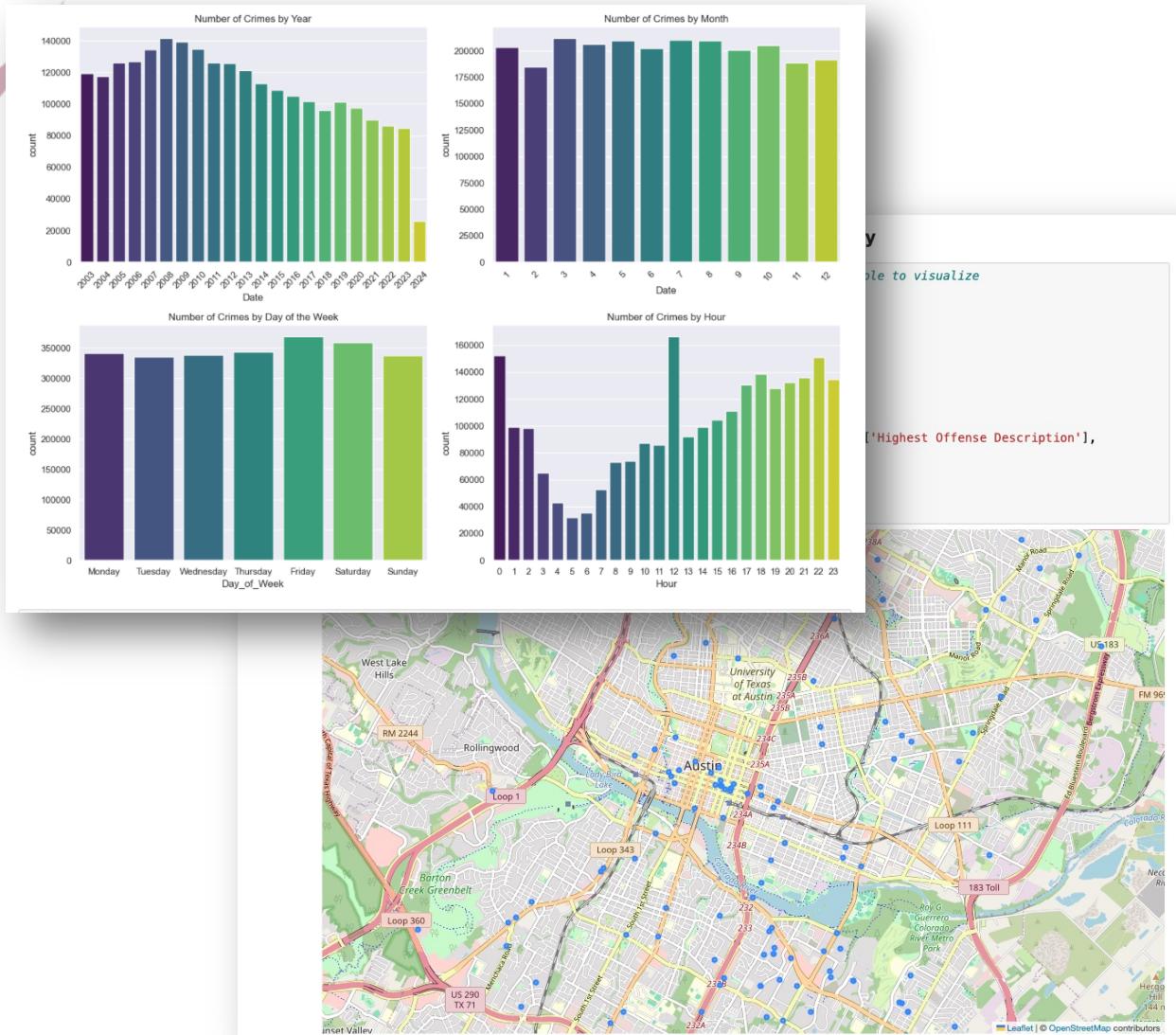
Testing with a sample of 0.01% of samples

In [49]:

```
1 sample_df = df.sample(frac=0.001)
2
3 # Create a map centered around the mean latitude and longitude
4 map_center = [sample_df['Latitude'].mean(), sample_df['Longitude'].mean()]
5 crime_map = folium.Map(location=map_center, zoom_start=10)
6
7 # Define colors for each cluster
8 cluster_colors = ['blue', 'green', 'red', 'purple', 'orange']
9
10 # Add markers for each crime location, colored by cluster
11 for index, row in sample_df.iterrows():
12     folium.CircleMarker(
13         location=[row['Latitude'], row['Longitude']],
14         radius=5,
15         color=cluster_colors[row['Cluster']],
16         fill=True,
17         fill_color=cluster_colors[row['Cluster']],
18         popup=row['Highest Offense Description']
19     ).add_to(crime_map)
20
21 # Display the map
22 crime_map
23
```

Out[49]:

It works on the notebook ... exporting to Tableau



+ a b l e a u®



Thank you for your attention!