Why is Vietnam's traffic so scary to foreigners?

Hung Nguyen (1022029)

I. Motivation

Vietnam's cities are growing fast, driven by economic development and urbanization. This rapid expansion has placed a burden on the road systems, resulting in heavy traffic congestion. The dominance of motorbikes, a seemingly symbol of Vietnamese culture, further adds to the dynamic and complex nature of the traffic situation. Additionally, many drivers in Vietnam are often perceived as having poor driving norms, some are even described as reckless. This creates an environment that, while fascinating to observe [1], can feel chaotic and even scary to outsiders navigating the streets for the first time.

To help the audience understand more about the situation, this project aims to demonstrate the traffic situation and its characteristics using an interactive data dashboard (**find the source and the final dashboard in Appendices section**). This dashboard contains visualization and in-depth analysis of the traffic conditions and related causes.

Overall, the main motivation is to inspire discussions about finding better ways to manage traffic, make roads safer, and improve transportation for everyone. Ultimately, this project hopes to achieve this goal by driving efforts to create solutions that help cities, not just those of Vietnam, to be more sustainable, smarter, and more comfortable to live in.

II. Expected Results

Throughout the project, we are guided by the following questions:

- What is the general perception of outsiders to the current traffic situation?
- How bad is the traffic situation in Vietnam in truth?
- What are the main causes of the current traffic situation?

By addressing these questions, this project aims to help the audience better understand the awareness surrounding traffic issues in Vietnam, the key challenges, their root causes, the unique nature of traffic in the country, and the answer to the question, "Why is Vietnam's traffic so scary to foreigners?"

III. Themes

This section explores the possible structure of themes and subthemes to be included in the visualization, which will be important in guiding the methodology for the rest of the project.

Based on the questions defined previously, we can define a rough flow of the story for our project.

What is the general perception of people of Vietnamese traffic? To answer this question, we can extract the overall sentiment of social media users on the traffic situation in Vietnam using natural language processing (NLP) techniques. Investigating the perception of foreigners on Vietnamese drivers and traffic helps the audience to have a clear picture of the situation, basically answering the "what?"

How bad is the traffic situation in Vietnam in truth? From that, we try to address the "How?" by analyzing traffic data, such as congestion patterns and accident rates to provide an objective view of the scale of the problem. This allows us to move beyond perception and give the audience a clear, data-driven understanding of just how challenging the traffic situation is in Vietnam.

What are the main causes of the current traffic situation? Finally, we address the "Why?" - the causes of the issues. To achieve this, we explore the issues related to motorbikes and high population density. Additionally, we can talk briefly about the nature of the drivers to understand how the human factor plays a big role in worsening the perception of foreigners to Vietnam's traffic.

Based on this narrative, the project is divided into these themes structure:

1. Foreigners' Overall Sentiment on Vietnamese Traffic
2. The Current Traffic Situation
    a. Congestion Visualization
    b. Traffic Accidents
3. Causes
    a. Motorbikes
    b. Population Density
    c. The Human Factor

Each theme is further divided into sub-themes that include detailed explanations and the data/tools used for analysis in the following section.

IV. Approach

This section presents the methodology to address the subthemes. For each subtheme, we detail the **Data**, **Tools**, and **Interactions** of the visualizations associated with it. Based on the narrative defined in the previous section, we have the following structure:

1. **Foreigners' Overall Sentiment on Vietnamese Traffic:**
   The goal is to help the audience see whether the overall negative sentiment about Vietnam's traffic is valid. It also helps to understand the source of problems.
    a. **Visualize sentiment of foreigners on Vietnamese's traffic through discussions:**
    Data:
    - Source: Reddit threads with discussions about Vietnam's traffic [2-14].

- Contains 1690 comments and posts to be reviewed.
- Note: The data was obtained by searching using neutral keywords on the Google search engine, avoiding biasing the results to negative threads.

Tools:

- Use Python Reddit API Wrapper (PRAW) [15] to get comments from threads.
- Obtain the most common n-gram terms (bigrams and trigrams).
- Perform sentiment classification (negative/positive) using VADER [16].
- Perform topic modelling using K-Means and Hierarchical clustering. Clustering results are evaluated using the Calinski-Harabasz index.
- Use Word Cloud to visualize common n-gram terms.

Interactions: The user will be able to explore common keywords, view sentiment analysis, and interact with topic clusters by hovering over the data points to better understand public perception of Vietnam's traffic situation.

2. **The Current Traffic Situation:**
   The goal is to help the audience understand the situation behind Vietnam's traffic better - *How is it bad?*
   a. **Congestion Visualization:** Create dynamic visualizations showing the cumulative traffic congestion in cities: traffic density heat maps of different hours of a typical day.

      Data:

      - Source: Kaggle [17] - latitude and longitude of traffic status for a segment reported at different dates and times of Ho Chi Minh city - one of the busiest cities of Vietnam.
      - Provider is Research group from HCM University of Technology, which implies high quality and completeness.
      - The data is very complex, containing multiple datasets mapping different aspects and requires a lot of preprocessing.

      Tools:

      - Use Python to preprocess the data by mapping nodes (representing a driver at a certain latitude and longitude) to their status.
      - Then, we can compute cumulative sums of the traffic of 30-minute periods from the beginning to the end of the day.
      - We can then detect traffic jams by dividing the current speed of the node by the maximum speed (capped at 60 km/h, typical speed limit in Vietnam) to determine traffic jams at that current timestamp.
      - The map data is visualized using Folium [18].

      Interactions: The users can change the timestamps to get the traffic density at that certain point. They can also hover over data points to see the speed at which vehicles are moving, which helps identify the traffic jams.

      Data: Images/videos of the typical traffic congestion in busy cities of Vietnam.

      Tools: Images are available on Google, and are simply rendered out.

      Interactions: Users can click through images to view different scenes of typical traffic congestion in busy Vietnamese cities.
   b. **Traffic Accidents:** Create a visualization displaying the traffic accidents and casualties happening in Vietnam on a time scale.

      Data: Statista [19], Statista [20] - the annual number of deaths/accidents caused by

traffic accidents.

Tools: Plot the accidents and casualties through each year using interactive line charts using Plotly.

Interactions: Users can hover over the line chart for detailed data points and zoom in on specific years to understand trends in accidents and casualties over time.

3. **Causes**

   The goal is to help the audience understand the possible causes behind the chaotic traffic in Vietnam.

   a. **Motorbikes:**

      i. **Why are motorbikes a problem?**

         When there is a high proportion of small vehicles like motorcycles, they can have a more significant influence on road congestion than bigger vehicles. This is because more drivers create more problems, as they tend to weave and cut others, creating more complex traffic patterns.

         Data: Display video simulation [21] to understand how traffic jams are formed.

         Tools: Simple video player, rendered using iframe.

         Interactions: The audience can play the video to visualize how motorbikes contribute to traffic congestion and observe how drivers' movements create complex patterns on the road. They can pause, rewind, or fast-forward the video to focus on specific moments.

      ii. **Ownership of motorbikes:**

         Data: Statista [22] - the percentage of preferred transportation modes in Vietnam.

         Tools: Plot the percentage of people preferring each vehicle with an interactive bar chart using Plotly.

         Interactions: Users can hover over the bar chart to view the exact percentage of people preferring each transportation mode to understand the dominance of motorbikes.

         Data: Statista [23] - countries with the highest ownership of motorbikes in Asia.

         Tools: Plot the percentage of households owning motorbikes for each country with an interactive bar chart using Plotly.

         Interactions: Users can hover over the bar chart to view the exact percentage of households owning motorbikes in each country.

   b. **Population density:** Display population density in certain areas.

      Data:

      - Source: WorldPop [24] - spatial distribution of Vietnam population in 2020.
      - The data is gathered by multiple research groups, which implies high quality.

      Tools:

      - Because there can be data points, we want to perform interpolation to smooth the density into larger regions. This was done using h3-py [25].
      - The data is mapped into a JSON-like format, which is plotted into a choropleth mapbox using Plotly.

      Interactions: Users can hover over each of the hexagonal regions to view the population density in that area. They can zoom in and out of the map to explore

different regions of Vietnam and analyze how population density varies across the country

c. **The Human Factor:**
Vietnamese drivers often exhibit poor driving norms, with behaviours like road rage, running red lights, and ignoring pedestrian right-of-way, particularly among motorcyclists. This reckless driving is worsened by the traffic structure, where motorcyclists weave between lanes or go against traffic to avoid congestion and a lack of road safety education that fosters an "everyone-for-themselves" mentality. The heavy reliance on motorbikes, which allow for more flexible movement, further encourages risk-taking behaviours as drivers navigate through tight spaces or accelerate past red lights to save time.
Data: Video evidence [1] from Youtube.
Tools: Simple video player, rendered using iframe.
Interactions: The audience can play the video to observe real-life examples of reckless driving behaviours, such as weaving through traffic and running red lights. They can pause, rewind, or fast-forward the video to focus on specific moments.

To put things together into a complete dashboard, the plan is to build a web page using Streamlit [26], which functions like a slideshow but with interactive plots to allow more diverse visualizations & interactions, which can potentially drag the audience's attention better. The website should have basic interactions (depending on each visualization, as described above) and minimal annotation text for context. This should allow the audience to explore and understand the data themselves even without a presenter.

V. Usability

I believe that the urban researchers and the general audience will gain a comprehensive understanding of Vietnam's traffic situation through the interactive dashboard. Using various charts and maps, they will learn how traffic congestion fluctuates across different regions and how these patterns are influenced by factors like motorbike ownership, population density, and road safety norms. Users can also analyze the impact of traffic accidents and fatalities to understand the importance of driving safely.

I believe that when performing tasks such as exploring congestion patterns over time or analyzing traffic trends alongside demographic information, researchers will be able to identify key areas where interventions may be most effective. These tasks will help them to identify critical areas, such as regions with high accident rates or areas with insufficient traffic education.

In addition, I aim for my tool to be engaging for general users interested in learning more about the traffic situation in Vietnam. This project can be taken further by integrating with API endpoints to be updated daily. I will know our tool is effective when we see frequent engagement with features such as exploring specific traffic patterns, interacting with data visualizations, and obtaining feedback/discussions from users on their ability to make better decisions or policy recommendations based on their findings.

VI. Time Relevance

Time is an important aspect of the project. It is always important to observe the data over the temporal aspect, which is an effective communicator of changes, trends, and fluctuations, which are essential for discovering patterns in data.

In this particular topic, by integrating temporal aspects into information visualization, we can observe how traffic patterns evolve, revealing key insights into congestion, flow dynamics, and behavioral changes of drivers across different times of the day. This also can help to highlight critical points in data (such as when traffic bottlenecks form) over a long and complex timestamp. Here, I select the following aspects to be visualized with respect to time.

- Traffic congestion: Visualizing fluctuations over a typical weekday/weekend can demonstrate the critical times when congestion is highest. With an animated chronological map, we can effectively understand how congestion builds up and highlights bottlenecks in traffic.
- Traffic accidents: By visualizing accident frequency and location over time, we can identify patterns related to specific times of day or days of the week when incidents are more frequent, potentially highlighting the correlation between this and traffic congestion. One can also see whether there has been a change in the system that leads to fewer traffic accidents by observing this data over a larger relative time span.
- Production of motorbikes: With demand comes supply, so visualizing the production volume helps us to understand the consumption level of motorbikes in Vietnam over time. This visualization can highlight how increases in motorbike usage relate to congestion issues, especially in densely populated urban areas.

It is also possible to combine temporal data with interactive plots, which can be an even more effective communicator. This should be designed with a focus on user interaction with temporal data. For example, by implementing time-based sliders or interactive timelines, users can adjust the time frame and view traffic data at specific moments, allowing them to observe traffic intensity at a chosen point in time.

VII. Space Relevance

Spatial factors are crucial in understanding Vietnam's traffic dynamics, as location heavily influences congestion, accidents, and flow. For instance, by mapping population density heatmaps, we can identify areas where chaos is most prominent. Identifying such urban centres, where narrow streets and high motorbike usage, can help to identify sources of congestion. Similarly, traffic congestion can be mapped to pinpoint critical chokepoints, especially during peak hours.

These can be done using choropleth map boxes and color mapping to indicate densities. These maps can allow users to gain knowledge of the board picture and identify critical points in the data with *pre-attentive processing*, a very important aspect of visualizations. The map also allows zooming in for users to explore certain regions on the map. Upon hovering

the data points, a tooltip with more specific information will appear to inform the users of the exact details.

Integrating spatial data in interactive maps or filters enhances analysis and information clarity, providing a quick and deep understanding of traffic patterns for users at a glance.

Appendices

- **Visualization Dashboard:** https://vietnamese-traffic.streamlit.app/.
  Please contact me at hung.6.nguyen@aalto.fi immediately if the website is not up - sorry for the inconvenience.
- **Source Code:** https://github.com/Hungreeee/Explorative-Viz-Project

References

[1] WorldAviation 4K, "Crazy yet Organized Traffic in Vietnam Ho Chi Minh," YouTube, Nov. 3, 2024. [Online]. Available: https://www.youtube.com/watch?v=fm0cX7P6PSw. [Accessed: Nov. 3, 2024].

[2] kudivn, "I'm fed up with Vietnam traffic," Reddit, [Online]. Available: https://www.reddit.com/comments/bviwz5. [Accessed: Nov. 3, 2024].

[3] Lancescs, "Vietnam traffic," Reddit, [Online]. Available: https://www.reddit.com/comments/10vokml. [Accessed: Nov. 3, 2024].

[4] Duytune, "What do you think of Vietnam's traffic?," Reddit, [Online]. Available: https://www.reddit.com/comments/uuw4sl. [Accessed: Nov. 3, 2024].

[5] zachwilly, "Absolutely in awe of Vietnamese driving," Reddit, [Online]. Available: https://www.reddit.com/comments/16sjcbv. [Accessed: Nov. 3, 2024].

[6] minhduc24, "Will Vietnam ever become civilized in traffic," Reddit, [Online]. Available: https://www.reddit.com/comments/1csgm19. [Accessed: Nov. 3, 2024].

[7] *Unknown*, "Is anyone rethinking Vietnam traffic culture now that cars are everywhere?," Reddit, [Online]. Available: https://www.reddit.com/comments/1eojg9h. [Accessed: Nov. 3, 2024].

[8] tientutoi, "Vietnam has a relatively low traffic casualty rate in Southeast Asia, ranking better than many other countries in the region.," Reddit, [Online]. Available: https://www.reddit.com/comments/1apjlja. [Accessed: Nov. 3, 2024].

[9] *Unknown*, "Why are Vietnamese drivers such cn#ts?," Reddit, [Online]. Available: https://www.reddit.com/comments/18ms0iq. [Accessed: Nov. 3, 2024].

[10] strokkur66, "Comparing traffic in Thailand and Vietnam," Reddit, [Online]. Available: https://www.reddit.com/comments/187bmst. [Accessed: Nov. 3, 2024].

[11] PrincipleLazy3383, "Why is Vietnamese driving so bad?," Reddit, [Online]. Available: https://www.reddit.com/comments/159x06b. [Accessed: Nov. 3, 2024].

[12] Darkmaster85845, "I just simultaneously saw and had my first traffic accident in Vietnam," Reddit, [Online]. Available: https://www.reddit.com/comments/1at29fj. [Accessed: Nov. 3, 2024].

[13] mojotarts, "Vietnam doesn't deserve the convenience of it's traffic lights," Reddit, [Online]. Available: https://www.reddit.com/comments/17hfqfb. [Accessed: Nov. 3, 2024].

[14] Dull-Two-8962, "Is nearly getting hit by a vehicle in vietnam a daily occurrence?," Reddit, [Online]. Available: https://www.reddit.com/comments/1eqmzno. [Accessed: Nov. 3, 2024].

[15] B. Boe, "PRAW: The Python Reddit API Wrapper," GitHub repository, [Online]. Available: https://github.com/praw-dev/praw. [Accessed: Nov. 3, 2024].

[16] C. Hutto and E. Gilbert, “VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text”, ICWSM, vol. 8, no. 1, pp. 216-225, May 2014.

[17] HCM University of Technology, "Traffic Flow data in Ho Chi Minh City, Viet Nam," Kaggle, [Online]. Available: https://www.kaggle.com/datasets. [Accessed: Nov. 5, 2024].

[18] R. Story, "Folium: Python Data, Leaflet.js Maps," GitHub repository, version 0.18.0, [Online]. Available: https://github.com/python-visualization/folium. [Accessed: Nov. 8, 2024].

[19] Number of deaths caused by traffic accidents in Vietnam from 2013 to 2023 (in 1,000s) [Graph], General Statistics Office of Vietnam, June 29, 2024. [Online]. Available: https://www.statista.com/statistics/986123/vietnam-number-deaths-traffic-accidents/. [Accessed: Nov. 6, 2024].

[20] Number of traffic accidents in Vietnam from 2013 to 2023 (in 1,000s) [Graph], General Statistics Office of Vietnam, June 29, 2024. [Online]. Available: https://www.statista.com/statistics/986115/vietnam-total-annual-traffic-accidents/. [Accessed: Nov. 6, 2024].

[21] J. Miller, "How Traffic Jams Are Formed," YouTube, Nov. 3, 2024. [Online]. Available: https://www.youtube.com/watch?v=Sy9v6U63Nd0. [Accessed: Nov. 3, 2024].

[22] Leading modes of transportation among people in Vietnam in 2023 [Graph], Standard Insights, April 5, 2023. [Online]. Available: https://www.statista.com/statistics/1337061/vietnam-most-used-modes-of-transportation/. [Accessed: Nov. 6, 2024].

[23] Share of households owning a motorcycle in Asia as 2023, by country [Graph], The Thaiger, February 24, 2023. [Online]. Available:

https://www.statista.com/statistics/1373225/asia-household-share-owning-a-motorcycle-by-country/. [Accessed: Nov. 6, 2024]

[24] WorldPop, School of Geography and Environmental Science, University of Southampton; Department of Geography and Geosciences, University of Louisville; Departement de Geographie, Universite de Namur, and Center for International Earth Science Information Network (CIESIN), Columbia University, "Global High Resolution Population Denominators Project," 2018. [Online]. Available: https://dx.doi.org/10.5258/SOTON/WP00671. [Accessed: Nov. 11, 2024].

[25] Uber Technologies, "h3-py," GitHub repository, 2024. [Online]. Available: https://github.com/uber/h3-py. [Accessed: Nov. 11, 2024].

[26] Snowflake Inc., "streamlit," GitHub repository, 2024. [Online]. Available: https://github.com/streamlit/streamlit. [Accessed: Nov. 3, 2024].