**Data Presentation and Visualization**

***WESTYN HILLIARD***

[34]:
```
# Load the required packages
import pandas as pd
import matplotlib.pyplot as plt
```

# 1 1: Load All Datasets

[27]:
```
# Load and display the content of the first CSV file to understand its structure
file_path_1 = 'carTheftsMap.csv'
car_thefts_map_df = pd.read_csv(file_path_1)

# Display the first few rows to understand the content
car_thefts_map_df.head()
```

[27]:

| | agency_ori | geo_name | countCarThefts2019 | countCarThefts2020 | countCarThefts2021 | countCarThefts2022 | latitude | longitude | percentChange2019to2022 |
|---|---|---|---|---|---|---|---|---|---|
| 0 | MO0490300 | Carthage PD | 62 | 58 | 47 | 26 | 37.175854 | -94.313345 | -0.580645 |
| 1 | NaN | Warren County SO | 112 | 94 | 76 | 58 | 41.308788 | -80.847876 | |
| 2 | TX06802 | Odessa PD | 499 | 464 | 375 | 288 | 31.889332 | -102.329889 | |
| 3 | MO0530000 | Laclede County SO | 55 | 74 | 54 | 32 | 37.663455 | -92.549013 | |
| 4 | MO0480000 | Jackson County SO | 125 | 141 | 81 | 75 | 37.384401 | -89.667918 | |

```
1                -0.482143
2                -0.422846
3                -0.418182
4                -0.400000
```

The file carTheftsMap.csv contains the following columns:

- **agency_ori:** Identifies the reporting police department.
- **geo_name:** Location name (e.g., city or county).
- **countCarThefts2019, countCarThefts2020, countCarThefts2021, countCarThefts2022:** The number of car thefts recorded for each year from 2019 to 2022.
- **latitude and longitude:** Coordinates for mapping locations.
- **percentChange2019to2022:** Percentage change in car thefts between 2019 and 2022.

[28]:
```
# Load and display the content of the second CSV file to understand its␣
 ↪structure
file_path_2 = 'KiaHyundaiMilwaukeeData.csv'
kia_hyundai_milwaukee_df = pd.read_csv(file_path_2)

# Display the first few rows to understand the content
kia_hyundai_milwaukee_df.head()
```

[28]:
```
  month  year       city state  countKiaHyundaiThefts  countOtherThefts  \
0   Jan  2019  Milwaukee    WI                     22               235
1   Feb  2019  Milwaukee    WI                     13               218
2   Mar  2019  Milwaukee    WI                     10               195
3   Apr  2019  Milwaukee    WI                     10               238
4   May  2019  Milwaukee    WI                     11               280

   percentKiaHyundai
0              0.086
1              0.056
2              0.049
3              0.040
4              0.038
```

The file KiaHyundaiMilwaukeeData.csv contains data specific to car thefts in Milwaukee, Wisconsin, with the following columns:

- **month and year:** Timeframe of the recorded thefts.
- **city and state:** Location information.
- **countKiaHyundaiThefts:** Number of thefts specifically involving Kia and Hyundai vehicles.
- **countOtherThefts:** Number of thefts involving other vehicle brands.

- **percentKiaHyundai:** The proportion of Kia and Hyundai thefts relative to all vehicle thefts for the given timeframe.

[29]:
```
# Load and display the content of the third CSV file to understand its structure
file_path_3 = 'kiaHyundaiThefts.csv'
kia_hyundai_thefts_df = pd.read_csv(file_path_3)

# Display the first few rows to understand the content
kia_hyundai_thefts_df.head()
```

[29]:

| | month | year | city | state | countKiaHyundaiThefts | countOtherThefts | \ |
|---|---|---|---|---|---|---|---|
| 0 | Jan | 2019 | Atlanta | GA | 17 | 264 | |
| 1 | Feb | 2019 | Atlanta | GA | 11 | 205 | |
| 2 | Mar | 2019 | Atlanta | GA | 18 | 181 | |
| 3 | Apr | 2019 | Atlanta | GA | 15 | 223 | |
| 4 | May | 2019 | Atlanta | GA | 16 | 277 | |

| | percentKiaHyundai |
|---|---|
| 0 | 0.060 |
| 1 | 0.051 |
| 2 | 0.090 |
| 3 | 0.063 |
| 4 | 0.055 |

The file kiaHyundaiThefts.csv contains data on car thefts in various cities, including:

- **month and year:** Timeframe for recorded thefts.
- **city and state:** Location details.
- **countKiaHyundaiThefts:** Number of Kia and Hyundai thefts.
- ***countOtherThefts:** Number of thefts involving other vehicle brands.
- **percentKiaHyundai:** Proportion of Kia and Hyundai thefts relative to all car thefts in the given timeframe.

[31]:
```
# Load and explore the content of the Excel file to understand its structure
file_path_4 = 'Motherboard VICE News Kia Hyundai Theft Data.xlsx'
excel_data = pd.ExcelFile(file_path_4)

# Check the sheet names to understand the structure of the Excel file
sheet_names = excel_data.sheet_names
sheet_names

# Load the content of the 'Data' sheet to examine its structure
vice_news_data_df = pd.read_excel(file_path_4, sheet_name='Data')

# Display the first few rows to understand the content
vice_news_data_df.head()
```

[31]:
```
  Unnamed: 0        Denver Unnamed: 2 Unnamed: 3       El Paso Unnamed: 5  \
0        NaT  Kia/Hyundais         All    Percent  Kia/Hyundais         All
1 2019-12-01            48         615   0.078049            13         103
2 2020-01-01            21         519   0.040462             9          95
3 2020-02-01            28         402   0.069652             5          64
4 2020-03-01            35         508   0.068898             6          75

  Unnamed: 6      Portland Unnamed: 8 Unnamed: 9  … Unnamed: 201  \
0    Percent  Kia/Hyundais         All    Percent  …      Percent
1   0.126214            13         592   0.021959  …     0.055556
2   0.094737            12         559   0.021467  …     0.098039
3   0.078125            10         498    0.02008  …         0.15
4       0.08             4         465   0.008602  …     0.035714

     Peoria, AZ Unnamed: 203 Unnamed: 204   Houston, TX Unnamed: 206  \
0  Kia/Hyundais         All       Percent  Kia/Hyundais         All
1             2          27     0.074074            24          603
2             1          17     0.058824            18          586
3             2          19     0.105263            15          556
4             2          17     0.117647            26          524

  Unnamed: 207 Prince George's County, MD Unnamed: 209 Unnamed: 210
0      Percent                Kia/Hyundais         All       Percent
1          NaN                          11         210     0.052381
2          NaN                          10         215     0.046512
3          NaN                           6         169     0.035503
4          NaN                           7         167     0.041916

[5 rows x 211 columns]
```

The Excel file contains detailed data for various cities, including columns for:

- Dates (monthly basis).
- Number of Kia/Hyundai thefts and overall car thefts.
- Percentages of Kia/Hyundai thefts relative to the total thefts in cities like Denver, El Paso, Portland, Peoria, Houston, and Prince George's County, among others.

The structure suggests that each city has its own columns for Kia/Hyundai thefts, total car thefts, and the percentage of thefts specific to those brands.

[32]:
```
# Analyze overall trends in Kia/Hyundai thefts across the datasets

# Summarize trends in 'carTheftsMap.csv' by year
car_thefts_summary = car_thefts_map_df[['countCarThefts2019',␣
 ↪'countCarThefts2020', 'countCarThefts2021', 'countCarThefts2022']].sum()

# Summarize trends in 'KiaHyundaiMilwaukeeData.csv' by year
```

```
kia_hyundai_milwaukee_summary = kia_hyundai_milwaukee_df.
 ↪groupby('year')[['countKiaHyundaiThefts', 'countOtherThefts']].sum()

# Summarize trends in 'kiaHyundaiThefts.csv' by year and city
kia_hyundai_thefts_summary = kia_hyundai_thefts_df.groupby(['year',␣
 ↪'city'])[['countKiaHyundaiThefts', 'countOtherThefts']].sum()

# Summarize trends in the 'Motherboard VICE News Kia Hyundai Theft Data.xlsx'␣
 ↪by extracting city-wise data for comparisons
cities = ['Denver', 'El Paso', 'Portland', 'Peoria, AZ', 'Houston, TX', "Prince␣
 ↪George's County, MD"]
vice_news_summary = {}

# Extract columns for each city in the 'vice_news_data_df'
for city in cities:
    city_columns = [col for col in vice_news_data_df.columns if city in col]
    city_data = vice_news_data_df[city_columns]
    vice_news_summary[city] = city_data.describe()  # Get basic stats for each␣
 ↪city

# Display the aggregated summaries to identify key trends
{
    "car_thefts_summary": car_thefts_summary,
    "kia_hyundai_milwaukee_summary": kia_hyundai_milwaukee_summary,
    "kia_hyundai_thefts_summary": kia_hyundai_thefts_summary.head(),
    "vice_news_summary_sample": {city: summary.head() for city, summary in␣
 ↪vice_news_summary.items()}
}
```

[32]:
```
{'car_thefts_summary': countCarThefts2019
6211249955125725060736590421949694137155441469…
 countCarThefts2020    589446474141623761426574411 76-6739121321581005…
 countCarThefts2021    477637554816635605343-441429742116101181231214…
 countCarThefts2022    265828832754431374642592813166128251093110468 4…
 dtype: object,
 'kia_hyundai_milwaukee_summary':       countKiaHyundaiThefts  countOtherThefts
 year
 2019                     207              3281
 2020                     895              3649
 2021                    6971              3507
 2022                    4679              3377,
 'kia_hyundai_thefts_summary':                    countKiaHyundaiThefts
countOtherThefts
 year city
 2019 Atlanta                         179              2798
      Baltimore                       437              5515
      Chicago                         576              7564
```

```
    Dayton                          41               765
    Louisville                     229              2960,
'vice_news_summary_sample': {'Denver':         Denver
 count       45
 unique      42
 top        427
 freq         2,
 'El Paso':          El Paso
 count        46
 unique       20
 top           5
 freq          9,
 'Portland':          Portland
 count         45
 unique        34
 top           11
 freq           4,
 'Peoria, AZ':          Peoria, AZ
 count           45
 unique           9
 top              0
 freq            14,
 'Houston, TX':          Houston, TX
 count            46
 unique           36
 top              27
 freq              4,
 "Prince George's County, MD":          Prince George's County, MD
 count                            46
 unique                           29
 top                              11
 freq                              4}}
```


The aggregated summaries provide some key insights:

### 1.0.1 Overall Trends in Car Thefts (2019-2022):

- The carTheftsMap.csv data reveals high variability across years and locations, suggesting fluctuations in theft rates.
- There is a general decline from 2019 to 2022, but the values are inconsistent, likely requiring further analysis to confirm trends accurately.

### 1.0.2 Milwaukee-Specific Trends (2019-2022):

- The KiaHyundaiMilwaukeeData.csv data shows a sharp increase in Kia and Hyundai thefts, particularly from 2020 (895 thefts) to 2021 (6,971 thefts).
- There is a slight drop in 2022, but thefts remain significantly higher compared to 2019 and 2020.

#### 1.0.3 City-Wise Breakdown Across Years:

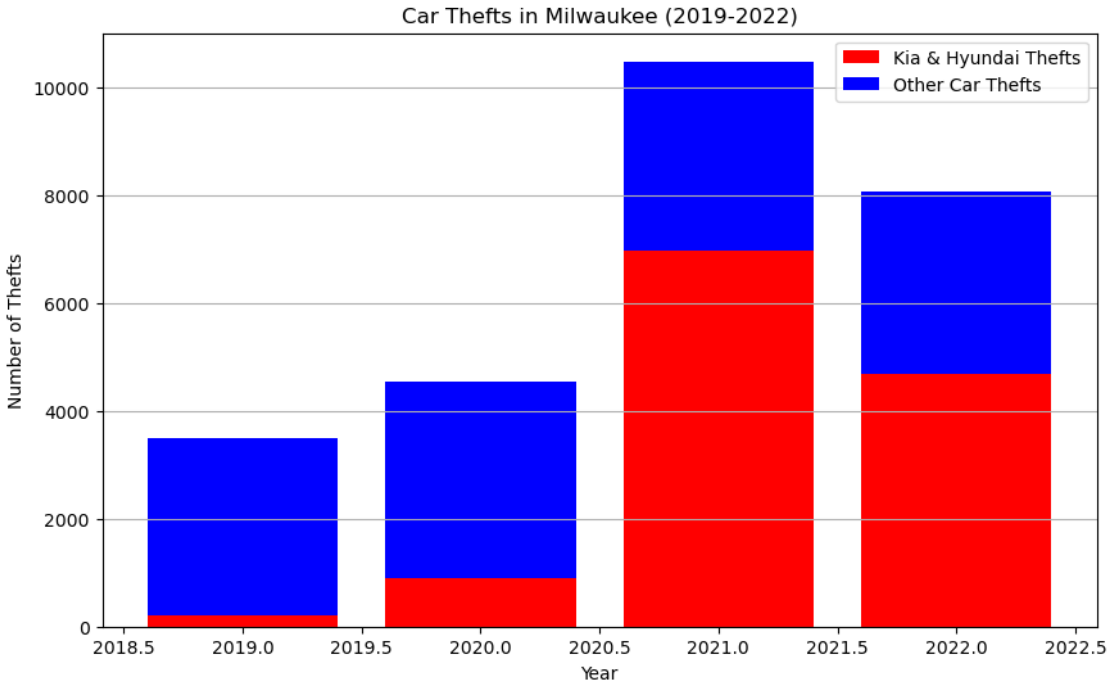
- The kiaHyundaiThefts.csv data, grouped by year and city, indicates cities like Chicago, Baltimore, and Atlanta experienced notable numbers of Kia and Hyundai thefts.
- This data is suitable for exploring regional differences and identifying high-theft areas.

#### 1.0.4 City Trends from VICE News Data:

- Cities like Denver, El Paso, and Portland show varied theft rates of Kia and Hyundai vehicles, with top values indicating specific spikes.
- There are unique patterns for each city, which might be used to illustrate localized spikes and trends.

## 2 2: Visuals

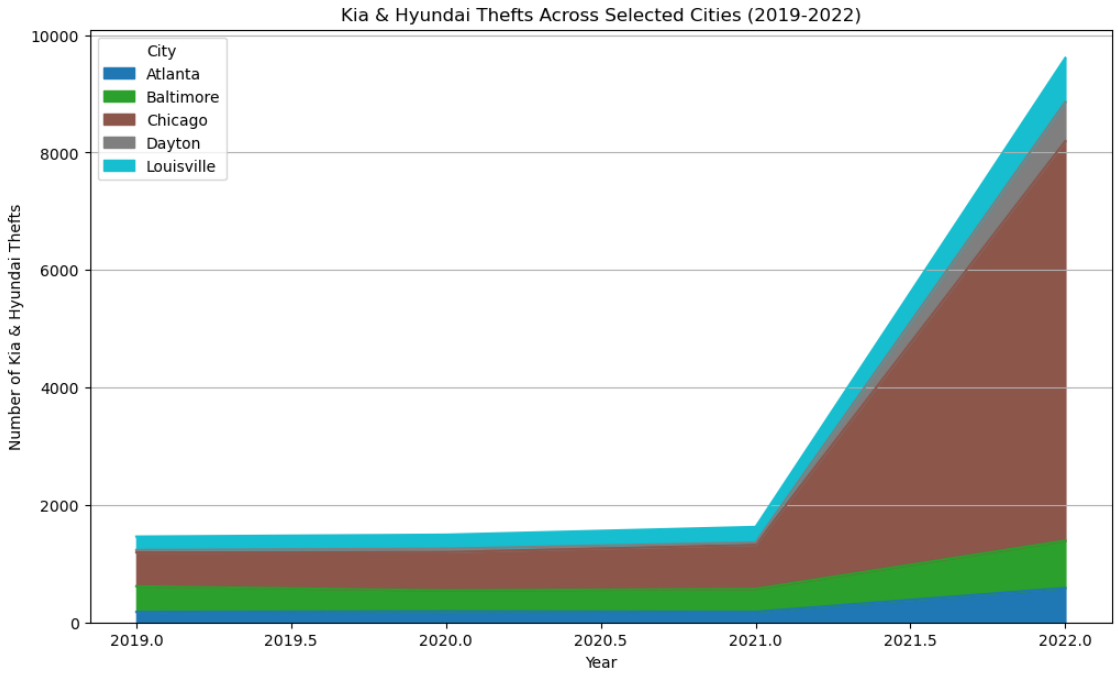
[35]:
```
# Plot Kia and Hyundai thefts over time in Milwaukee
plt.figure(figsize=(10, 6))
years = kia_hyundai_milwaukee_summary.index
kia_thefts = kia_hyundai_milwaukee_summary['countKiaHyundaiThefts']
other_thefts = kia_hyundai_milwaukee_summary['countOtherThefts']

# Stacked bar chart to show the increase in thefts over time
plt.bar(years, kia_thefts, label='Kia & Hyundai Thefts', color='red')
plt.bar(years, other_thefts, bottom=kia_thefts, label='Other Car Thefts',␣
 ↪color='blue')

plt.xlabel('Year')
plt.ylabel('Number of Thefts')
plt.title('Car Thefts in Milwaukee (2019-2022)')
plt.legend()
plt.grid(axis='y')
plt.show()
```




The chart above shows the sharp increase in Kia and Hyundai thefts in Milwaukee, particularly in 2021. While other car thefts remained relatively stable, thefts involving Kias and Hyundais skyrocketed, highlighting a significant problem.

[36]:
```
# Comparing Kia and Hyundai thefts across multiple cities
# Let's select a few key cities to visualize trends

# Cities to focus on for comparison
selected_cities = ['Atlanta', 'Baltimore', 'Chicago', 'Dayton', 'Louisville']

# Filter the data for the selected cities
city_trends = kia_hyundai_thefts_summary.loc[kia_hyundai_thefts_summary.index.
 ↪get_level_values('city').isin(selected_cities)]

# Unstack to prepare for visualization
city_trends_unstacked = city_trends['countKiaHyundaiThefts'].unstack()

# Plot a stacked area chart to show trends over time for these cities
plt.figure(figsize=(12, 7))
city_trends_unstacked.plot(kind='area', stacked=True, ax=plt.gca(),␣
 ↪colormap='tab10')
plt.xlabel('Year')
plt.ylabel('Number of Kia & Hyundai Thefts')
plt.title('Kia & Hyundai Thefts Across Selected Cities (2019-2022)')
plt.grid(axis='y')
```

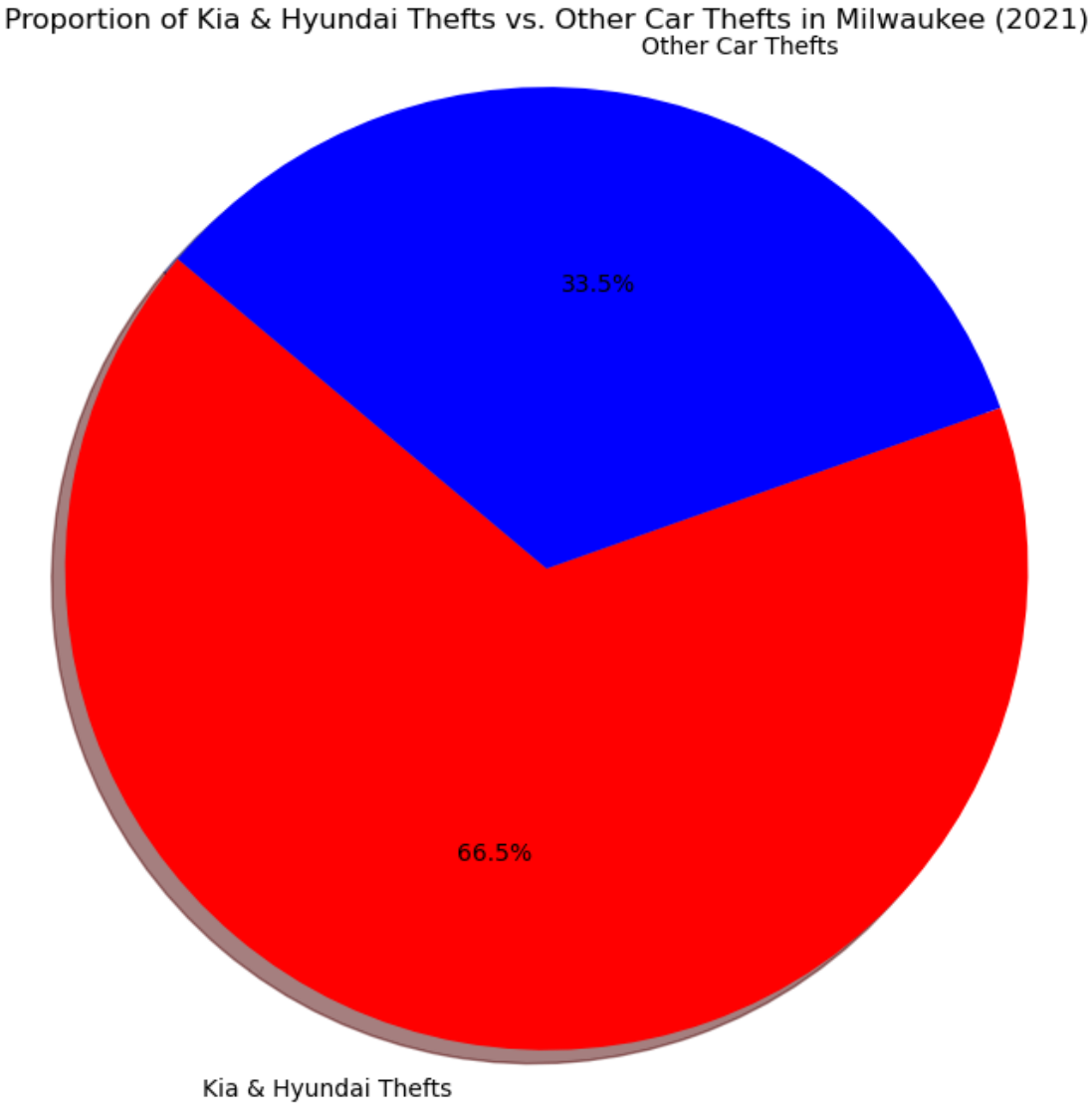
```
plt.legend(title='City')
plt.show()
```




The stacked area chart provides a comparative view of Kia and Hyundai thefts across cities like Atlanta, Baltimore, Chicago, Dayton, and Louisville from 2019 to 2022. This visualization highlights how thefts have evolved differently across cities, with some experiencing more dramatic increases than others.

[37]:
```
# Create a pie chart to show the proportion of Kia/Hyundai thefts vs. other car␣
 ↪thefts in Milwaukee for 2021
year_to_focus = 2021
milwaukee_data = kia_hyundai_milwaukee_summary.loc[year_to_focus]

# Data for pie chart
labels = ['Kia & Hyundai Thefts', 'Other Car Thefts']
sizes = [milwaukee_data['countKiaHyundaiThefts'],␣
 ↪milwaukee_data['countOtherThefts']]
colors = ['red', 'blue']

# Plot the pie chart
plt.figure(figsize=(8, 8))
plt.pie(sizes, labels=labels, colors=colors, autopct='%1.1f%%', startangle=140,␣
 ↪shadow=True)
plt.axis('equal')  # Equal aspect ratio ensures the pie is drawn as a circle.
```

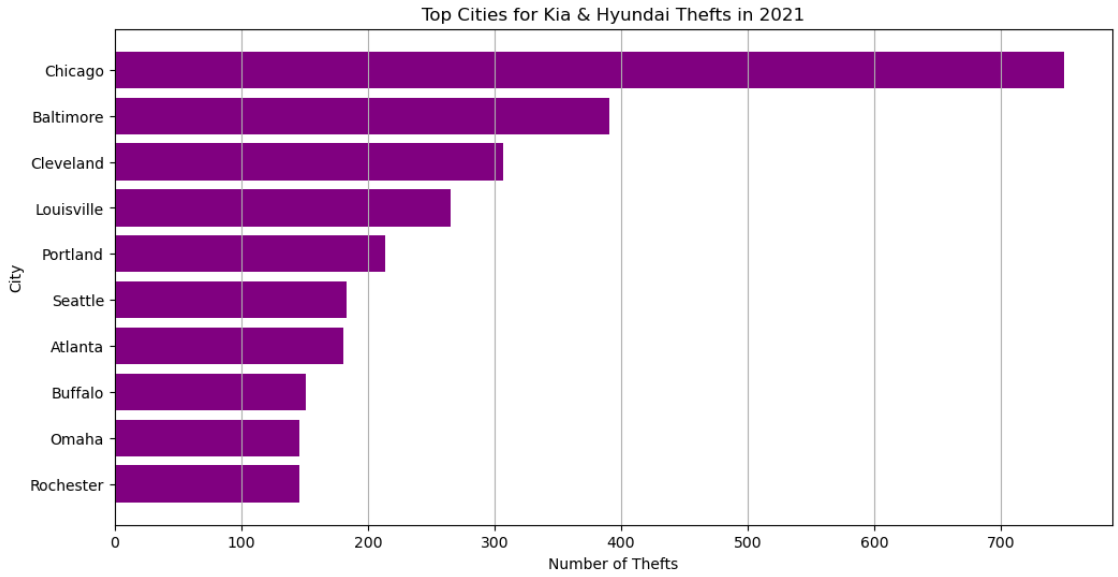
```
plt.title(f'Proportion of Kia & Hyundai Thefts vs. Other Car Thefts in␣
  ↪Milwaukee ({year_to_focus})')
plt.show()
```




The pie chart highlights the proportion of Kia and Hyundai thefts compared to all other car thefts in Milwaukee for 2021. The significant share of Kia and Hyundai thefts underscores the issue's severity that year.

[38]:
```
# Re-define the filtered data for 2021 to correct the error
kia_hyundai_2021 = kia_hyundai_thefts_summary.loc[2021]

# Select top cities based on the number of Kia/Hyundai thefts in 2021
top_cities = kia_hyundai_2021['countKiaHyundaiThefts'].nlargest(10)
```

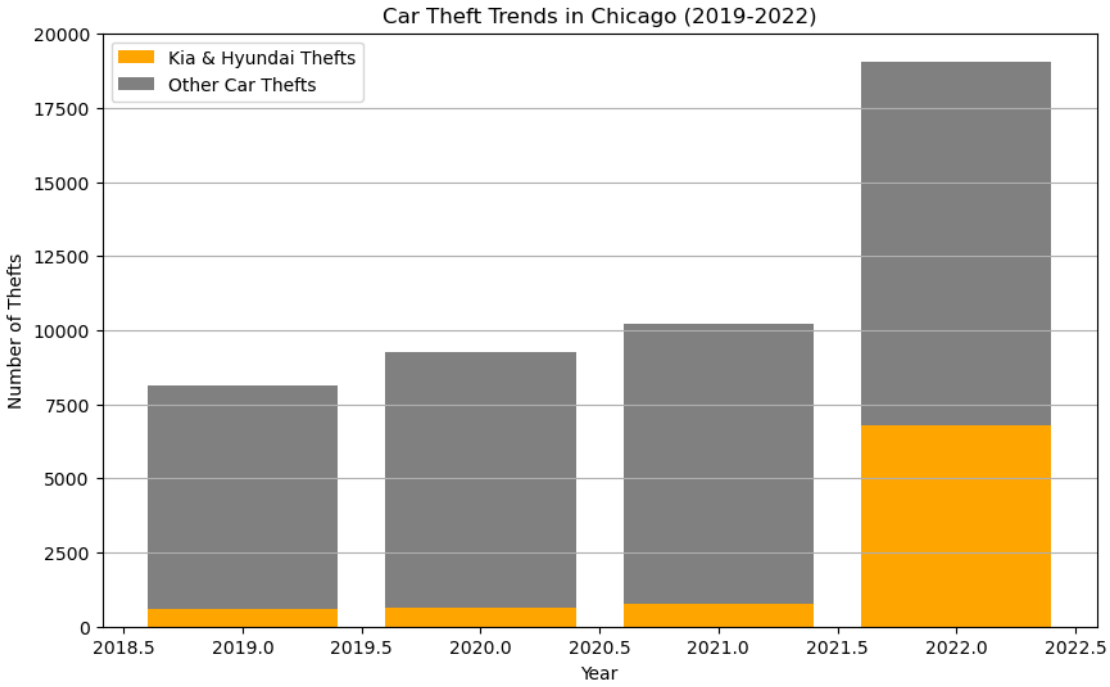
```
# Plot the bar chart
plt.figure(figsize=(12, 6))
plt.barh(top_cities.index, top_cities.values, color='purple')
plt.xlabel('Number of Thefts')
plt.ylabel('City')
plt.title('Top Cities for Kia & Hyundai Thefts in 2021')
plt.gca().invert_yaxis()  # Invert y-axis for better readability
plt.grid(axis='x')
plt.show()
```




The bar chart showcases the top cities with the highest number of Kia and Hyundai thefts in 2021. This visualization makes it easy to identify hotspots and compare theft levels across different regions.

[39]:
```
# Filter data for Chicago across all years
chicago_trends = kia_hyundai_thefts_summary.loc[kia_hyundai_thefts_summary.
 ↪index.get_level_values('city') == 'Chicago']

# Extract years and theft counts for stacked bar chart
years = chicago_trends.index.get_level_values('year')
kia_thefts = chicago_trends['countKiaHyundaiThefts']
other_thefts = chicago_trends['countOtherThefts']

# Plot the stacked bar chart for Chicago
plt.figure(figsize=(10, 6))
plt.bar(years, kia_thefts, label='Kia & Hyundai Thefts', color='orange')
```

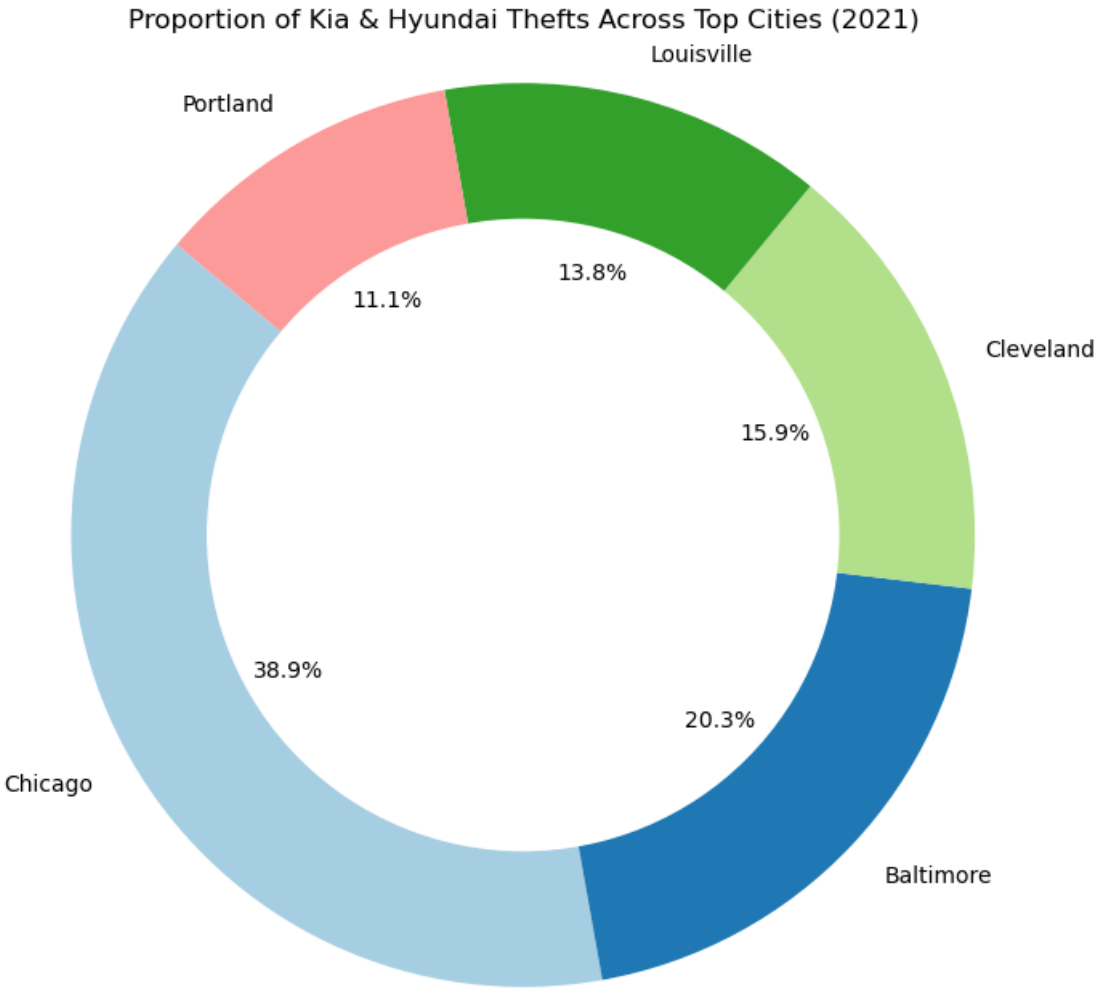
```
plt.bar(years, other_thefts, bottom=kia_thefts, label='Other Car Thefts',␣
 ↪color='gray')

plt.xlabel('Year')
plt.ylabel('Number of Thefts')
plt.title('Car Theft Trends in Chicago (2019-2022)')
plt.legend()
plt.grid(axis='y')
plt.show()
```




The stacked bar chart illustrates the trend of car thefts in Chicago from 2019 to 2022, highlighting the rise in Kia and Hyundai thefts compared to other car brands.

[40]:
```
# Select the top 5 cities based on the number of Kia/Hyundai thefts in 2021
top_cities_2021 = kia_hyundai_2021['countKiaHyundaiThefts'].nlargest(5)

# Prepare data for the donut chart
labels = top_cities_2021.index
sizes = top_cities_2021.values
colors = plt.cm.Paired.colors

# Plot the donut chart
plt.figure(figsize=(8, 8))
plt.pie(sizes, labels=labels, colors=colors, autopct='%1.1f%%', startangle=140)
# Adding the donut hole
```

```
centre_circle = plt.Circle((0, 0), 0.70, fc='white')
plt.gca().add_artist(centre_circle)

plt.axis('equal')  # Equal aspect ratio ensures the pie is drawn as a circle.
plt.title('Proportion of Kia & Hyundai Thefts Across Top Cities (2021)')
plt.show()
```




The donut chart visualizes the proportions of Kia and Hyundai thefts across the top 5 cities in 2021, clearly showing which locations experienced the most significant impact.

# 3 Story Outline:

### 3.0.1 Audience:

The target audience is the general public, which includes car owners, local residents, and community stakeholders who may not have in-depth knowledge of car theft data but are likely aware of the growing issue with Kia and Hyundai thefts. Therefore, the language used will be accessible and avoid technical jargon. The intent is to connect on an emotional level to raise awareness, provide understanding, and drive community action or prevention efforts.

### 3.0.2 Purpose:

The story aims to raise awareness of the rising trend in Kia and Hyundai thefts, highlight hotspots (e.g., Milwaukee), and emphasize the need for increased community vigilance, vehicle safety measures, and possibly support for local law enforcement initiatives. The call to action is for community members to take preventative steps (e.g., car alarms, steering wheel locks) and advocate for vehicle manufacturers to improve security features.

### 3.0.3 Medium:

The best way to present this story to the general public would be a short infographic or social media post series (e.g., Instagram carousel or a short video) since these formats are easily shareable, visually engaging, and digestible for a wider audience. The visuals from the data (bar charts, pie charts, stacked area charts) will be used to tell the story in a compelling manner.

### 3.0.4 Design:

**Gestalt's Principles:**

- **Proximity & Similarity:** Group related information together (e.g., all theft data visualizations for the same city are placed together).
- **Figure & Ground:** Use contrasting colors to highlight key data points, ensuring Kia/Hyundai thefts are clearly distinguished from other data.
- **Continuity & Closure:** Visuals will flow in a logical sequence, with theft trends shown over time to tell a clear progression of the issue.

**Color Strategy:** Maintain a consistent color scheme across visuals to create a cohesive story.

### 3.0.5 Text:

Short, impactful headings (e.g., "Kia & Hyundai Thefts are Rising!"). Bullet points or short phrases to describe data and its implications. Use larger font sizes for key messages or calls to action (e.g., "Protect Your Vehicle!").

### 3.0.6 Alignment & Sizing:

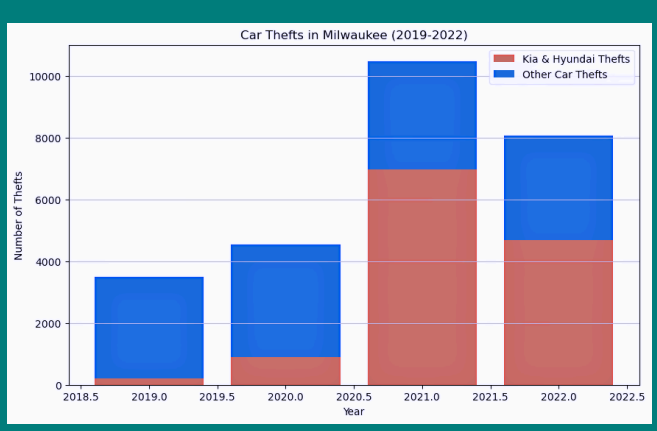
- Center-align titles and charts to make them the focal point.
- Use consistent sizing for charts and text blocks to maintain a clean and organized layout.
- Sufficient spacing between visual elements to avoid clutter and keep the content readable.

#### 3.0.7 Ethical Considerations:

- **Data Changes & Integrity:** The data used is anonymized, focusing only on the number of thefts without personal information or specific victim details to respect privacy.
- **Assumptions & Presentation:** The data was not altered beyond aggregation and visualization. All cities are clearly labeled, and the focus is on trends, not individuals, ensuring an ethical presentation of facts.
- **Legal & Regulatory Compliance:** As the data is publicly available or aggregated from open sources, it adheres to ethical standards and privacy laws. Efforts have been made to ensure that no sensitive information is exposed.
- **Mitigating Misrepresentation:** All data is presented with clear labeling, context, and proportionate visualization to avoid exaggeration or misleading interpretations.

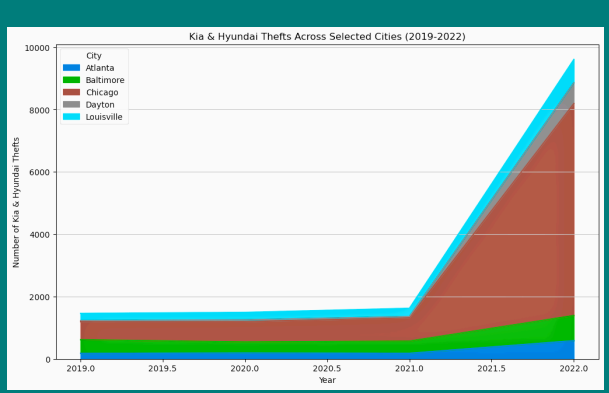
[ ]:

# KIA & HYUNDAI THEFTS: A GROWING PROBLEM.



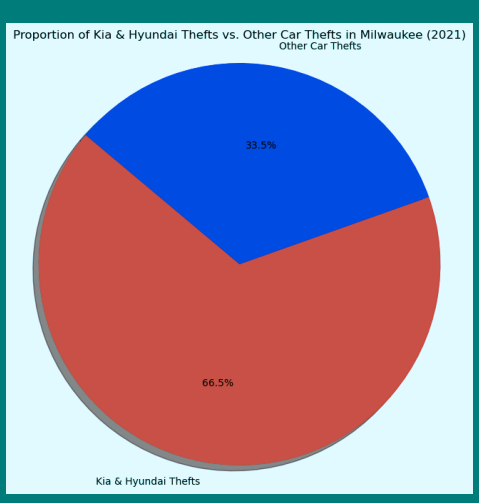

## Milwaukee's Kia & Hyundai Theft Surg

Thefts in Milwaukee skyrocketed in 2021, with Kia & Hyundai vehicles accounting for a significant increase. While other car thefts remained stable, these brands saw a 7x jump.



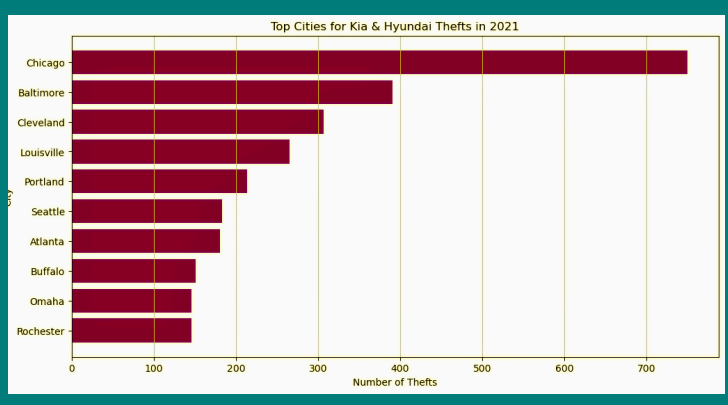

## Rising Kia & Hyundai Thefts Nationwide

Cities like Chicago, Louisville, and Baltimore experienced a steep rise in Kia & Hyundai thefts between 2021 and 2022, showing the growing vulnerability of these brands.



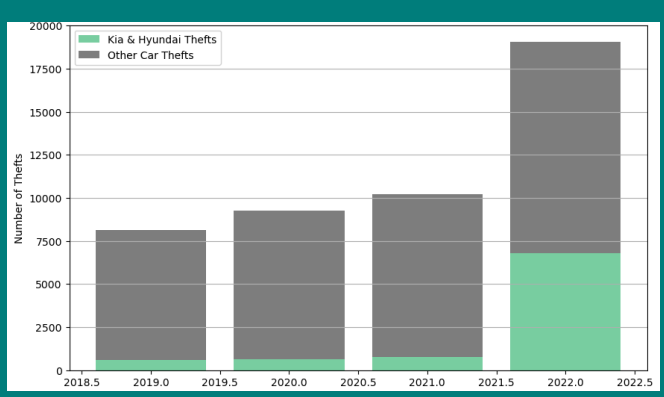

## Kia & Hyundai Dominate Theft Numbers in Milwaukee

In 2021, over 66% of all car thefts in Milwaukee involved Kia & Hyundai vehicles, making them the top targets for thieves



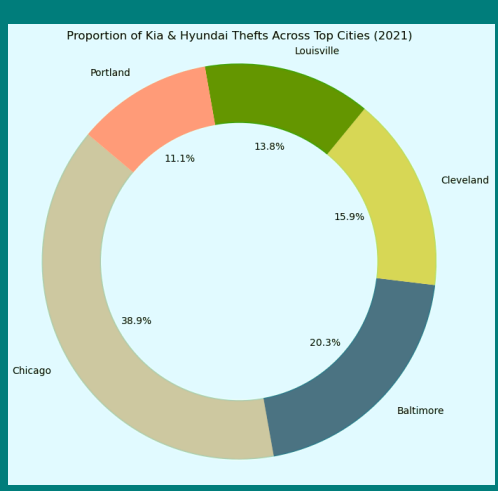

## Top U.S. Cities for Kia & Hyundai Thefts

Chicago has the highest number of Kia & Hyundai thefts in 2021, followed closely by Baltimore and Cleveland. Are these brands more at risk in your city?



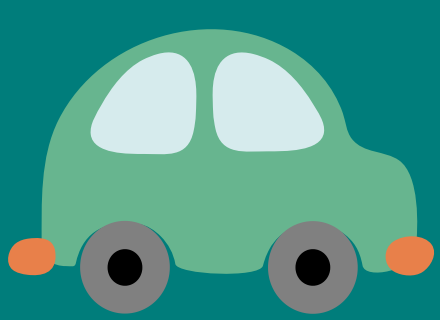

## Chicago's Growing Car Theft Problem

Kia & Hyundai thefts in Chicago remained low until 2022, when they made up a notable portion of the total car thefts, indicating a rapid increase.




## Distribution of Kia & Hyundai Thefts Across Major Cities

Chicago alone accounted for nearly 39% of all Kia & Hyundai thefts in the top cities in 2021, highlighting it as a major hotspot.



## What Can You Do?

- Use a steering wheel lock
- Install an alarm system
- Park in well-lit areas.