Safe Streets For Cyclists

Visualizing the risk index of streets based on Google Street View and historical bike collision data.

Team 64 :
Guangyue Cao,
Yue Huang,
Ting Li,
Kensen Zehao Tan,
Yifan Zhu,
Zhuohan Zhang

Georgia Tech

Motivation

The trending bike movement in the United States is advocated by a large population who is seeking a healthy and green lifestyle, especially in big cities like San Francisco. With its plan for net zero emissions in 2050, San Francisco is the second most bike friendly city in the United States [2, 20].

Meanwhile, as the bicycle becomes a more prevalent mode of transportation, **the rate of bicyclist injury has increased drastically by 21% from 2013 to 2017[20].** Even though there is political will and cycling initiatives across the US to improve policy environment, cyclists are still extremely vulnerable when compared to other road users.

To boost safety for cyclists and improve the civil infrastructure, designers need to know what features and characteristics of the biking environment directly relates to the accident rate.

Civil projects like this are expensive in Bay Area. To make matters worse, there is a lack of a sufficient tool that transportation professionals can use to support their actions. **There is an urgent need to identify and prioritize the area where changes would result the greatest improvement in safety.**

"Creating safe streets is a critical responsibility shared by all stakeholders."

Our Approach

Current practice Use historical data to construct a map with high injury rates [5,15,19]; Little visibility into root causes of accidents; Areas with more traffic are more likely to be sampled [3]; Lack of risk assessment in under-visited areas; Use crowdsourcing [7, 16, 18] is inefficient

Our solution, Safe Street for Cyclist, will provide the key contributors of accidents and identify where the stakeholders should focus their resources and guide them with budget decisions. While the platform provides infinite support to planners, it serves as educational pathway and provides rich insight into San Francisco for rest of the world.

Our goal is to utilize various machine learning models with historical bicycle collision.

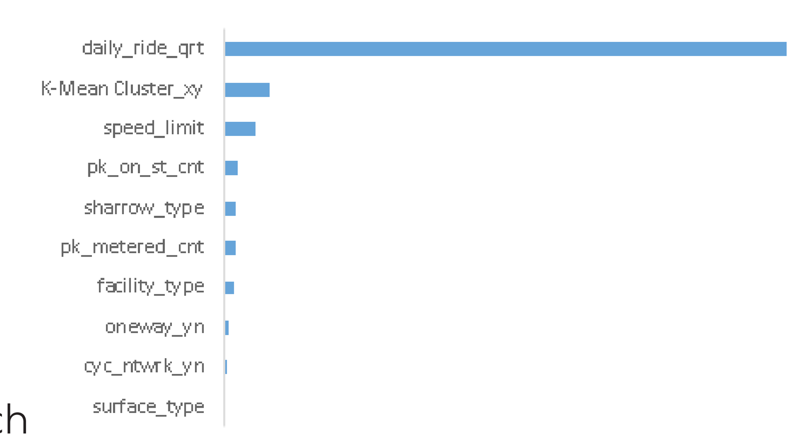
1. identify key risk factors associated with dangerous streets. [5,6,18]
2. Develop and assign a composite risk score [6,18] to streets in San Francisco, California.
3. Provide more effective insights for bike-friendly street improvements.

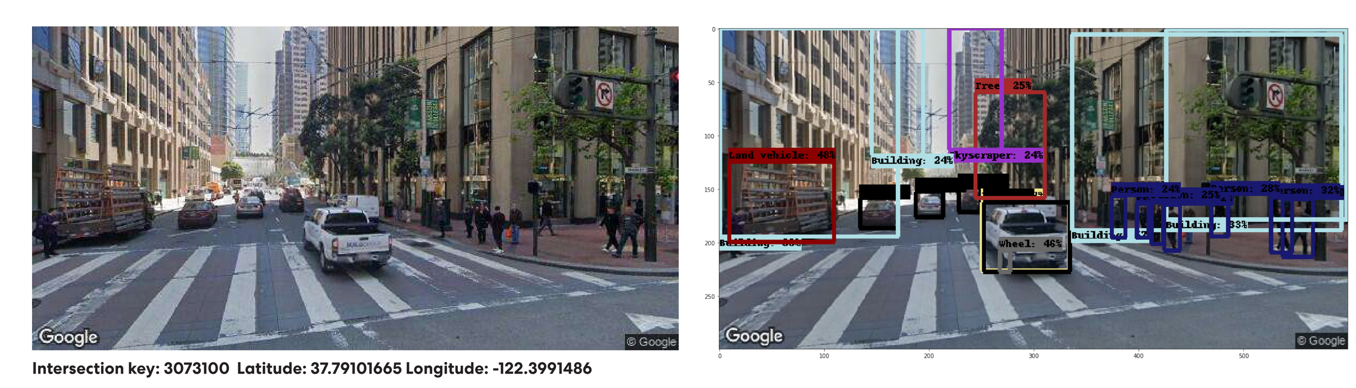
Data Gathering

The data collection process begins with literature research and compiling a list of relevant datasets. The original datasets are then transformed to suit our needs. The bike collision datasets are from two pubic data resources: Transportation Injury Mapping System (TIMS) and TransBase. We gathered bike collision data of San Francisco from 2008 to 2012. We aggregated this collision dataset to street segment level, and then joined it to street infrastructure dataset.

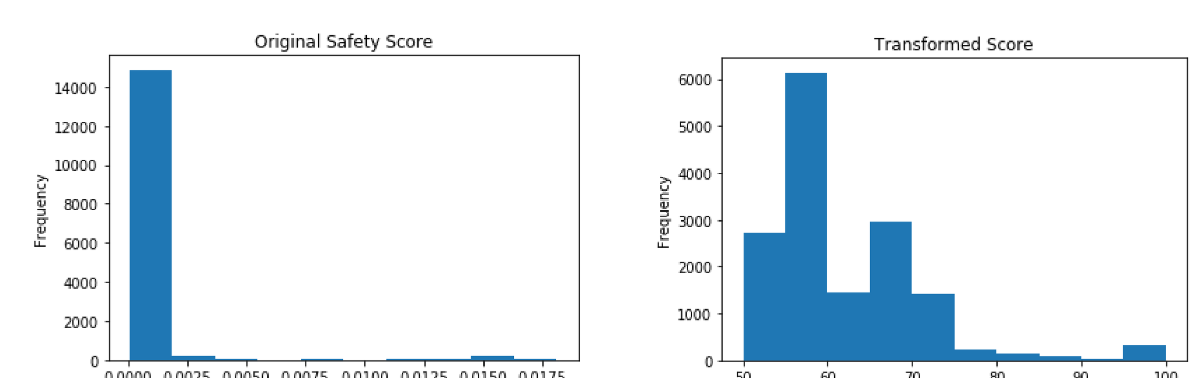
Since both datasets are public, they contain some redundant information and have large amount of missing data. The cleaning process is done with Python Pandas, NumPy and SQLite3 to minimize the chance of human error on our end. After a thorough review of the original datasets, we removed columns that are not meaningful to street demographics. We used KNN for imputation to replace the missing values with substitutes from KNN.

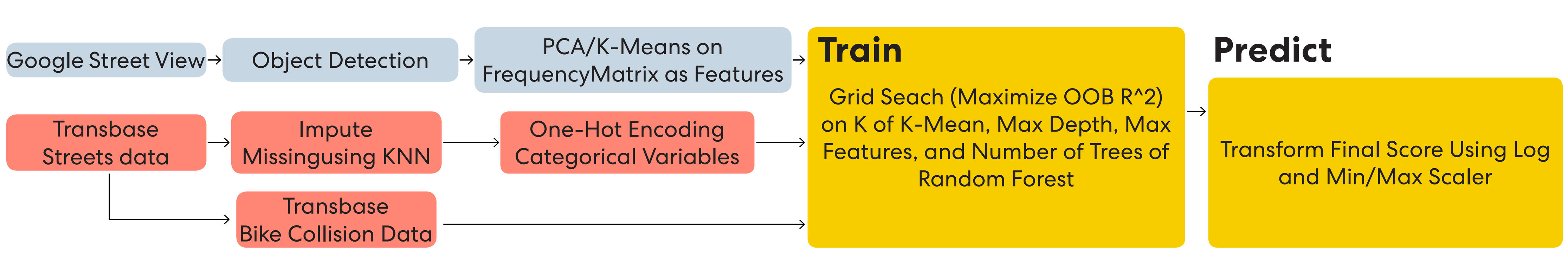
To get image data for object detection later, we extract a list of latitude and longitude pairs for each intersection in San Francisco and a list of coordinates for each collision. Using the Google Street Views API, we collected over 9,147 images from latitude-longitude coordinates for intersections and over 3,852 images for collision locations.





The image above is an example of objective detective model.





Computation

Tensorflow object detection model trained on Google's Open Image dataset to identify possibly relevant objects and street characteristics on all the Google Street Views images collected.

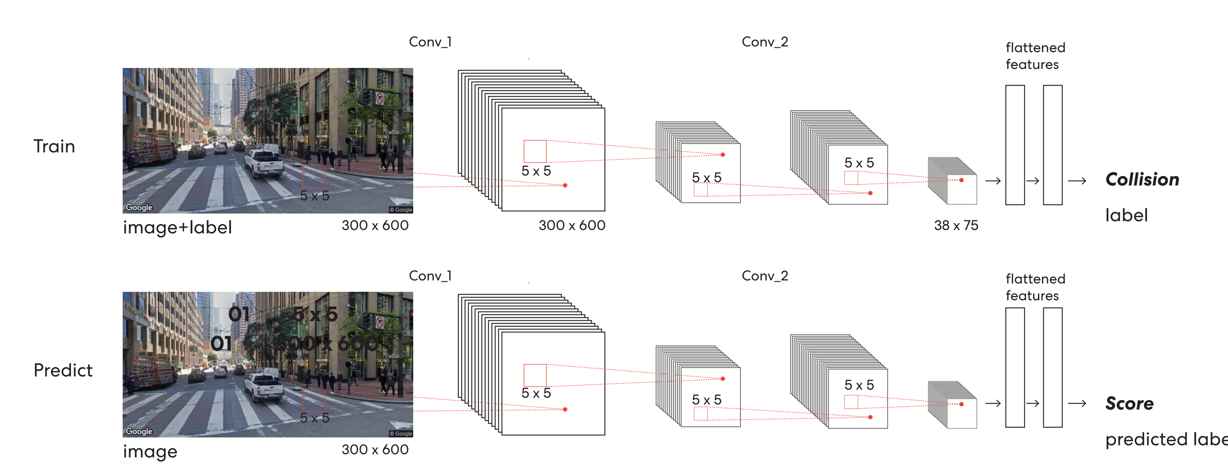
- For each image, we detect 20 objects, ranked by number of occurrence and probability score.
- We were able to cluster intersections using K-means into custom groups on those street segments.
- Combine Clusters into other existing street features in the TransBase dataset.

Then developed a baseline Random Forest model using the features collected to identify street features that are highly correlated with a street segment being more dangerous than others.

- Random Forest is selected because some of the independent variables were highly skewed
- The resulted Random Forest is able to achieve a 65% out-of-bag R-squared on our safety metrics, and the most significant feature is Estimated number of daily ridership per quarter mile from street segment. Going forward, we will explore more feature transformation and selection methods and compare to this baseline model.

Using Convolutional neural networks(CNN) to identify the image similarity or dissimilarity.

Powerful because of its ability to comprehensively capture visible features when analyzing image data.



Do it in an automated process for feature extraction by learning the visual feature representation using CNN.

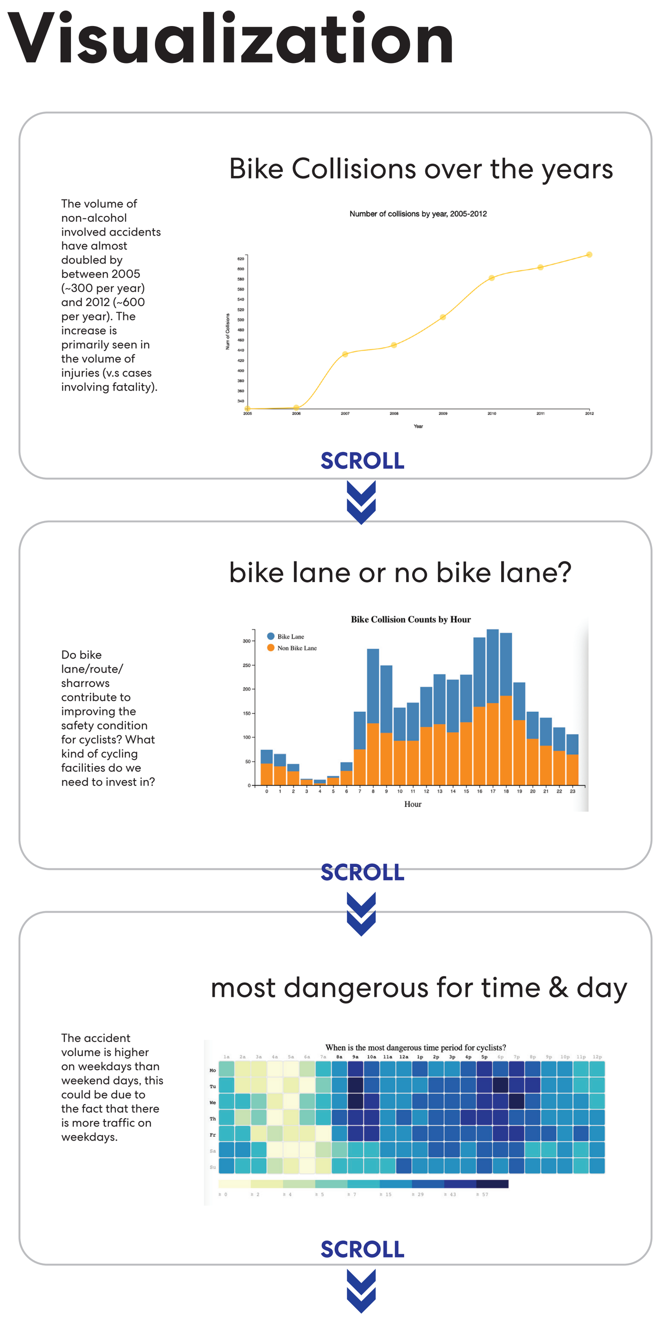
We apply several layers of filters to transform the image data from pixel representations to the complimentary feature dataset that we need.

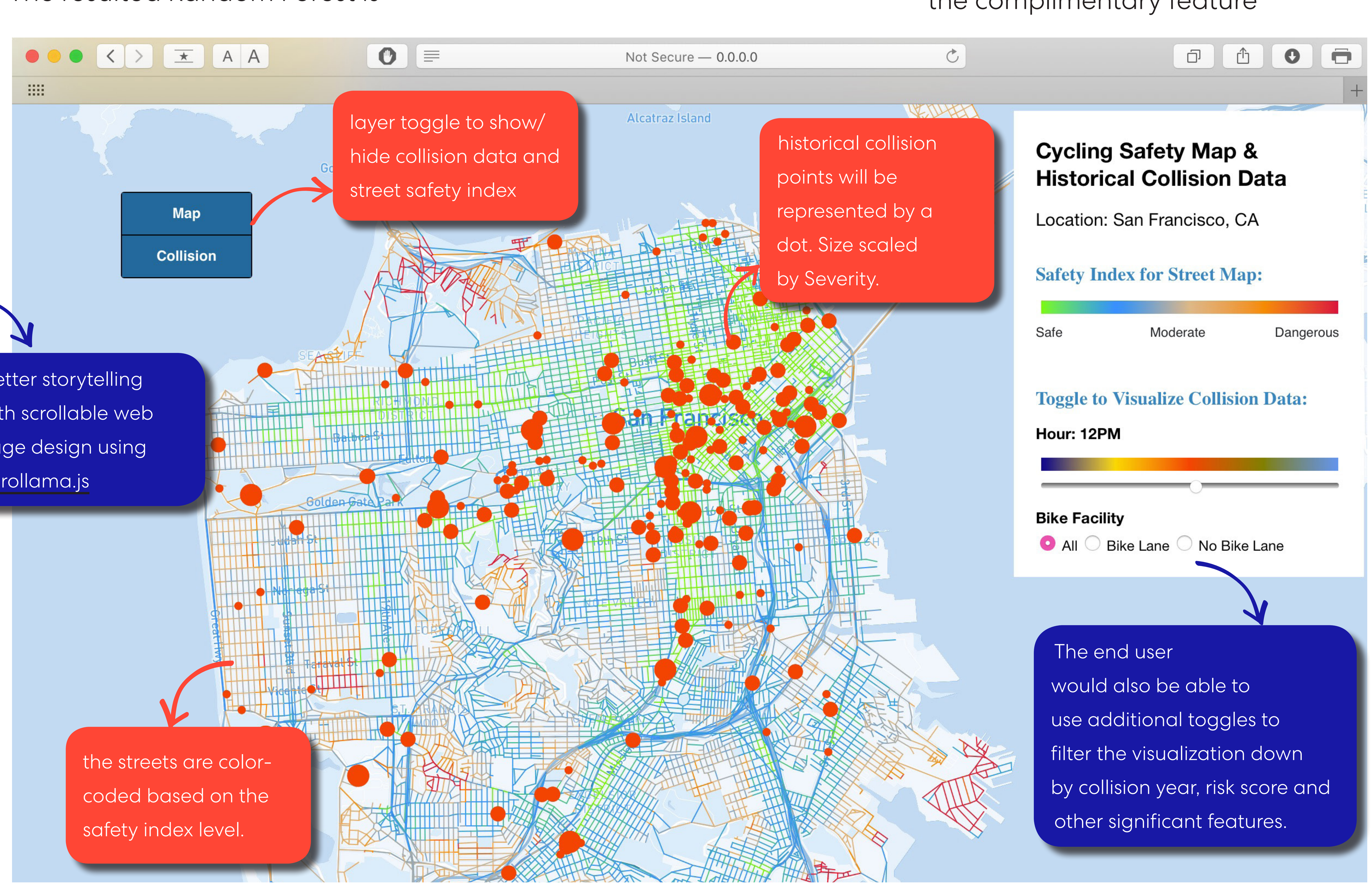
We aim to use these complimentary features in conjunction with those from the TransBase dataset to make prediction.

The model needs a lot of training data with corresponding labels to give feedback to the network to optimize effective feature representation.

We will try to use CNN as a learned representation of visual features which we can use to interactively explore the TransBase data.

Visualization





Evaluation/Validation

Random Forest: 65% out of bag R-square.

CNN:

$$Accuracy = \frac{1}{N}\sum_{i=1}^{N} 1(\hat{y}_i = y_i)$$

$$Macro-F1 = \frac{1}{C}\sum_{c=1}^{C} F1_c = \frac{1}{C}\sum_{c=1}^{C} \frac{2 \times Precision_c \times Recall_c}{Precision_c + Recall_c}$$

We sent the survey to 20 volunteers that are cyclists and urban designers. The response is positive with average score 8 that they would recommend to their friends/colleagues.