Group 19:

Khyathi Balusu

Pooja Dawada

Hannah Li

Rachel Meade

**Project Goals**

**Description**

Using a dataset of traffic collisions in the Los Angeles area between January 2010 and July 2019, we looked for patterns in when and where collisions occur, as well as the demographics of people involved in collisions.

**Importance of Problem**

Isolating the most dangerous traffic circumstances is the first step to making the streets of Los Angeles safer for drivers and pedestrians. The data indicates not only the location of each collision, but also the city council district and neighborhood council district for each incident. The inclusion of this information facilitates direct action from local government and municipalities responsible for the most dangerous areas. Los Angeles can use this information to better allocate resources to redesign or modify the most dangerous intersections in the city, and to better police dangerous areas that become dangerous and specific times.

Beyond helping to direct governmental action, this analysis would be helpful for drivers in the Los Angeles area. With the knowledge of when and where collisions occur frequently, they can better choose their routes, avoiding dangerous areas, intersections, or times if they so desire. Rideshare companies like Uber and Lyft, and taxi and car services could also make use of this analysis, programming their navigation systems to avoid high-frequency collision areas whenever possible. For companies which rely on driving to generate revenue, better information on what conditions increase likelihood of collision represents an opportunity to better protect their drivers, not only from dangers that might affect their wellbeing, but also from potential delays that could harm revenues or company reputation.

Although they make up a very small percentage of the dataset, we also have information on drivers who were involved in more than one collision. Along with the place and time collision information, insurance companies may be able to improve pricing policies for specific risk profiles.

**Exploratory Analysis**

We began the analysis process by looking first at where accidents occur. Of all of the location information included in the data, he broadest units of location divide the city into Council Districts. More accidents occur in district 12 than the other district, followed closely by districts 9 and 14. On a smaller scale, every collision is also assigned a certified neighborhood council, which represents a smaller local area. Each of these city Council Districts and Neighborhood Council districts is clearly defined, and maps of each are available on the city’s website (www.lacity.org). The data also provides zip codes, reporting district, and census tracts for each collision, but there the level of information is very similar to either Council District or Neighborhood District. We used Council and Neighborhood Districts in our analysis because we could find excellent maps of them, and because they seem to represent more consistent geographic areas than zip code, for example.

Apart from the Council Districts and Neighborhood districts, the dataset also divides the area of Los Angeles into 21 distinct areas. Both “Area Name” and “Area ID” referred to these areas; the first represents each area by name, and the latter with numerical areas numbers. Of these, the highest number of collisions occurred in the 77th Street area, followed by the Southwest, and Wilshire.

Although this broad information might be useful for city planners, it is unlikely that individual drivers will be able to avoid entire areas of the city at will. They may, however, have the power to avoid the most dangerous intersections. With this in mind, we found the intersections at which the highest numbers of collisions occurred. Some of these intersections had over 400 unique collisions – meaning that on average they had almost one collision every week over the time frame captured in the dataset.

Continuing the exploratory analysis, we examined the timing of accidents, including year, month, day of the week, and time. Overall, more collisions occurred in later years in the data set than earlier years. For example, more collisions occurred in 2017 and 2018 than in 2010 and 2011. The highest number of collisions occurs on Friday, followed by Thursday. The highest number of collisions occurred in the 5:00pm hour and was followed by the surrounding hours, between 3:00pm and 7:00pm. This is likely not a surprise to anyone who has been stuck in an afternoon rush hour before.

We also explored the demographics of people involved in collisions. Men are in more collisions than women. People between the ages of 20 and 30 are in the highest number of collisions, followed by people 30-40. People of Hispanic descent are in the highest number of collisions, followed by people identified as being of white descent. This does not mean, however, that people of Hispanic descent are in proportionally more accidents than people identified as white. According to the US Census Bureau, Hispanic or Latino people are 48.7% of the population of Los Angeles, while 28.4% are white, non-Hispanic. This suggests that people of different descent are in collisions at similar rates.

Within the age data, there were unexpectedly high numbers of people identified as 99 years old. Within the Los Angeles Police Department, it is standard procedure to use 9999 for certain values in a report when they are unknown or referred from an outside law enforcement agency (LAPD Online). Based on the disproportionately high occurrences of collisions involving people with 99 years of age, we believe that police officers or administrators may be occasionally applying the same guidance to age. If, when a person's age cannot be determined, police departments report age as 99, this could explain the relatively higher numbers of collisions that involve people aged 99 years.

We also found that of all the collisions over the eight-year period recorded, only 954 of them, or 0.2% involved the same DR Number. No DR Number was associated with more than 2 collisions.

**Solution and Insights**

The first goal was to uncover patterns in where collisions occur. Within each Council District, we found the Neighborhood Council responsible for the areas in which the highest number of collisions occur. For example, within council district 12, more collisions occur in the area which belongs to Neighborhood Council number 89 than any other Neighborhood Council district. Neighborhood 89 is the East Hollywood Neighborhood Council, which is led by president Arasele Torrez. This is precisely the kind of information that Los Angeles residents can use to demand better traffic management and policing from their heads of local government.

Another primary objective was to look for patterns in when collisions occur. The number of collisions appears to be increasing between 2010 to 2017. Between 2010 and 2018, there was a 26% increase in collisions, but during the same period, the city’s population only grew by an estimated 5% (US Census). However, 2018 may indicate a more hopeful outlook for the city, as it had a slightly lower number of collisions compared to 2017. With research beyond the data, we found that Los Angeles has invested heavily in reducing the number of collisions, especially those that involve pedestrians and cyclists. The main component of this investment is in a program called Vision Zero which began implementation in 2017, with phases 2 and 3 implemented in 2018. The city credits this program for a reduction in collisions within the main investment areas.

Further patterns in when collisions occur were uncovered in the relationship between area of the city and time. The places where collisions occur most frequently change depending on the time of day. For example, between 11:00pm and 2:59am, more accidents occur in Hollywood. In the early mornings (3:00am to 6:59am) and later evenings (7:00pm to 10:59pm), more accidents occur in the 77th Street Area, and from 9:00am to 12:59pm, more accidents occur in West LA.

To meet our final objective, we identified demographic patterns among people involved in traffic collisions. Although there are distinct trends in the ages of people who are involved in collisions overall, the times at which people of different age demographics are involved in collisions are not uniform. For example, people in their teens and 20's are in proportionally more collisions between 9:00pm and 4:00am than they are overall. It is the opposite for people in their 50's: more collisions that involve this age demographic occur between 5:00am and 5:00pm. They are involved in proportionally fewer collisions between 6:00pm and 4:00am. At a citywide level, the pattern of collisions among younger people late at night could indicate a relationship between the number of collisions and drinking and driving. The city should investigate this further, as it may be able to reduce the number of collisions by investing in a more robust police presence in the areas where people tend to go out late at night and drink.

In addition to providing utility to the City of Los Angeles, we believe that these insights could improve navigation systems and individual driving habits. If navigation applications (including those used by rideshare and car services) can better balance the quickest path from one place to another with information about particularly dangerous intersections, perhaps traffic can be reduced in the worst intersections, leading to a lower rate of collisions without re-designing them. This could create economic benefits for car service and rideshare companies by helping to protect the well-being of drivers and improve the accuracy of estimated arrival times. This could also be a good stop-gap solution to reduce collisions until the city has time to address the design of its most dangerous intersections.

**Sources**

www.census.gov/quickfacts/losangelescitycalifornia

www.lapdonline.org/lapd\_manual/volume\_5.htm

www.lacity.org

www.kaggle.com/cityofLA/los-angeles-traffic-collision-data