Carjacking Analysis in Chicago: Visualizing Trends for Policy Implications

PPHA 30538 Final Project Group 28 - Surya Hardiansyah and Astari Raihanah

2024-12-07

Group 28 Members:

- Astari Raihanah (CNetID: astari) - GitHub username: astari1007
- Surya Hardiansyah (CNetID: sur) - GitHub username: suryahardiansyah
- Both members are from Section 2 of Professor Ganong's lecture (Monday and Wednesday, 10:30-11:50 AM).

Research Question

This project examines the temporal and spatial patterns of carjacking incidents across Chicago neighborhoods from 2001 to 2024. The goal is to identify carjacking hotspots and trends, offering actionable insights for policymakers and insurance companies to design fair and effective crime prevention strategies.

Approach

Data Sources

We utilized datasets from the Chicago Data Portal, including:

1. **Carjacking Data**: Provides details of carjacking incidents, including date, time, and location coordinates (2001-2024)
2. **Neighborhood Boundaries**: Geospatial data (GeoJSON) to associate carjackings with specific neighborhoods.

Data Preparation and Analysis

Key Steps

1. **Data Retrieval**: Carjacking reports retrieved from the Chicago Data Portal using API pagination stored locally as a CSV file. Neighborhood Boundaries is downlaoded as a GeoJSON file via a single API request and stored locally.
2. **Data Preparation**: The already structured datasets required no cleaning, and spatial join merged carjacking incidents with neighborhood boundaries using coordinate variables.
3. **Data Aggregation**: Dataset is grouped by Year, Month and Time of Day (Morning, Afternoon, and Evening) for temporal trend analysis.
4. **Visualization**: By static choropleth map and static line chart, identifying patterns by neighborhood and trend over time.
5. **Interactive Dashboard**: Using Shiny to built dynamic filtering of data by neighborhood and date range by choropleth map.
6. **NLP Sentiment Analysis**: Text data on carjacking, auto insurance, and policies were collected using SerpAPI to scrape news articles and blog posts with queries like "Chicago car insurance policy", "auto insurance Chicago", "carjacking auto insurance Chicago", and "auto insurance Chicago car theft." The data was analyzed using spaCy and TextBlob to extract sentiment scores, including polarity (positivity/negativity) and subjectivity (degree of bias or opinion).
7. **Storage and Reproducibility**: All raw and processed data files are stored in a local 'data' directory. Documentation ensures reproducibility of the analysis and visualizations.

Challenges

1. **API Limitations**: The Chicago Data Portal limits API downloads to 1,000 rows per request, hence we used pagination to retrieve all 22,192 records.
2. **Data gaps**: Missing or incomplete records, such as missing coordinates, required exclusion (144 data rows per November 30th, 2024).

Static Plots

1. Choropleth Map: Carjacking Incidents by Neighborhood

This map visualizes carjacking incidents by neighborhood. Some neighborhoods such as Austin, West Garfield Park, East Garfield Park, Englewood, and West Englewood have higher carjacking counts. These neighborhoods should be prioritized for policy interventions, such as increased patrolling or community safety programs.

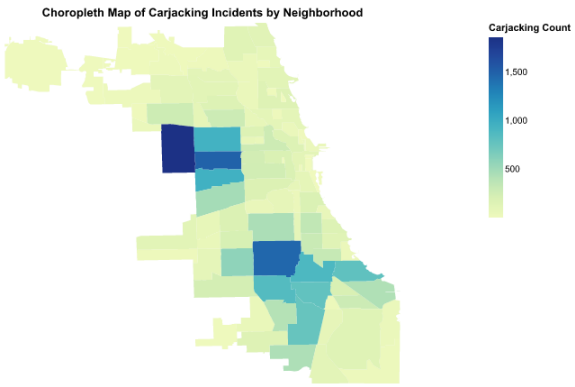

Figure 1: Choropleth Map of Carjackings

2. Line Chart: Carjacking Trends Over Time

This chart highlights trends in carjacking incidents over time. A sharp increase is evident around 2020, peaking in 2021. This trend may reflect broader societal disruptions, such as those caused by the COVID-19 pandemic and economic turmoil. Policymakers could use this data to explore the impact of external events on crime rates.

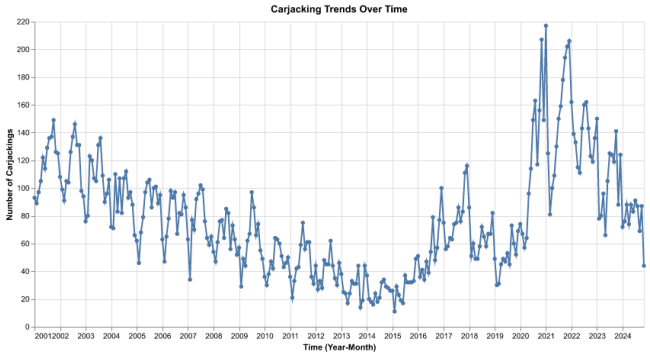

Figure 2: Carjacking Trends Over Time

3. NLP Analysis

Collected text data on Chicago auto insurance, carjacking, and insurance policy using web scraping and APIs and Applied NLP for sentiment analysis: Subjectivity and Polarity. The sentiment analysis of auto insurance-related text data shows a predominantly neutral-to-negative sentiment regarding carjacking impacts on premiums.

NLP Analysis: Polarity and Subjectivity

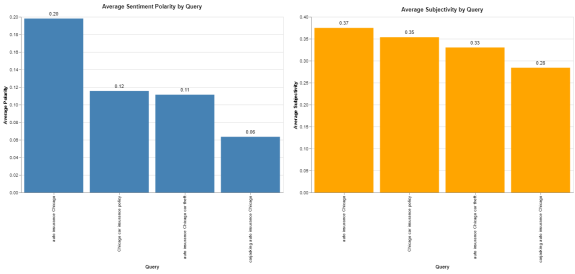

Figure 3: Average Polarity and Subjectivity by Query

Findings from sentiment analysis charts reveal that "auto insurance Chicago" is discussed more positively compared to terms directly related to carjacking, reflecting a gap in addressing public concerns about the impact of crime on insurance.

Shiny Dashboard

The **Interactive Carjacking Dashboard** enables users to: 1. Explore carjacking patterns dynamically using a **Choropleth Map** filtered by neighborhood and date range. 2. Visualize temporal patterns through a **Line Chart** with customizable filters based on neighborhoods and time for precise analysis.

Click here to open the Shiny app in the browser.

Strengths

- **Dynamic Exploration**: Offers flexibility to investigate specific neighborhoods or time periods.
- **Policy Relevance**: Combines spatial and temporal data along with sentiment analysis for actionable insights.

Weaknesses

- **Performance**: Processing large datasets slows responsiveness
- **Data Limitations**: Missing records in the dataset can affect accuracy.

Policy Implications

1. Neighborhood Hotspots: Targeted Interventions

Neighborhoods like Austin, West Garfield Park, East Garfield Park, Englewood, and West Englewood are carjacking hotspots, requiring focused interventions.

- **Policy Actions**: Increase police patrols during high-risk hours, implement community safety programs like neighborhood watch and workshops, improve infrastructure like installing street lighting and surveillance cameras, and invest in socioeconomic support and community engagement like job training and education.

2. Temporal Patterns: Addressing Crime Spikes

The sharp increase in carjackings during 2020-2021 and patterns in seasons and times of day highlight actionable trends. This trend may be attributed to disruptions caused by the COVID-19 pandemic, economic uncertainty, and resource reallocation during social unrest. Incidents are higher in the evening and winter months

- **Policy Actions**: Allocate law enforcement resources during high-crime months and times, and prepare for crime surges during external crises such as pandemics or economic crises.

3. Insurance Policies: Data-Driven Solutions

Car insurance often unfairly penalizes residents of high-crime areas. Policies should prioritize fairness.

- **Policy Actions**: Use risk-based pricing focused on individual factors, incentivize safety measures such as giving discounts for anti-theft devices, require transparency in how crime stats influence premiums, and collaborate with law enforcement to lower risks and premiums.

Future Directions

Future work aims to integrate additional datasets, such as socioeconomic indicators, traffic patterns, and law enforcement reports, to provide deeper insights into the factors influencing carjacking trends. Enhancements to the dashboard will focus on improving interactivity and scalability, enabling broader accessibility and more dynamic exploration of spatial and temporal crime patterns.

Conclusion

Combining spatial, temporal, and sentiment analysis provides a comprehensive understanding of carjacking trends in Chicago. By addressing hotspots, temporal patterns, and public sentiment, this project offers valuable insights for crime prevention and insurance reform.