CS591 L1 Data Mechanics

Reducing the number of car accidents

By Ivan Uvarov and Stephanie Chiao

Introduction

Our goal in this project was to use the City of Boston's data to find a relationship between car accidents and other factors. Our first step was to find areas where car accidents are most common. Then we found a correlation between street lights, traffic signals, MBTA stops, and car accident locations. We can use this data to find ways to decrease the number of car accidents throughout Boston.

Data Sources

We used car accidents database from Massachusetts Department of Transportation. It contained information about all registered car accidents in 2013: their location, type of crash, weather information etc. Dataset is 3 years old, but it shouldn't affect the results because locations of street lights and traffic signals don't change very often.

We also used 3 datasets from Boston Open Data website: traffic signals, MBTA stops, and transportation districts.

Lastly, we used streetlight locations dataset from the City of Boston website.

Implementation

We implemented a wide range of algorithms to analyze the data we have, but not all of them gave us useful results. For example, we used k-means clustering algorithm to find clusters of crimes, but we didn't use this data. We also calculated p-value for some correlations, but it didn't give us any meaningful result.

Results

Data By Districts

We grouped all of our data by transportation districts (there are 6 of them in Boston) to get an overview of all information available to us:

| District	Car Accidents	Traffic Signals	MBTA stops	Street Lights
1	450	107	274	10203
2	371	120	609	11322
3	445	180	347	14314
4	343	79	268	8167
5	984	84	260	10897
6	1107	261	215	18147

Analyzing Accidents

Then we performed an analysis of car accidents. For every accident we calculated how many streetlights were near it, how many traffic signals and how many MBTA stops. The results can be seen in graphs on the right.

As can be seen, there's a strong correlation between the number of streetlights and car accidents and between the number of traffic signals and car accidents.

MBTA Stops

We ran various algorithms to analyze locations of MBTA stops and find a correlation with locations of car accidents. We didn't get strong evidence to support our assumptions.
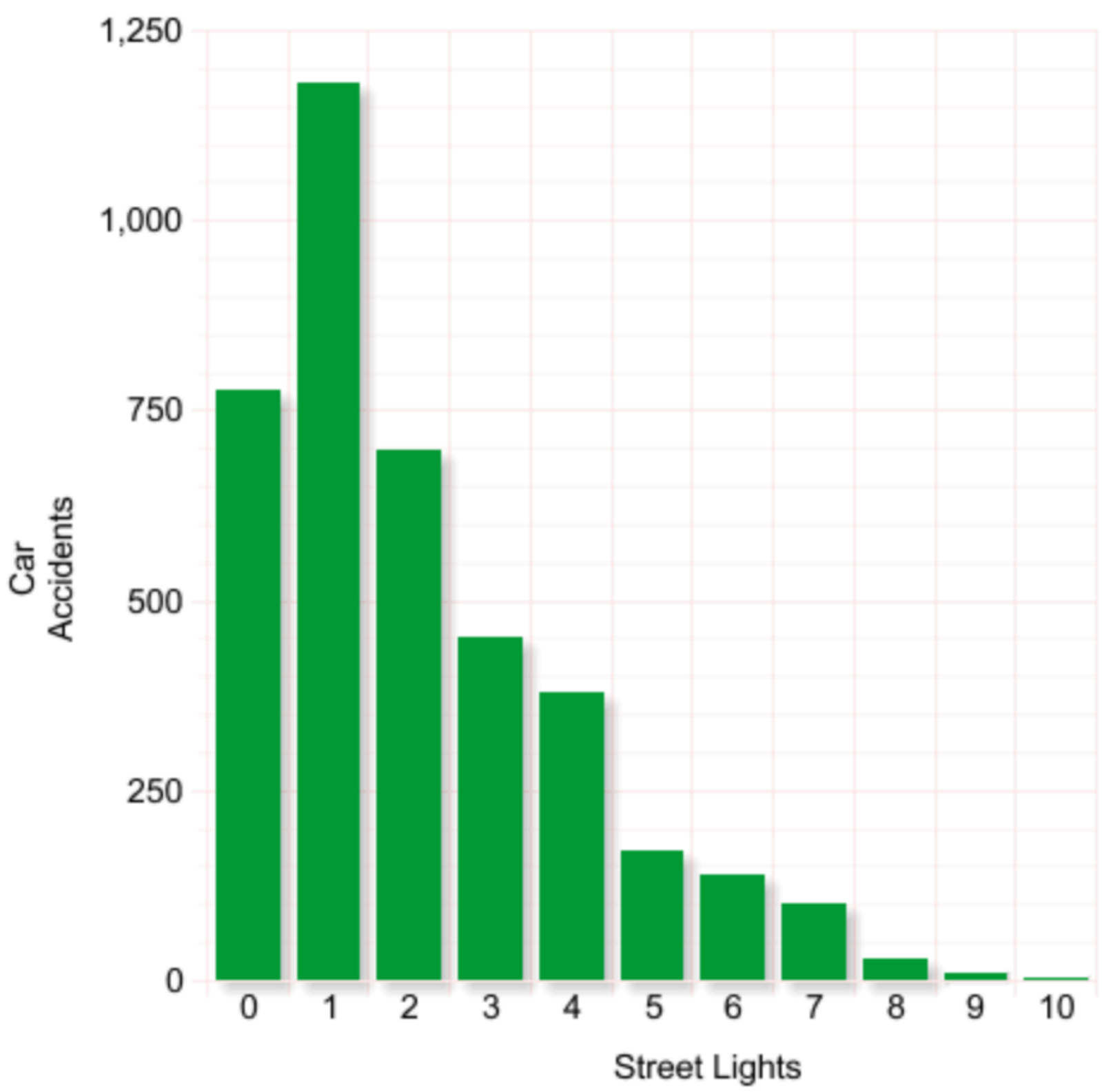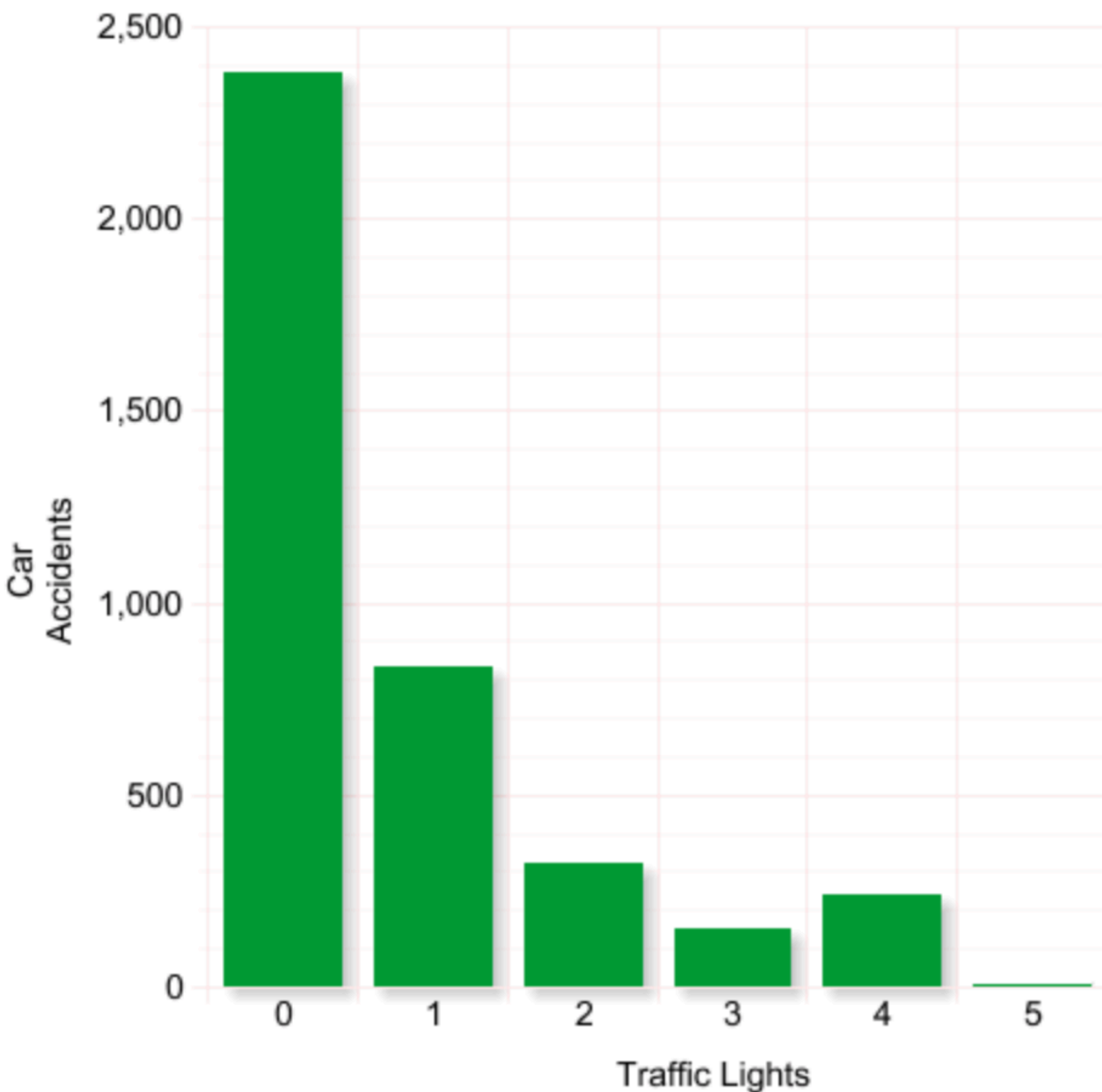
Future Work

If we were to continue to work on this project, we would find more relevant data that can influence the number of car accidents. Additionally, we would analyze time and weather conditions for past car accidents.