US Traffic Accidents: A Data-Driven Approach to Understanding Road Risks

Anshu Reddy Dhamana (018172141), Brendan Chao (014142726), Sri Gopi Sarath Gode (018191537), Srithareddy Devireddy (018175456), and Vinuthna Papana (018176106)

***Abstract*—In the United States, traffic accidents are a major challenge to public safety and urban planning, which require proper knowledge in order to gain a comprehensive understanding of all of the contributing factors. This project uses a rich dataset on U.S. traffic accidents from 2016 to 2023 to create interactive, data-driven dashboards through Power BI. The dashboard provides insights into accident patterns based on location, time, weather conditions, and severity, which allows city planners and urban development to gain deep understanding of all factors. There are also additional features, including geographic visualizations, time-based charts, and drill-through features, which allow users to analyze accidents on all levels.**

Our data preparation involved transforming the raw dataset through means such as standardizing zip codes, and creating additional time-based attributes like day of the week and month. The dashboards give insights into trends, such as states that are most accident prone, and slicers enable deeper analysis within states, cities, or even zip codes. This project serves to use interactive dashboards as a potential method to help inform policymakers and urban planners with a user friendly analytics mechanism. In the future, there could be methods that allow for use of machine learning prediction on accident data and the use of live data for real-time analysis.

***Index Terms*—traffic accident analysis, PowerBI, data visualization, data preparation, data understanding**

I. Introduction

In the United States, public safety and infrastructure management are both greatly impacted by traffic accidents. Reducing accident rates and strengthening emergency response systems need an understanding of the variables that affect accident severity, such as weather, road conditions, and traffic patterns. The objective of this project is to examine data on traffic accidents from 2016 to 2023, with an emphasis on changes in accident severity, assessing the influence of weather and road conditions, and investigating the geographic distribution of incidents throughout the nation. Through this analysis, we aim to offer insightful information that will help emergency responders, the lawmakers, and urban areas planners make data-driven decisions that will improve public safety.

Our goal is to identify key trends that lead to an increase in traffic accident in the United Sates through the US Traffic Accident Dataset. This dataset provides key attributes such location, time, weather conditions, severity, and traffic point of interests. Using Power BI and the dataset, we have created interactive dashboards that bounce from page to page seamlessly, to help make data analysis simpler and more accessible for even those with no background in data.

The dashboards allow users to dive into traffic accident data across many different attributes provided in the dataset, including geographic-based and time-based analysis. Features, such as drill-through, allow for users to explore the dataset at many possible levels. Our dashboards help identify areas prone to accidents, time of day that accidents occur more often, and much more.

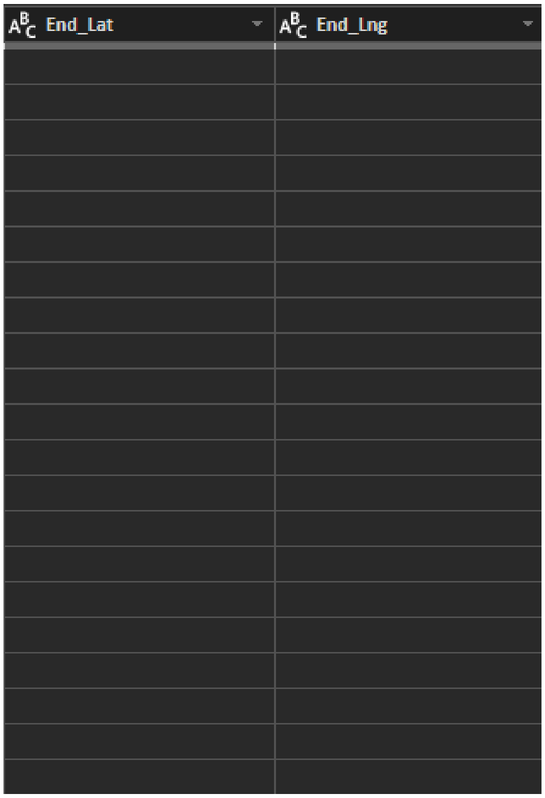
This project can allow for a better understanding of traffic accidents and inspire alternate methods for enhancing road safety. This project seeks to help policy makers take actions that will improve roadways and urban planning in the United States. We hope that through our deliverables we have found a potential solution that could help mitigate traffic accidents in the United States.

II. Methods - Data Understanding

This project uses a dataset from Kaggle that contains over 7 million entries of traffic accident data, covering a wide geographic area and time range. In relation to our project goals, this dataset contains data on accident severity, ranging from 1 (low risk) to 4 (high risk, likely fatal), helpful for understanding impact and risk factors involved in the area. The dataset also includes information like temperature, humidity, visibility, road conditions, which help us understand any external environmental factors that can influence accident severity. Additionally, we have attributes like city, state, time of day, and date which can give additional insights into location and time-based trends.

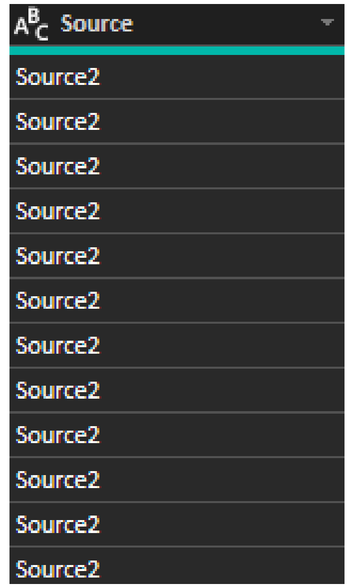
III. Methods - Data Understanding

A significant amount of data preparation was required for the project, including cleaning the dataset by filling in any missing values, formatting date and time columns, and manipulating attributes like weather and geographic location information. This was done using tools such as Python and PanDas. New variables including accident time, day, month, year, and categories for accident severity were created via feature engineering. Visualization technologies such as Power BI are utilized to create interactive dashboards that show accident hotspots and trends over time in order to effectively explain the data. By identifying the elements that lead to serious incidents, the research will also produce important insights and provide data-driven suggestions to increase road safety. By highlighting relevant patterns in traffic accidents, the project also seeks to provide important information to



End_Lat	End_Lng

Fig. 1. End_Lat and End_Lng columns.



Source
Source2
Source2
Source2
Source2
Source2
Source2
Source2
Source2
Source2
Source2
Source2
Source2

Fig. 2. Source Column

stakeholders such as emergency responders, lawmakers and city planners.

This dataset contains 46 columns of data: ID Source Severity StartTime EndTime StartLat StartLng EndLat EndLng Distance(mi) Description Street City County State Zipcode Country Timezone AirportCode WeatherTimestamp Temperature(F) WindChill(F) Humidity(%) Pressure(in) Visibility(mi) WindDirection WindSpeed(mph) Precipitation(in) Weather_Condition Amenity Bump Crossing Give_Way Junction No_Exit Railway Roundabout Station Stop Traffic_Calming Traffic_Signal Turning_Loop Sunrise_Sunset Civil_Twilight Nautical_Twilight Astronomical_Twilight. It is a feature-rich dataset that spans across 7 years from 2016 to 2023. 46 columns coupled with 7 million rows makes this dataset total up to over 3 gigabytes of CSV data.

Our first adjustment was to remove the End_Lat and End_Lng columns from the dataset. This is a mostly blank column that shows the ending latitude and ending longitude of the each accident. We felt that we could already derive enough information from the starting latitude and longitude, as well as the distance column, in order to feel confident enough to delete these two columns seen in Fig. 1.

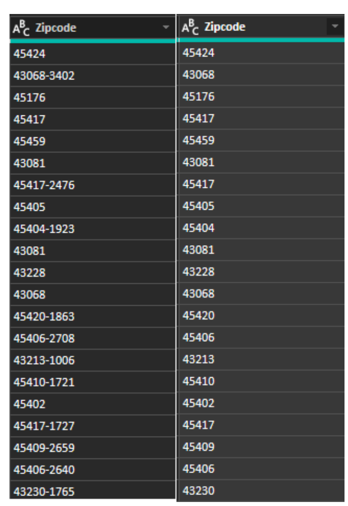
In addition, we remove additional columns that we felt didn't provide any additional context and were not useful for the overall project. This included the source column, which only contained ”Source1” and ”Source2” and provided no additional information on the actual source of the information as seen in Fig 2.

Another column we removed was the airport column. This provided information on local airports which we didn't believe provided any context on the actual accident itself.

For data transformation, we wanted to standardize our zip code format. There were both 5-digit and 9-digit formats in the dataset and we wanted to transform them all into the 5-digit format. To achieve this, we used Microsoft Power Query's ”Truncate” and took only the head 5 numbers of each zip code. This helps us with analyzing areas of the same zip code without any potential conflicts. This can be seen in Fig 3.

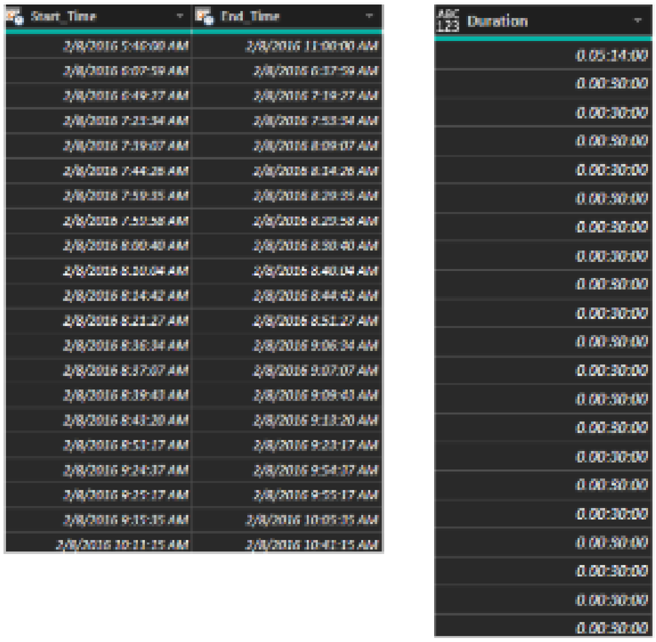
Another transformation was to take the start and end time of the accident, and through feature engineering, create a column that represented the entire duration of the accident, which was not provided in the original dataset. To achieve this, we subtract the start time from the end time of the accident through Power Query once again. The columns, as well as the resulting column for duration can be seen in Fig. 4.

Lastly, we created additional time-based columns derived



Zipcode	Zipcode
45424	45424
43068-3402	43068
45176	45176
45417	45417
45459	45459
43081	43081
45417-2476	45417
45405	45405
45404-1923	45404
43081	43081
43228	43228
43068	43068
45420-1863	45420
45406-2708	45406
43213-1006	43213
45410-1721	45410
45402	45402
45417-1727	45417
45409-2659	45409
45406-2640	45406
43230-1765	43230

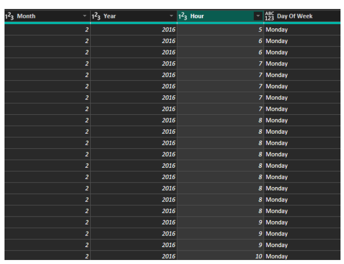
Fig. 3. Zipcode Column, before and after data transformation



Start_Time	End_Time
2/8/2016 5:46:00 AM	2/8/2016 11:00:00 AM
2/8/2016 6:07:59 AM	2/8/2016 6:37:59 AM
2/8/2016 6:49:27 AM	2/8/2016 7:19:27 AM
2/8/2016 7:23:34 AM	2/8/2016 7:53:34 AM
2/8/2016 7:39:07 AM	2/8/2016 8:09:07 AM
2/8/2016 7:44:26 AM	2/8/2016 8:14:26 AM
2/8/2016 7:59:35 AM	2/8/2016 8:29:35 AM
2/8/2016 7:59:58 AM	2/8/2016 8:29:58 AM
2/8/2016 8:00:40 AM	2/8/2016 8:30:40 AM
2/8/2016 8:10:04 AM	2/8/2016 8:40:04 AM
2/8/2016 8:14:42 AM	2/8/2016 8:44:42 AM
2/8/2016 8:21:27 AM	2/8/2016 8:51:27 AM
2/8/2016 8:36:34 AM	2/8/2016 9:06:34 AM
2/8/2016 8:37:07 AM	2/8/2016 9:07:07 AM
2/8/2016 8:39:43 AM	2/8/2016 9:09:43 AM
2/8/2016 8:43:20 AM	2/8/2016 9:13:20 AM
2/8/2016 8:53:17 AM	2/8/2016 9:23:17 AM
2/8/2016 9:24:37 AM	2/8/2016 9:54:37 AM
2/8/2016 9:25:17 AM	2/8/2016 9:55:17 AM
2/8/2016 9:35:35 AM	2/8/2016 10:05:35 AM
2/8/2016 10:11:15 AM	2/8/2016 10:41:15 AM

Duration
0.05:14:00
0.00:30:00
0.00:30:00
0.00:30:00
0.00:30:00
0.00:30:00
0.00:30:00
0.00:30:00
0.00:30:00
0.00:30:00
0.00:30:00
0.00:30:00
0.00:30:00
0.00:30:00
0.00:30:00
0.00:30:00
0.00:30:00
0.00:30:00
0.00:30:00
0.00:30:00
0.00:30:00

Fig. 4. Duration column, along with start and end time columns



Month	Year	Hour	Day Of Week
2	2016	5	Monday
2	2016	6	Monday
2	2016	6	Monday
2	2016	7	Monday
2	2016	7	Monday
2	2016	7	Monday
2	2016	7	Monday
2	2016	7	Monday
2	2016	7	Monday
2	2016	8	Monday
2	2016	8	Monday
2	2016	8	Monday
2	2016	8	Monday
2	2016	8	Monday
2	2016	8	Monday
2	2016	8	Monday
2	2016	8	Monday
2	2016	8	Monday
2	2016	9	Monday
2	2016	9	Monday
2	2016	9	Monday
2	2016	10	Monday

Fig. 5. Time based columns derived from Start time column

from the start time of the accident. These include month, year, monthYear, hour, and day of week columns. We use these columns as they can be useful to slice across different time based features on accident data in order to gain better understanding of any influences that can be derived from time-based situations, such as season, weekday/weekend, time of day, holidays, as well as many more. These columns can be seen in Fig 5.

IV. Results - PowerBI

After preparing all of our data, we can begin data visualization and dashboard creation in PowerBI. PowerBI is a powerful data visualization tool introduced by Microsoft. It allows us to easily create dashboards using their feature-rich toolset. To best explain our dashboard, we will explain the process from the point of view of a person who is using our PowerBI report for the first time.

Upon using the dashboard, the user will arrive first on the the title dashboard. This dashboard simply serves as a hub for going deeper into our other dashboards. On the top, we have the title in large font that reads "Exploring Traffic Accident Data from the US". This gives the user a brief, yet decent understanding of what our project dashboard will cover. There

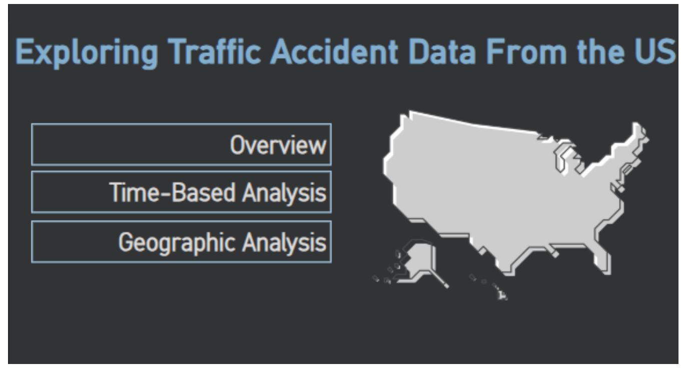


Fig. 6. Title Dashboard

is also a figure of the United States and three buttons. These buttons serve to help users navigate between dashboards with ease. Our title slide can be seen in Fig. 6.

A. Overview

First, let's click on the Overview button. Clicking this brings the user to an overview dashboard. The main idea behind this dashboard is to create a simple dashboard that can give more general statistics across the United States. Figure 7 gives us a view of what the user sees.

On the left hand side, we have a multi-row card in PowerBI that shows us general statistics. This includes accident count, average severity, and average accidents per month. These are statistics that can inform the user on nationwide statistics as well as just give general information on the dataset as a whole.

On the top there is a line chart that shows accident count by month. This is a time series that allows tracking of time across the year that accidents most frequently occur. The time series is bucketed by month-year which gives us a better understanding compared to a year-based chart but a better curve compared to a daily chart. From these results, we can scroll left to see an uprise in accidents towards the end of the dataset from 2021 to 2023, as seen in Fig. 8.

Next, starting from the left, we have accident count by state represented with a bar chart. Our dataset, as mentioned before contains, data to much finer grain, such as city, zip code, street. But in the overview, we just wanted to provide

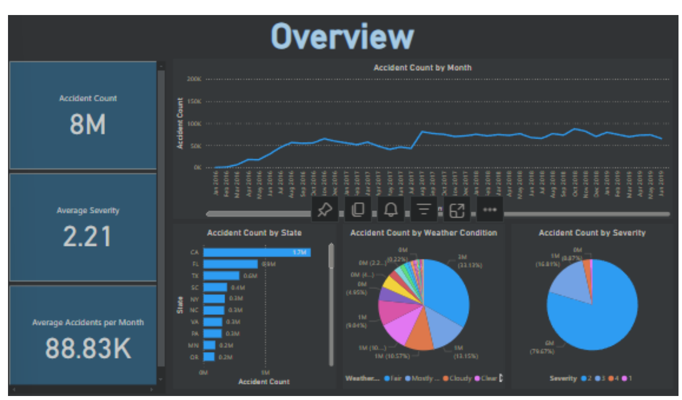


Fig. 7. Overview Dashboard

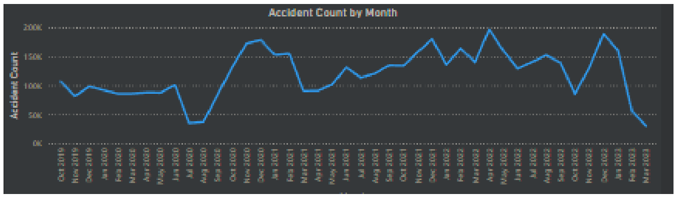


Fig. 8. Accidents by Month, from 2019 to 2023

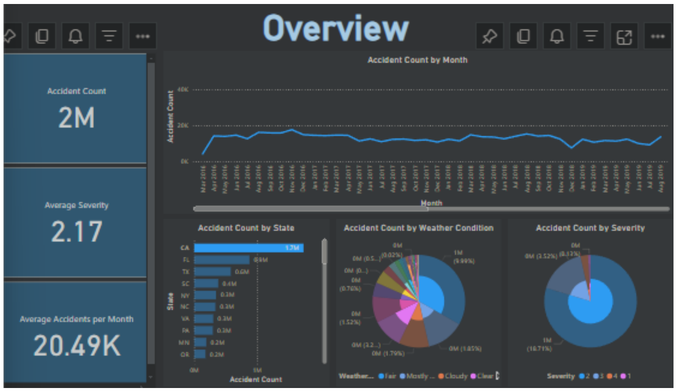


Fig. 9. Overview dashboard, updated after selecting California in the bar chart.

much more general information and not to dive too deep into the finer grain. As seen in Fig. 9, the state that has the most reported accidents within the timeframe of the dataset is CA, or California with almost 2 million entries. An interesting addition from PowerBI is that it allows the user to select a state in the bar chart and it acts as a slicer for the other data in the dashboard. The updated dashboard can also be seen in Figure 9.

In the middle, we have a pie chart showing accident count by weather condition. The pie chart is useful for viewing proportions and provides percentages in order to help users in that direction. From this chart, we derive that almost a third of all accidents in the time frame occurred during fair weather. We cannot say fair weather causes more accidents however, as fair weather could just be the most commonly occurring weather condition. But we can, once again slice by weather condition and see that by selecting fair weather, the average severity of accidents goes down from 2.21 to 2.13, which could indicate fair weather accidents are not as severe as other weather conditions. The updated dashboard can be seen in Figure 10.

Lastly, for our overview, we have accident count by severity as seen in Figure 11. Severity in our data is a scale of 1 to 4, which is basically a scale that is used to somewhat accurately describe how severe the accident was, with 4 being almost certainly fatal. From this chart, we can see almost no accidents are given a 1 on the severity scale, amounting to less than 1 percent of all total accidents. We can also see a majority of accidents are of severity level 2, totaling almost 80 percent of the data. Severity level 4 accidents are seemingly not as common, amounting to around 2.5 percent of the data. However, 2.5 percent of the data is totaling over 200,000 accidents that were almost all certainly fatal. This total puts the

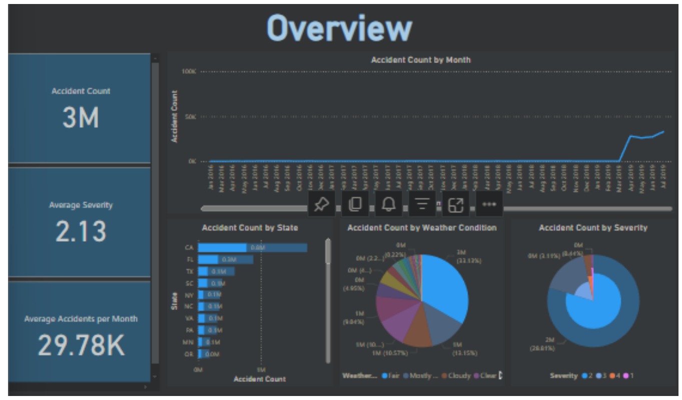


Fig. 10. Overview dashboard, updated after selecting Fair Weather in the middle pie chart.

data in a much harsher context than of which can be viewed just from observing the pie chart.

B. Time-based Dashboard

Next, let's go over the time-based dashboard. This dashboard contains data visualizations as well as time-based slicers that allows us to observe data across time and draw conclusions from there. The entire dashboard can be seen in Figure 12.

At the top is our accident count, similar to the card from the overview dashboard. This just shows the accidents that occured based on any slicers we chose and updates it in the PowerBI card. Next, we have a severity gauge that shows the average severity of accidents that occurred. The gauge is a range from 1 to 4 and the bar gives us a visual indicator of how severe accidents were in the time frame.

Next to the severity gauge, we have weather averages that give more context into how different weather variables were across these times. These averages are all computed using PowerBI measures which basically act as aggregation queries for our dataset. These weather attributes include average temperature, wind chill, windspeed, precipitation, visibility, humidity, and pressure. These key insights will help drive important decisions in order to mitigate the effects of weather on traffic conditions.

Below, we have the main exhibit of the dashboard, which is the accident map. This map shows points based on where

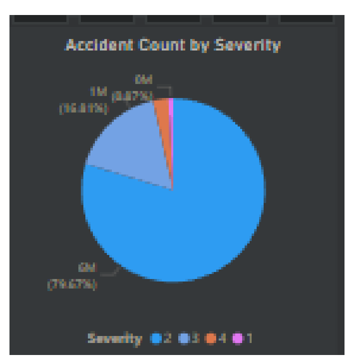


Fig. 11. Severity pie chart

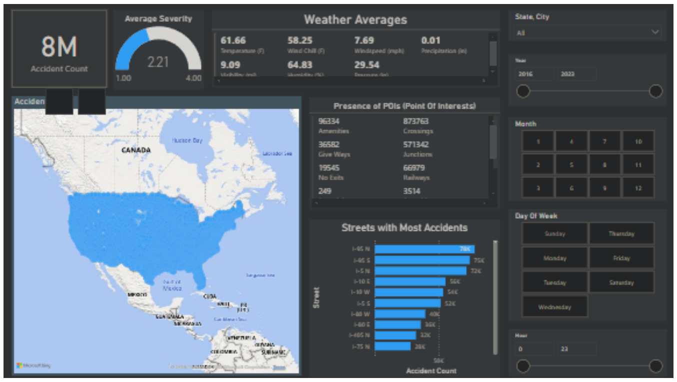


Fig. 12. Time-based dashboard

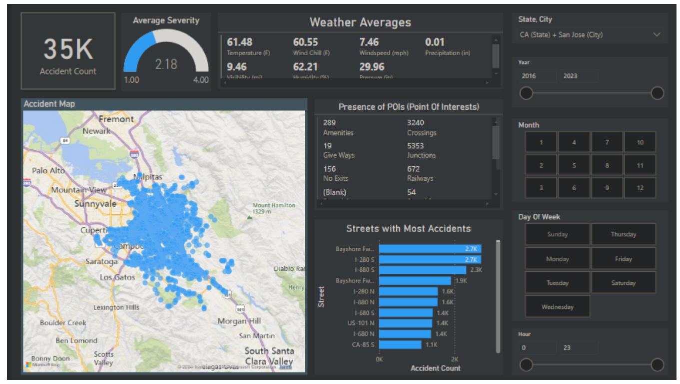


Fig. 14. Dashboard when selecting San Jose

accidents occurred on an actual map provided by PowerBI. This is done by providing our starting latitude and longitude to the map as values. It also updates based on any slicers we may choose to include. The map will also automatically zoom into the key areas that are set by slicers. This map, in general, gives the user a tool to guide them visually as they explore this dashboard. Next, we have traffic point of interests. These are different features that were present in each accidents.

This includes crossings, junctions, railways, etc. These show the total of each type and display them in a multi-row card. Understanding the point of interests allows decision makers to make decisions on the best way to navigate the use of point of interests as well as fixes that would lessen the harm of certain areas.

The last chart shows streets with the most accidents and allows us to see which streets have the most accidents during different time periods in the year. Selecting a street here also allows us to slice across both street and time dimensions at the same time, giving the user more actionable insights into the data. As shown in Figure 13, when selecting the top street, I-95 N, the map updates to show all accidents across the interstate.

Next, the main point of this dashboard is our slicers. First, we have a simple hierarchical slicer for state and city. This can be useful for more local decision makers to help mitigate accidents on streets within their communities. In Figure 14, we can see the updated dashboard based on selecting San Jose, California, in our state, city slicer. The map along with all other visualizations updates to display information based on accidents occurring in San Jose.

Next, the time based slicers allow us to view trends based on time. First, there is a year slider. This allows the user to view accidents occurring within particular timeframes based on year. Below that, there is a month slicer. This slicer uses tiles to select months for the slicer. The user can also select several months which could be useful for seasonal data such as the holidays, as seen in Figure 15, where we select December and January together on San Jose.

Lastly, we have a day of week slicer, which allows the user to filter by different days across the week. This could useful for finding trends along the weekend/weekdays on different streets in different cities/states. Below, at the bottom right corner, there is an hour slicer, to track information on accidents occuring at a specific time of the day. Any of these slicers can be combined in order to generate even more insights into the dataset.

C. Geographic Dashboard

Next, let's look at the geographic dashboard. Users can get here by going back to the title dashboard and selecting the button for Geographic Analysis.

On arrival, many of the features that were seen in the time-based dashboard can still be seen in the geographic dashboard. This dashboard can be seen in Figure 16.

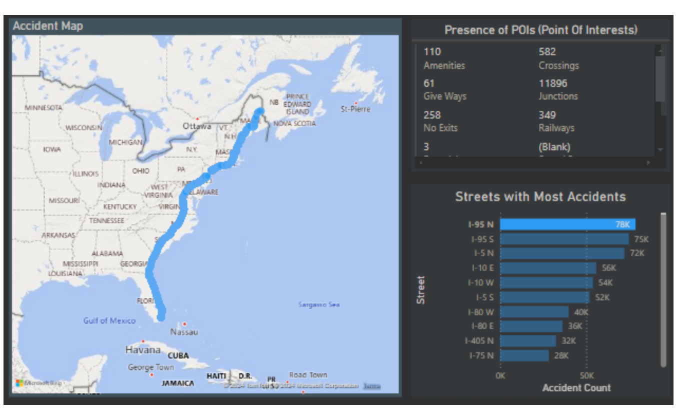


Fig. 13. Map with I-95 N selected

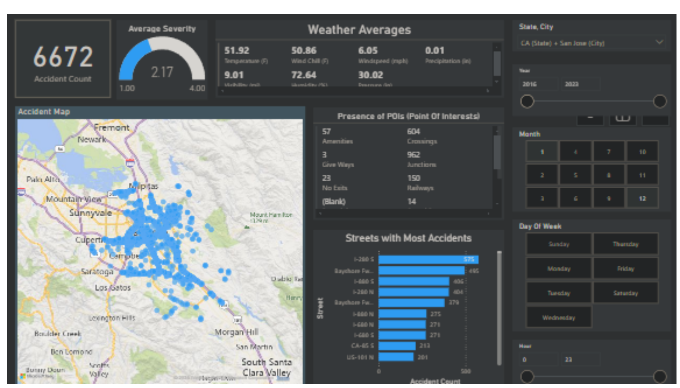


Fig. 15. Time-based Dashboard, with December and January selected, based in San Jose

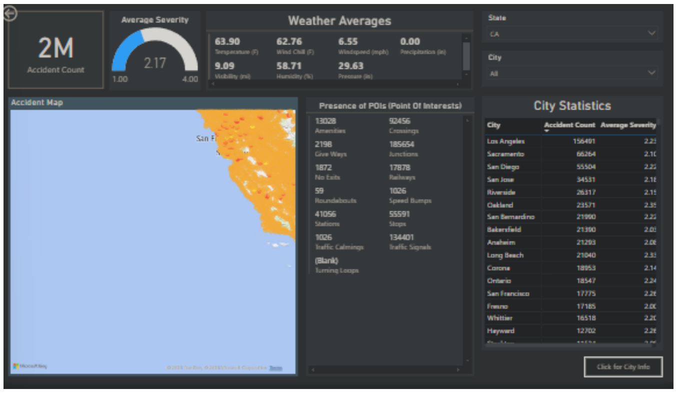


Fig. 16. Geographic Dashboard

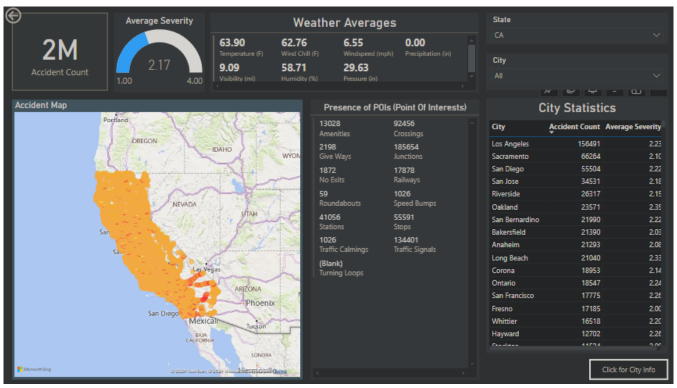


Fig. 17. Geographic Dashboard after selecting California

There are, however, key differences. First off, on the map, we have color coded accidents by severity. This could be useful for visually seeing streets and areas where accidents are more severe than other areas. The scale ranges from yellow, to represent low severity, to red, which represents higher severity.

Next, the time-based slicers were also removed as they are not necessary for the more geographic based dashboard. They are instead replaced by a table that shows cities, their accident counts, as well as average severity of each city. The table allows users to sort by different attributes in order to find cities with more accidents/more severe accidents than others. Figure 17 shows what the dashboard as a whole looks like after selecting the state of California. The chart, when sorted by accident count, shows that Los Angeles has the highest accident count of all cities.

This dashboard gives a slightly more drilled down view of the dataset compared to that of the overview slide. In order to get more in-depth analysis, we can select a city on the table and click the button at the bottom for more info on individual cities. This button uses PowerBI's drill-through capabilities to send the user to another dashboard which views the data on a finer grain.

D. City Dashboard

Clicking the button brings the user to an individual city dashboard. The dashboard is built based on which city the user selected on the previous Geographic Dashboard. For example, if the user selected Los Angeles, their dashboard would look like the one in Figure 18.

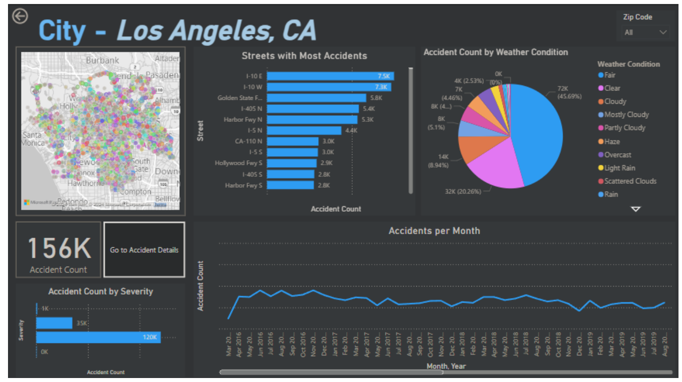


Fig. 18. City Dashboard with Los Angeles selected

First off, this page has a title text box. The title is "City - Los Angeles, CA" with Los Angeles being changed with any city selected in the previous dashboard, as well as CA being replaced by the corresponding state. This is done through PowerBI's text boxes and the feature that allows us to include variables as values in the text box, similar to a formatted string in Python. The back button on the top left also allows the user to easily go back to the previous geographic dashboard and select another city if they desire.

Next, we have a map centered on the city involved. The points are randomly colored as we needed to include the ID of each accident as a legend in order to add another feature discussed later.

Below is another card showing the aggregate accident count, seen on all other slides thus far. This card is very useful for getting a sense of the overall picture of the area.

Below that is a bar chart showing accident count by severity. Before, on the overview dashboard, this was displayed a pie chart. But, due to size constraints, we decided to change this to a bar chart. As a bar chart we can order by severity and it turns the chart into a chart more similar to that of a histogram. Clicking a bar also slices the other visuals on the slide as seen in figure 19, where we select severity level 3.

The first visual next to the map shows the streets with the most accidents. Because this is also city-based, the data

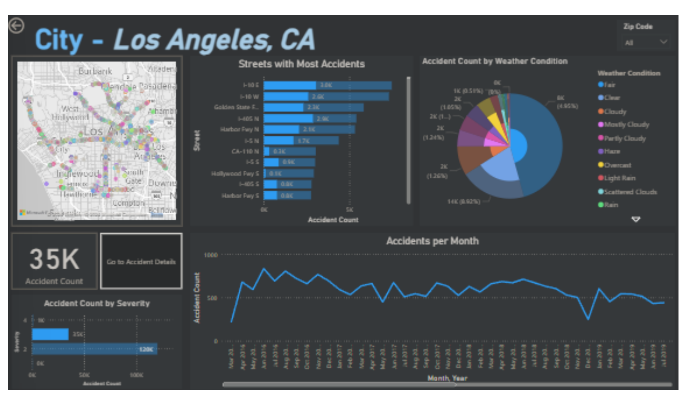


Fig. 19. City Dashboard for Los Angeles with Severity Level 3 selected

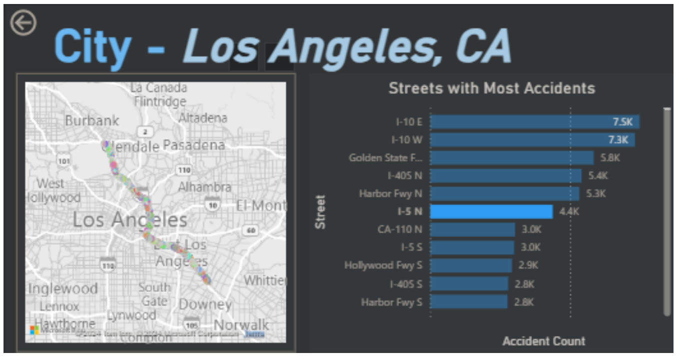


Fig. 20. City dashboard map in Los Angeles with I-5 N selected

will only include parts of the street/highway that is within the bounds of the city. This is particularly useful for city planners to work with only information on roads within their particular regions. For example, if the user selects Interstate 5 Northbound, a road famous for connecting northern and southern California, the map will only show the section of Interstate 5 in Los Angeles, as seen in Figure 20.

Next to the bar chart, is a pie chart showing accident count by weather condition. This visual helps users understand what weather conditions may impact driving in the area and also gives the user a good sense of the weather conditions that frequently occur through this particular city/street. For example, in Los Angeles, almost half of the weather conditions in traffic accidents are accounted as "Fair" weather. This is higher than that of the entire United States as seen in the Overview dashboard. Selecting "Fair" also allows for other insights. For example, even though almost half of the accidents in Los Angeles occur under fair weather, only around a quarter of accidents of level 3 and 4 severity occurred under fair weather. This could interpret as fair weather being less dangerous in comparison to other weather conditions. This dashboard with "Fair" weather selected can be seen in figure 21.

At the bottom, we have a time series showing accidents per month. Time series allows the users to track trends across time within each city and help try to correlate any events that could explain spikes in the data. For example, we can select

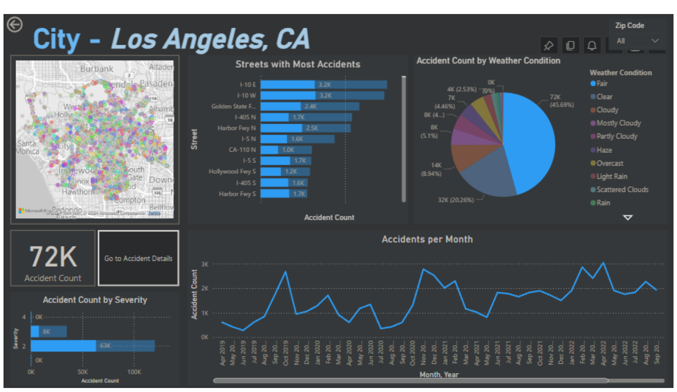


Fig. 21. City dashboard for Los Angeles, with "Fair" selected for Weather Condition

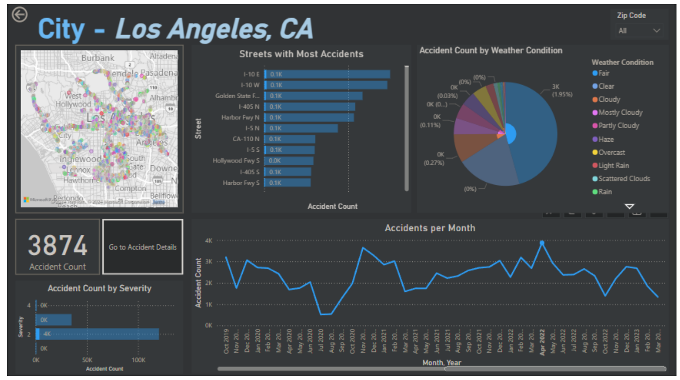


Fig. 22. City dashboard after selecting the most accident-dense month in the time-series, Apr 2022

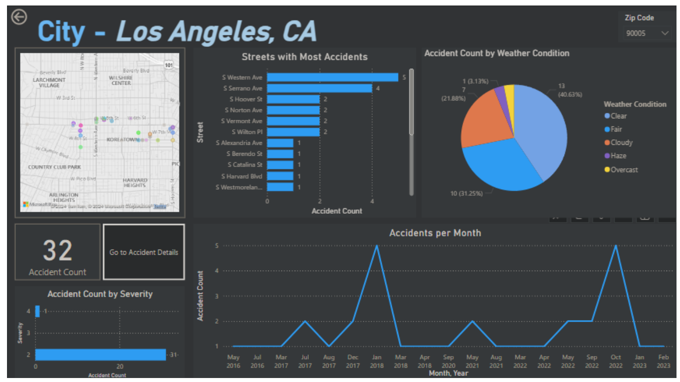


Fig. 23. Dashboard for Los Angeles after selecting "90005" as Zip Code

"Apr 2022" which appears to be the highest point in our chart. Doing so updates the other charts and map to give insight into traffic accident occurrences that take place in the month. Figure 22 shows the dashboard after selecting "Apr 2022" in the time series chart.

Another inclusion on this dashboard is the ability to slice the dashboard by zip code. This helps local districts and community members to look up data on accidents happening within their own communities. By selecting a zip code, the data on the rest of the dashboard transforms, and the map will automatically zoom into the section of the map pertaining to that particular zip code. For example, Figure 23 shows what the dashboard looks like when selecting zip code "90005".

Lastly, let's go back to colored points on the map. The points on the map are colored in order to include each accident ID as an attribute of each accident on the map. This allows the user to select a point on the map for more detailed information on the current slide, as well as, drill-through to our final dashboard that gives details on each individual accident. Clicking the button that reads "Go to Accident Details" after selecting a point on the map will bring the user to the "Accident Details" dashboard. This can be seen in Figure 24.

E. "Accident Details" Dashboard

When entering the Accident Details dashboard, This flex dashboard provides detailed information on each specific accident and is useful for studying particular variables surrounding

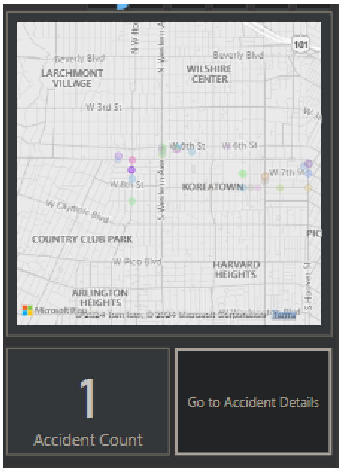


Fig. 24. Selecting a point on the map of the City Dashboard lights up the button below

Fig. 25. Interactive flex title for the "Accident Details" dashboard

an accident. Users can explore data such as the exact location, time, severity, weather conditions of an accident. By creating this dashboard, users can see the data at the finest grain possible given from our dataset.

Like the City Dashboard, the Accident Details dashboard starts with a title text box and has a variable that changes to the corresponding accident ID number of each accident. The number doesn't mean much to the user, but it allows us to ensure each individual point is unique and loaded in properly. The title can be seen in Figure 25.

Below the title and to the left, we have important time and location information of the accident. This data includes: street name, city, county, state, zipcode, datetime, day of the week, and timezone of the accident. As the first piece of information the user encounters, it provides a insightful overview, providing crucial details about the accident. The data is displayed in a PowerBI card column, aligned to the right for easy reading as seen in Figure 26.

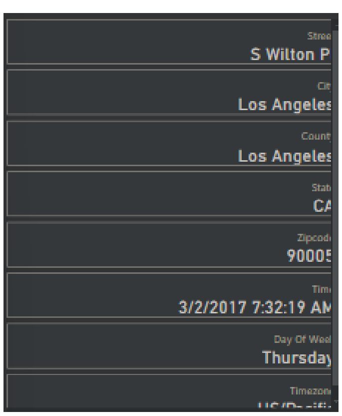


Fig. 26. Card containing accident time and location information

Fig. 27. Latitude and longitude of the accident

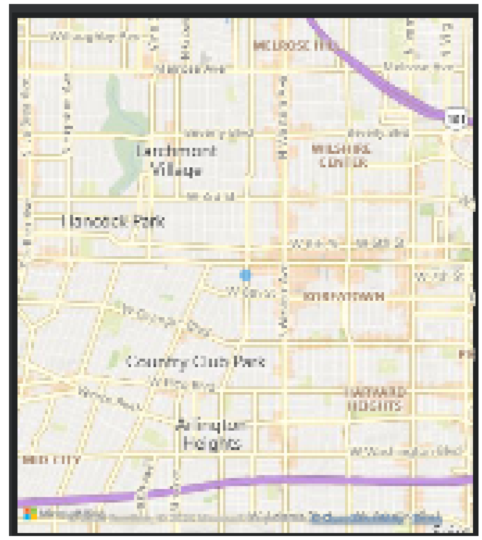


Fig. 28. Map with accident on PowerBI

To the right, there is also information on the latitude and longitude of the accident as seen in Figure 27, which provides precise geographical coordinates for the particular incident. Combined with the time and location details on the left, this section provides a comprehensive overview of the accident's location and time, allowing users to know exactly where and when the incident occurred.

To the far right, we pinpoint the exact location of the accident on a PowerBI provided map. This is done using the latitude and longitude to provide exact geographical coordinates of the accident. By placing the point on the map, we can provide context on the surrounding area, such as whether the accident occurred on a highway or if the intersection may be particularly confusing. Mapping the point also allows users to contextualize areas they may not be familiar with. For the example here in Figure 28, users may not recognize S Wilton Pl when just reading the name of the street, but can recognize the area on the map as it is next to Koreatown in Los Angeles.

In the center of this graphic is the severity gauge. This gauge ranges from 1 to 4 and gives context into how severe the accident is compared to others in the dataset. The clear design of the gauge allows users to easily understand the impact of the accident. The gauge is also color coded, similar to the map on the Geographic dashboard. Yellow represents low severity while red represents high severity. This accident here in Figure 29 is of severity level 4 and in bright red, giving context into the seriousness of this particular incident.

Lastly, at the bottom of the page, we have weather and other environmental factors listed. These include humidity, precip-

Fig. 29. Severity Gauge on the "Accident Details" dashboard

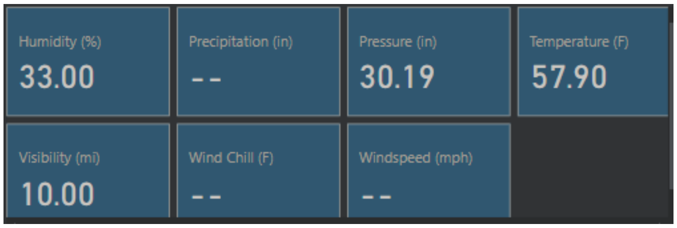


Fig. 30. Weather and environmental context card for each accident

itation, pressure, temperature, visibility, wind chill, and wind speed. This section allows users to visualize environmental conditions that affected this particular incident which could provide insights and potential correlations between weather factors and accident occurrence. Displaying this data in a structured multi-row card allows the user to explore the context of each accident in regards to weather. This card in PowerBI can be seen in Figure 30.

V. Discussion

Our goal for this project was to explore traffic accident data in the United States, and provide data-driven dashboards that can show the dataset's characteristics, insights, and patterns. The insights that can be derived from this dashboard hopefully will improve safety measures across the United States. The use of dashboards can hopefully be used to bring insights and analytics into other fields as well.

A. Key Findings

The analysis highlights the role environmental conditions play in road accidents, supporting Bessen's (2019) research [1], which identifies how adverse weather—like precipitation and reduced visibility raises the likelihood of accidents. Our inclusion of weather factors, such as humidity, wind speed, and temperature, aligns with studies that explore the complex relationship between climate and road safety, reinforcing the importance of these variables in accident analysis.

Our dashboard's ability to pinpoint accident severity within specific contexts draws from the methodologies outlined by Krishna et al. (2023) [2], who apply machine learning models to analyze accident severity. While this project does not yet incorporate predictive models, the underlying data structure and framework set the stage for future integration of such techniques, paving the way for more dynamic, data-driven insights.

Viewing accident severity and location details aligns with the National Highway Traffic Safety Administration's (2020) [3] traffic safety report, which talks about how local-level data is used in identifying accident hotspots. By utilizing map visuals and severity gauges, our dashboard allows users to focus their efforts in areas where they are most needed, such as improving roads in high-risk areas.

The World Health Organization's (2018) [4] report on road safety also shows the impact of traffic accidents in high-risk areas as well. This paper relates to how our project has a wide range of use cases and applications. Our project is a valuable tool that can aid public safety officials, community members, etc. and help them gain insights into trends and guide efforts to improve infrastructure.

Lastly, the dashboard could potentially be improve with the help of machine learning, as touched upon by Moosavi (2024) [5].

B. Future Directions

This project not only serves as a tool for descriptive analytics but also could lead to more predictive and prescriptive models. Future versions could integrate machine learning algorithms to predict accident probabilities under different conditions, as suggested by Krishna et al. (2023) and Moosavi (2024) [2], [5]. We could also enhance the dashboard with more granular information such as heavy traffic to aid urban planning and public safety.

Building on top of existing research and branching out into more modern techniques could allow this dashboard to become a main hub to tackle the problem that is imposed by traffic. By embracing the new age of machine learning and its capabilities, we can mitigate the issues of traffic not just in the United States, but worldwide.

Acknowledgment

The authors would like to thank Professor Shih Yu Chang and TA Sourab Rajendra Saklecha for their guidance and support throughout this project, through the process of the data analytics pipeline and data visualization.

References

[1] J. E. Bessen, "The impact of weather conditions on road accidents," National Bureau of Economic Research, Tech. Rep., 2019, accessed on [Insert Date]. [Online]. Available: https://www.nber.org/papers/w24235

[2] U. V. Krishna, S. Sudhakaran, S. Sanju, E. Vignesh, and R. Kaladevi, "Accident prediction and analysis using machine learning models," in *Proceedings of the International Conference on Intelligent Data Communication Technologies and Applications (ICIDCA)*, 2023. [Online]. Available: https://ieeexplore.ieee.org/document/10099532

[3] N. H. T. S. Administration, "Annual traffic safety report," https://www.nhtsa.gov/press-releases/2020-traffic-crash-data-fatalities, 2020, accessed on [Insert Date].

[4] W. H. Organization, *Global Status Report on Road Safety*. Geneva: World Health Organization, 2018. [Online]. Available: https://www.who.int/publications-detail/global-status-report-on-road-safety-2018

[5] S. Moosavi, "Recent advances in traffic accident analysis and prediction: A comprehensive review of machine learning techniques," *arXiv*, 2024. [Online]. Available: https://arxiv.org/abs/2406.13968