Paper ###-2020

Lyft Jacking: Analysis of Lyft's Open Source Autonomous Driving Data

John Tomaselli, Chidananda s.a. Nayak, and Jorge Vargas Amezcua,

University of North Carolina Wilmington

ABSTRACT

Proliferation of autonomous vehicle testing is no longer a dream, but a reality. Dozens of companies actively test platforms on live city streets that navigate many of the same hazards as human drivers (Scale AI Inc. , 2020). Contributing to this ecosystem first entails understanding the basic inputs to driving, one of which is object identification. This paper seeks to lay the groundwork for future research by characterizing the general patterns of one aspect of object recognition observed in a dataset collected on public roads. Using the Lyft Level 5 Autonomous Vehicle dataset, different camera calibrations are studied to identify their effect on object detection. With object annotations varying by as much as 16.3% between different calibration bins, select groups of camera calibrations do appear to be affecting object recognition.

INTRODUCTION

Autonomous vehicles already occupy American roads. According to pcmag, well over 1.7 million miles of autonomous driving occurred during 2018 in California alone (Marvin, 2019). Supporting further development of these vehicles by a wider audience, several academic and industry groups recently released datasets containing both images and supporting metadata from select sets of their autonomous vehicle fleets (Scale AI Inc. , 2020). Lyft joined these groups on July 23, 2019 by releasing their Level 5 Autonomous Vehicle dataset (Vincent, 2019). The depth and extent of Lyft's data enables detailed examination of many different aspects of autonomous driving that allow for Lyft's vehicles to safely drive.

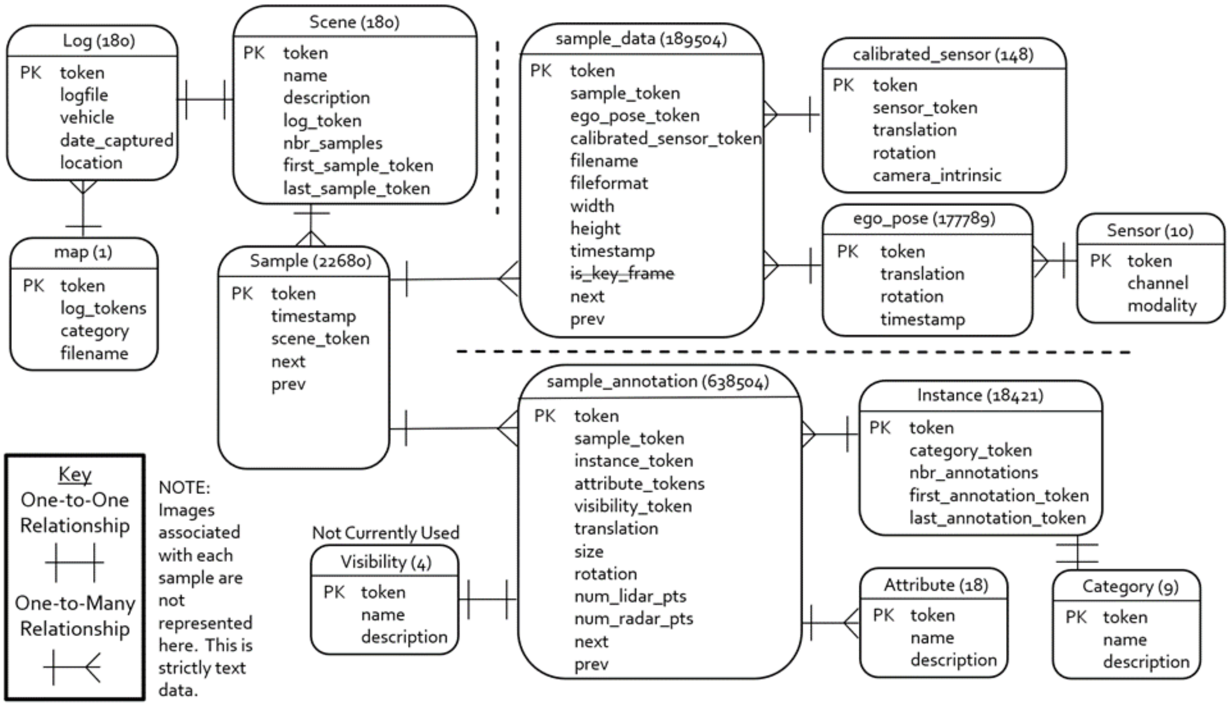
Fundamentally, this project's objective is to leverage that depth and extent to identify optimal patterns for object recognition, a critical aspect of safe driving. Focusing narrowly on camera calibrations, the reader will see how the focal length and optical center of an image return varying mean number of object annotations in the Lyft Level 5 Autonomous Vehicle dataset. If a subset of camera calibrations in this dataset allows for greater object detection relative to other camera calibrations under similar circumstances, this could indicate a set of optimized calibrations across the wider population of Lyft's Fleet.

Note that an extended version of this paper, with extra figures, can be found at this project's github page. All SAS code to support the below analysis can be found in the Code folder on this project's Github page.

DATA SET AND DEFINITIONS

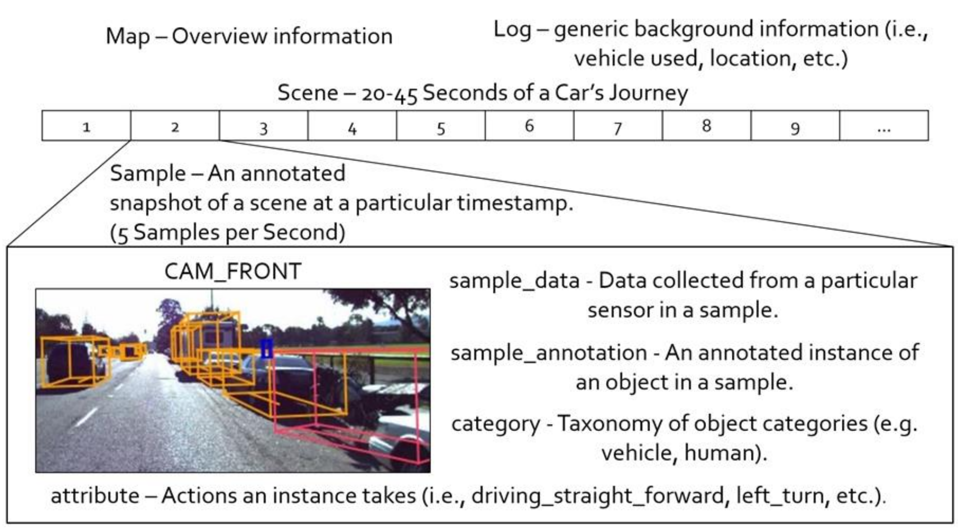
Lyft's Level 5 Dataset organizes data according to the nuScenes data format (Lyft Inc. , 2019). 13 individual schema in JSON format allow access to different types of metadata. The entry relationship diagram shown in figure 1 below displays how these schemas logically divide data to allow both one-to-one and many-to-one relationships. Forming these relationships permits relatively flexible analysis of patterns across data from different short videos of vehicles driving called scenes.

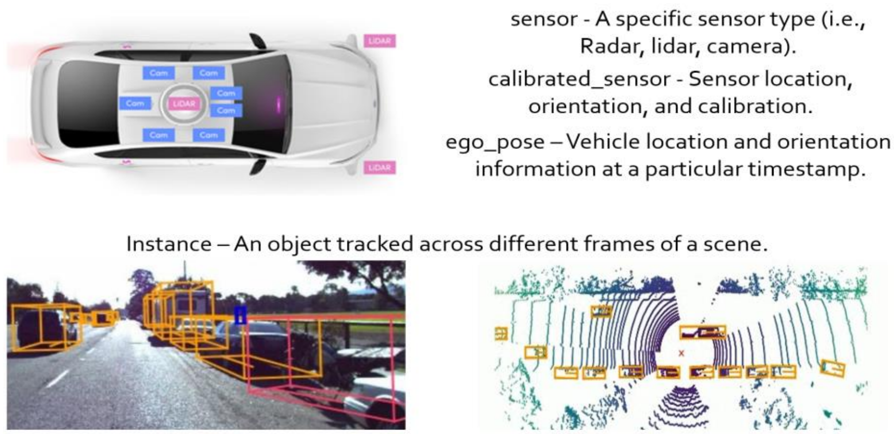
Figure 1: Lyft Level 5 Entry Relationship Diagram (ERD)



Scenes form the basis for analysis. In the Lyft Level 5 Dataset, these short driving videos, showing approximately 25 seconds of a Lyft vehicle's journey, consist of a set of snapshots at specific timestamps known as samples. From each sample, the Lyft Level 5 Platform catalogues recognized objects using the *sample_annotation* and associated schema. Conversely, the components of the Lyft Level 5 Platform used to collect a sample, such as the vehicle's location and the sensor being used, are aggregated by the *sample_data* schema. Figure 2 below goes into more detail concerning the definition of each schema.

Figure 2: Lyft Schema Definitions





SENSOR CALIBRATION INFLUENCE

Examining how different camera calibrations affect object detection first requires exploring response and predictor variables. The only viable response variable for this project proves to be the *nbr_annotations* variable in the *Instance* schema. This variable fluctuates based on the total time an object is detected by the Lyft Level 5 Platform over the course of a scene. Predictor variables for camera calibrations are aggregated in the *camera_intrinsic* variable of the *calibrated_sensor* schema. This variable forms a 3x3 matrix contains the *focal length* and *optical center* variables, the values of which sufficiently vary to allow analysis (MathWorks Inc. , n.d.).

Controlling for unintentional variations with the three predictor variables, *focal_length*, *cx*, and *cy*, is challenging. Figure 3 shows how vehicle use fluctuated considerably over the 30 days for which data is available. When aggregated, the locations in which the *nbr_annotations* variable were captured, as displayed in figure 4, appears to match the complete track used by Lyft. However, examination of the *nbr_annotations* associated with each vehicle reveals that vehicles with 5 or more recorded days of driving cover a much larger portion of the track than vehicles with fewer than 5 days of driving. As a final consideration, each Lyft Platform contains 7 camera settings with 1 camera setting, *CAM_FRONT_ZOOMED*, having different dimensions than the other 6. These location, date, and size variances must be considered when determining the optimal camera calibrations.

Figure 3: Vehicle Variation by Date

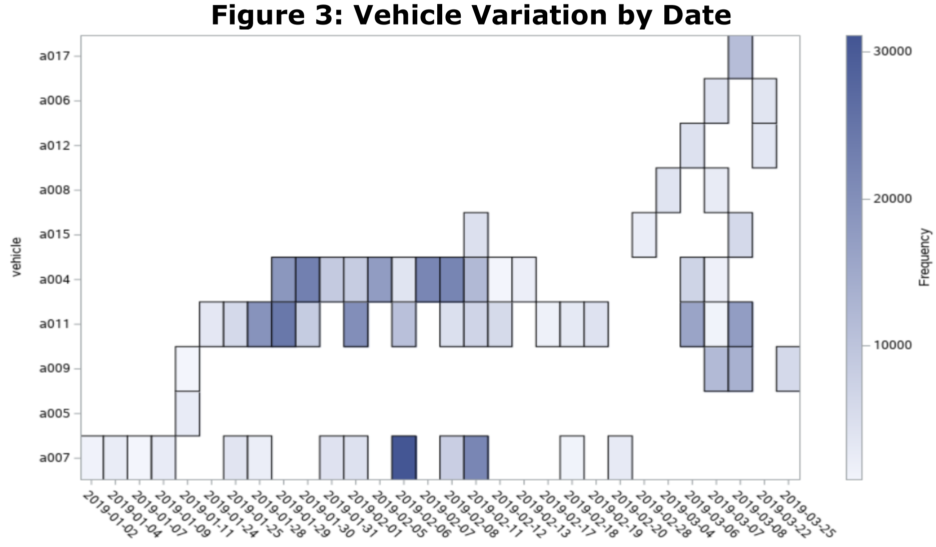
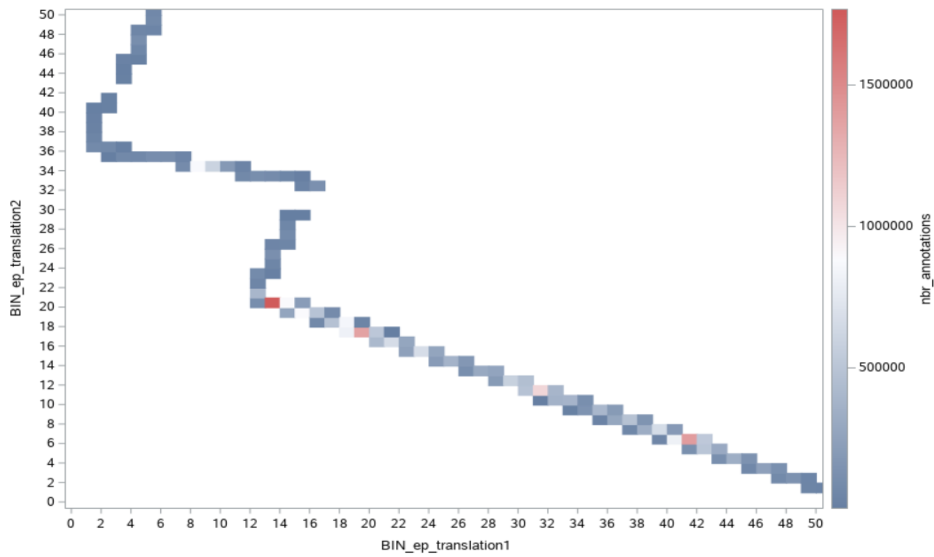


Figure 4: Lyft Dataset Total Annotations vs Vehicle Map



Initial examination of the data shows why controls are necessary. Our first attempts to find an optimal set of variables to produce the maximum *nbr_annotations* returned low adjusted R-squared values, indicating a poor fit to the data. However, high significance is associated with each identified variable as demonstrated by p values below 0.05. Subsequent HPBIN results show that two different clusters of *focal length*, *cx*, and *cy* values exist due to the *CAM_FRONT_ZOOMED* camera setting. Removing the *CAM_FRONT_ZOOMED* setting still produces a low adjusted R-squared value, but shows significantly different numbers of responses for distinct ranges of focal length. Figure 5 further develops this observation with the average *nbr_annotations* for each *focal_length* range appearing to vary by as much as 14.86% of the lowest observation. Figure 6 and figure 7 respectively reinforce this observation with the cy and cx variables observing variances of 15.39% and 16.30%, respectively.

Figure 5: Mean Number of Object Annotations per Focal Length Bin

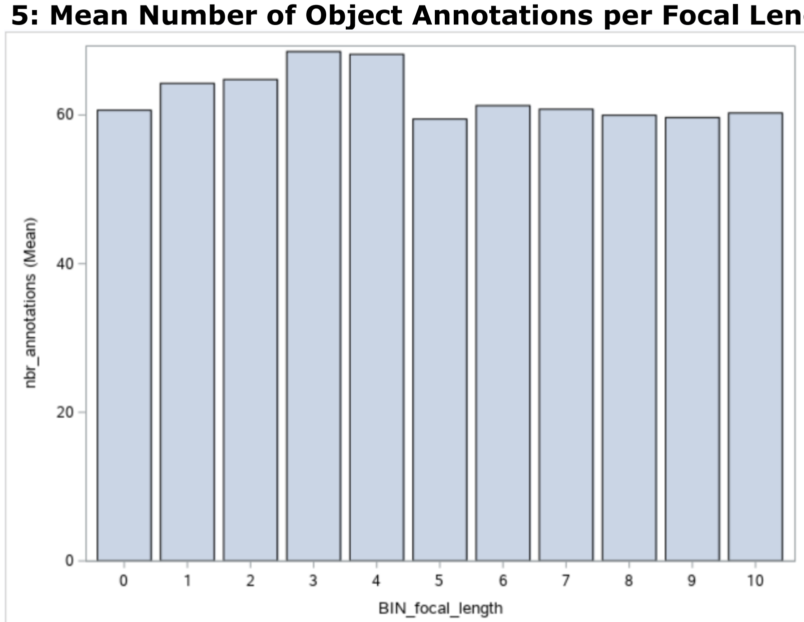


Figure 6: Mean Number of Object Annotations per cy (Optical Center) Bin

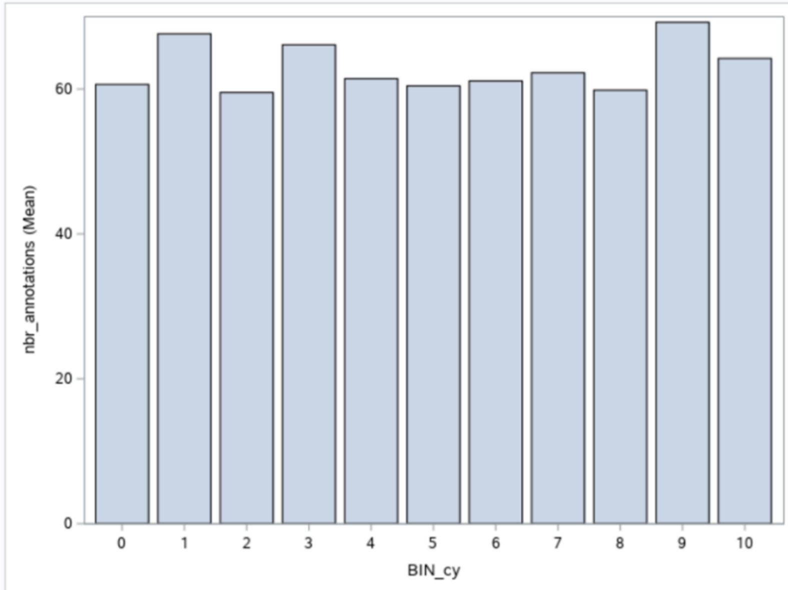
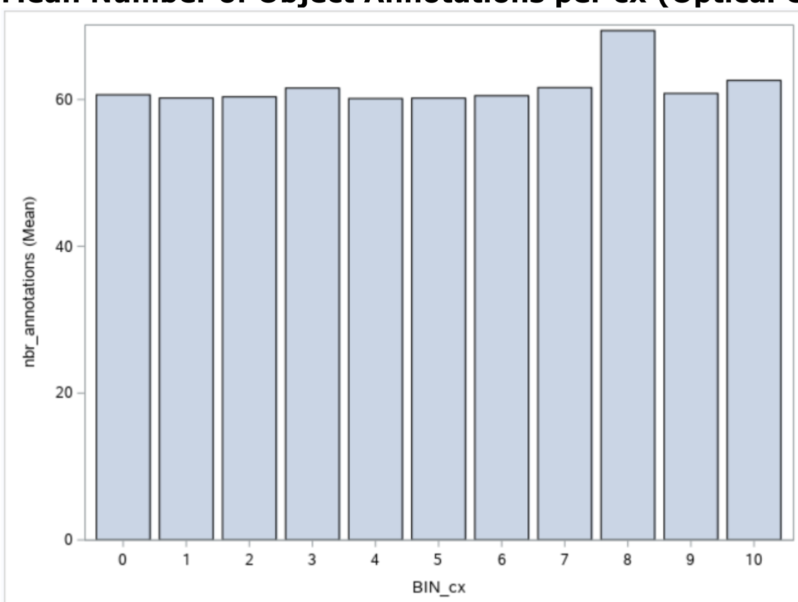


Figure 7: Mean Number of Object Annotations per cx (Optical Center) Bin



Thankfully, even with relatively poor fitting data, reasons exist to lend confidence to variables with higher *nbr_annotations*. Controlling for the date, vehicle, and location using the REPEATED statement in PROC GENMOD produces the results observed in figure 8. These results attribute significance to the *focal_length, cx,* and *cy* variables due to the observed low p values. Examining the two-factor interactions amongst the predictor variables using PROC GLMSELECT, found in figure 9, reinforces the above evidence, showing especially synergistic effects between the *focal_length* and *cx* variables. Moreover, figure 9 also shows how select bins of *focal_length, cx,* and *cy* variables continually produce higher t values, indicating greater overall responses. These observations all appear to lend credibility to the idea that certain sets of *focal_length*, *cx*, and possibly *cy* values produce higher average values of *nbr_annotations* and, therefore, optimizing object detection.

Figure 8: PROC GENMOD Repeated Analysis of Camera Calibration Variables

Analysis Of GEE Parameter Estimates						
Empirical Standard Error Estimates						
Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept	492.6091	35.6612	422.7144	562.5038	13.81	<.0001
focal_length	-0.5810	0.0393	-0.6580	-0.5040	-14.79	<.0001
cx	0.0990	0.0095	0.0804	0.1176	10.44	<.0001
cy	0.0483	0.0154	0.0181	0.0786	3.13	0.0017

Score Statistics For Type 3 GEE Analysis			
Source	DF	Chi-Square	Pr > ChiSq
focal_length	1	218.59	<.0001
cx	1	108.91	<.0001
cy	1	9.81	0.0017

Figure 9: PROC GLMSELECT Repeated Analysis of Camera Calibration Variables

Parameter Estimates			
Parameter	Estimate	Standard Error	t Value
Intercept	59.267726	0.244797	242.11
BIN_fo*BIN_cx*BIN_cy 7 8 3	15.919524	0.715434	22.25
BIN_fo*BIN_cx*BIN_cy 3 3 4	14.857738	0.718729	20.67
BIN_fo*BIN_cx*BIN_cy 7 6 5	14.270092	0.729281	19.57
BIN_fo*BIN_cx*BIN_cy 4 3 6	12.820105	0.716809	17.88
BIN_fo*BIN_cx*BIN_cy 3 7 6	13.289088	0.744722	17.84
Parameter	Estimate	Standard Error	t Value
Intercept	59.267726	0.244797	242.11
BIN_focal_len*BIN_cx 7 8	32.285081	1.657493	19.48
BIN_focal_len*BIN_cx 4 3	12.820105	0.716809	17.88
BIN_focal_len*BIN_cx 5 3	24.628059	1.617804	15.22
BIN_focal_len*BIN_cx 6 4	17.914198	1.276252	14.04
BIN_focal_len*BIN_cx 6 8	16.468274	1.189067	13.85

CONCLUSION

This examination of the Lyft Level 5 Autonomous Vehicle dataset found patterns that strongly correlate with object recognition. A small group of focal lengths and optical centers appear to produce significant differences in the number of object annotations recognized. Both individually and, when combined, synergistically, these variable bins contribute to the identification of greater numbers of objects. Controlling for lengths of roadway driven, date, and camera size successfully demonstrated consistent effects associated with each predictor variable.

Future research on this dataset might proceed with aspects of objects that were observed, but not fully scrutinized. Figure 4, when separated into distinct days and times, might be a good starting point for analyzing traffic along the provided route. Moreover, we largely ignored the *CAM_FRONT_ZOOMED* camera setting in order to focus on the other 6 camera settings. Examination of this setting and comparison to data derived from the other 6 settings may produce interesting results. Finally, investigating the inflection points either directly after an object is detected or before an object is no longer detected proved beyond the scope of this project. With blank *num_lidar_pts* and *num_radar_pts* variables, insufficient data was available to give confident answers as to the limits of when objects are recognized or fail to be recognized. If future versions of the Lyft Level 5 Autonomous Vehicle dataset include meaningful values with these variables, investigation of these inflection points may be warranted.

REFERENCES

Lyft Inc. . (2019, July 23). *Lyft Level 5 AV dataset and nuScenes devkit tutorial*. Retrieved from Github: https://github.com/lyft/nuscenes-devkit/blob/master/notebooks/tutorial_lyft.ipynb

Lyft Level 5. (2019). *Lyft Level 5 Dataset*. Retrieved from Lyft: https://level5.lyft.com/dataset/

Marvin, R. (2019, March 1). *Which Self-Driving Cars Put in the Most Fully Autonomous Miles?* Retrieved from PC Magazine: https://www.pcmag.com/article/366797/which-self-driving-cars-put-in-the-most-fully-autonomous-mil

MathWorks Inc. . (n.d.). *What Is Camera Calibration?* Retrieved from MathWorks: https://www.mathworks.com/help/vision/ug/camera-calibration.html

Scale AI Inc. . (2020, January 6). *Open Datasets*. Retrieved from Scale: https://scale.com/open-datasets

Vincent, L. (2019, July 23). *Unlocking Access to Self-Driving Research: The Lyft Level 5 Dataset and Competition*. Retrieved from Medium: https://medium.com/lyftlevel5/unlocking-access-to-self-driving-research-the-lyft-level-5-dataset-and-competition-d487c27b1b6c

ACKNOWLEDGMENTS

We would like to thank our mentor, Dr. James Blum, for his guidance in moving through this project.

We also thank Lyft Incorporated (Inc.) and the Lyft Level 5 Team for releasing the Level 5 dataset as well as the corresponding tutorial and SDK. The Lyft Level 5 Autonomous Vehicle dataset and associated materials remain the property of Lyft, Inc. and are licensed under version 4.0 of the Creative Commons Attribution-NonCommercial-ShareAlike license (CC-BY-NC-SA-4.0). The v1.02-train dataset used was not altered before project commencement, but was significantly altered through merging, sorting, and renaming both schema as well as individual variables in preparation for answering our questions.

RECOMMENDED READING

- *Lyft Dataset SDK (https://github.com/lyft/nuscenes-devkit)*
- *Lyft Level 5 AV dataset and nuScenes devkit tutorial (https://github.com/lyft/nuscenes-devkit/blob/master/notebooks/tutorial_lyft.ipynb)*

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

John Tomaselli
john.tom.abc@gmail.com