TravelAware: Your Commute Safety Assistant

Pranav Agarwal
pranava2@uci.edu

David Assio
dassio@uci.edu

Serena Elizabeth Alvarez
sealvare@uci.edu

Satvik Bhatnagar
satvikb@uci.edu

Sebastian Algharaballi-Yanow
salghara@uci.edu

Samar Vikas Save
svsave@uci.edu

Abstract

TravelAware is a navigation application that integrates crime data with route planning to provide safer travel options through Los Angeles County. While traditional navigation systems solely focus on minimizing travel time and distance, TravelAware incorporates historical crime statistics to generate route recommendations that balance safety and efficiency based on user preference. This application processes Los Angeles crime data to calculate risk scores for potential routes while considering user demographics and time-of-travel in its analysis and presents it in an intuitive web application that allows adjustment based on user preference between faster and safer routes. We demonstrated TravelAware's effectiveness through practical use cases and discussed the technical implementation details as well as the ethical considerations regarding safety-based routing. Our work demonstrates the feasibility of incorporating public safety into navigation systems while understanding the complexities of implementing a technology responsibility in historical urban environments such as Los Angeles.

1 Overview & Project Goals

The navigation of dense urban areas such as Los Angeles County can be a challenge, especially to those unfamiliar with the area. While traditional navigation systems excel at finding the fastest routes, they overlook a critical concern for travelers: personal safety and the risk of encountering crime or accidents along the way.

TravelAware addresses this gap in navigation technology. Our product strengthens conventional routing by incorporating safety-based recommendations, using historical crime data from Los Angeles County to generate a "risk score" for each potential route. This score helps users understand the likelihood of encountering crime during their journey. Although TravelAware was originally designed to serve all travelers—whether on foot, by car, recreational vehicle, or public transportation—this initial iteration focuses specifically on travel by automobile.

2 Background Information

While vast databases of crime data exist online, it remains largely separate from everyday navigation tools. Current navigation systems prioritize speed and shortest distance, focusing almost exclusively on traffic conditions and time optimization. This overlooks the fact that many travelers would willingly accept a slightly longer journey in exchange for increased safety assurance. Despite the abundance of crime statistics available, navigation systems rarely incorporate these safety factors into their route calculations. This disconnect between the available safety data and route planning presents an untapped opportunity to enhance the travel experience - especially for those who prioritize personal security over saving a few minutes on their journey.

2.1 Dataset

Using the "Crime Data from 2020 to Present" dataset sourced from the Los Angeles Police Department (LAPD) and the City of Los Angeles, we developed the basis for our product. This dataset is a representation of the crime in the city of Los Angeles from 2020 to present and is a transcription from the original written crime reports. This dataset in particular interested us because it had exact geolocation data with both the longitude and latitude of the incident recorded and is updated biweekly. It must be noted that there are some location fields with missing data that are recorded at ((0°, 0°) that we removed before beginning.

For this dataset, we also make several assumptions to ensure that our data accurately reflects the crime occuring in the area and can provide a correct assessment of risk. First we assumed that this dataset is a reflection of the real-world and represents the entirety of the reported crime taking place in Los Angeles County. We additionally assumed that the longitude and latitude recorded is accurate to where the crime took place. Finally, we assume this dataset is not manipulated in any way from the real written crime reports by the LAPD.

2.2 Exploratory Data Analysis

As we started out EDA, we understood what the different variables stood for and their contribution to the dataset. We then identified the key variables in our dataset that would be crucial to the problem that we are trying to solve. We examined a number of crucial pieces of data that improved our understanding of Los Angeles' safety and crime trends. The DATE OCC and TIME OCC variables showed us when the crimes happened, helping us see patterns across different times. The AREA and AREA NAME variables helped us locate pre-defined areas on the map where crimes happened the most and where it was safer in the city. The Crm Cd and Crm Cd Desc variables told us what kind of crimes were reported, helping us sort and weigh different types of crimes in the dataset. This categorization was reflected in our dataset in the form of a new variable which we called "Category". The Vict Age and Vict Sex variables provided insights into the demographics of crime victims which allowed us to analyze if certain groups were more vulnerable in specific areas or times. Finally, the LAT and LON variables gave us the exact geo-coordinates where crimes happened, which we needed to accurately plot these incidents and use them in our routing system.Together, these variables gave us an almost complete picture of crime in Los Angeles, gaining insights about the safety across the city's many neighborhoods and streets.

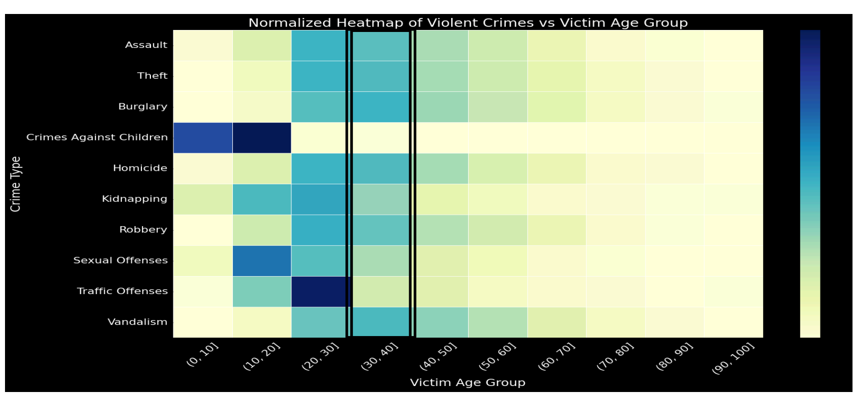
Our initial data analysis reviewed a crucial distinction needed to identify high-crime areas and impact zones for our algorithm to process. Looking at crime rates across Los Angeles County, we discovered that among the top 10 areas, Central (Downtown) Los Angeles stood out significantly, with approximately 70,000 reported crimes during the most recent two-week period in our dataset (see Visualization A in the appendix).

Not only did considering this exploration allow us to learn more about the crime density of Los Angeles County and help make more connections between population density and number of crimes committed throughout a particular area, but, most importantly, this data led us to wanting to further investigate the Central Los Angeles area and create a visual representation for how our product would demonstrate and handle a large amount of crimes in a primarily dense area with several cross streets.

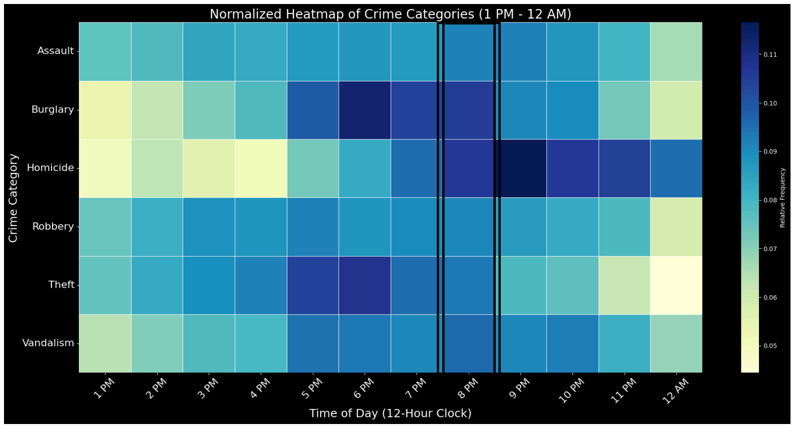
Therefore, we decided to expand on this image and deploy the usage of the mapbox API with our dataset to explore the exact cross streets of crimes in Central Los Angeles, in addition to wanting to obtain a better "feel" for our data by exploring the types of crimes that are being reported in and around Central Los Angeles.

Shown above is the “LA Live” entertainment center, a densely populated area within Central Los Angeles that hosts multiple sporting events and concerts throughout the year, being known as a popular area in Los Angeles County for visitors and residents to commute to for a night out with friends or a significant other. This is also an area where a large majority of Los Angeles’ working class commutes to and from, further adding to the point that Central Los Angeles would be an integral location for our algorithm to process. As seen in the visualization, the majority of crimes within the cross streets of LA Live are classified as burglary, fraud, theft, and assault, crimes that informed us of the importance that our algorithm would need to successfully weigh the impact of severe crimes like these when processing a given route. It was also beneficial to view our data within a proper map format, as it confirmed that the coordinates of said crimes within our dataset was accurate and our data as a whole could properly be used to scale up our application from a basic mapbox API integration to a full system that uses mapbox to visualize the routes our algorithm would process.

Building upon these revelations, we delved deeper into how the different profile variables affect the types of crimes an individual is most likely to encounter. This exploration was important for developing a more personalized approach to our safe routing algorithm.



We first examined the combined effect of age and gender on crime victimization patterns. Our analysis revealed that certain age groups are more susceptible to specific types of crimes. For instance, individuals between 30 and 40 years old showed a higher likelihood of being victims of crimes such as Assault, Theft, Burglary, Homicide, Robbery, and Vandalism. This insight allows our algorithm to adjust risk assessments based on the user's age profile.



Next, we investigated the impact of time of day on crime occurrences. Our visualization above clearly demonstrated that certain crimes are more prevalent at specific hours. For example, we found that Homicide and Burglary rates peak around 8 PM. This analysis enables our algorithm to factor in time-based risk when calculating the safest routes.

Interestingly, when we attempted to include the day of the year into our solution, we encountered limitations in the current dataset. We found no significant or consistent patterns in crime rates across different months or specific dates (Visualization B in appendix). The proportions of different types of crimes remained relatively consistent across months(Visualization C in appendix). This lack of clear yearly patterns and limited occurrences of the exact dates led us to conclude that, with the current data, incorporating the day of the year into our algorithm would not significantly enhance its accuracy.

We do acknowledge that the day of the year is a very important factor to consider in crime analysis and route safety. However, the current state of the dataset doesn't allow us to incorporate this factor into our solution effectively. If real-time data or more comprehensive historical data becomes available in the future, the time-of-year factor could become a valuable input for our system, potentially enhancing its predictive capabilities.

By combining these insights - the user's age and gender, the time of travel, and historical crime locations - we can assign appropriate weights to different types of crimes along potential routes. This multi-faceted approach allows us to provide users with personalized, safer path recommendations based on their specific profile and travel circumstances. This comprehensive analysis demonstrates the complex interplay of various factors in determining crime risks in urban environments. As we move forward, we remain open to incorporating additional data sources that could further refine our predictive capabilities and ultimately provide even safer routing options for users in Los Angeles.

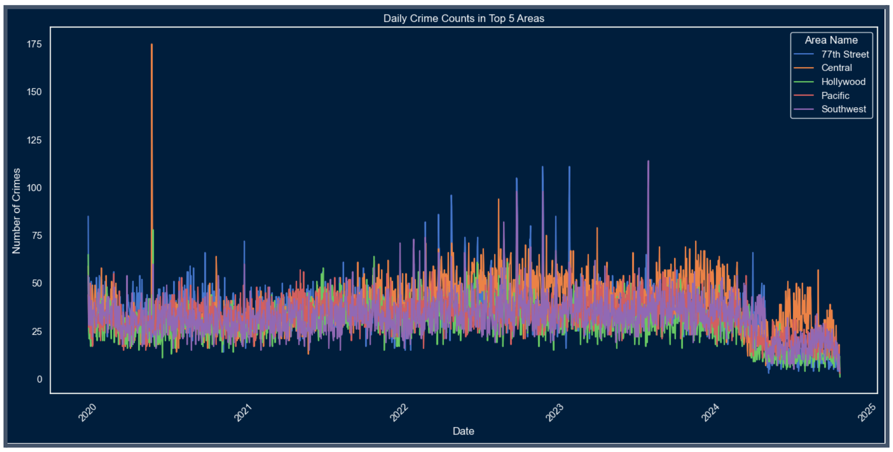
3 Methods

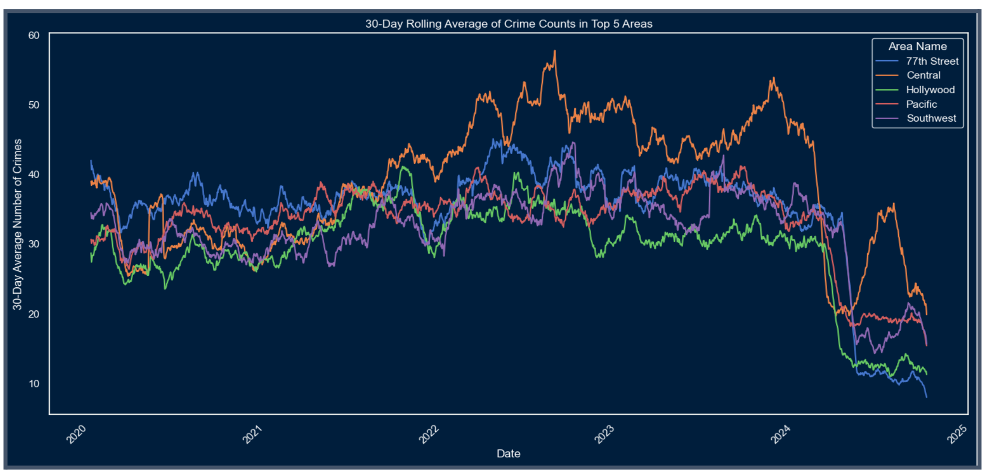
Describe the approach to the problem and how you've planned on solving it. You may consider including links to Github or other implementation specific details as necessary.

3.1 Handling peaks in the data

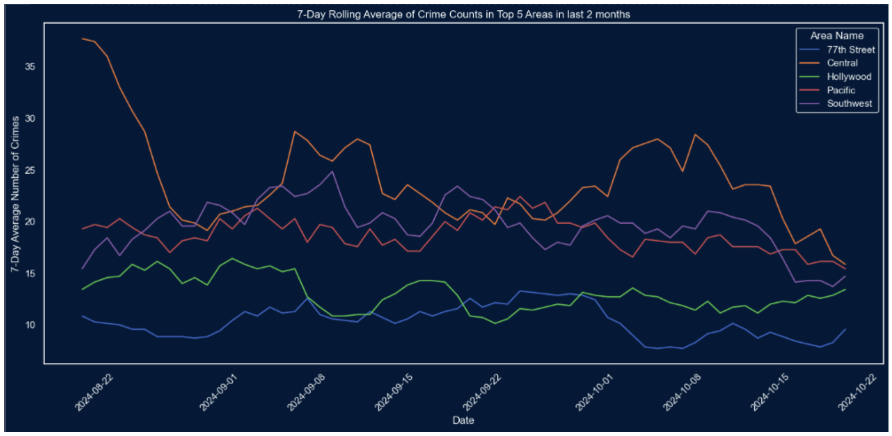
The data often includes one-off events that result in an unusually high number of crimes on a single day. These anomalies do not accurately reflect the overall safety or danger of the area and create large spikes in the data, which can skew the calculation of a meaningful crime score. Below is a plot of the data from the past five years for the top five areas with the highest crime rates: Central, 77th Street, Pacific, Southwest, and Hollywood.

The data appears highly volatile, with numerous peaks that make it difficult to interpret when calculating the crime score for a region. To address this, we apply the concept of a rolling mean. For example, with a 30-day rolling mean, each data point is replaced by the average of the previous 30 days. This approach smooths the data, providing a more continuous and meaningful representation. Below is the plot of the same data, smoothed using a 30-day rolling mean.





The impact of one-off peaks is significantly reduced, ensuring that the crime score reflects the average crime rate of an area rather than being skewed by anomalies. Additionally, we've decided to provide users with as much data as possible to help them make more informed decisions. Users will have access to the 7-day rolling mean for the past two months, allowing them to see both the longer-term average and how safe or dangerous the area has been over the last week.




```
average weekly crimes in last 2 months AREA NAME
77th Street    10.306452
Central        23.995392
Hollywood      13.029954
Pacific        18.753456
Southwest      19.629032
Name: 7-Day Average, dtype: float64
average crimes reported in last 7 days AREA NAME
77th Street     9.571429
Central        15.857143
Hollywood      13.428571
Pacific        15.428571
Southwest      14.714286
Name: 7-Day Average, dtype: float64
```

	Area Name	Top Category	Count
0	77th Street	Theft	55
1	Central	Theft	77
2	Hollywood	Theft	75
3	Pacific	Theft	98
4	Southwest	Theft	85

3.2 Tradeoffs or Alternative Approaches

Our selected approach of combining rolling means for crime data analysis with location-based risk assessment offers several advantages over alternative methodologies. The primary alternative approaches we considered included using raw crime counts without temporal smoothing, which would have been computationally simpler but would have made our system highly susceptible to anomalous events and statistical outliers. Our rolling mean approach provides more stable and reliable risk assessments while still preserving meaningful trends in the data. We also considered implementing a purely distance-based routing system with crime avoidance, which would have potentially created routes that completely circumvent higher-crime areas. Instead, our current approach allows for a more nuanced balance between safety and practicality, particularly important given our ethical considerations regarding economic impact on local businesses.

As a whole, the combination of 7-day and 30-day rolling means in our final implementation proves particularly effective as it captures both immediate safety trends and longer-term patterns. Lastly, our approach's integration with existing mapping services ensures routing reliability while adding the dimensionality of safety considerations rather than attempting to build a routing system from scratch, which is beyond the scope of us Data Scientists.

4 Prototyping

The development of the application was completed in two separate sections: the backend was completed in python with flask, and the frontend in the Angular framework. The backend consists of the code for the data ingestion, route generation, top crimes sorting, and the path preference calculation while the frontend includes the display features that the end user sees in the User Interface.

4.1 Backend

The backend application was developed using Python with Flask with `pandas` and `Pytorch`. It handles several main functions: data ingestion and preprocessing, route generation, crime analysis, and route preference calculations.

We preprocess the data by standardizing the geolocation data, and converting the dates into datetime format. Route generation is accomplished with use of the Mapbox API, which converts the user inputted locations into geographic coordinates and generates potential routes between these points.

The safety and crime analysis is then completed by calculating the crime score for each potential route by using the historical crime data within a 160-meter radius around the path and weighing the crime by severity. Crime severity weights were determined through a survey of 20 participants who rated various crime types on a scale of 1-10 based on their perceived severity when traveling through an area (see appendix section 8.3 for additional details). Severe crimes like homicide and assault carry higher weights (10 and 9 respectively) while crimes considered less severe such as vandalism (3) or fraud (2) carried lower weights.

We incorporated the user demographics and preferences with a scoring system which identifies the top 3 most relevant crime types for the user's specific profile (age, gender, and travel time) and applies a 20% higher weight to those categories. The final risk score is then adjusted accordingly.
Route preference scores are calculated using the formula:

$$Path\ Preference\ Score_p = (p * \frac{1}{d}) + ((1 - p) * \frac{1}{c})$$

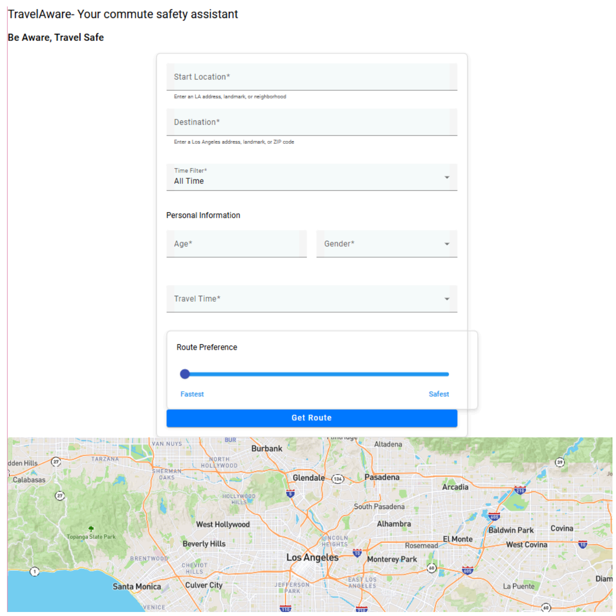
Where:

- p ∈ [0,1] represents the user's preference between time and safety (1 for safest routes, 0 for fastest)
- d represents route duration in seconds
- c represents the route's crime score, calculated as $RP = \sum f_i * s_i$
    - f_i= frequency of crime category i
    - s_i = severity weight of crime category i

The final scores are normalized by dividing by the sum of all route preference scores, ensuring they sum to 1 which allows for comparison between routes.

4.2 Frontend

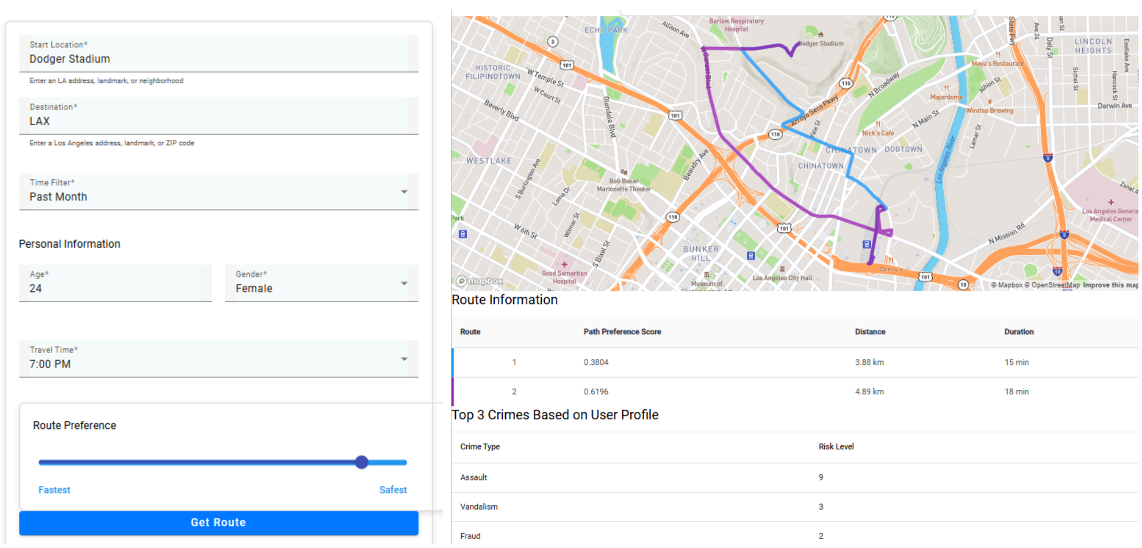
The frontend of the application was developed in Angular, which is a Typescript development platform for building web applications. It is connected to the flask backend script where it ingests the data and map information and displays it to the User. The UI is made up of four main components: the map display, the user input fields, the route information table, and the top crimes table. When first launching the app, a user will only see the map that is centered in Los Angeles with no routes and user input fields:



A view of the User Interface when launching the application

The user is prompted to fill in the following fields before a route is generated: Start Location, Destination, Time Period of Crimes committed Filter, Personal Information (including Age, Gender, and Travel Time), and Route Preference. Some of these fields have restrictions on what can be entered, such as Age not being a negative number, or the Starting and Ending Locations not remaining empty, and will not allow a route to be generated if these fields are not imputed correctly.

After the user correctly enters their information, the 'Get Route' button then can be pressed and will show it is loading. This feeds the user information to the backend application where the calculations for the routes and crimes based on demographic can be calculated and sent back to the frontend. Once the calculations are completed, the routes will appear on the map and the Route Information and Top Crimes Based on Profile Tables will appear under it:



Based on a trip from Dodger Stadium to LAX over the past month, these are the generated routes for a 24 year old female traveling at 7:00pm (with a safe-route preference score of 88).

4.3 Running the Application

To run this application, first clone the GitHub repository. Following that, upload the Modified crime dataset (the Los Angeles crime dataset with the crime categories added) into the project folder. Then, start the flask backend by navigating into the project directory, and running the app with the command `python3 app.py`. Finally, navigate to the angular-frontend folder in the directory and launch the Angular frontend with `ng serve`. The app then can be accessed at http://localhost:4200/. More information on how to run the app can be found in the README file in the repository.

4.4 User Considerations

Designed for the Los Angeles Area

This app is only designed to work for the Los Angeles area. While routes can be generated for areas outside the Los Angeles area, we have included a preference in the code to prefer areas around the center point of Los Angeles. This means that while other addresses can be included and will generate routes that populate the map, it will take much longer than designed to calculate the route, and will only include crime data for the Los Angeles area. This will result in the possibility of inaccurate route preference scores. So for the purpose of this prototype, the authors would strongly discourage the use of this app for routes that are not within the bounds of Los Angeles county.

API Generated Routes

We would also like to stress that our application does not choose the routes based on crime- it simply calculates the crime of the areas in the routes generated by the Mapbox API. For this reason, generated routes will not route differently based on crimes in certain areas and therefore does not avoid travelling in certain areas.

Limitations of Route Generation

Additionally, while most start and end destination combinations that were tested generate two routes, there are some instances where only one route is generated. This could be because the start and end locations are close to each other and do not have that many routes to take in general. Alternatively, it could be that the routes that would be generated would not be considered as ‘unique enough’ to the API to differentiate them as separate routes. Nonetheless, this is something for the user to consider when using and is an aspect that should be improved in future versions.

5 Ethical Considerations

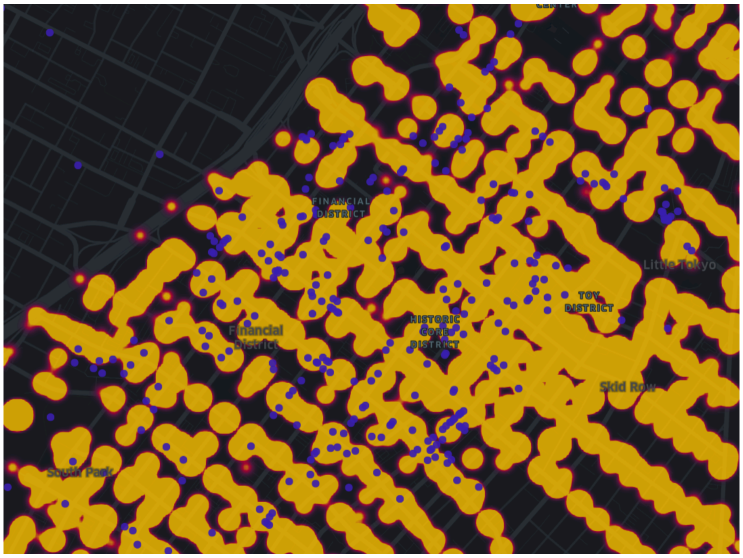
Los Angeles County, a sprawling and diverse metropolitan region, faces significant challenges related to public safety and economic vitality. TravelAware operates across the entirety of LA County, using data-driven insights to provide safer navigation options to its users. However, Central LA, a key urban hub, serves as a useful case study to illustrate the app’s impact on high-risk areas.

In Central LA, 94% of 320 sampled businesses are located within 180 meters (approximately one block) of high-risk crime zones. These businesses—spanning restaurants, cafes, gift shops, clothing stores, and more—are heavily reliant on foot traffic to sustain operations. The spatial overlap of high-risk zones and business districts underscores the complexity of balancing safety and economic stability in urban environments. While highlighting safety concerns, crime data can inadvertently create economic barriers, steering potential customers away from vulnerable neighborhoods.

TravelAware acknowledges that its safety-driven design must account for these broader implications, ensuring that its functionality supports both individual users and the communities they navigate through.

5.1 Addressing Why It Matters

The implications of diverting travelers from high-risk areas extend beyond personal safety considerations. Many small businesses, especially in economically marginalized neighborhoods, depend on steady foot traffic to accumulate profits. Reducing visitor flow through these areas risks diminishing revenues and exacerbating systemic challenges like poverty and underinvestment.



Additionally, the over-policing of low-income neighborhoods compounds these issues. Communities frequently classified as high-risk face stigmatization, limiting their access to resources and economic opportunities. TravelAware recognizes that simplistic avoidance of these areas could unintentionally worsen inequalities, alienating residents and businesses.

This dynamic illustrates the dual challenge faced by TravelAware: providing safer navigation options while avoiding harmful socio-economic consequences for vulnerable communities.

5.2 Potential Future Improvements: Designing for Balanced Outcomes

Although TravelAware currently focuses on providing personalized and safety-informed routes, there are opportunities to expand its functionality in the future to address the socio-economic challenges described above. One potential improvement involves integrating a balanced routing approach. Instead of completely avoiding high-risk zones, TravelAware could identify safer paths within these areas. This design would enable travelers to navigate through less vulnerable sections of high-risk neighborhoods, supporting local businesses while maintaining user safety. By incorporating such a feature, the app could better balance individual safety with community economic activity.

Another future improvement could involve providing users with additional contextual information, such as the presence of nearby businesses or safety tips tailored to specific neighborhoods. These enhancements would empower users to make informed travel decisions, fostering greater transparency and trust in the app.

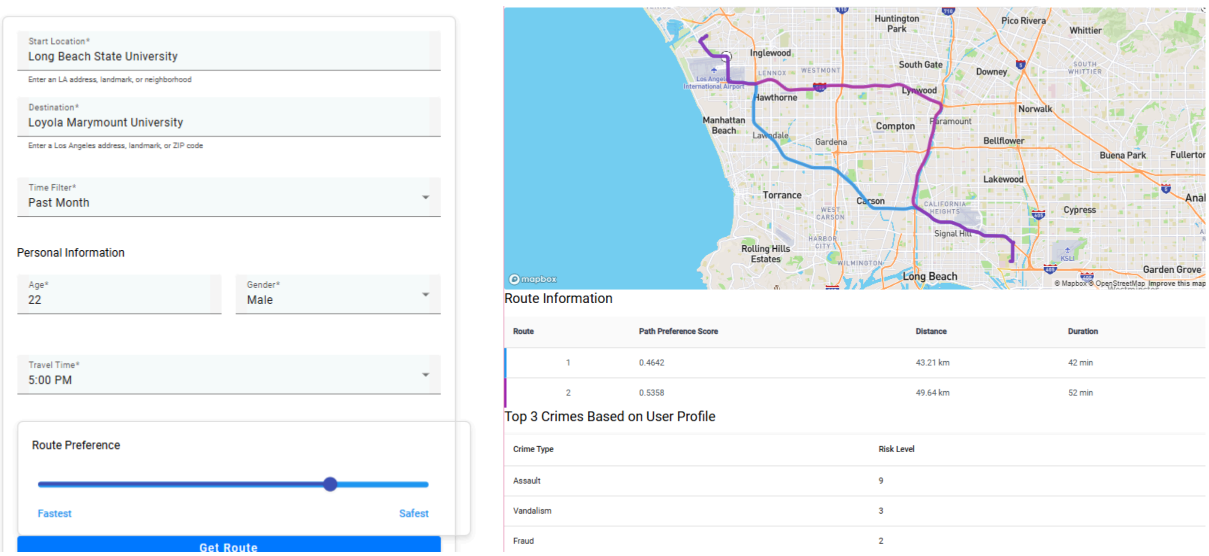
5.3 Potential Future Integration: Weighted Safety-Impact Modeling

TravelAware could also adopt a weighted safety-impact model in the future to evaluate the broader consequences of its routing recommendations. This model would incorporate variables such as income levels, crime severity, and policing intensity to calculate a dynamic "negative impact score" for neighborhoods. Such a score would help assess not only the safety of a route but also the potential economic harm to communities caused by diverting foot traffic away from certain areas.

By assigning higher weights to low-income neighborhoods, the model could reflect their increased vulnerability to economic disruption. While maintaining the app's focus on safety, these future enhancements could help mitigate unintended socio-economic consequences and promote equitable outcomes for all communities within Los Angeles County.

6 Results

Since our project was product-oriented, the concrete results of our work comes in the form of the successful demo of our application. To help digest the below results, consider the following situation: You are a college sports data analyst that travels from arena to arena to complete your assigned game coverages. You check your schedule for the week and see that on Saturday, you have been assigned to a college basketball game at Long Beach State University (LBSU) at 2pm, and another college basketball game at Loyola Marymount University (LMU) at 7pm. Understanding that basketball games take about 2 hours to complete and being that the two schools are located across the county, you conclude that covering both games given the time constraints would be difficult, especially considering typical weekend Los Angeles traffic. You mull it over while covering the game at Long Beach State and decide that you want to boot up TravelAware to determine which route would get you to LMU in a quick but safe manner, since getting into a car accident or driving along a route that may be subject to historical carjackings would make you late for the second game.



Adjusting the TravelAware route preference to lean toward a preference for safety while still considering speed implications, you see that TravelAware recommends you take the second available route, which sacrifices about 10 minutes of commute time but is the safer route based on the application's algorithm. Once the game at Long Beach State finishes, you then decide to take this route out of caution, and since the first game ended slightly earlier than expected, giving you the necessary breathing room to account for the additional 10 minutes that the safer route would require.

7 Conclusion

While our current solution effectively processes crime data and generates safety-weighted routes, several areas present opportunities for enhancement. First, our reliance on historical crime data, while providing valuable insights, could be augmented with real-time incident reporting to offer more immediate risk assessments. Integration with police dispatch systems or crowd-sourced safety reports could provide users with up-to-the-minute safety information.

Additionally, the current geographical limitation to Los Angeles County, while allowing for focused development and testing, suggests a clear path for scalability. Expanding TravelAware to other metropolitan areas would require establishing partnerships with local law enforcement agencies to access standardized crime data, developing automated data preprocessing pipelines to handle variations in crime reporting formats across jurisdictions, and calibrating our risk assessment algorithms to account for regional differences in crime patterns and severity.

Furthermore, our analysis revealed that the relationship between time of year and crime patterns could not be effectively incorporated due to limitations in the current dataset. Future iterations could benefit from more comprehensive temporal data collection and analysis, potentially incorporating seasonal trends and special event impacts on safety patterns.

As a whole, as cities continue to digitize their crime data and improve reporting standards, TravelAware's framework positions it well for expansion beyond Los Angeles, and our project demonstrates an excellent building block towards the larger goal of a commercialized solution.

8 Appendices

8.1 Team Member Contribution Statements

Pranav Agarwal : Discussion of problem statement with the team and Discussing what the outcome of the project should be, Development of the rolling mean methodology to combat peaks; Utilizing mapbox and fetching routes, Contribution to final paper, helping team whenever required

Serena Elizabeth Alvarez : Project Purpose + Statement (Presentation 1), Assumptions + Complexity + Narrative Structure (Presentation 2) , Project Overview + Demo (presentation 3), UI (Frontend) Development, Overall Presentation Aesthetics, Overall Project Development Coordination, Contributions to final paper

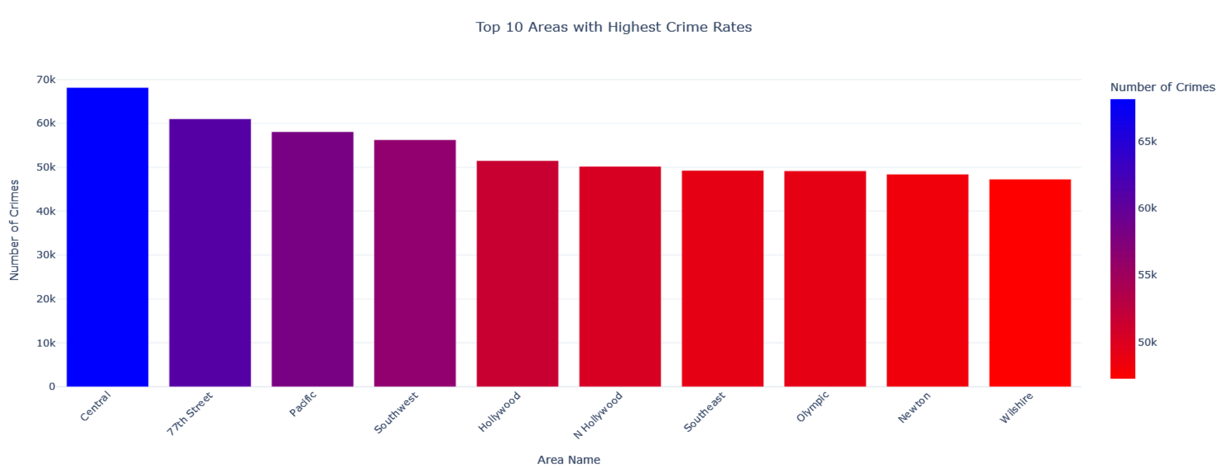
Sebastian Algharaballi-Yanow : Technical Stack + Roadmap (Presentation 1), Interactive Temporal + Interactive Mapbox Visualizations (Presentation 2), "What was Accomplished" + Supporting EDA + Demo Example (Presentation 3), Final Paper (Some of EDA + Results + Conclusion + Additional Methods Details)

David Assio : Analysis of data structure + discussion of problem background (Presentation 1), Assistance in mathematical algorithm formation + created/distributed survey for data collection (Presentation 2), EDA of businesses in high crime areas + addressing ethical issue/feedback (Presentation 3), Contributions to final paper

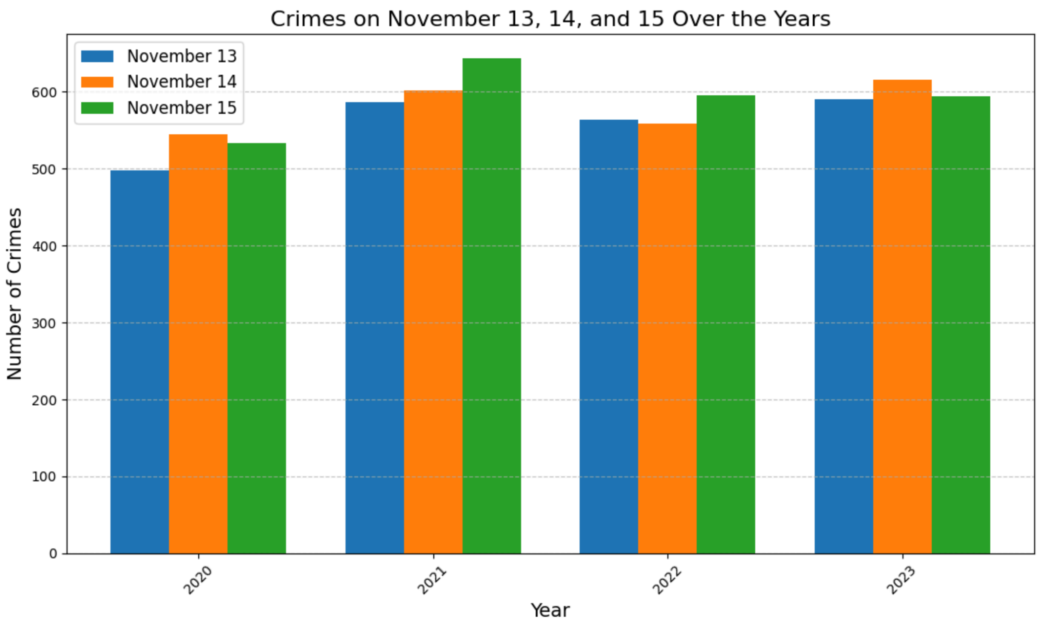
Satvik Bhatnagar : Original Project Idea + Data Sourcing + Dataset Category (Presentation 1), Data Cleaning + Safety Cutoff + Algorithm Exploration + Mathematical Algorithm Formation (Presentation 2), Path Preference Score Implementation + Mathematical Justification of Algorithm (Presentation 3), Development of Path Preference Score, Suitability Score, Crime Category Survey, Contributions to final paper

Samar Vikas Save : Dataset Exploration, Temporal Analysis and Visualizations (Presentation 1), Profile Variables Analysis and Geo-Coordinate Filtering Approach (Presentation 2), Coding and Integration for Profile based filtering, addressing feedback (Presentation 3), Contribution to drafting the Final Paper

8.2 Additional EDA



Visualization A: Top 10 Areas with Highest Crime Rates



Visualization B: Top Crimes for Sample Dates over the Years

Visualization C: Pie Chart of Crime Distribution by Month

8.3 Additional Methods Details

To establish a quantitative measure of risk assessment, we developed a severity scoring system based on a survey of 20 participants. Each crime category was rated on a scale of 1 to 10, with 10 representing the highest level of perceived danger. The results revealed a clear hierarchy of risk perception, with violent crimes such as homicide (10), assault (9), kidnapping (9), and sexual offenses (9) receiving the highest severity scores. Property crimes and non-violent offenses were generally perceived as less threatening to travelers, with fraud and trespassing both receiving scores of 2. This weighted scoring system enables our algorithm to more accurately assess route safety by considering not just the frequency of crimes in an area, but also their relative severity and potential impact on traveler safety, in addition to confirming that our ideas of "risk" would be confirmed by a group of individuals outside of our project group.

Crime Category	Avg. Severity Score
Homicide	10
Assault	9
Kidnapping	9
Sexual Offenses	9
Crimes ag. Children	8
Robbery	8
Weapon Offenses	8
Burglary	7

Crime Category	Avg. Severity Score
Criminal Threats	6
Theft	6
Traffic Offenses	4
Public Disorder	4
Order Violation	4
Lewd Letters/Calls	3
Fraud	2
Trespassing	2