

U.S. Traffic Accident Mapping and Prevention based on Clustering and Factor Analysis

Team 034 : Lisa Li, Zixi Chen, Yiwei Sun, Junyu Chen, Shunxian Wang, Shuqi Wang

Background and Motivation

Problem:

- About 5.25 million accidents yearly in America, with more than 38,000 people die in crashes. We want to analyze the pattern of accidents thus minimizing harm.

Why it is important:

- Analyzing and visualizing road accidents can help us understand accidents occurrence pattern better, so that we can better tackle accidents by setting up speed cameras or intensifying police presence where risk is high, etc.

Approaches

Data Cleaning
and Exploratory
Analysis

Data Visualization & Interaction

Produced a interactive, three layer
webpage through D3 which includes
further data insights about time and
severity

K-Means Clustering
Finding Accidents
Hot-spots

Random Forest
Model to Predict
Severity Level

Comparing former approaches, we offer a website
with integrated traffic accidents information and
specific clustering center analysis to assist related
department make better decision.

Data

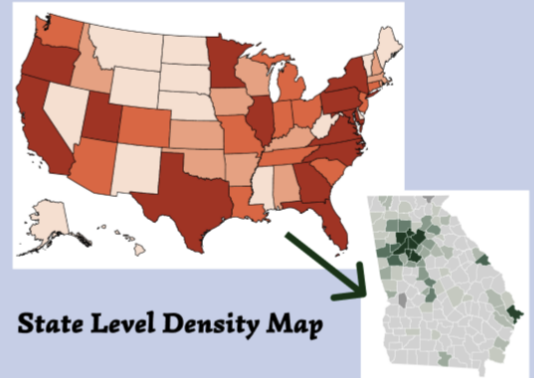
Data Resource:

- U.S. Traffic Accident dataset - Kaggle
- www.kaggle.com/sobhanmoosavi/us-accidents

Characteristic:

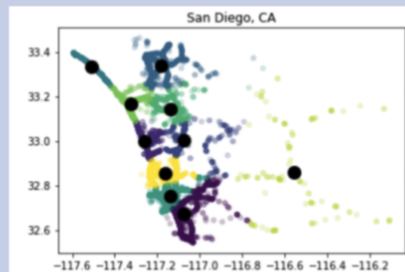
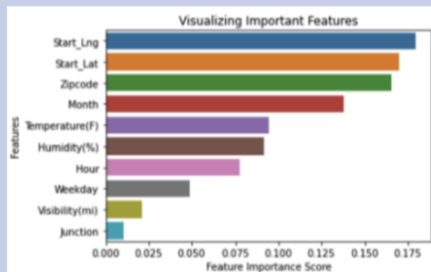
- To scale down the data, we only picked attributes that we were interested in, like location, weather, time, and severity.
- Cleaned data have 1,516,064 records (528MB).

U.S Level Density Map

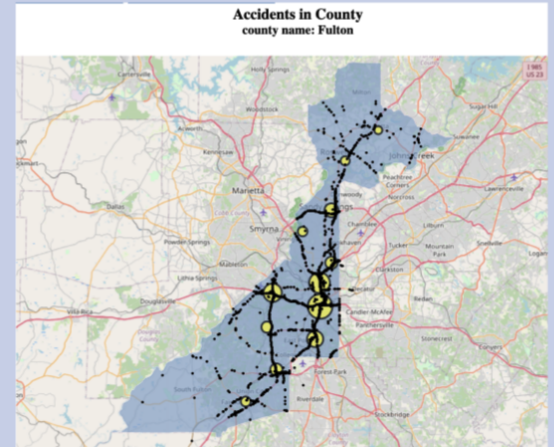


State Level Density Map

Experiment and Results



County Level Accidents and Cluster Center Map



Evaluation

We measured the success of our project by mapping our clusters onto a map and seeing how realistic they are. Also, we did severity analysis to what features are more important to predict severity. In the future, we can see if our recommendations actually reduced the number of incidents.

Results

We delivered a website that users can see from country to county level information, and use it to place new police patrol stations or install traffic signs in order to reduce the amount of accidents that take place.

Methods Comparing Others

Other methods are mainly research papers with images or a simple mapping of all traffic accidents.

Why Work?

Using our front-end website, users can easily identify problematic areas and see their correlation to the local roads. Users can also see the breakdown by hour and severity, providing a specific course of action.

What's New?

Previous methods did not do this on a county scale or integrate all levels together, and they did not investigate the correlation between accidents and time of day.

County Level Descriptive Analysis

