# **COGS 108 - Group 043 - Project Proposal**

# **Group Members**

* Devin Junanto / A15754079
* Gabriela Shirley / A16095578
* Natalia Sudjana / A16112711
* Shih Gau Peng / A16059733

# **Research Questions**

1. What would be the major contributing factor to traffic accidents?
2. Are there any correlations between visibility and the severity of the accident?
3. Does the presence of stop signs contribute well to the car crash preventions?

## **Background and Prior Work**

According to WHO’s infographics on road safety facts1, nearly 1.25 million people die in road crashes each year, on average 3,287 deaths a day. An additional 20-50 million are injured or disabled. These facts make road crashes the leading cause of death among young people aging from 15 to 29, and it costs $518 billion globally. Unless action is taken, road traffic injuries are predicted to become the fifth leading cause of death by 2030.

Factors that contribute to traffic accidents can be grouped into some categories, such as human errors, engine failures, or hazardous environments. In one of the articles in the Accident Analysis & Prevention journal2, “Modeling traffic accident occurrence and involvement”, they model the frequency of accident occurrence and involvement of some factors, such as horizontal curvature, lane, shoulder and median widths, urban/rural, and the section’s length. Several models were also developed to account for the demographic characteristics of the driver, like age and gender. This analysis concludes that heavy traffic volume, speeding, narrow lane/shoulder width, a larger number of lanes, urban roadway sections, and reduced median width will likely increase the risk of accidents. Additionally, the models also indicated that female drivers experience more accidents in heavy traffic volume, reduced median width, narrow lane width, larger number of lanes whereas male drivers tend to be involved in traffic accidents while speeding. Age-wise, young and elder drivers experience more accidents than middle-aged drivers.

Given that prior work has been done in comparing the factors of lane width, age, and gender to the likelihood of accidents, we would like to examine further in different factors and try to explain if there are some correlations among them. The factors that we would focus on are weather conditions, visibility, and the presence of stop signs or regulatory signs in general. We will be using a dataset we found on *Kaggle* that shows statistics of 3 million road accident records in the United States from 2016-20193. The dataset will cover multiple factors that might or might not significantly lead to accidents, the factors include time (start and end), location (street, city, state, country, etc), weather condition (temperature, wind chill, humidity, visibility, wind direction, wind speed, precipitation), presence of amenity, bump, crossing, junction, railway, roundabout, station, stop signs, traffic signal, and period of day (day or night). Upon further observations, we hope to come up with a conclusion of which variable has a greater influence on the accident occurrence that might help to prevent it from happening.

References (include links):

1. <https://www.who.int/violence_injury_prevention/road_safety_status/2013/facts/en/>
2. <https://www.sciencedirect.com/science/article/abs/pii/S0001457599000949>
3. <https://www.kaggle.com/sobhanmoosavi/us-accidents>

# **Hypothesis**

1. We hypothesize that the weather condition is most likely to cause an accident. Although people are more alert in bad weather conditions, we believe that there are unforeseeable variables that might cause accidents.
2. We hypothesize that the lower level of visibility the severity level of the accident will likely be higher since visibility plays a crucial role when it comes to driving.
3. We hypothesize that the presence of stop signs, or any regulatory signs in general, would more likely prevent car accidents as they reinforce traffic laws at all times. People who tend to break the traffic law would more likely have a higher chance of accidents.

# **Data**

We will be using a US traffic accident dataset which can be found in this link: <https://www.kaggle.com/sobhanmoosavi/us-accidents>.

This dataset covers 49 states of the United States with 3.0 million records and is collected from February 2016 to December 2019, using several data providers, including two APIs which provide streaming traffic event data. These APIs broadcast traffic events captured by a variety of entities, such as the US and state departments of transportation, law enforcement agencies, traffic cameras, and traffic sensors within the road-networks. Since our biggest question for this research is what would be the greatest factor of traffic accidents, we found the above dataset covers what we need. With filesize of 1.12 GB, this dataset observes 3.0 million records of countrywide traffic accidents (rows) with 49 variables (columns), including: severity, start time, end time, distance, street, temperature, visibility, wind speed, bump, junction, stop sign, period of day, and many more. We are certain that the analysis result from this dataset would be as accurate and reliable as possible as we are using big relevant data.

# **Ethics & Privacy**

This dataset is being distributed only for Research purposes, under Creative Commons Attribution-Noncommercial-ShareAlike license (CC BY-NC-SA 4.0). When downloading this data, we agreed to use this data only for non-commercial, research, or academic applications. Since the dataset doesn’t include personal information, the privacy of the people contributing to the dataset is secured. Furthermore, the dataset doesn’t exclude any specific group since it simply holds records of accidents. We included a citation of the dataset at the end under the Reference tab.

**Team Expectations**

* We agree to respect one another and have an open mind
* We agree to do our job as best as we can and continuously support each other
* If we are unable to meet, we would catch up and actively be in the discussions
* We will communicate regularly to not only be sure of what we are doing but also to check up on each other
* We agree to always stay involved and committed to doing the project.

# **Project Timeline Proposal**

|  |  |  |  |
| --- | --- | --- | --- |
| **Meeting Date** | **Meeting Time** | **Completed Before Meeting** | **Discuss at Meeting** |
| 1/22 | (online) | Read & Think about COGS 108 expectations; brainstorm topics/questions | Determine best form of communication: make group chat on Whatsapp, share google docs for the proposal |
| 1/24 | 5 PM | Brainstorm topics/questions | Discuss and decide on final project topics: demographic with income, factors that increase song popularity, factors that cause traffic accidents; discuss hypothesis; begin background research |
| 1/29 | 5 PM | Do background research on topic | Discuss ideal dataset(s) and ethics; draft project proposal |
| 1/31 | 8 PM | Edit, finalize, and submit proposal (Gabriela, Natalia, Shih Gau); Search for datasets (Devin) | Discuss Wrangling and possible analytical approaches; Assign group members to lead each specific part |
| 2/14 | 6 PM | Import & Wrangle Data (Gabriela; EDA (Natalia) | Review/Edit wrangling/EDA; Discuss Analysis Plan |
| 2/23 | 12 PM | Finalize wrangling/EDA; Begin Analysis (Shih Gau, Gabriela) | Discuss/edit Analysis; Complete project check-in |
| 3/13 | 12 PM | Complete analysis; Draft results/conclusion/discussion (Devin) | Discuss/edit full project |
| 3/19 | Before 11:59 PM | NA | Turn in Final Project & Group Project Surveys |

**References**

* Moosavi, Sobhan, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, and Rajiv Ramnath. “[A Countrywide Traffic Accident Dataset](https://arxiv.org/abs/1906.05409).”, 2019. <https://www.kaggle.com/sobhanmoosavi/us-accidents>
* Moosavi, Sobhan, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, Radu Teodorescu, and Rajiv Ramnath. ["Accident Risk Prediction based on Heterogeneous Sparse Data: New Dataset and Insights."](https://arxiv.org/abs/1909.09638) In proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM, 2019.
* “Infographics on Global Road Safety 2013.” World Health Organization, World Health Organization, 15 Oct. 2018, [www.who.int/violence\_injury\_prevention/road\_safety\_status/2013/facts/en/](http://www.who.int/violence_injury_prevention/road_safety_status/2013/facts/en/)
* Abdel-Aty, Mohamed A., and A.Essam Radwan. “Modeling Traffic Accident Occurrence and Involvement.” Accident Analysis & Prevention, Pergamon, 18 July 2000, www.sciencedirect.com/science/article/abs/pii/S0001457599000949