Analysis Of US Accidents Dataset

Phase 1

Group 18

Date: January 28, 2020

Members: Xiaohu(69363179)
Fangyuan Cai(24322202)
Xuyang Lin(74342826)
Qihao Wang(90739839)

Database Management Systems
Unversity of Florida

Contents

1 Introduction

Traffic accidents keep happening around us. According to *Road Accidents in the U.S. - Statistics & Facts*, there were some 6.3 million fatal, injury, and property damage crashes that occurred in the U.S. in 2015. However, the number of accidents had increased nearly 40 times which had been reached 264 million in 2018. Hence, how to reduce traffic accident's rate is an essential problem for human's daily life. According to the statistic number, we know that many factors would contribute to traffic accidents such as weather condition, road condition even our states. Then, we could analyse the relationship between different factors and the number of accidents. At the same time, we could come up with some solutions which might reduce the accidents number.

After searching from the internet, we found a data set which perfect fit our expect. The dataset which contains the traffic accidents number and some external condition information which is related to accidents will be introduced in the following segment.

With the support of motivation and dataset, we are willing to build a traffic accidents analysis system. We mainly look for the relationship between the accident and various external factors, through these relationships, on the one hand, we could we can provide a reference for how each factor affects the occurrence of the accidents, on the other hand, we could help more rational allocation of transportation resource.

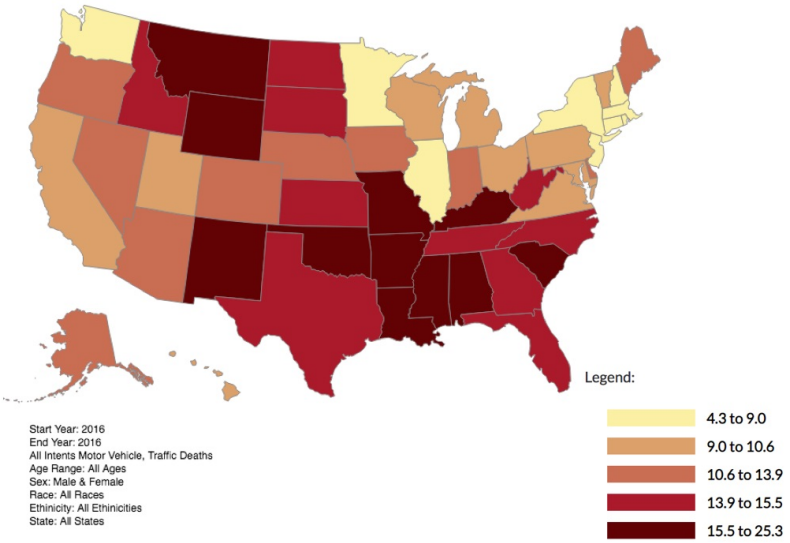

figure1:U.S. state accident rate distribution
resource: https://www.acsh.org/news/2018/08/10/most-dangerous-drivers-ranked-state-age-race-and-sex-13300

2 Dataset

The dataset is mainly about traffic accidents and relative information from February 2016 to December 2019 for the Contiguous United States.

Data source: https://www.kaggle.com/sobhanmoosavi/us-accidents

Since our main purpose is to find the relationship between number & severity of accidents and factors about time, location, weather, and traffic signs/facilities nearby, some columns are filtered as they are irrelative to our analysis (e.g., zip code of accident location).

Five types of columns are selected including severity, time, location, weather, and traffic signs / facilities nearby. Selected columns and some columns' explanation are listed below:

1. Severity
Shows the severity of the accident, a number between 1 and 4, where 1 indicates the least impact on traffic (i.e., short delay as a result of the accident) and 4 indicates a significant impact on traffic (i.e., long delay).

2. Time
We use start time (date and hour) of the accident in local time zone as time of the accident because duration is not relative to our analysis.

3. Location
Columns are Start_Lat (latitude of accident), Start_Lng (longtitude of accident), City, County, and State.

4. Weather
1) Sunrise_Sunset
Shows the period of day (i.e. day or night) based on sunrise/sunset.
2) Precipitation(in)
Shows precipitation amount in inches, if there is any.
3) Weather_Condition
Shows the weather condition (rain, snow, thunderstorm, fog, etc.).
Others
4) Weather_Timestamp, Temperature (F), Wind_Chill (F), Humidity (%), Air_Pressure (in inches), Wind_Direction, Visibility (in miles), Wind_Speed(mph)

5. Traffic Signs / Facilities
These columns indicates whether there is corresponding traffic signs / facilities in a nearby location.
These columns are Amenity, Bump (speed bump or hump), Crossing, Give_Way, Junction, No_Exit, Railway, Roundabout, .Station (bus station, train station, etc.), Stop, Traffic_Calming, Traffic_Signal, Turning_Loop

3 Software Solution

3.1 Functionalities

We first could write the SQL code which can query the data information.

Then we can write the different functions which can use those different SQL code to query the information we want. While in the front-end. We can call those function to let the data display on the screen.

There are two types of members in the system, User and Administrators. Administrators can manage the information of resource and users. Users can use the system to query data.

Users:
1.Users need to log into the system to view the information from the database. However, during the first time, user should register first.
2.Users could get information about the relationship between the number of car accident or severity and the external conditions.
3.Users could view the detail information of specific car accident.
4.Users can have an intuitive image of the data through visual charts.
5.Users can request to change the information in the database.
However, the users can not change the data individually. The only way is to ask the administrates to change the data .

Administrators:
1.Administrators could re-write the basic information of users, like the name and password.
2.Administrators could increase the number of data resources through upload resource files (csv file).
3.Administrators could change or delete the data information in the database.
4.Administrators could add new search function in the website.

3.2 Development Environment

We design a web to show the accidents information in the USA.
For database,we use oracle.
For web front-end language,we use html+css+javascript
For web backend language,we use php
For web server,we use apache

4 Queries

1. Which state has the largest/smallest number of traffic accidents in total?
2. What is the ranking of the number of accidents during certain hours in a day or certain months in the whole year for all states?
3. Which time period of a day has the largest number of accidents?
4. Which weather condition (a combination of wind direction, wind power, humidity, etc.) has the largest number of accidents?
5. Which traffic sign/facility appears at an accident the least (Which traffic sign/facility most probably reduces the possibility of accidents)?
6. What are the numbers of accidents of different severities in during certain hours in a day or certain months in the whole year?
7. What are the numbers of accidents of different severities in each state?

An example: from query 2, we can provide rankings of the number of accidents in different months, which could help relative departments and organizations allocate resources to prevent and handle accidents for different locations and time periods.