Vartan Pashayan

12/19/2024

Final Project Report

1. Introduction

(a) Vehicle Collision Research Case Study

This project explores motor vehicle collision data to analyze patterns and trends in crashes in New York City, focusing on factors that contribute to these events and how external conditions such as the COVID-19 pandemic have impacted crash rates.

(b) Research Question and Synopsis

Research Question: What conclusions can we draw to correctly address vehicle crashes using crash data? What times, cars, and reasons are associated with the highest motor vehicle collisions?

Synopsis: This project investigates motor vehicle crash data from NYC to uncover critical insights into when, where, and under what conditions crashes are most likely to occur. By analyzing this data, we aim to identify actionable patterns that can help improve road safety and inform public policy. Additionally, the project examines the effects of the COVID-19 pandemic on crash rates and severity.

(c) Research Questions

- What conclusions can we draw to correctly address vehicle crashes using crash data?
- What times, cars, and reasons are associated with the highest motor vehicle collisions?
- How did the COVID-19 pandemic influence crash patterns and rates in NYC?

---

2. Data Collection

(a) Data Source

The data for this project was collected from the **NYC OpenData Motor Vehicle Collisions - Crashes dataset**, which is maintained by the New York City Police Department (NYPD). The dataset is publicly available and regularly updated.

Rensselaer Polytechnic Institute

Research shows that traffic accidents decreased by nearly half during the two-month period at the start of the COVID-19 pandemic when residents of most states were under a stay-at-home order.

(b) Data Collection Approach

The first dataset was pre-collected and made available by the NYPD. It contains detailed information about motor vehicle collisions, including crash times, locations, contributing factors, and the types of vehicles involved.

The second dataset involved scraping sciencedaily.com to find the conclusions from the study done by Rensselaer Polytechnic Institute

(c) Number of Data Sources

Two data sources data source was used for this project: the NYC Open Data Motor Vehicle Collisions dataset and Louisiana Department of Transportation.

(d) Data Volume

The dataset includes over 1,000,000 samples spanning a five-year period, offering a robust dataset for analysis.

(e) Changes and Challenges

Initially, I planned to analyze data from multiple cities to provide a comparative analysis. However, the size and complexity of each city's dataset led me to focus exclusively on NYC to ensure a more in-depth and manageable study. Additionally, I supplemented the analysis with web scraping to retrieve metrics from studies on pandemic traffic trends.

---

3. Data Cleaning

(a) Changes to the Data

The dataset required significant cleaning before analysis.

- Removed entries with missing or "Unspecified" values in key columns, such as contributing factors and vehicle types.
- Filtered data to include only crashes occurring after 2016 for consistency.
- Converted date columns to datetime format for proper time series analysis.
- Removed irrelevant columns to streamline the analysis process.

(b) Original Data Format

The raw data was in CSV format, containing a mix of numerical and categorical variables, as well as missing and inconsistent values.

(c) Processed Data Format

The cleaned dataset was stored as a pandas DataFrame in Python for ease of manipulation and analysis.

(d) Effects of Cleaning

Cleaning the data reduced the total number of samples by removing incomplete or irrelevant entries. However, this ensured the reliability and accuracy of the analysis. The cleaned dataset maintained sufficient volume for meaningful statistical analysis and visualization.

---

4. Data Analysis

(a) Techniques and Findings

Techniques

- **Descriptive Statistics:** To summarize data trends.
- **Time Series Analysis:** To evaluate crash patterns over time and identify changes during the pandemic.
- **Categorical Analysis:** To identify top contributing factors and vehicle types involved in crashes.
- **Percent Change Calculations:** To measure the reduction in crashes during the pandemic compared to pre-pandemic periods.
- **Web Scraping:** To retrieve supplementary data on crash patterns from external studies.

Findings

- The most common contributing factors were **driver inattention/distraction**, **following too closely**, and **failure to yield the right-of-way**.
- **Covid-19** did not change the top 3 contributing factors leading to crashes.
- Sedans and SUVs and passenger vehicles (ex. Taxis) were the vehicle types most often involved in crashes.
- **Covid-19** did not change the top 3 vehicle types that were involved in the crashes.
- Crash rates peaked during evening rush hours (4 PM – 6 PM) and late at night (11 PM – 1 AM).
- The COVID-19 pandemic significantly reduced crash frequencies due to lockdowns and reduced road usage.
- Calculations revealed a **62.95% drop in crashes** during the pandemic compared to pre-pandemic levels.
- Average monthly crashes decreased from **12,941.04 (pre-pandemic)** to **5,677.79 (pandemic)**, a drop-off of **7,263.25 crashes per month**.
- Seasonal patterns persisted, with winter months showing higher crash rates due to adverse weather conditions.

Supplementary studies showed similar trends nationwide:

- Traffic accidents decreased by **47%** in Louisiana (March-May 2020).
- Accidents involving injuries decreased by **46%**.
- Accidents where an ambulance was called decreased by **41%**.
- Fatal accidents showed no significant decrease.
- An estimated **$21 billion** reduction in car crash costs occurred nationally during this period.
- Demographic impact: Smaller reductions were observed among males, nonwhite drivers, and individuals aged 25–64.

---

5. Visualizations

(a) Types of Visualizations

1. **Bar Charts:** To display the top contributing factors and vehicle types involved in crashes.
2. **Time Series Plots:** To analyze trends and fluctuations in crash frequencies over time.
3. **Pandemic vs. Pre-Pandemic Comparisons:** To illustrate the significant reduction in crashes during the pandemic.

(b) Visualization Details

- **Bar Chart (Contributing Factors):** Showed the top 10 contributing factors, highlighting "Driver Inattention" as the leading cause.
- **Bar Chart (Vehicle Types):** Highlighted that sedans, SUVs, and taxis were the most common vehicle types involved in crashes.
- **Time Series Plot:** Illustrated the number of crashes per month from 2016 to 2024, showing a sharp decline during the pandemic.
- **Drop-Off in Crash Counts:** Highlighted the average 47% reduction in crashes during the pandemic period, quantified and visualized.

---

6. Conclusion

(a) Observations and Conclusions

- **Key Patterns:** Crashes are most frequent during rush hours and late-night hours, with driver behavior being a critical factor.
- **Pandemic Impact:** The COVID-19 pandemic led to a significant reduction in crash rates, reflecting changes in mobility patterns and reduced road activity.

- **Vehicle Involvement:** Sedans and SUVs and Taxis, being the most common vehicle types on the road, are disproportionately involved in crashes.
- **Supplementary Insights:** Web-scraped metrics supported the findings, showing consistent reductions in crash rates nationwide during the pandemic.

(b) Impact of Findings

The insights from this project can guide policymakers and urban planners to:

- Implement targeted measures to reduce driver inattention and aggressive driving behaviors.
- Enhance traffic enforcement during peak crash hours.
- Prepare contingency plans to maintain safety during crises such as pandemics.

These findings emphasize the importance of data-driven strategies for improving road safety in urban environments and adapting to changing conditions such as public health emergencies.

4. Future Work

(a) Given more time, what would you do in order to further improve your project?

- If I was given more time to work on this project, I would analyze more patterns, like where crashes occur, delve deeper into seasonality charts and trend lines. I would also explore which ages and sexes are involved in crashes and if neighborhoods with less cars have less car crashes.

(b) Would you use the same data sources next time? Why yes or why not?

- Yes, I did not find any other data sources that came close to the detail and depth of information that the NYPD Crash Data provided. It was a lot to work with, but it was manageable. There is so much more information and features in this dataset that I have not explored. It gives me room to explore more.