Midterm QLIK Project

Project Title: Trends of Collisions in New York City

Team members:

1. **Anh Diep Chung**
2. **Haomin Gu**
3. **Prajwal Guntnur Anantha**
4. **Sangayya Rudrayya Hiremath**
5. **Tingyu Long**

Northeastern University: College of Professional Studies

ALY6070 – Communication and Visualization for Data Analytics

Professor - Jack Bergensen

Week 4 – 6th February 2025

Introduction

This study leverages the "Motor Vehicle Collisions - Crashes" dataset spanning from 2012 to 2024, sourced from Data.gov [1]. The main goal of this report is to analyze traffic collisions in New York City, identifying key patterns, trends, and high-risk areas. The report specifically aims to answer the following critical questions:

1. Where are the primary collision hotspots in NYC?
2. Total number of people injured each year due to vehicle collisions.
3. Total number of people killed each year due to vehicle collisions.
4. Which types of vehicles are most frequently involved in collisions?

To analyze these questions, the following variables are utilized:

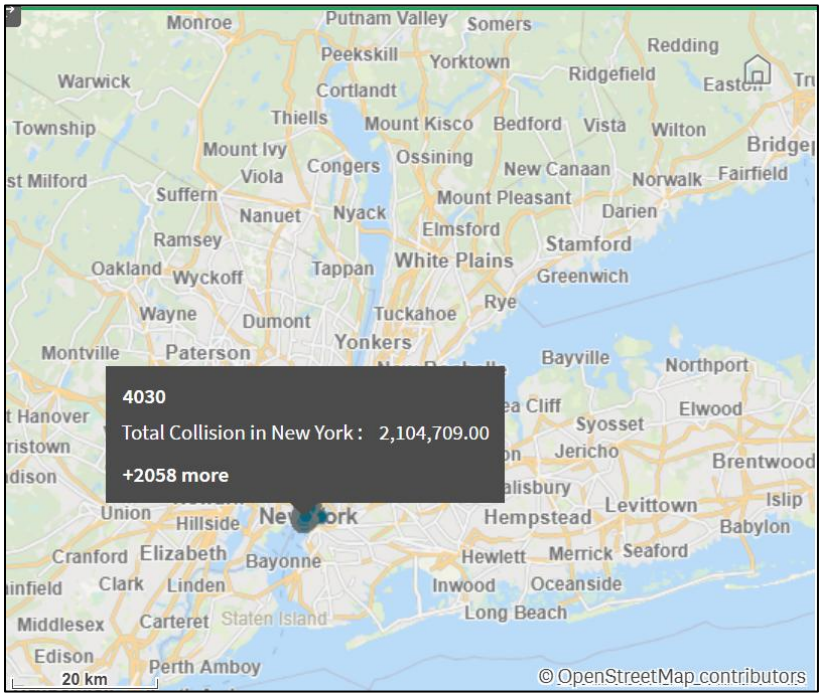
1. Collision Factors: Vehicle Types
2. Impact Metrics: Number of Person Injured, Number of Person Killed and Total Collisions Recorded

The analysis conducted using QLIK, leveraging its capabilities to generate meaningful insights through visualization techniques. The primary methods include:

1. **Filtering and Grouping:** Data is grouped by, vehicle types, number of injured and and number killed to identify trends and patterns.
2. **Data Visualization:** Various graphical representations, including bar charts, geographical maps, and tree maps, are employed to illustrate findings effectively.

Analysis and Insights

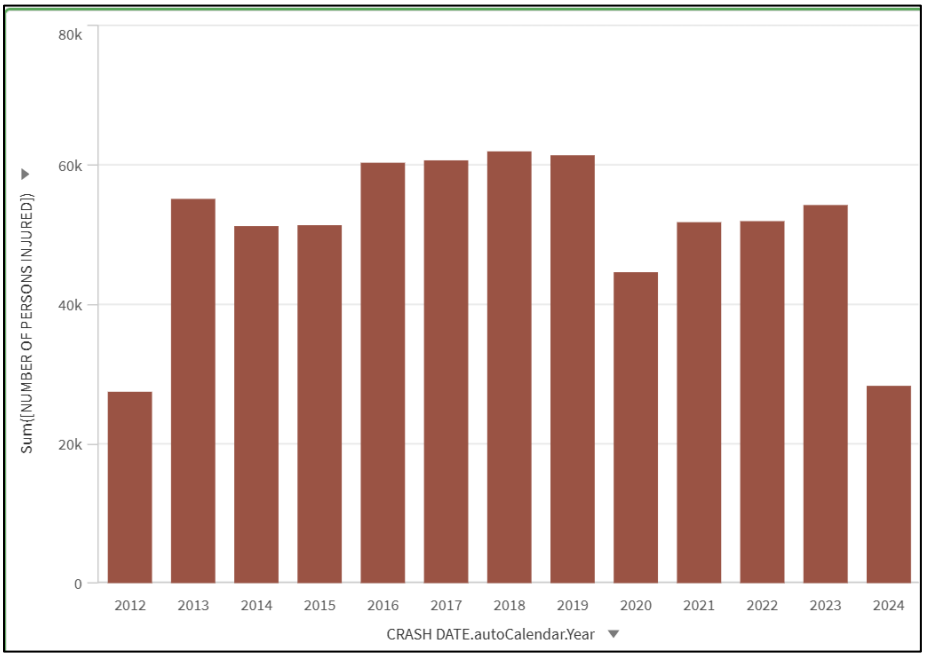
1. **Where are the primary collision hotspots in NYC?**



Insights: The geographic map above presents vehicle collision data in New York City. The number 4030 likely represents incidents occurring at a specific location or within a defined timeframe, while the total collisions of 2,104,709 indicate the cumulative number of recorded accidents across the city. The +2058 more suggests additional incidents that are not currently visible in the displayed view. With over two million reported collisions, the data highlights the significant traffic risk in NYC. The hotspots are primarily concentrated in Manhattan, Queens, and Brooklyn, as these areas experience high traffic volume and dense urban activity.

Design decisions: A geographic map is used in this case because it visually represents the distribution of accidents across various boroughs, such as Manhattan, Queens, and Brooklyn, making it easier to identify high-risk areas [2].

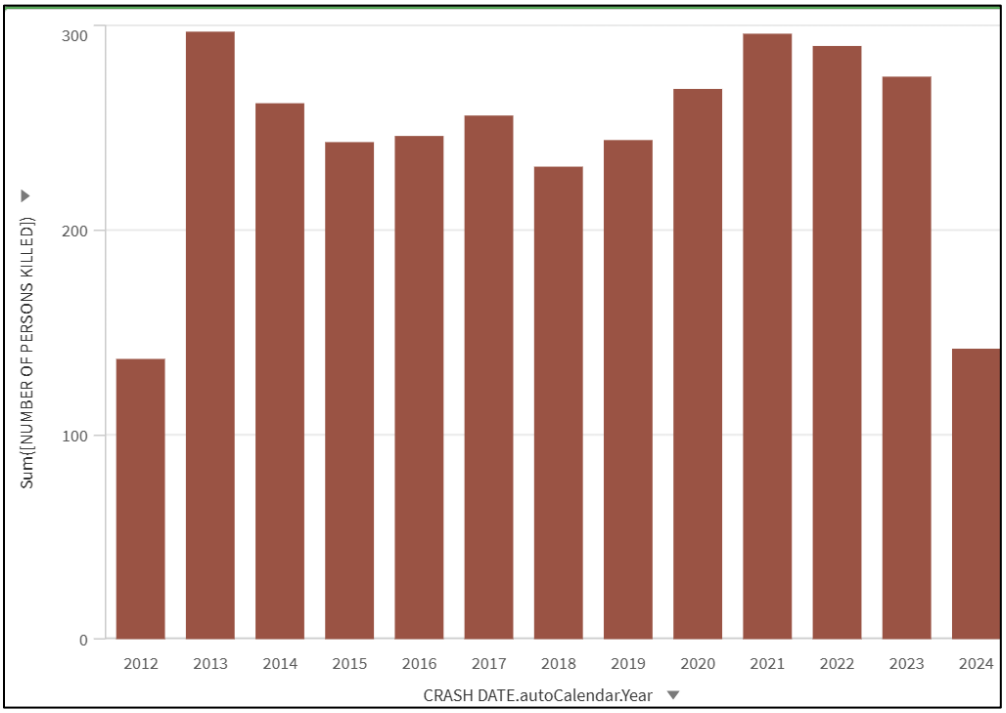
2. **Total number of people injured each year due to vehicle collisions.**



Insights: The bar chart illustrates the total number of individuals injured in vehicle accidents in New York City between 2012 and 2024. There was a significant rise in injuries from 2012 to 2013, which remained elevated from 2016 to 2019, reaching its highest point around 2018–2019. A noticeable drop in 2020 can likely due to decreased traffic during the COVID-19 lockdown, followed by a slow recovery in 2021–2023 as traffic conditions returned to normal. The decrease observed in 2024 might indicate incomplete data or the effects of enhanced traffic safety initiatives.

Design decisions: A bar chart is used in this case for its clear and straightforward representation of information. It effectively displays trends while maintaining simplicity, making it easy for viewers to follow and interpret.

3. **Total number of people killed each year due to vehicle collisions.**



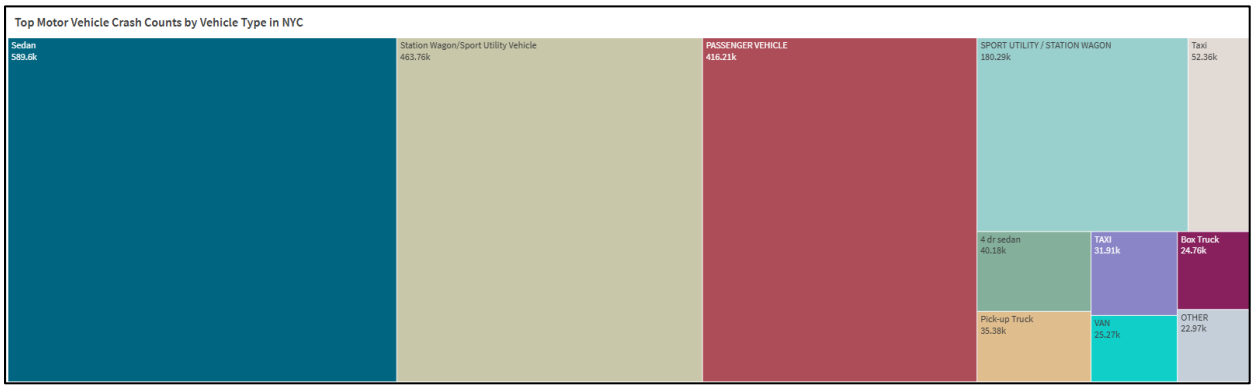
Insights: The bar chart shows traffic fatalities in New York City from 2012 to 2024, with peaks in 2013 and 2021 and consistently high numbers from 2014 to 2023. The 2013 spike may be due to increased traffic volume and the lack of Vision Zero safety measures, which were introduced in 2014 [3]. The 2021 rise is linked to post-pandemic traffic surges, increased speeding, and reckless driving. The 2024 decline could indicate incomplete data or improved safety measures.

Design decisions: For the dashboard above, a bar chart is used due to its simplicity and effectiveness in clearly presenting the information.

4. **Which types of vehicles are most frequently involved in collisions?**

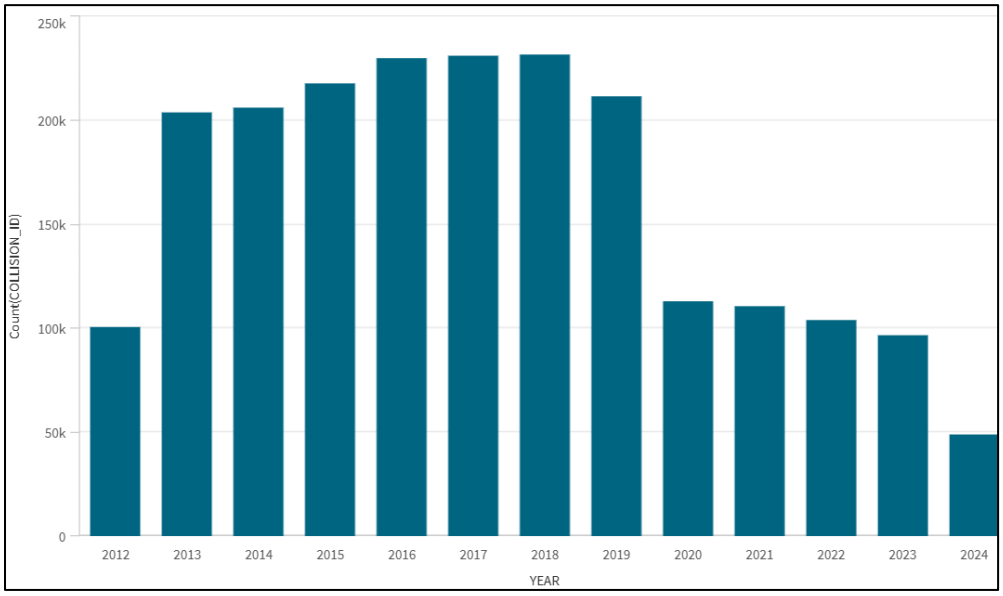
Insights: The tree map chart illustrates how motor vehicle crashes are distributed in NYC based on vehicle type, revealing that Sedans (589.6k) account for the highest number of incidents, followed by SUVs (463.76k) and passenger vehicles (416.21k), which reflects their significant

presence in urban traffic. Taxis (52.36k), box trucks (24.76k), and vans (25.27k) also play a role, underscoring the risks associated with commercial transportation. The involvement of pickup trucks (35.38k) and 4-door sedans (40.18k) suggests that size, weight, or common use in urban environments may be contributing factors.



Design decisions: A tree map is an effective visualization tool for displaying proportions through rectangles, allowing for a quick comparison of category sizes [4]. It is especially suitable for this dataset as it clearly highlights the relative contribution of each vehicle type to the total number of crashes.

5. **Collision trend over the years.**



Insights:

The bar chart depicts the number of collisions in New York City from 2012 to 2024. There is a noticeable upward trend in the number of collisions from 2012, peaking in 2016 and 2017, with consistent high levels through 2018. However, starting in 2019, the number of collisions declines sharply, with 2020 and 2021 reflecting significantly lower figures, likely due to reduced traffic volumes during the COVID-19 pandemic. The steady decline continues in subsequent years, with the lowest values in 2024, possibly because of improved road safety measures, changing traffic patterns, or incomplete data for the year.

Design decisions: A bar chart was chosen because it effectively compares collision counts over the years, making it easy to observe trends and fluctuations. Year is placed on the X-axis for chronological comparison, while the Y-axis displays the count of collisions, providing a clear and straightforward representation.

Conclusion

This analysis of NYC traffic collisions using QLIK uncovered important trends and risk factors. Manhattan, Queens, and Brooklyn stood out as major crash hotspots, with injuries and fatalities peaking between 2016 and 2019 before dropping in 2020, likely due to reduced traffic during COVID-19. Sedans, SUVs, and passenger vehicles were involved in the most crashes, reflecting their heavy presence on city roads. These findings highlight the ongoing need for traffic safety measures and better urban planning. Incorporating additional factors like weather and traffic signals in future research could help create more effective accident prevention strategies.

References

[1] Publisher data.cityofnewyork.us. (2025b, January 24). *City of New York - Motor vehicle collisions - crashes*. Catalog. https://catalog.data.gov/dataset/motor-vehicle-collisions-crashes

[2] Dodge, M., McDerby, M., & Turner, M. (2008). *Geographic visualization : concepts, tools and applications*. Wiley. https://doi.org/10.1002/9780470987643

[3] Vision Zero in NYC and Chicago. (2018). *The Brian Lehrer Show*.

[4] Lichenko, A. (2022). Treemap as a type of data visualization in the media. *Sinopsis: tekst, kontekst, medìa*, *28*(3), 156–163. https://doi.org/10.28925/2311-259x.2022.3.8