#### **Our Goals:**

Our project seeks to investigate intentional driver-directed fraud in the taxi industry through deceptive practices. Due to the time constraints of the project and for the sake of simplicity, we will be analyzing data for New York taxi cabs provided by the New York Limousine and Taxi Commission from the years 2009 to 2024 [1]. To determine route efficiency, we will be using a modified Dijkstra’s Algorithm proposed in *A Hybrid Link-Node Approach for Finding Shortest Paths in Road Networks with Turn Restrictions* [2] to determine the estimated shortest, safest, and most traffic-efficient paths to compare to the true paths taken. Data on traffic [3] and car accidents [4] will be gathered from smaller datasets, with the node edge graph of New York provided by Freudenberg Sealing Technologies [5]. Users may choose which paths are most important to them. Detecting Fraudulent Taxi Drivers: Overview [6] provided our team with guidance when determining which algorithms to implement, noting the advantages and disadvantages of various methods in obtaining similar research goals.

#### **Today’s Research and Our Innovation:**

Many investigations into the taxi industry exist in this day in age, usually implementing artificial neural networks to calculate travel efficiency or various patterns in demand. As proposed in *Taxi Demand and Fare Prediction with Hybrid Models: Enhancing Efficiency and User Experience in City Transportation* [7], it is possible to use Long Short-Term Memory Recurrent Neural Networks with a Mixture Density Networks to estimate the current demand for taxis and its effect on taxi fares at any given time. *One Model Fits All: Cross-Region Taxi-Demand Forecasting* [8] cooperates this finding, utilizing graph neural networks to expand the demand function further and predict demand in regions that lack historical data. Additional studies explore known types of fraud in the taxi industry. Emphasis has been put on understanding the driver’s decision process to maximize profits. *Optimizing Taxi Driver Profit Efficiency: A Spatial Network-Based Markov Decision Process Approach* [9] introduced a Markov Decision Process model to determine the most profitable routes for drivers based on fuel efficiency and traffic conditions, regardless of passenger expense. We have seen fraud detection in action in Iran, as *Fraud Detection System in Online Ride-Hailing Services* [10] utilized GPS data and DBSCAN clustering to identify and mark fraudulent ride-sharing trips.

*Optimizing Taxi Carpool Policies via Reinforcement Learning and Spatio-Temporal Mining* [11] uses a spatio-temporal neural network to optimize carpooling efficiency by paring passengers based on trip patterns. *Taxi dispatching strategies with compensations* [12] proposes action be taken to offer economic incentives to drivers to encourage them to take more time-efficient routes without compromising pay. Our study innovates through combining the principles of today’s current studies to seek inefficiencies beyond explainable reason as evidence of proposed fraud.

#### **What does it Matter:**

Taxis and taxi-like services are as vital today as ever. In a world dependent on motor vehicles for transportation, millions of people rely on taxi services yearly, trusting that they will be charged a fair going rate. However, as highlighted in *Reducing Inefficiencies in Taxi Systems* [13], drivers may inflate fares by prolonging trips, increasing idle time, and taking inefficient routes to maximize earnings, violating the trust of the customer. We aim to shine a light on the occurrences of such fraud and recognize patterns in where and how they occur to give riders information on what fares they should expect.

#### **Risks:**

It is impossible for us to consider all facets of the taxi industry. As denoted in *Understanding Taxi Travel Patterns* [14] and *Taxi apps, regulations, and the market for taxi journeys* [15], the taxi industry has always been affected by fluctuating demand, taxi hot-spots, and deregulation of taxi-like services through competitors. While we may consider these factors as reasons for discrepancies, it is infeasible to efficiently calculate the effect of all factors throughout time. We additionally run risks in simplifying traffic and safety analysis for the sake of time. We may unintentionally write of discrepancies because of these factors where it was not actually present. *Intelligent Traffic Management: A Review of Challenges, Solutions, And Future Perspectives* [16] proposes the use of machine learning algorithms to calculate theoretical traffic patterns and highlights the challenges of real-time data collection, data integration, and computational efficiency, serving as a caution for the costs of attempting to implement an inference model alongside our project’s main goal.

#### **Payoffs:**

We acknowledge that the risk in summarizing traffic and accidents over our timeframe may result in a less accurate depiction of truth and that we may be able to perform similarly if we used a machine learning model to replicate taxi trips. However, we believe the reduced time needed to process the summarized data outweighs the slight inaccuracy and accept Dijkstra’s algorithm as it forces taxi trips to remain on real streets.

#### **Costs, Work, and Timeframe:**

Given the rigid deadline of the project, it is impossible for this project to take longer than 40 days from the project proposal. Our goal is to have our data, computational algorithm, and initial static visualization finished by March 20th. We will then dedicate the remaining project time to the interactive component of the visualization and clean up any errors.

| **Plan of Activities** | | |
| --- | --- | --- |
| **Effort/Task** | **Lead** | **Timeline** |
| Collect and Clean Data | Omar | March 3rd - March 10th |
| Implement Dijkstra’s Algorithm | Cole | March 3rd - March 20th |
| Initial Visualization | Cameron | March 3rd - March 20th |
| Visualization Cleanup | Cameron | March 20th - April 11th |
| Add Interactive Component | Saif | March 20th - April 11th |
| Team Lead (Submits Papers and Presentation) | Omar | Per Assignment |

We believe that the best practice for our team to distribute the effort is to simultaneously have everyone contribute towards each task instead of assigning each member individual tasks. We have designated a lead for each task based on our skillsets. The lead is responsible for coordinating the team effort and ensuring the team completes tasks on time.

# References

|  |  |
| --- | --- |
| [1] | NYC Taxi and Limmosine Commission, "TLC Trip Record Data," 25 February 2025. [Online]. Available: https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page. |
| [2] | Q. Li, B. Y. Chen, Y. Wang and W. H. K. Lam, "A Hybrid Link-Node Approach for Finding Shortest Paths in Road Networks with Turn Restrictions," *Transactions in GIS,* vol. 19, no. 6, p. 915–29, 2015. |
| [3] | Department of Transportation (DOT), "Traffic Volume Counts," NYC Open Data , 27 May 2022. [Online]. Available: https://data.cityofnewyork.us/Transportation/Traffic-Volume-Counts/btm5-ppia/about\_data. [Accessed 25 February 2025]. |
| [4] | NYPD, "Motor Vehicle Collisions - Crashes," NYC Open Data, 24 February 2025. [Online]. Available: https://data.cityofnewyork.us/Public-Safety/Motor-Vehicle-Collisions-Crashes/h9gi-nx95/about\_data. [Accessed 25 February 2025]. |
| [5] | Freudenberg Sealing Technologies, "Street Network of New York in GraphML," Freudenberg Sealing Technologies, 2017. [Online]. Available: https://www.kaggle.com/datasets/crailtap/street-network-of-new-york-in-graphml. [Accessed 25 February 2025]. |
| [6] | Z. S. a. M. R. Al-Sudani, "Detecting Fraudulent Taxi Drivers: Overview.," *Al-Sadiq International Conference on Communication and Information Technology ,* p. 115–20, 2023. |
| [7] | K. S. Chou, K. L. Wong, B. Zhang, D. Aguiari, S. K. Im, C. T. Lam, R. Tse, S.-K. Tang and G. Pau, "Taxi Demand and Fare Prediction with Hybrid Models: Enhancing Efficiency and User Experience in City Transportation," *Applied Sciences,* vol. 13, no. 18, pp. 10192 - , 2023. |
| [8] | R. Ozeki, H. Yonekura, A. Baimbetova, H. Rizk and H. Yamaguchi, "One Model Fits All: Cross-Region Taxi-Demand Forecasting," in *Proceedings of the 31st ACM International Conference on Advances in Geographic Information System*, ACM, 2023. |
| [9] | X. Zhou, H. Rong, C. Yang, Q. Zhang, A. V. Khezerlou, H. Zheng, Z. Shafiq and A. X. Liu, "Optimizing Taxi Driver Profit Efficiency: A Spatial Network-Based Markov Decision Process Approach," *IEEE Transactions on Big Data,* vol. 6, no. 1, pp. 145-58, 2020. |
| [10] | K. Bakhshi, B. Bahrak and H. Mahini, "Fraud Detection System in Online Ride-Hailing Services," in *7th International Conference on Signal Processing and Intelligent Systems (ICSPIS)*, Tehran, Iran, 2021. |
| [11] | I. J. a. Z. Q. a. X. C. a. M. N. a. J. Ye, "Optimizing Taxi Carpool Policies via Reinforcement Learning and Spatio-Temporal Mining," 2018. [Online]. Available: https://arxiv.org/abs/1811.04345. [Accessed 25 February 2025]. |
| [12] | H. a. F. A. a. O. S. a. P. J. a. B. J. Billhardt, "Taxi dispatching strategies with compensations," *Expert Systems with Applications,* pp. 173-182, 2019. |
| [13] | B. P. Chenguang Zhu, Reducing Inefficiencies in Taxi Systems, 56th IEEE Conference on Decision and Control (CDC), 2017. |
| [14] | H. Cai, X. Zhan, J. Zhu, X. Jia, A. S. Chiu and M. Xu, "Understanding Taxi Travel Patterns," *Physica A,* vol. 457, no. C, pp. 590-597, 2016. |
| [15] | S. Harding, M. Kandlikar and S. Gulati, "Taxi apps, regulations, and the market for taxi journeys," *Transportation Research. Part A, Policy and Practice,* vol. 88, pp. 15-25, 2016. |
| [16] | R. Ravish and S. R. Swamy, "Intelligent Traffic Management: A Review of Challenges, Solutions, and Future Perspective," *Transport and Telecommunication,* vol. 22, no. 2, pp. 163-182, 2021. |