Vision Zero D.C. Analytics

Adam Peir | Justin Schulberg | Maynard Miranda | Ryan Doogan

Motivation and Introduction

What is the problem?

Washington D.C. has launched its Vision Zero Initiative which aims to reduce vehicle related crashes to zero by 2024. The current Vision Zero Website is sufficient for its stated purpose, but there is substantial opportunity to improve upon it. The dataset provided is large, robust, and ripe for application of analytical techniques, especially geographical clustering, time series analysis, and probability-based risk analysis.

Why is it important and why we should care?

This Vision Zero Initiative is important because every year, car accidents are among the top causes of death in Washington D.C. Our goal is to create an interactive and robust visualization dashboard and machine learning models to communicate the true impact of traffic-related crashes in Washington, D.C. to stakeholders who would be most interested in D.C.'s Vision Zero plan.

Our Data

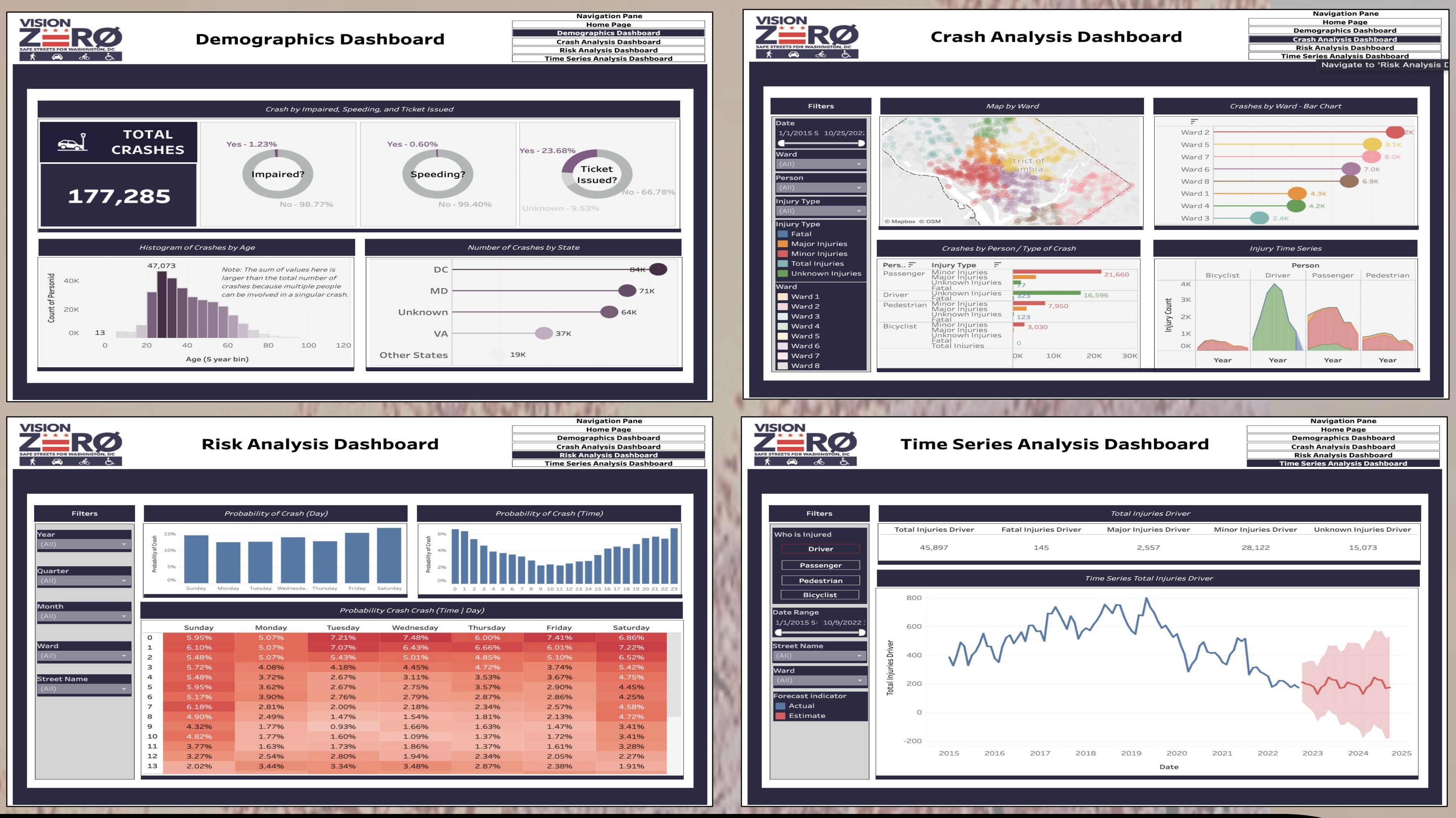
How was the data obtained?

The team programmatically pulled 270,000+ records of car crashes in D.C. using the OpenDataDC API. Data cleaning was performed using the pandas and numpy libraries. The data was pivoted into a more effective visualization structure, strings were reformatted, dates corrected into the correct date-time format, and duplicates were removed.

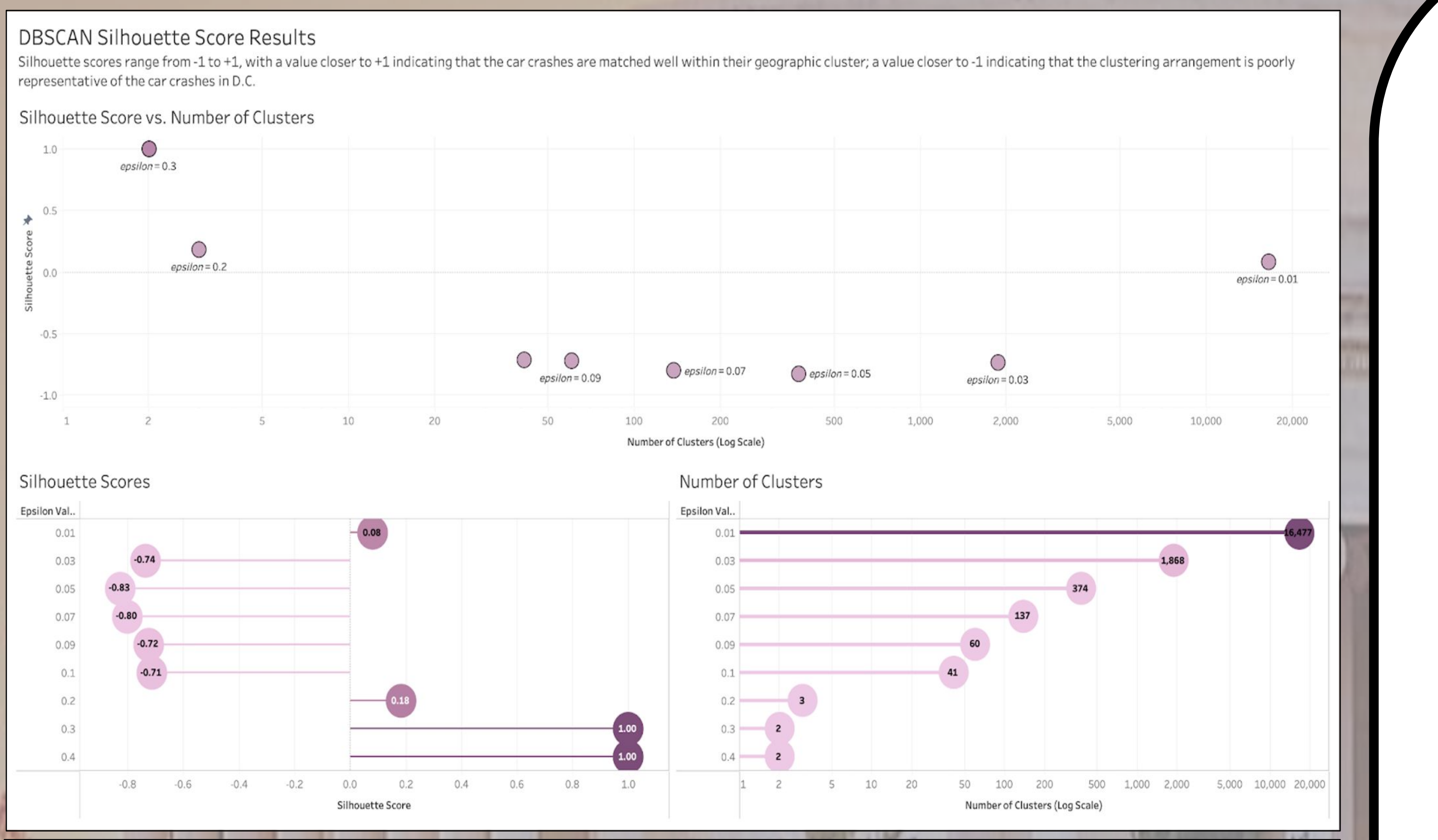
Data Description

- **Crashes_in_DC.csv** (270,000+ rows, 110MB): Represents the crash locations along the D.C. roadway blocks network.
- **Crash_Details_Table.csv** (720,000+ rows, 66MB): Contains demographic details for each crash.

Vision Zero D.C. Analytics Dashboard

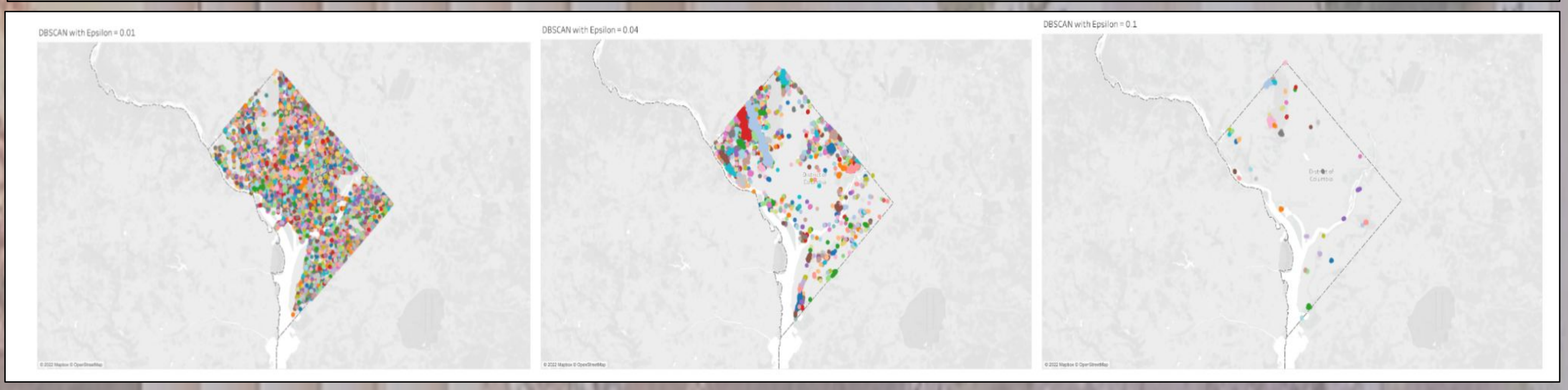


Density Based Spatial Clustering (DBSCAN)


DBSCAN Silhouette Score Results

Silhouette scores range from -1 to +1, with a value closer to +1 indicating that the car crashes are matched well within their geographic cluster; a value closer to -1 indicating that the clustering arrangement is poorly representative of the car crashes in D.C.

Epsilon Val..	Silhouette Scores	Number of Clusters
0.01	0.08	16,477
0.03	-0.74	1,868
0.05	-0.83	374
0.07	-0.80	137
0.09	-0.72	60
0.1	-0.71	41
0.2	0.18	3
0.3	1.00	2
0.4	1.00	2


DBSCAN with Epsilon = 0.01
DBSCAN with Epsilon = 0.04
DBSCAN with Epsilon = 0.1

Our Approach

The team created its own Vision Zero Analytics website that contains several dashboards and machine learning blogs.

Dashboard

- **Demographics Dashboard** | Breaks down crashes by rider impairment, whether the driver was speeding and/or issued a ticket, age, and state.
- **Crash Analysis Dashboard** | Visualizes crashes by ward, person, and type of injury sustained. The dashboard can be filtered by date, ward, person, injury type, and is fully interactive. Due to data limitations, this dashboard contains data from 2015 onwards only.
- **Risk Analysis Dashboard** | Uses Bayesian Statistics to determine the probability of car crashes given time, the likelihood of car crashes given day, and the possibility of car crashes per time given day. The dashboard can be filtered by year, quarter, month, ward, and street name. Due to data limitations, the dashboard only contains data from 2021 onwards only.
- **Time Series Analysis Dashboard** | Shows the actual and forecasted number of fatal, major, minor, and unknown injuries over time. Moreover, the dashboard can be filtered by month, year, and on who was injured - driver, pedestrian, bicyclist, and passenger. It can be filtered by who is injured, street name, and ward. Due to data limitations, the dashboard contains data from 2015 onwards only.

Machine Learning Blogs

- **K-Means Clustering Algorithm** | In this approach, we pre-set the number of clusters we would expect, and see how cleanly the car crashes separate into respective clusters. The rationale here is that DC has 8 pre-defined wards, so we can see if a 'K' value of 8 provides optimal separability in the data, or if a different value of 'K' does. We measure separability of the clusters by looking at an elbow plot of the sum of squared differences between clusters.
- **Density Based Spatial Clustering (DBSCAN)** | In this approach, we use DBSCAN, a density-based clustering approach. In this approach, we pre-set the maximum distance that points can be set apart in order to be clustered together. The benefit is that we do *not* need to pre-set the number of clusters we would expect. The rationale here is that it can be difficult to pre-determine the number of clusters needed like in K-Means Clustering; instead, looking at the density of points on a map could be a useful approach.

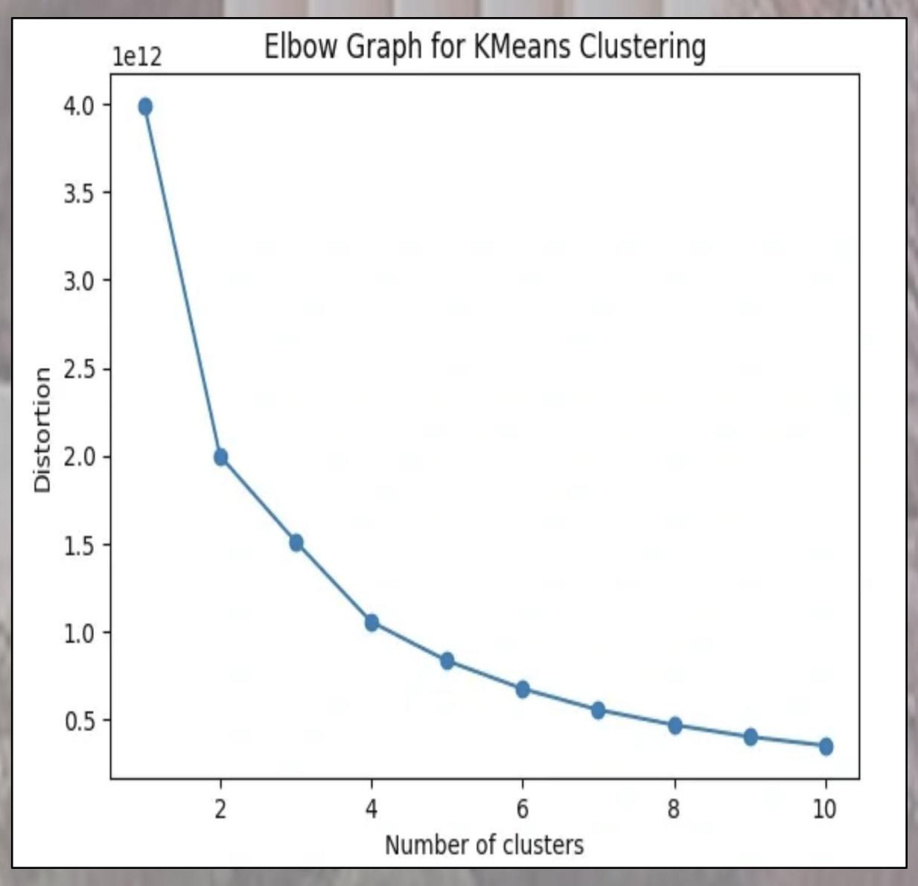
Why does our approach solve the problem?

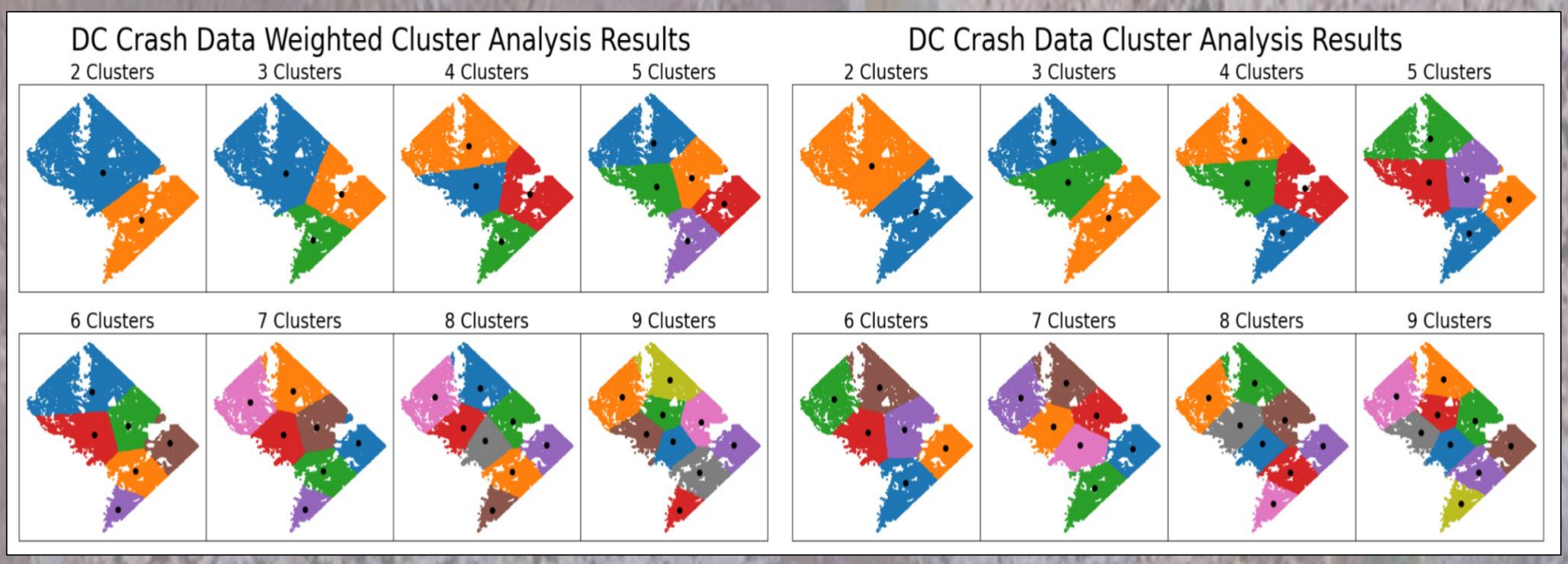
The D.C. Government's approach to the car crash problem is a considerable start to understanding the problem; however, with us providing more robust visualization offerings and making it interactive, accommodating different needs and stakeholders, plus providing relevant statistical and machine learning models, we firmly believe that our approach provides for a deeper understanding of the problem and would allow for a greater decrease in the number of car crashes in the city.

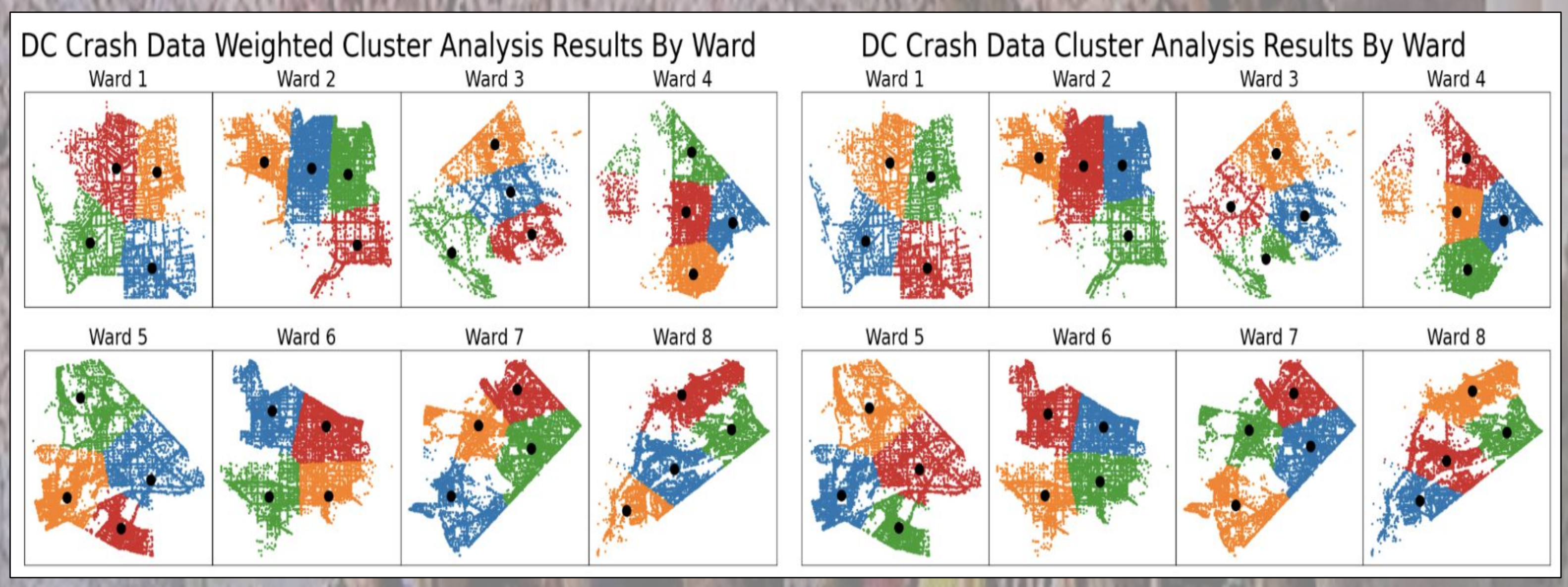
What is new in our approaches?

We provide interactive visualizations that allows users to analyze specific locations, dates, individuals injured, injury type, etc. We also provide statistical and machine learning models such as Time Series Analysis, Density Based Spatial Clustering (DBSCAN) and K-means Algorithm.

K-Means Clustering




DC Crash Data Weighted Cluster Analysis Results
DC Crash Data Cluster Analysis Results
2 Clusters
3 Clusters
4 Clusters
5 Clusters
6 Clusters
7 Clusters
8 Clusters
9 Clusters


DC Crash Data Weighted Cluster Analysis Results By Ward
DC Crash Data Cluster Analysis Results By Ward
Ward 1
Ward 2
Ward 3
Ward 4
Ward 5
Ward 6
Ward 7
Ward 8

Experiment and Result

How did we evaluate our approach?

Improving a city's predisposition to having a high car crash frequency is challenging and cannot be fixed overnight. The team has continuously evaluated its approach and provided new visualizations and machine learning analysis, as we have updated our priors. Our contact information is also provided on the website so that we can easily be reached for any questions, requests, or suggestions. The current approach of the city government has led to a significant decrease in the number of car crashes from 2015 up to the present (2022). We are confident that with the addition of our approach, interactive visualizations, and machine learning models, D.C. car crashes will further decrease, making the Vision Zero Initiative goal of 0 traffic-related fatalities achievable by 2024.

Dashboard and DBSCAN Results

This study evaluated the D.C. car crash dataset, and used interactive dashboards, machine learning models, and an easy-to-use website to help the D.C. government achieve its Vision Zero initiative (Website Link). Users are able to utilize the resources the team has created directly from the website to analyze common demographics and risk factors associated with increased risks of car crashes. From our analysis, we determined that some of these primary risk factors include, but are not limited to: those with ticket violation(s) are 13.45% more likely to get involved in a car crash; Saturday evening is the most dangerous time to drive; people aged 25-30 are more predisposed to being involved in a crash; drivers from Maryland contributed to more crashes than drivers from any other state (including Washington, D.C.); and crashes are most prevalent in Wards 2, 5, and 7. Lastly, despite the team's best efforts to use unsupervised clustering algorithms to look for geographic trends in the data, the data is unfortunately too dense to get meaningful results.

K-Means Results

Our results indicate that the K-Means algorithm, using both weighted and unweighted data, provides little value to our analysis of crash data. Optimal separability occurs at about 4 clusters, with limited improvement in performance beyond that, meaning that separability is only slightly better for 8 clusters (corresponding to the number of wards in D.C.) than for 4. The crash data is so dense relative to the size of each ward, that the results are rarely meaningful outside of specific subgroups exhibiting sufficient sparsity for meaningful patterns to emerge.

Weighting by the number of crashes occurring at a location does meaningfully change the cluster centers in situations where a disproportionate share of crashes in a region occur in a specific area. In Ward 3 for example, in the far northwest of the city, a heavy concentration of crashes occur along key road arteries. Weighting significantly changes the location of cluster centers, denoted by black dots.

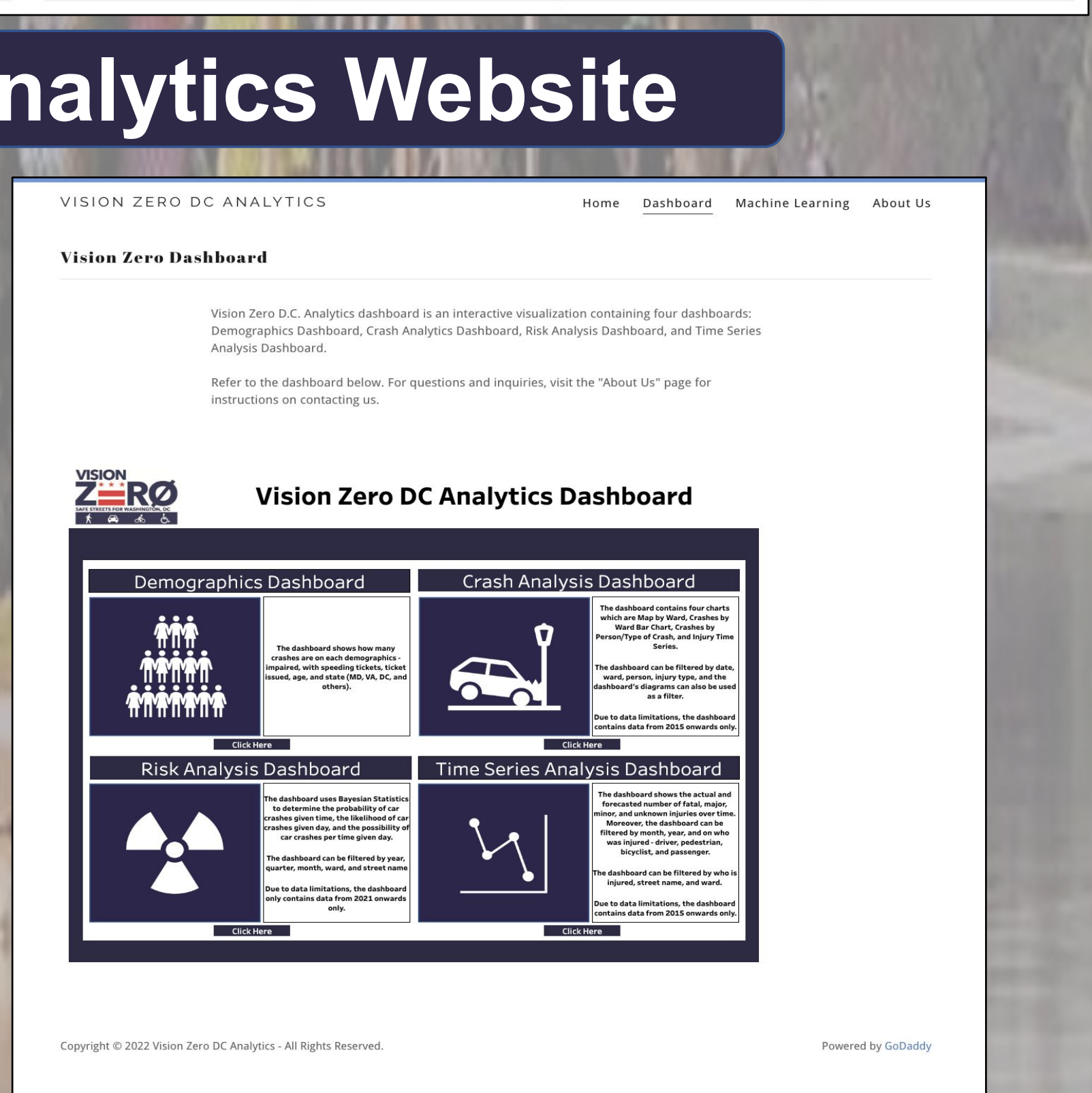
How does our method compare to others?

Our website is comparable to the official Vision Zero Website. The official website provides generally static and disconnected visualization tools to understand the context of the problem. In contrast, our website provides interactive visualizations which can accommodate different needs and stakeholders. Lastly, we provide statistical and machine learning algorithms to provide more useful insights to the stakeholders.

Vision Zero D.C. Analytics Website

Our Mission

Every year, car accidents are always among the top causes of death in Washington, D.C. **Our goal is to create an interactive and robust visualization dashboard to communicate the true impact of traffic-related crashes in Washington, D.C. to stakeholders who would be most interested in D.C.'s Vision Zero plan.**



Vision Zero Dashboard

Vision Zero D.C. Analytics dashboard is an interactive visualization containing four dashboards: Demographics Dashboard, Crash Analytics Dashboard, Risk Analysis Dashboard, and Time Series Analysis Dashboard.

Refer to the dashboard below. For questions and inquiries, visit the "About Us" page for instructions on contacting us.

Vision Zero DC Analytics Dashboard

Demographics Dashboard

The dashboard shows how many crashes are on each demographics - impaired, with speeding tickets, ticket issued, age, and state (MD, VA, DC, and others).

Crash Analysis Dashboard

The dashboard contains four charts which are Map by Ward, Crashes by Ward Bar Chart, Crashes by Person/Type of Crash, and Injury Time Series.

The dashboard can be filtered by date, ward, person, injury type, and the dashboard's diagrams can also be used as a filter.

Due to data limitations, the dashboard contains data from 2015 onwards only.

Risk Analysis Dashboard

The dashboard uses Bayesian Statistics to determine the probability of car crashes given time, the likelihood of car crashes given day, and the possibility of car crashes per time given day.

The dashboard can be filtered by year, quarter, month, ward, and street name

Due to data limitations, the dashboard only contains data from 2021 onwards only.

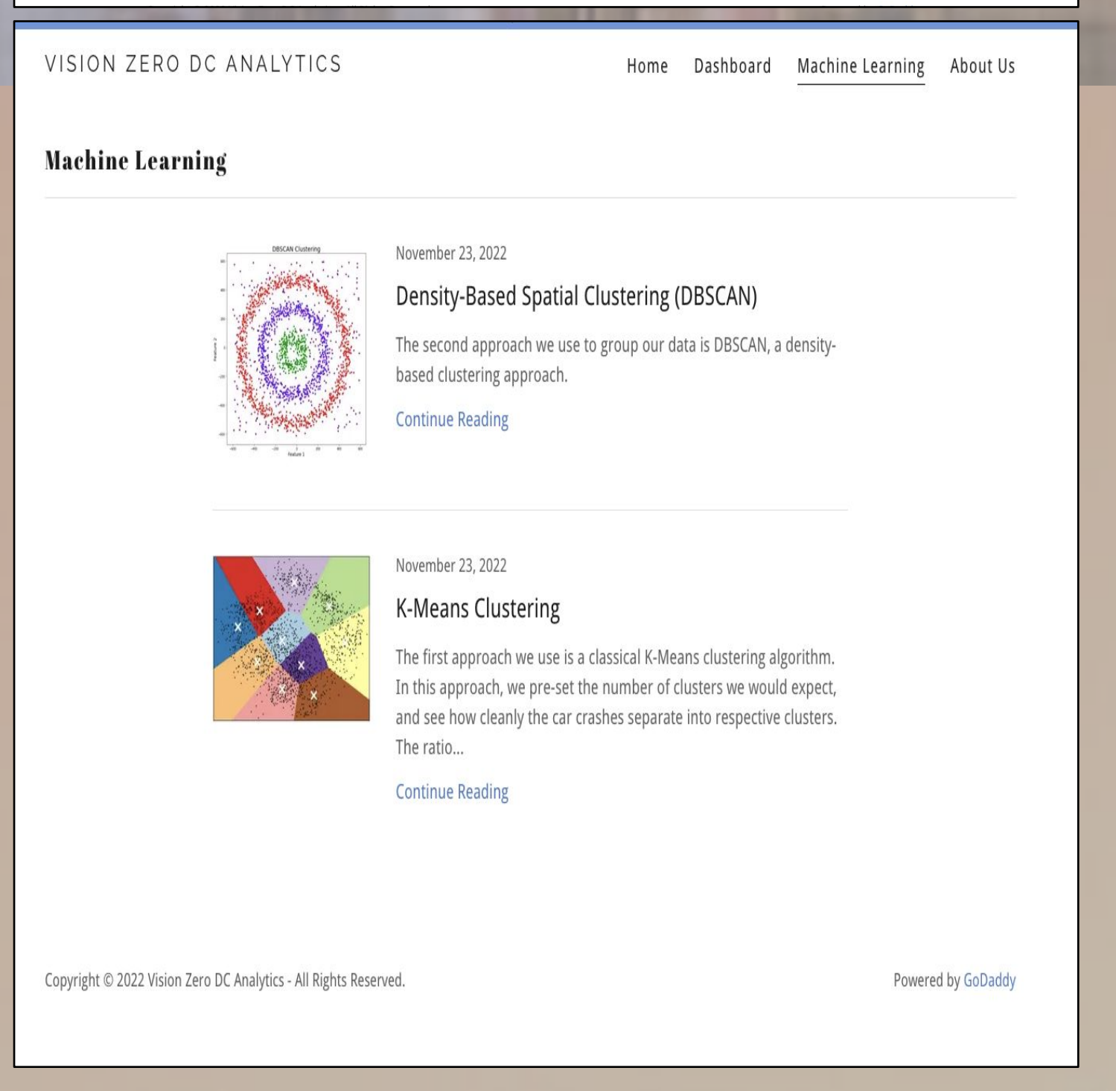
Time Series Analysis Dashboard

The dashboard shows the actual and forecasted number of fatal, major, minor, and unknown injuries over time. Moreover, the dashboard can be filtered by month, year, and on who was injured - driver, pedestrian, bicyclist, and passenger.

The dashboard can be filtered by who is injured, street name, and ward.

Due to data limitations, the dashboard contains data from 2015 onwards only.

Copyright © 2022 Vision Zero DC Analytics - All Rights Reserved. Powered by GoDaddy



Machine Learning

November 23, 2022

Density-Based Spatial Clustering (DBSCAN)

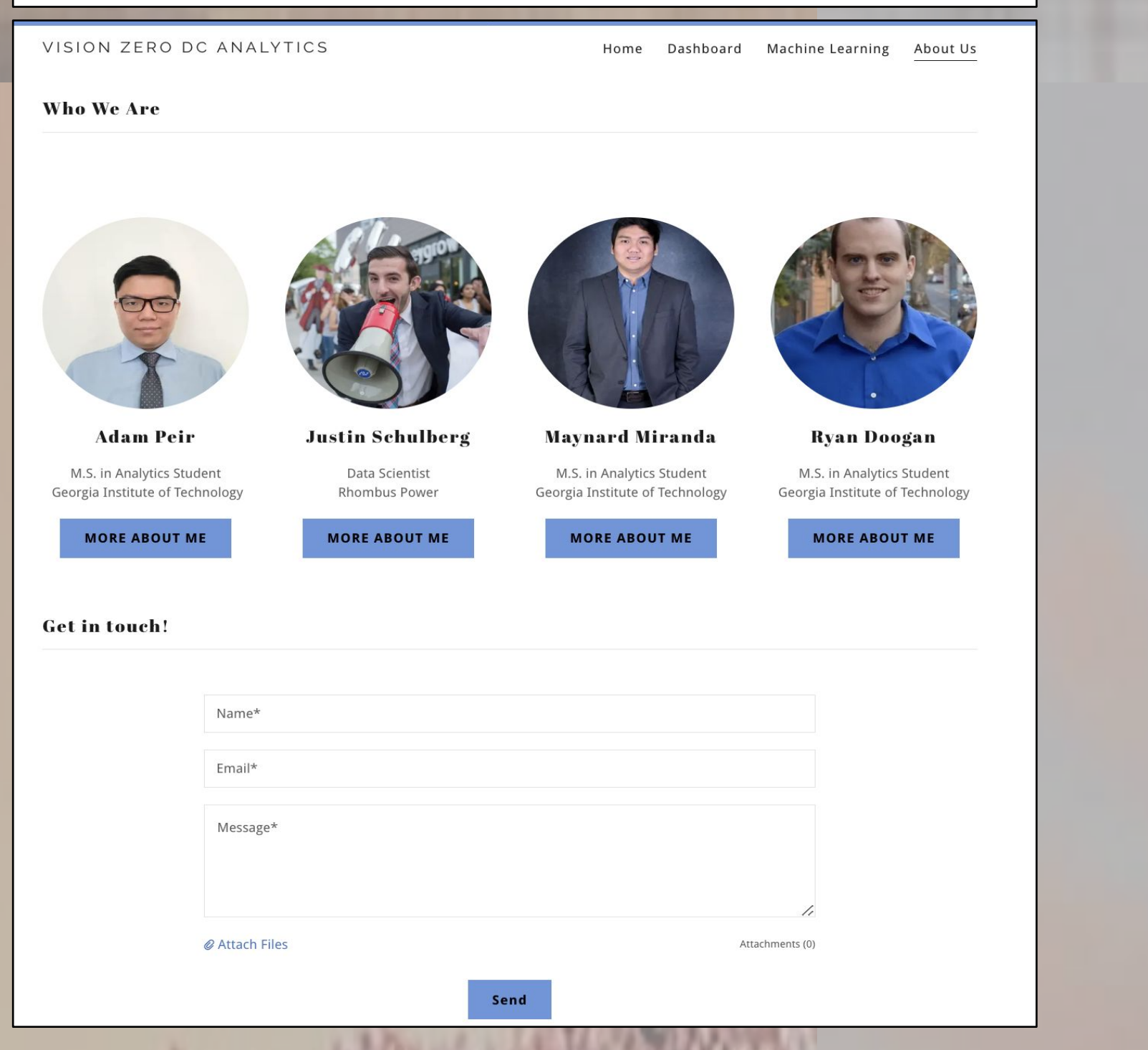
The second approach we use to group our data is DBSCAN, a density-based clustering approach.

November 23, 2022

K-Means Clustering

The first approach we use is a classical K-Means clustering algorithm. In this approach, we pre-set the number of clusters we would expect, and see how cleanly the car crashes separate into respective clusters. The ratio...

Copyright © 2022 Vision Zero DC Analytics - All Rights Reserved. Powered by GoDaddy



Who We Are

Adam Peir
M.S. in Analytics Student
Georgia Institute of Technology

Justin Schulberg
Data Scientist
Rhombus Power

Maynard Miranda
M.S. in Analytics Student
Georgia Institute of Technology

Ryan Doogan
M.S. in Analytics Student
Georgia Institute of Technology

Get in touch!

Name*

Email*

Message*