De La Salle University

De La Salle University - Manila
3rd Trimester, Academic Year 2024–2025

In partial fulfillment of the course
DATANVI

WatchMetro: An Interactive Dashboard for Metro Manila Traffic Incidents

Final Project Report

Submitted by:
Feliciano, Jan Robee E.
Sese, Alphonse Juanito T.

Submitted to:
Sir Joshua Manzano

This final report covers each phase of progress towards the completion of WatchMetro. The first phase contains the proposal of the project, the second phase details the prototyping and planning process, and the final phase discusses the intricacies towards completion as well as the details of the deliverables.

I. Project Proposal

"With the increasing concern that vehicular accidents are becoming more and more frequent over the years, many Filipino commuters are starting to incur road anxiety. Whether driving in private vehicles or taking public vehicles as their mode of transportation, some accidents are bound to occur regardless of how careful we can be. We would like to create an interactive dashboard that helps our target audience understand where and how frequent accidents may occur in certain areas and at certain times of the day. This can aid in planning routes ahead of time, as well as avoiding certain roads that may have a higher risk of accident during specific times of the day. The dashboard will be a valuable educational tool for commuters to better visualize the incidents recorded across different regions of Metro Manila."

A. Target Audience

The target audience of this project will be the mass commuters/general public that drive or are being driven along the streets of Metro Manila. As such, the background of these people may consist of a wide range: Public utility/TNVS/privately hired drivers, 9-5 workers, motorcycle drivers, etc. As such, one can also expect that their potential comprehension of this project to vary significantly as well. Furthermore, the problem of a possible technological gap between the project and a portion of the target audience may arise, as not all can be expected to withhold the necessary skills to be considered technologically literate. Nevertheless, this project aims to minimize its complexity in order for the target audience to fully understand the information that is aimed to be provided, regardless of literacy level and comprehension ability.

B. Visualization Problem

While it is ideal to construct roads that are well lit, visible, and spacious enough, not all roads are built and designed the same way. While some roads may be brightly lit during the morning, these same roads may become dark and hard to see during the night due to several factors such as broken road lamps or insufficient lighting. Certain months are also more prone to rain due to the weather in the Philippines as well as stronger rain due to typhoons, which can lead to variance in the occurrence of vehicular accidents compared to other times of the year. Given the initial paragraph at the start of this proposal, the main goal of this project is to help our target audience visualize which streets are more prone to vehicular accidents.

Similar to this reference that revolves around traffic congestion in Metro Manila, the goal of our project is to create a map that visualizes how often vehicular accidents happen in certain streets of Metro Manila by highlighting roads where vehicular accidents occur based on our data and emphasizing the frequency these accidents happen by saturating the colors of these roads based on several different factors.

Besides the main objective, a couple side goals of the project are the following:

- Allow the user to filter the data by time, providing visualization to the frequency of vehicular accidents during certain times in a 24 hour interval.
- Allow the user to filter the data by certain dates or date ranges, providing the ability to visualize how often vehicular accidents may occur during certain dates or periods of the year.

In planning road trips in advance, most people usually aim to optimize it by minimizing their travel time. By providing our target audience with this project, they are also able to take into account which roads within Metro Manila are more prone to accidents. This benefits not only our general audience, but also specific crowds within our audience such as beginner drivers, drivers of unfamiliar roads, and drivers with road anxiety that wish to minimize the risk of being involved in such accidents.

C. Dataset Description and Connections

Dataset: Manila MMDA Traffic Incident Data

The dataset was obtained via Kaggle, created by user Panji Brotoisworo. The dataset contains shared traffic data that is collected via the MMDA Twitter page. Using Twitter's/X's API, Mr. Brotoisworo created a python script that was able to parse through the account's tweets and get the relevant data. The source code used to achieve this may be viewed from the dataset author's github repository. Afterwards, each relevant tweet was then converted by Mr. Brotoisworo into a row or an entry that organizes the data into different columns in an MS Excel file. These columns were categorized into date, time, city, location, latitudinal and longitudinal coordinates, direction of road, type of accident, number of lanes blocked, vehicles involved, tweet's original text content, and the link to the tweet. Below are further details about each feature:

Date	Date when the accident occurred formatted in YYYY-MM-DD
Time	Time when the accident occurred formatted in 12 Hour Clock (e.g., 2:45 PM)
City	City where the accident occurred
Location	General location where the accident occurred (e.g., EDSA Shaw Tunnel)
Latitude	Latitude of the location where the accident occurred
Longitude	Longitude of the location where the accident occurred
High_Accuracy	Boolean value that states whether the location is of high accuracy (will not be used in the final dashboard)
Direction	The direction of the road in which the accident occurred (e.g., northbound)
Type	The type of accident that occurred (e.g., stalled vehicle, vehicular collision)
Lanes_Blocked	# of lanes blocked as a result of the accident
Involved	The type of vehicles involved
Tweet	The full text content of the MMDA Tweet
Source	The URL to the Tweet source

The dataset has a total of 17,313 entries spanning from August of 2018 up until December of 2020. Upon initial observation, the group has determined the coordinates of each data report to be accurate under the use of random sampling. Continuous and possibly further rigorous testing may be needed in order to create security in regards to data accuracy.

II. Prototyping and Planning

Attached are the following designs proposed for the WatchMetro project

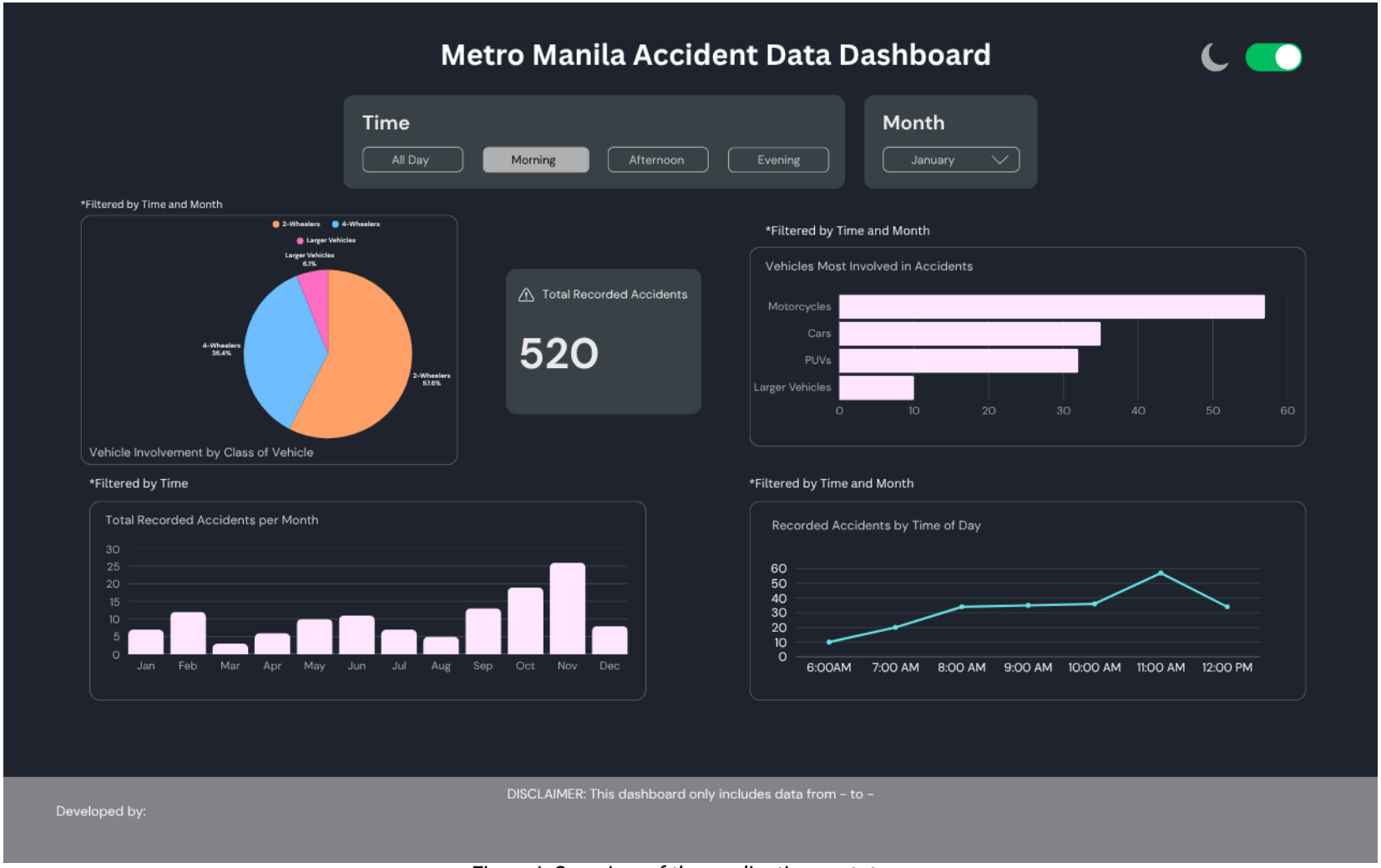
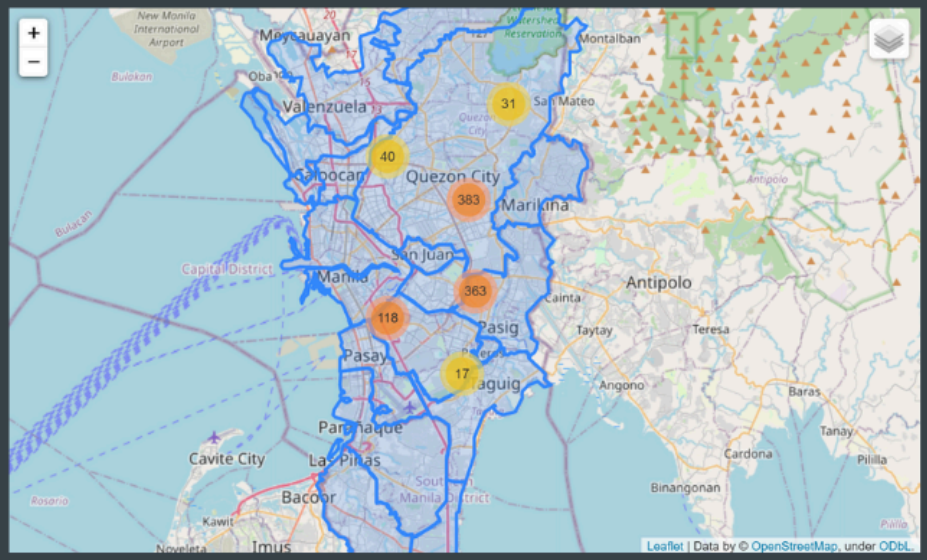


Figure I: Overview of the application prototype

Map Overview of Accidents



Developed by:

DISCLAIMER: This dashboard only includes data from – to –

Figure II: Proposed map visualization for the project

As we aim to maximize the amount of information our target audience would be able to absorb from this project, we must be able to identify the key components of our data in order to simplify the project and for them to fully understand what the project tries to convey. Nonetheless, we must keep in mind that an overcomplexity or excessive amounts of data may become complicated for the everyday Filipino to comprehend, thus preventing us from reaching this goal. Hence, a balance should be considered between the process of maximization and simplification. As we dive further into the analysis of vehicular accident data, here are a list of points that we want to consider upon creating this project:

A. Visualization Idiom Choices:

While it may be interesting to integrate more intricate charts and graphs into this project, it may not be in the best interests for the target audience's comprehension. Hence, we propose to use simpler visualizations for our data, such as the following:

A. Total Accidents Recorded per Month

Idiom	Bar Chart
Data	X-Axis: Month of the recorded accident (Separated into 12 bars representing each month) Y-Axis: Numerical value of accidents recorded
Channels	Color for the bars: -Light mode: #505050 -Dark mode: Light Neon Pink (#ffe8ff)
Task	Compare trends

B. Total Accidents Recorded by Time of Day

Idiom	Line / Dot Plot
Data	X-Axis: Time of day during the vehicular accident Y-Axis: Numerical value of accidents recorded
Channels	Color for the line and dots: -Light mode: #505050 -Dark mode: Cyan (#6ce5e8)

Task	Identify peaks

C. Vehicles Most Involved in Accidents

Idiom	Bar Chart
Data	X-Axis: Numerical value of accidents recorded Y-Axis: Type of vehicle involved in accident (Separated into x amount of bars)
Channels	Color for the bars: -Light mode: #505050 -Dark mode: Light Neon Pink (#ffe8ff)
Task	Analyze involvement

D. Vehicle Involvement in Accidents by Class of Vehicle

Idiom	Pie Chart
Data	Classification of vehicle/s recorded in an accident (Separated into x amount of partitions)
Channels	Colors to be used: -Light Mode: ● #6A5ACD ● #FF8C69 ● #4682B4 -Dark Mode: ● #FF6EC7 - 2-Wheeler ● #6EC1FF - 4-Wheeler ● #FFA36C - Larger Vehicles
Task	Compare involvement

E. Spatial visualization of the recorded accidents

Idiom	Spatial Map
Data	Location where vehicular accident/s occur
Channels	Color for the entries: #2881CA

Task	Locate hot-spots

B. Interactivity Justification and Techniques:

All visualizations of this project are mainly filtered by the optional filters featured on top of the dashboard. The “Time” filter consists of four buttons, where clicking one button filters the data into certain timeframes during a 24 hour period. The “Month” filter is a drop-down menu, consisting of the 12 months that the user is allowed to filter the data by.

We can argue that all visualizations are interconnected with one another, since a change in a single filter above the dashboard has the capability of affecting the amount of data entries each visualization will take into consideration. For example, if we were to filter the time by morning, the dashboard and spatial map will only take into consideration data entries that fit the category. Therefore, the visualizations turn into "Total Recorded Accidents in the Morning", "Vehicle/s Most Involved in Accidents in the Morning", "Total Recorded Accidents per Month in the Morning", etc. Additional application of the month filter further specifies the data based on the details of accident reports that the user wants to idealize.

What we can infer from this is that the visualizations provided can be described as the “characteristics” of our dataset and the filters that we apply affect these so-called characteristics based on the specifications that the user wishes to apply. An exception to this general interactivity can be seen in the bar chart of total recorded accidents per month. As this chart will always organize the data by the month of its recording, this chart only gets affected by the time filter.

As stated above, the main source of interactivity with this application will come from the filters provided. These filters help specify the information provided such that it meets what the user aims to learn about. Other notes of interactivity can be seen on the top-right corner of the dashboard, where it contains a button that controls how the dashboard is displayed. This button controls whether the user prefers to use the dashboard with a lighter interface, or a darker one such as provided in the prototype.

Another source of interactivity comes from the map overview that is provided. Users are able to zoom in and zoom out of the map, wherein data can cluster into certain areas if zoomed out a certain amount whilst data can also disperse into specific locations if the map is zoomed in enough. These clusters are differentiated by normal data entries by a modification on the hue/saturation of the mark based on the number of reports in a single cluster as well as a numerical value on the mark itself. This interaction can help the user generalize the data based on the location of accidents or it can help make the data’s location more specific, depending on what the user wants to visualize. This kind of

interaction aids in the project by allowing the user to determine the frequency of vehicular accidents in locations regardless whether they want to observe a general area or a certain location. If the user wishes to zoom in on a specific location, they'd be able to see the amount of data entries. On the other hand, general areas show frequency of vehicular accidents by the hue/saturation change in the mark, as well as the numerical value of the cluster.

C. Potential Modifications

Any potential modifications that may possibly be put into consideration can come from the following bullet points:

- A change in the channel/colors of the graphs
- A modification in the pre-existing filter options
- Additional filter options for the dashboard
- Changing how the data entries are presented on the spatial map

III. Final Project

A. Exploratory Data Analysis and Dataset Preprocessing

Data accuracy in the project proposal phase was described to be accurate in terms of the longitudinal and latitudinal data. However, it would not hurt to implement exploratory data analysis to further examine the dataset and clean up on some of its irregularities.

One of the particular details that users may want to see is the distribution or frequency of the involvement in accidents by each classification of vehicle. As the size of road vehicles range from scooter-type motorcycles to significantly larger and heavier cargo trucks, it would be beneficial to the users to understand how often accidents may occur with their respective vehicles in comparison to other vehicles.

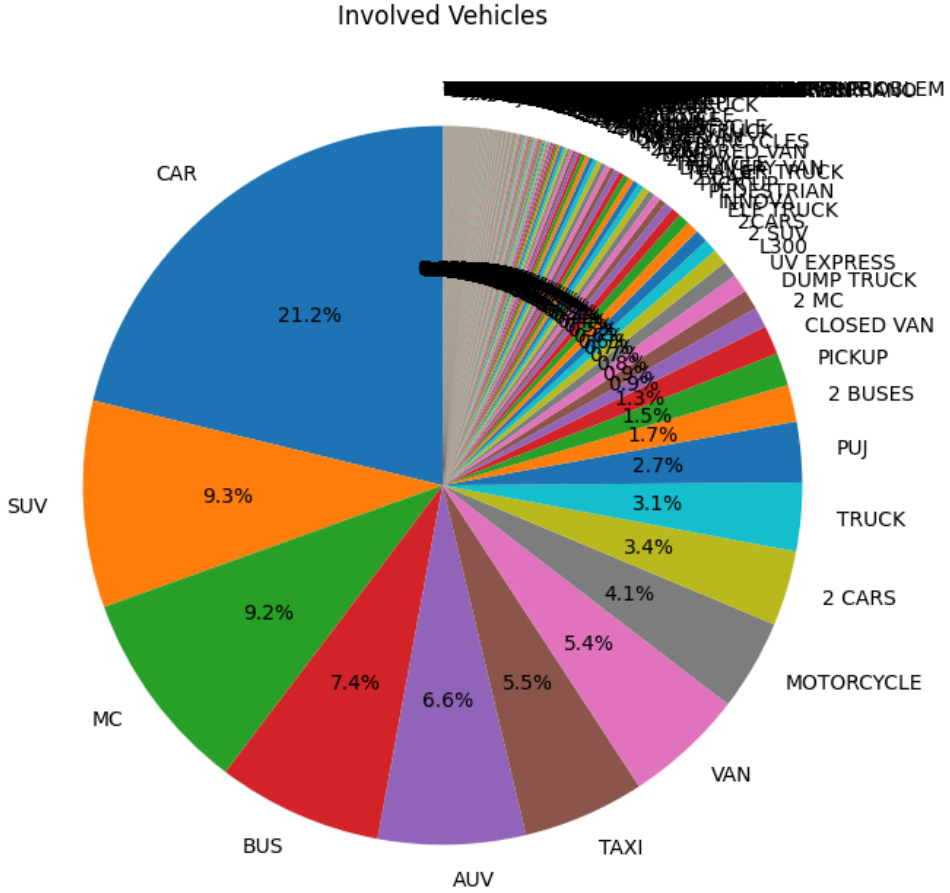


Figure III: Pie chart of vehicle involvement in accidents based on the "Involved" column of the dataset

From the initial chart, we can clearly see a significant quantity of unique reports that makes the use of a pie chart ill-advised. This is due to the uniqueness of the reports that contain two or more vehicles. Also seen in the chart is the inconsistent categorization of vehicles reported. SUVs and AUVs that fall under the category of Cars are separated from their general category. There is also the separation in distribution of MC and Motorcycles, which should only fall under the same group. Whether these are due to inconsistencies in the production of the automated or manual reports, it would be best to categorize the involvement into the most concerned categories. The final categories of such vehicles involvement, upon consideration, were organized into the following:

Cars, Motorcycles, Buses, Vans, Taxis, Trucks, Jeeps, and Others

**Upon further consultation of the project the pie chart was to be changed into a horizontal bar chart for better comprehension.*

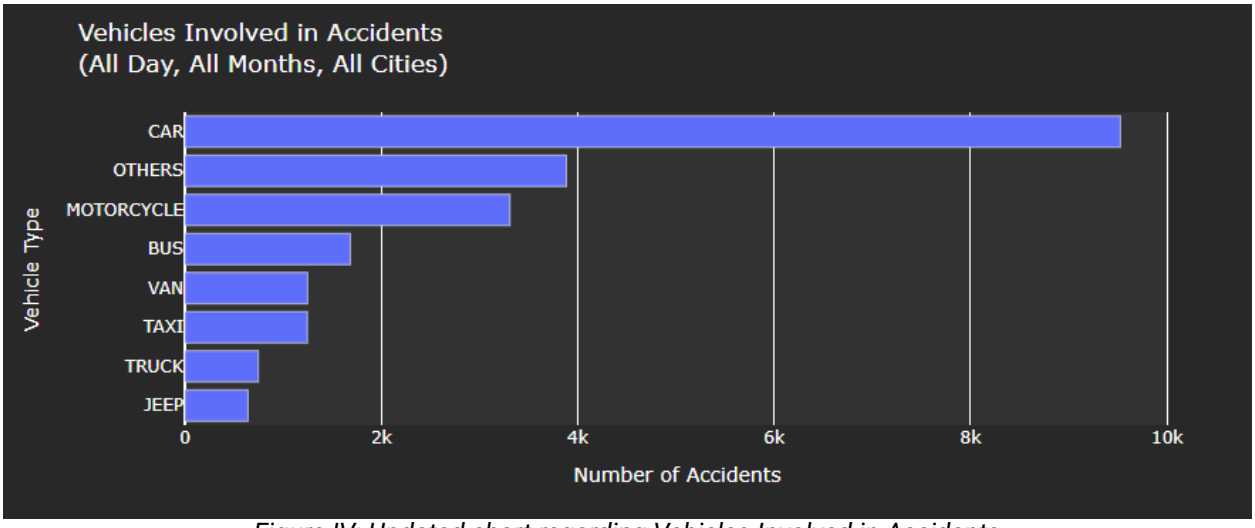


Figure IV: Updated chart regarding Vehicles Involved in Accidents

Upon observation of the dataset, we can also conclude that the columns for City, Location, and Type of Accident are key components as well towards the creation of this project. Exploratory data analysis shows us that there are null values for some of these corresponding columns. We are to remove these as manual input of city and location would be too tedious of a process despite the accuracy of the data's coordinates. Furthermore, accidents reported without specific details are too difficult to verify one-by-one. It is also worth noting that there are data entries wherein the column for "Time" has a null value. We have decided to keep these, albeit they are only shown in the general overview of the map visualization, or rather when the "Time of Day" feature is filtered as "All Day".

```
# Check for nulls
df.isnull().sum()
```

	0
Date	0
Time	122
City	187
Location	23
Latitude	0
Longitude	0
High_Accuracy	0
Direction	857
Type	57
Lanes_Blocked	687
Involved	432
Tweet	0
Source	0

dtype: int64

Figure V: the amount of null values detected per column of the dataset

The main vision of the project was to help other people visualize vehicular accidents within Metro Manila. However, we must put into consideration how the definition of vehicular accidents may differ from one person to another. For one person, it may be defined as any incident that negatively affects the flow of road traffic. Others may define this as incidents that occur due to the misjudgement of fellow road users. For this project, we have decided to align with the latter definition. Hence, we have decided to filter the data further such that each data entry present clearly and precisely fits the definition of a vehicular accident.

```
[ ]  # See all values in Type
     df['Type'].unique()
```

```
'STALLED CLODE VAN DUE TO MECHANICAL PROBLEM',
'STALLED MILITARY TRUCK', 'HIT AND RUN ACCIDENT',
'ON GOING MANILA WATER LEAK REPAIR',
'STALLED DELIVERY TRUCK DUE TO MECHANICAL PROBLEM',
'STALLED TANKER DUE TO FUEL LEAK',
'STALLED WING VAN TRUCK DUE TO MECHANICAL PROBLEM',
'VEHICULAR ACIDENT',
'STALLED DELIVER BAN DUE TO MECHANICAL PROBLEM',
'STALLED AUV DUE TO EMPTY FUEL',
'STALLED CAR AND MC DUE TO MECHANICAL PROBLEM',
'STALLED DELIVERY VAN DUE TO MECHANICAL',
'STALLED AUV DUE TO MECHANICAL',
'STALLED MIXER DUE MECHANICAL PROBLEM',
'STALLED TRAILER BUS DUE TO MECHANICAL PROBLEM',
'STALLED SCHOOL BUS DUE TO MECHANICAL PROBLEM',
'SUSPECTED DEAD PERSON', 'STALLED BUS DUE TO TIRE PROBLEM',
'STALLED WATER TANKER DUE TO MECHANICAL PROBLEM',
'VEHICHULAR ACCIDENT', 'VEHICULAR FIRE INCIDENT',
'STALLED CLSED VAN DUE TO FLAT TIRE',
'STALLED CLOSED VAN DUE TO STUD PROBLEM',
'STALLED ELF TRUCK DUE TO STUD PROBLEM', 'STALLED WING VAN',
'STALLED CAR DUE TO DOUBLE FLAT TIRE',
'STALLED MINI DUMP TRUCK DUE TO FLAT TIRE',
'STALLED CLOSE VAN DUE TO BATTERY PROBLEM', 'MISALIGNED BARRIER',
'STALLED CLOSE VAN DUE TO MECHANICAL TROUBLE',
'STALLED TRAILER TRUCK  DUE TO MECHANICAL PROBLEM',
'VEHICULAR ACCCIDENT',
'STALLED TANKER TRUCK  DUE TO MECHANICAL PROBLEM',
'STALLED TRUCK DUE TO A FLAT TIRE',
'ONGOING DPWH BEAUTIFICATION PROJECT',
'STALLED L300 VAN DUE TO MECHANICAL PROBLEM',
'ONGOING MMDA TEC LANE MARKING',
'INSTALLATION OF COLORED WATER GALLON FOR BIKE LANE',
'ONGOING DPWH ASPHALTING', 'STALLED MOTORCYCLE DUE FLAT TIRE',
'ONGOING MOTORCADE', 'STALLED COASTGUARD BUS DUE TO FLAT TIRE',
'QUEZON CITY LGU ONGOING RFID STICKER INSTALLATION',
'MMDA TEC ONGOING BARRIER MAINTENANCE WORK',
'MMDA MPCG ONGOING MRT WALL PAINTING',
'STALLED TRUCK DUE A FLAT TIRE', 'STALLED VAN DUE TO A FLAT TIRE',
'STALLED TRUCK DUE TO  MECHANICAL PROBLEM',
'STALLED PUJ DUE TO A FLAT TIRE',
'STALLED TRUCK  DUE TO FLAT TIRE',
'STALLED TRUCK DUE TO MECHANICAL  PROBLEM', 'A TREE HAS FALLEN',
```

Figure VI: Unique values obtained from the type of accident reported

For context, the original dataset contained 17313 entries. An estimated 3900 entries revolve around vehicles stalling due to mechanical problems. A few other incidents reported in the dataset include ongoing projects of the Department of Public Works and Highways (DPWH) such as concrete blocking, lane marking, road asphalting/patching, and more. While such incidents may affect the flow of traffic comparable to the effects of vehicular accidents, these events can't be classified as accidents based on our criteria and definition. Hence, in order to make the dataset as conducive towards the goal of this project, the only incidents that remained were filtered with the use of words such as "collision" and "accident". This does not significantly affect the integrity of our dataset as it still leaves us with 12633 data entries.

The final note that we would like to mention from the EDA is that the date range of the dataset starts from August 20, 2018 up until December 27, 2020.

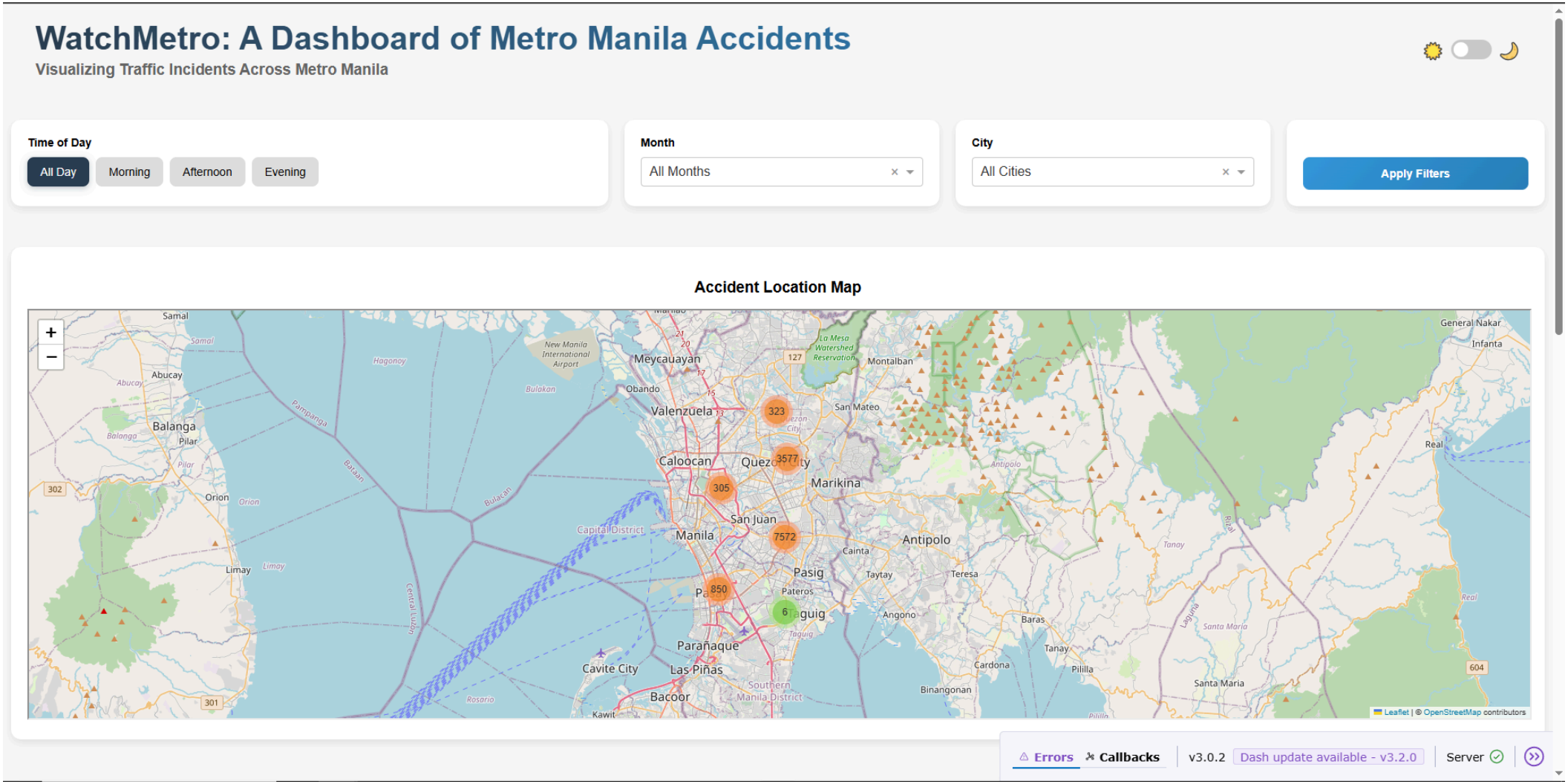


Figure VII: Part 1 of the dashboard application

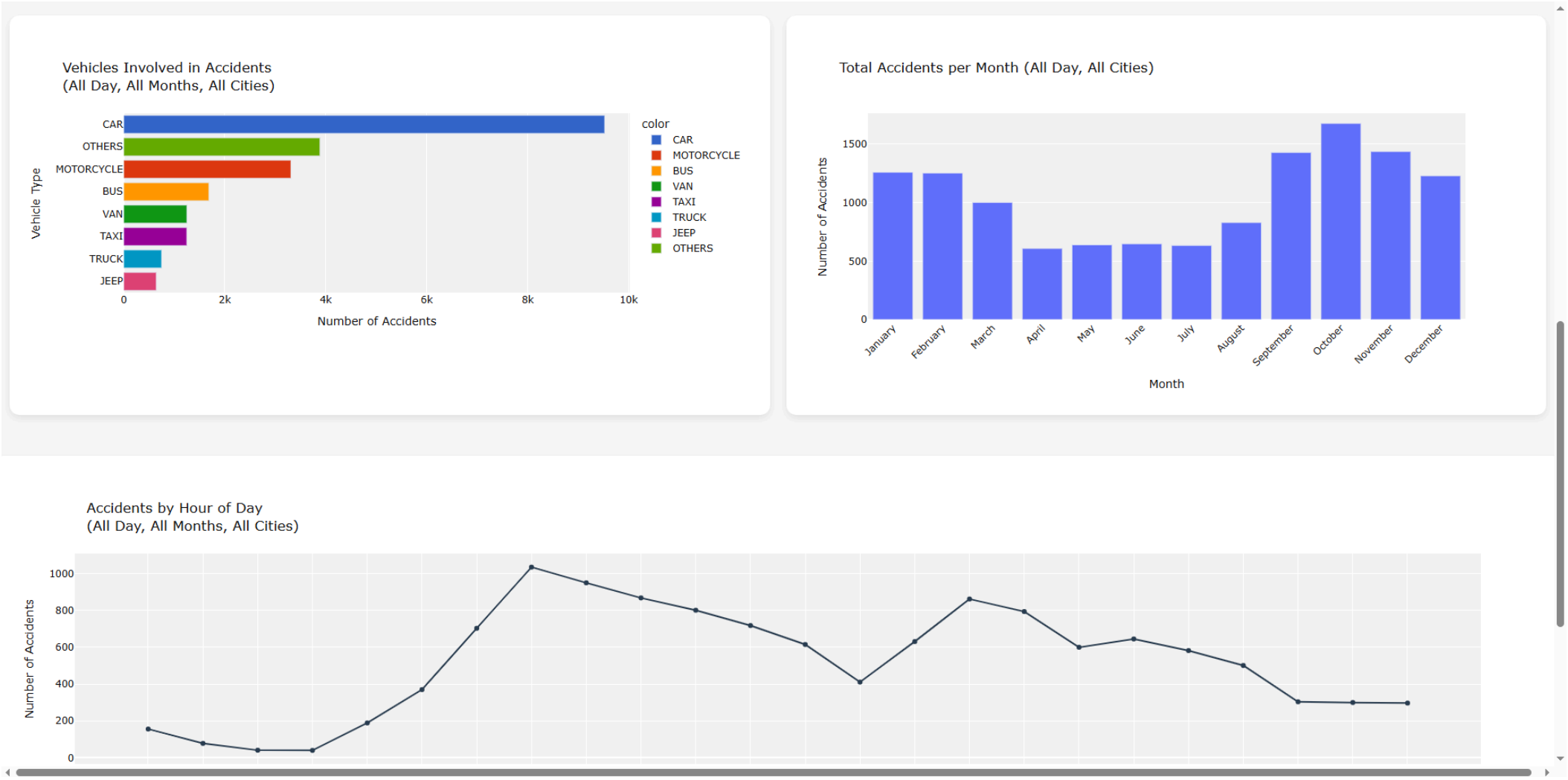


Figure VIII: Part 2 of the dashboard application

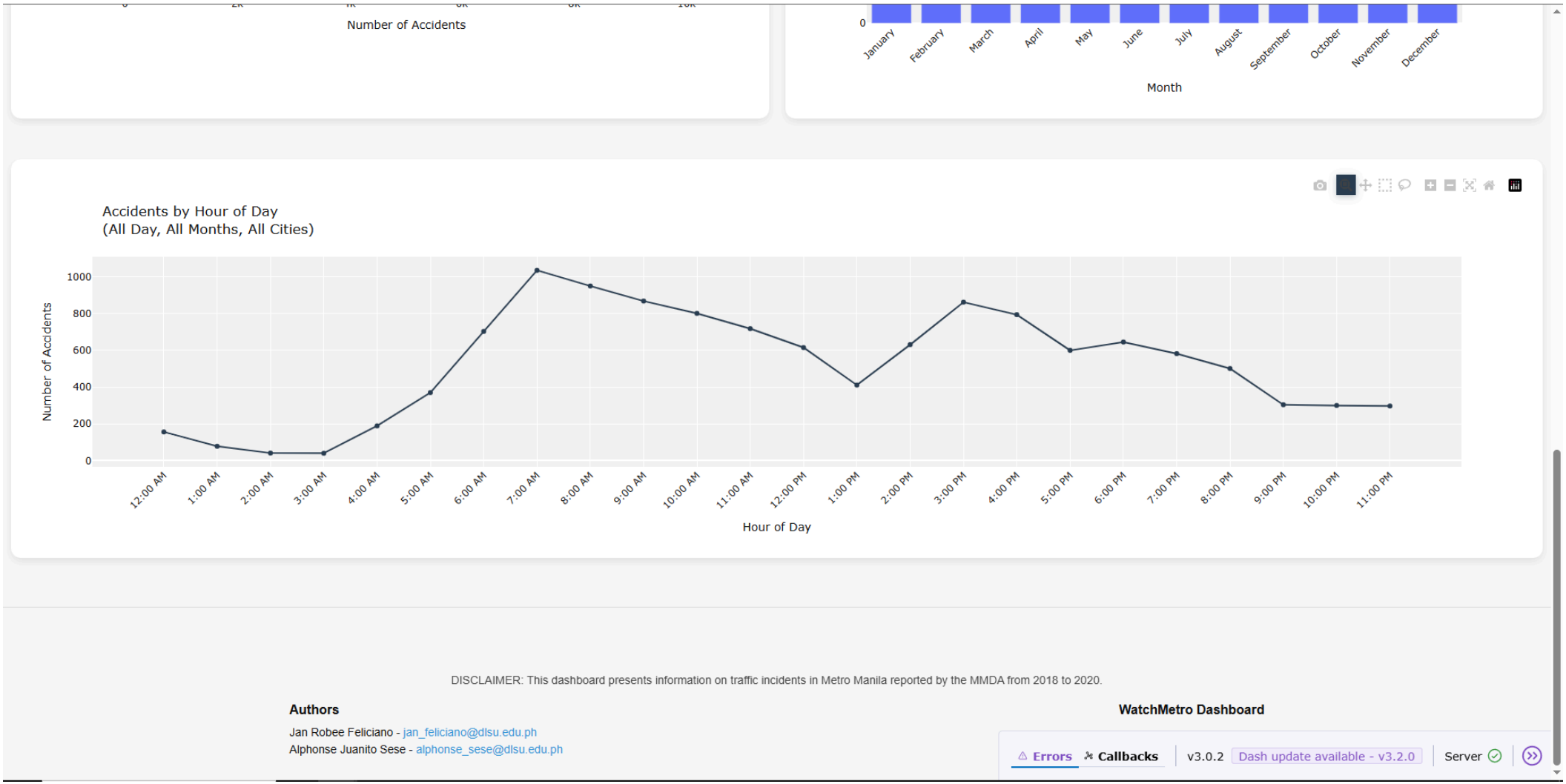


Figure IX: Part 3 of the dashboard application

B. Final Output

Production of the project was successfully completed and the application functions as it was intended to. A few notes that we would like to add, including from what can be observed in the screenshots of the application above, are the following:

1. As one zooms in and out of the spatial visualization, one may observe the grouping and breaking of several clusters of data. This is dependent on how zoomed in/out the person sets the visualization at. The colors of these clusters are dependent on the number of entries contained within each cluster. The more entries that are clustered into a single group, the darker or more intense the color of the cluster becomes. It can also be observed that the fewer entries there are in a group, the lighter the color of the cluster appears, as can be seen from the group of 6 data entries in Figure VII around the Taguig City area.

2. The first filter for the data is titled “Time of Day”. It allows the user to filter the data whether they want to see vehicular accidents that happen in the morning, afternoon, evening, or within the dataset’s entirety. The criteria for the morning filter is reports that happen from 06:00 am-11:59 am; The criteria for the afternoon filter is reports that happen between 12:00 pm-05:59 pm; And the evening filter gathers data that was reported either from the time range of 06:00 pm-11:59 pm or between 12:00 am-05:59 am. This filter affects all 3 charts below.

3. The “Month” filter simply filters the data by the user’s month of choosing. This only affects the horizontal bar chart of *Vehicles Involved in Accidents* and the Dot/Line Graph of *Accidents by Hour of Day*, as the vertical bar chart generalizes the monthly data based on time of day and city of choosing.

4. The “City” filter is similar to the month filter, however it filters by city of the user’s choosing. Furthermore, this filter affects all 3 charts located in the dashboard. It is worth noting that not all cities are included here, as the dataset is restricted to major roads of Metro Manila that were arbitrarily selected by the dataset’s author. Hence, there are no accidents reported for cities such as Las Piñas and Muntinlupa.

5. Once the desired filters are selected, clicking apply filter regenerates the spatial map. Instant regeneration upon filter selection was not implemented due to impracticality in performance should the user desire to apply multiple filters.

6. The first chart provided is the horizontal bar chart of *Vehicles Involved in Accidents*. As mentioned in the EDA and preprocessing part of the paper, the pie chart was to be changed upon initial consultation. Instead, to better visualize the data, this chart was implemented. Furthermore, the colors corresponding to the distributions of the pie chart were substituted here to allow the user to better distinguish the data distribution between the vehicle groups. This chart is affected by all three filters.

7. The second chart is a vertical bar chart of *Total Accidents per Month*. This chart is filtered by the time of day and city filter, as the graph in itself is already a distribution of the data by month.

8. The third chart is a dot/line graph of *Accidents by Hour of Day*. This graph helps track the amount of vehicular accidents that occur during specific time intervals. The graph helps identify peak times or time intervals wherein vehicular accidents are most likely to occur. This graph is filtered by all three filters.

9. A convenient quality of life option is present on the top-right corner of the application. Clicking on this button lets the user choose whether they want to use the application in light mode,or in dark mode. The figures provided above are what the application appears in light mode.