Exploring top contributing factors in injury and fatal crashes and hourly seasonality and spatial difference in total crashes

Yuanyuan Wu
Division of Nutritional Sciences,
Cornell University

Introduction

With the NYC OpenData Motor Vehicle Collisions dataset, we were able to track and analyze the crashes happened in NYC for more than ten years. Insights from this dataset are essential for regulation and directions for safe driving. The research questions we had are:

•What are the major contributing factors for fatal crashes?
•Is there any seasonality for the crashes within a day (hour in the day)?
•Are the geospatial distribution of the crashes across different time of the day comparable?

Methods

Colab notebooks were used for this task. The notebook was attached to this poster for reproduction of any of the figures below. We employed data preprocessing, data visualization, time series analysis (TSA) and geospatial analysis in this project.

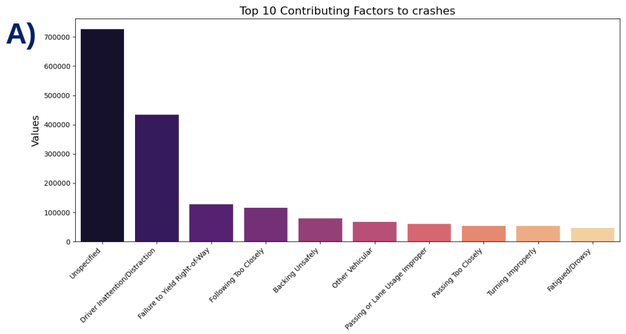
Results

One section of a poster should present the results. Often the results can be depicted with graphs, such as for an experiment, or with drawings such as with a design.

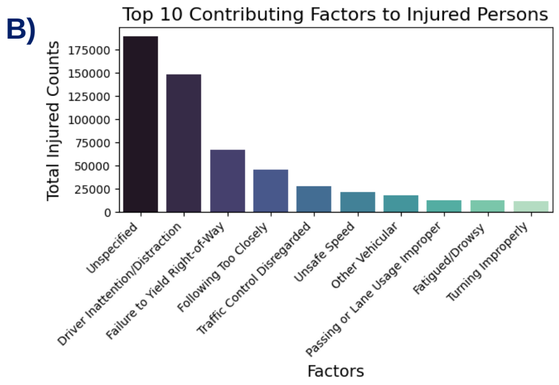
A. What are the major cause for fatal crashes?

We analyzed the top 10 contributing for total crashes, in total crashes reported (Figure 1A), and crashes in two severities: injury related crashes (Figure 1B) and fatal crashes (death reported) (Figure 1C).

A)



B)



C)

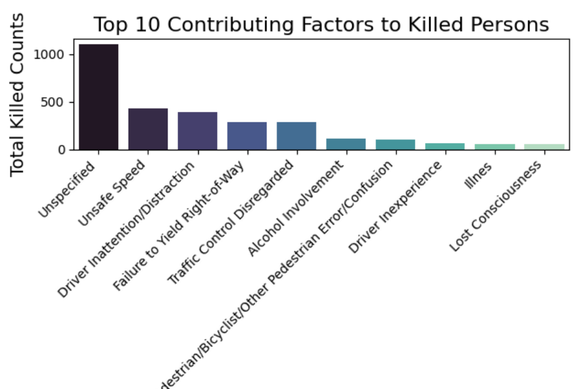


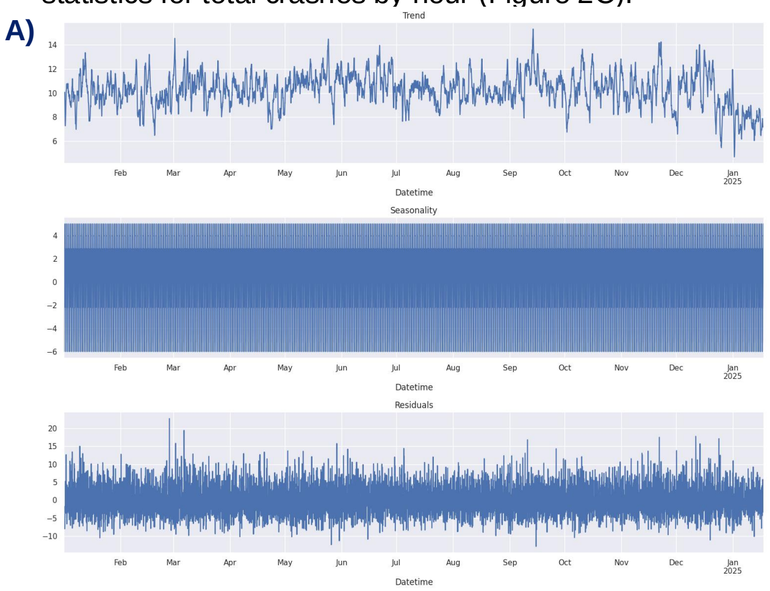
Figure 1. Top 10 contributing factors for total crashes, injury related crashes, and fatal crashes.

We found that the top three contributing factors for total and injury related crashes are the same, and in descending order they are : Driver Inattention, Failure to Yield Right-of-Way, and Following too closely. However, for crashes that resulted in death, there much less occurrence of this type of crashes, and the top three contributing factors are: Unsafe Speed, Driver Inattention, and Failure to Yield Right-of-Way.

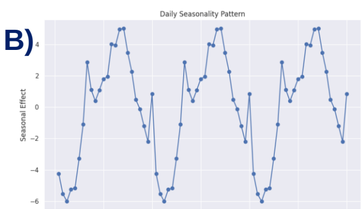
B. Is there any seasonality for the crashes within a day (hour in the day)?

Seasonality in months for the crashes have been detected. We would like to ask if there is any daily seasonality pattern for the crashes. Indeed, we found seasonality in daily pattern (Figure 2A). A closer look at the daily seasonality (Figure 2B) revealed that peak crashes happen near 4-5 pm. This seasonality resembles the statistics for total crashes by hour (Figure 2C).

A)



B)



C)

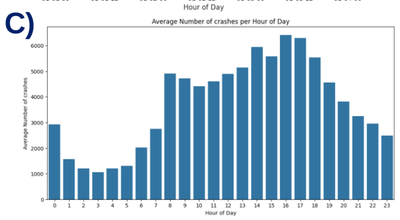


Figure 2. Time series analysis for daily pattern (24 hour period).

C. Are the geospatial distribution of the crashes across different time of the day comparable?

We found the geospatial distribution of the crashes across 4 daily timeframe (0-6 am, 6am -12pm, 12pm-6pm, 6pm-0 am) are highly comparable (Cosine Similarity Between Timeframes all > 0.99).

Conclusion:

Unsafe speed is a leading contributing factor for fatal crashes, with driver inattention and failure to yield right-of-way major contributing factors for all crashes. There is a daily peak of crashes happen around 16-17 pm. And the geospatial distribution of the crashes are not affected by time of accident in the day.

Acknowledgments

Part of the script was reused from 'EXPLORER / ACCELERATOR - Transportation Data Science Project (TDSP)' notebook. We appreciated this opportunity to learn and to be certified.