C2SMART
CONNECTED CITIES WITH SMART TRANSPORTATION
A USDOT University Transportation Center

New York University
Rutgers University
University of Washington
The University of Texas at El Paso
City College of New York

COVID-19's Effect on Transportation: Developing a Public COVID-19 Data Dashboard

June 2022

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No.	2.Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle COVID-19's Effect on Transportation: Developing a Public COVID-19 Data Dashboard		5. Report Date June 2022 6. Performing Organization Code
7. Author(s) Kaan Ozbay, Jingqin Gao, Joseph Chow, Fan Zuo, Hang Dong		8. Performing Organization Report No.
9. Performing Organization Name and Address Connected Cities for Smart Mobility towards Accessible and Resilient Transportation Center (C2SMART), 6 Metrotech Center, 4th Floor, NYU Tandon School of Engineering, Brooklyn, NY, 11201, United States		10. Work Unit No. 11. Contract or Grant No. 69A3551747119
12. Sponsoring Agency Name and Address Office of the Assistant Secretary for Research and Technology University Transportation Centers Program U.S. Department of Transportation Washington, DC 20590		13. Type of Report and Period Covered Final Report, 4/01/2020-3/31/2022 14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

The COVID-19 outbreak has dramatically changed travel behavior in cities across the world. With changed travel demand, economic activity, and social-distancing/stay-at-home policies, transportation systems have experienced an unprecedented shift in demand and usage. Since the start of the pandemic, the C2SMART research team has been collecting data and investigating the impact of COVID-19 on mobility and sociability, including: passenger travel and freight traffic trends; mode shift and usage based on various policies; effect of social distancing policies on transit use and emissions; sidewalk, crosswalk, and intersection crowd density; and effect of COVID-19 Policies on Transportation Systems. Leveraging open data from multiple sources, this project features both traditional and innovative techniques, such as data mining and visualization, agent-based traffic simulation model and real-time computer vision technique, to help researchers and transportation authorities understand and observe the impact of the pandemic on transportation.

17. Key Words	18. Distribution Statement Public Access

19. Security Classif (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No of Pages 60	22. Price

Form DOT F 1700.7 (8-69)

COVID-19's Effect on Transportation: Developing a Public COVID-19 Data Dashboard

Principal Investigator: Kaan Ozbay
New York University
0000-0001-7909-6532

Co-Principal Investigator: Jingqin Gao
New York University
0000-0002-1718-2432

Fan Zuo
New York University
ORC-ID

Co-Principal Investigator: Joseph Chow
New York University
0000-0002-6471-3419

Hang Dong
New York University
ORC-ID

C2SMART Center is a USDOT Tier 1 University Transportation Center taking on some of today's most pressing urban mobility challenges. Some of the areas C2SMART focuses on include:

Urban Mobility and Connected Citizens

Urban Analytics for Smart Cities

Resilient, Smart, & Secure Infrastructure

Disruptive Technologies and their impacts on transportation systems. Our aim is to develop innovative solutions to accelerate technology transfer from the research phase to the real world.

Unconventional Big Data Applications from field tests and non-traditional sensing technologies for decision-makers to address a wide range of urban mobility problems with the best information available.

Impactful Engagement overcoming institutional barriers to innovation to hear and meet the needs of city and state stakeholders, including government agencies, policy makers, the private sector, non-profit organizations, and entrepreneurs.

Forward-thinking Training and Development dedicated to training the workforce of tomorrow to deal with new mobility problems in ways that are not covered in existing transportation curricula.

Led by New York University's Tandon School of Engineering, **C2SMART** is a consortium of leading research universities, including Rutgers University, University of Washington, the University of Texas at El Paso, and The City College of NY.

Visit c2smart.engineering.nyu.edu to learn more

C2SMART CONNECTED CITIES WITH SMART TRANSPORTATION

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

Acknowledgements

In addition to the funding support from C2SMART, some of the researchers were supported by NYU's Summer Undergraduate Research Programs. Traffic camera data used in this project was extracted from https://webcams.nyctmc.org/ and was originally compiled by the New York City Department of Transportation (NYCDOT). This study (#IRB-FY2020-4638) is reviewed by the University Committee on Activities Involving Human Subjects (UCAIHS) at New York University and it was determined that it does not involve human subjects as defined by 45 CFR part 46.102.

Executive Summary

The COVID-19 outbreak has brought profound changes to almost every aspect of transportation. With changed travel demand, economic activity, and social distancing/stay-at-home policies, transportation systems have experienced an unprecedented shift in demand and usage. Leveraging open data from multiple sources, this project developed a public all-in-one data dashboard (https://c2smart.engineering.nyu.edu/c2smart-data-dashboard/) with interactive analytics and visualizations which features both traditional and innovative techniques, such as data mining, agent-based traffic simulation model and computer vision for real-time detection, to help researchers and transportation agencies understand and observe the impact of the pandemic on transportation. This project aims to develop results-oriented approaches to aid decision-making and enhance the shared knowledge and resiliency of transportation systems in the face of future public health emergencies.

The COVID-19 Data Dashboard contains the following sub-boards/components:

I. Mobility Data Dashboard Integrated with Multiple Data Sources

This interactive mobility data dashboard consolidates public data sources to track the mobility trends on transportation systems in New York City (NYC) and Seattle. The mobility board presents multi-data views in terms of vehicular traffic volume, corridor travel time and citywide traffic speed, transit and paratransit ridership, micromobility usage, freight traffic, and risk indicators like reported crashes, pedestrian and cyclist fatalities, and speeding tickets. This sub-dashboard supports scenario analyses with these metrics in timeseries or comparisons to understand the temporal and spatial aspects of the data. This platform is updated regularly and continues to evolve with the addition of new data, metrics, and visualizations.

II. Sociability Board Based on Computer Vision

This project also developed a continuous, real-time pedestrian detection framework that uses public traffic camera feeds and deep-learning-based computer vision techniques to analyze sidewalk, roadway densities and monitor social distancing patterns in urban areas. This framework allows us to capture “perishable” data on vulnerable road users and vehicle flows without any additional infrastructure investment. It also provides data that assist with answering novel questions such as how often pedestrians maintained the recommended “6 feet” of social distance with the unfold of the COVID-19 pandemic. In addition, system deployment cost estimation and lessons learned are summarized. Since the project methodology relies on pre-existing deployed ITS infrastructure, implementation and maintenance are low cost and highly scalable. The system deployment cost, based on a 3-year estimate for the proposed pedestrian detection system with 68 cameras, ranges from $500 to $1700 per year, depending on the usage of local servers or cloud-based services.

Compare with the current usage of traffic cameras that often used for monitoring and passive detection for incident management, the benefits of the proposed approach include: 1) build on existing resources (e.g., CCTV cameras) to reduce agency costs, 2) no additional safety risk and labor cost for field checks as it uses remote approach and reference-free distance approximation, 3) enable real-time object detection for different road user classes to assist system and local monitoring, 4) account for COVID-19 impacts such as risk indicator on social distancing and 5) can be easily adopted to other cities/states and expanded to other use cases such as traffic counts, incident detection, and curb lane monitoring.

III. Agent-based Simulation Virtual Testbed

As a part of the COVID-19 research efforts, C2SMART's MATSim-NYC agent-based simulation virtual testbed was utilized to provide predictions of network performance and emission evaluation for phased reopening of COVID-19 in NYC. The findings are summarized in the COVID-19 data dashboard.

IV. COVID-19 Travel Survey

An infographic summarizing the findings from the C2SMART COVID-19 travel survey with 2022 respondents is also presented in the COVID-19 Data Dashboard. This is an online survey focusing on travel trends under the impact of COVID-19, which was administered from July to October 2020. The objective of the survey is to look at how different disadvantaged population groups, especially people with disabilities and low-income households, were affected by the changes as a result of COVID-19 as the crisis may magnify the existing systemic inequities and produce new mobility disparities for certain disadvantaged populations. The survey data can help us understand what the real needs for disadvantaged population groups are, in the event of a pandemic. This, in turn, will aid the cities and transportation agencies in assessing a more equitable planning process to address their short-term and long-term trip needs.

V. White Papers

Over the course of the COVID-19 Data Dashboard's development, the team has published multiple white papers detailing this research. All white papers available for open access download.

Overall, the interactive data dashboard has successfully provided a public platform for visualizing and examining short-term and long-term passenger travel and behavior changes, the effect of social distancing policies on transit use and mobility patterns, including the decrease and recovery of vehicular traffic and transit ridership, changes in speeding behavior and reported crashes, bike usage, and parking and freight impacts in response to the pandemic. It will continue to serve as a data visualization hub post-pandemic.

Table of Contents

List of Figures

List of Tables

1. Introduction

The outbreak of the novel coronavirus COVID-19 and the ensuing restrictions in most cities around the world have resulted in unprecedented changes in the daily life of people. Most of these changes are unpredictable, for example, it was the first time ever that traffic congestion disappeared across many city streets. With little information on similar historic pandemics, collection of perishable mobility, safety, and behavior data and learning from them becomes the key for decision-makers [1].

Since transportation acts as a bridge connecting people to their places of interest, it is important to develop timely indicators to track transportation changes due to COVID-19 that allow decision makers to work on effective strategies immediately actionable in the current environment. Although many individual efforts have been made by agencies and organizations to publish open data on traffic, such as ridership or volumes, there is a need to have a data visualization "hub" that gather the data into one platform timely and translate the information into easy-to-understand visualizations and statistics.

Since the start of the pandemic, the research team has been collecting data and investigating the impact of COVID-19 on mobility and sociability, including:

- Passenger travel and freight traffic trends
- Mode shift and usage (e.g., increase in micromobility usage)
- Effect of social distancing policies on transit use and emissions
- Sidewalk, crosswalk, and intersection crowd density
- Effect of COVID-19 Policies on Transportation Systems
- Behavior changes due to less congested roadways/transportation systems

Data is critical to understanding the impacts and needs in times of crisis. For example, mobility is one good indicator of the effectiveness of Nonpharmaceutical interventions (NPIs) such as social distancing policies during the outbreak and reveals the recovery of the cities. However, simply collecting data is not enough. Data visualization is one of the best tools to understand the data and communicate findings in constructive ways and visualizing various types of data during the COVID-19 pandemic can help us to fast track the changes and develop effective strategies.

The objective of the research is to develop an all-in-one data dashboard with interactive analytics and visualization of various traffic-related data sources. The data dashboard aims to allow the users to explore the impact that COVID-19 has had on transportation systems in affected cities with a focus on New York, the epicenter of coronavirus in the U.S. Information from Seattle and some cities in China were also explored as time lag between different cities may offer some insights into how travel patterns evolve in the context of COVID-19.

To achieve this goal, mobility data from multiple open sources, including transit and paratransit ridership, vehicular traffic volumes, shared bike system usage, freight volume, taxi and for-hire vehicles trend, crash records, and speeding tickets, were collected. The data collected in the dashboard can be used in various ways. For example, examine short-term passenger travel trends and behavior changes due to social distancing policies on transit use and mobility patterns, such as the sharp decrease in vehicular traffic and transit ridership, changes in bike usage, parking and freight impacts in response to the pandemic.

In addition, through the fast-evolving pandemic, one term that everyone has quickly become familiar with is “social distancing.” Social distancing refers to efforts including avoiding mass gatherings, closing public places, and keeping a sufficient distance (at least 6 feet) between people to reduce disease spread by minimizing the physical distance and frequency of human contacts [2]. Although social distancing orders are mandated, it is not clear that how people are responding to these policies as people may still go outside for essential activities (e.g., work, grocery, etc.). In such a context, investigating the crowd density and actual frequency of social contact between people becomes crucial to measuring the effectiveness of the policy and reducing the chances of community transmission [3]. Moreover, when the pandemic begins to ease, volumes of people traveling within a city will start to increase gradually. It is also possible that the pandemic may lead to changes in how people prefer to travel within urban areas, potentially eschewing typically crowded modes such as public transportation or car-sharing, and instead choosing more solitary modes, such as walking [3]. This shift in preference may result in higher levels of interactions on dense city sidewalks. These behavioral changes and their impacts on urban environments are unpredictable and such data is usually not collected.

Early pedestrian detection programs used manual methods, such as data collection sheets, clickers, or pushbutton-based detection. In recent years, improvements to sensors have enabled the presence and movement of pedestrians to be detected continuously, more accurately, and across a greater geographic area. This has been accomplished with a variety of existing and emerging detection technologies, including Bluetooth and Wi-Fi, infrared sensors, microwave, radar, radio beam, thermal, vision-based, smartphone-based, and LiDAR sensors [4].

To shorten the gap of collecting precise social distancing information, this study developed a data acquisition framework and a deep-learning approach with a pre-trained convolutional neural network to process public video data to quantify social distancing effects on urban street. Object detection and distance approximation between pedestrian pairs are applied to traffic camera videos at multiple NYC and Seattle locations to analyze local social distancing patterns. Real-time traffic camera data collected in NYC was used to demonstrate the feasibility and validity of the approach. The quantified heterogeneity in terms of pedestrian density, social distance distribution, temporal variations can be used to inform residents of the potential risk of exposure to an urban environment and assist in evaluating the effectiveness of relevant public interventions [3].

Moreover, the crisis can magnify inequities in the existing transportation system as well as produce new mobility disparities for certain disadvantaged populations. For example, for people with disabilities, the effects of the pandemic-induced restrictions are compounded by a lack of available alternative transportation modes [5]. As a part of the efforts to assess the CVOID-19 impacts on transportation systems and travelers' behavior, this study also highlights an agent-based simulation model's results to predict the impact of NYC's phased reopening strategies and findings from an online COVID-19 travel survey that was administered by the C2SMART center from July to October 2020. The survey has two phases and collected a total of 2,022 responses with a focus on people with disabilities (27% of the respondents) and old population (13% of the respondents).

2. Dashboard Architecture

The core of the dashboard tool relies on data automation pipeline and data analytics algorithms. The data automation allows an up-to-date platform with minimal infrastructure maintenance efforts and can be customized for solid performance on large scale data processing, machine learning and graphic-intensive analysis. The ingestion instances inside the platform consume the raw data stream and evaluate it based on data accuracy, timeliness, validity and granularity. Once the evaluation is completed, the outcome data will be dispatched to the internal data warehouse/online tools (e.g., google sheets) for further analysis. The data dashboard is capable of performing real-time data analytics and visualization. Figure 1 illustrates the overall architecture of the dashboard.

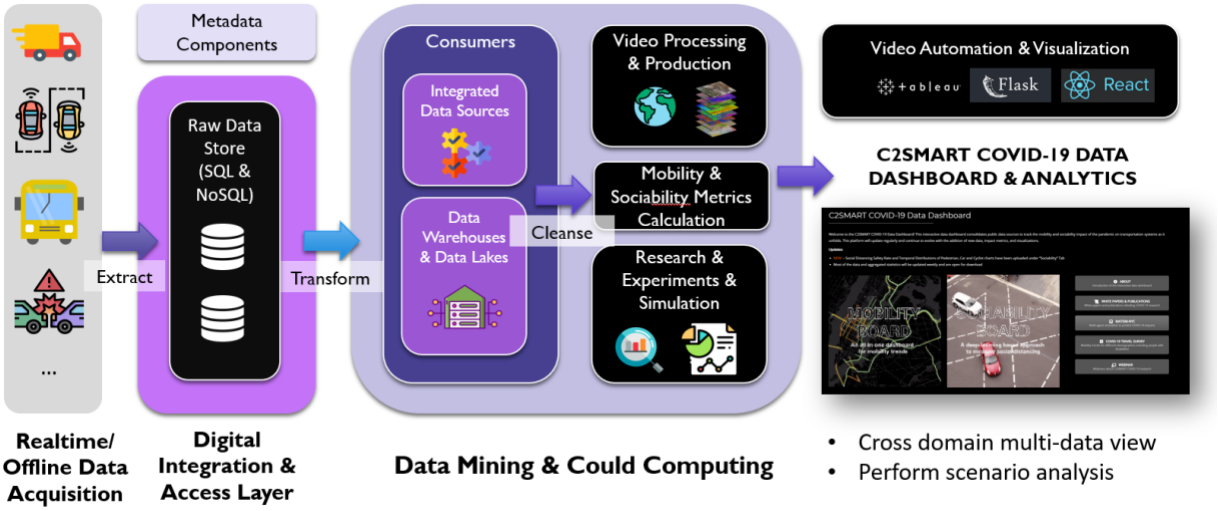


Figure 1: Overall Dashboard Architecture

2.1 Mobility Board Architecture

The mobility data dashboard is a collection of interactive visualizations based on mobility data from multiple data sources. The visualizations are built using Tableau Public and updated through an automated data pipeline.

Raw data are fetched from open data sources. Each data source has a corresponding script that is responsible for downloading the data, transforming it to the desired format, and uploading the processed results to Google Sheets or other storage engines. Most of the scripts are executed automatically in a weekly routine. The others are executed manually for advanced analysis. Tableau Public automatically synchronizes with Google Sheets and displays the result. As a data analysis and visualization tool, Tableau can further transform the data and enable rich interaction. The processed data may also be used for other research or visualization purposes.

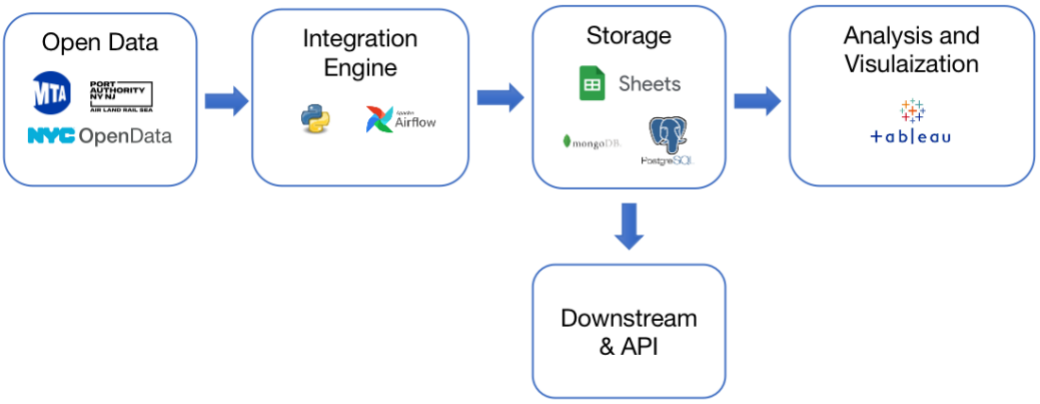


Figure 2: Mobility Board Architecture

2.2 Visualization Tool Comparison

When choosing a visualization solution, the first decision is between developing an application from scratch and using SaaS (Software as a Service) such as Tableau. A Self-developed application offers the highest flexibility and customization but requires time and technical skills to build. Scalability is also a consideration. When the volume of data grows, the application will get more complex and harder to maintain. Existing software like Tableau, comparatively speaking, is easy to get started and can significantly shorten the development cycle. This software can access common databases and file formats and can load relatively large amounts of data efficiently. This means that the back end is no longer required. At the same time, they can take on some of the data analysis work which simplifies the pre-processing scripts. They allow for the rapid creation of rich data interactions without writing code.

The major drawback is that they are less flexible. It would be difficult to optimize or extend further when the desired functionality exceeds what is supported. In our case, SaaS meets the need in the current stage, so we chose to use SaaS to build data visualizations quickly and cost-effectively.

We compared three common visualization software: Tableau, Power BI, and Carto. In terms of visualization, we have a variety of data types, including time series and geographic data. Line charts, bar charts, heat maps, various maps, and many other types of visualization may be involved. Carto DB focuses on maps only and does not meet the demand. Another consideration is data connection and access control. Both Tableau and Power BI can connect to mainstream data sources and formats. Since we tend to use the free version, data used for visualization will be made public in either of the software. However, this is acceptable because the data must be public to enable front-end interaction. Privacy issues can be avoided by loading only the data needed for visualization into the software. In addition, in Tableau Public, although users can download the transformed data, they cannot access the original Google Sheets, which also helps to control data access to some extent.

As a result, we choose Tableau Public to build the visualizations. Besides what is mentioned above, it has some other advantages. First, it supports Javascript API and is therefore easy to embed into web pages using pure front-end code. Second, it supports editing on multiple systems and platforms, including Windows, Mac, and browsers, making it easier to collaborate. Last, with many tutorials and a large community, it is easy to learn. It also has some limitations, for example, it supports loading up to one million rows of data. When exceeding this limit, the preprocessing scripts need to compress the data. In addition, it is based on Canvas and DOM rendering on the web and does not support GPU acceleration. When the size of data increases, performance may drop. Thus, it is not recommended to render too many charts on the same page.

Table 1 Visualization Tool Comparison

	JavaScript Libraries	**Tableau**	**Power BI**	**Carto**
Flexibility	High	Medium	Medium	Medium
Scalability	Good	Good	Good	Good
Data Access Control	High	Medium	Low	Low

Development Time	High	Low	Low	Low
Cost	Free	Tableau Public with limitation	Free version with limitation	Free version with limitation

2.3 Data Automation

2.3.1 Extract, Transform, and Load (ETL)

Data Automation is a data pipeline that consists of three basic components: Extract, Transform, and Load (ETL). Extract is the process of extracting data from multiple sources. In this process, we fetch open data by calling the APIs or reading existing files and databases. Transform is the process of transforming data to the desired structure. In our case, data is often aggregated into greater temporal and spatial granularity to reduce network time and rendering time or grouped by year to compare the changes over time. Load is the process of loading the transferred data into a Database or Data Warehouse. For visualization purposes, most of the data is loaded to Google Sheets which is connected to Tableau. They may also be used for other research purposes and may be stored in databases like MongoDB and PostgreSQL, which are more flexible and supports richer data formats and larger amounts of data.

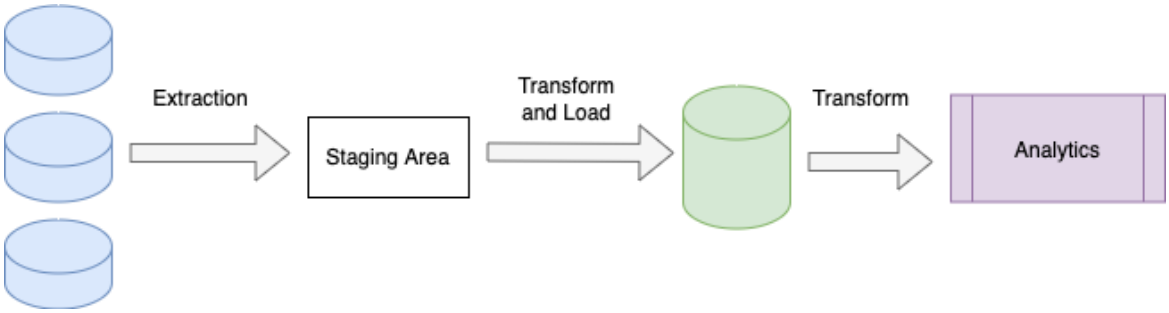


Figure 3: Extract, Transform, and Load (ETL) Process

Automation can help to scale up the project when the number of datasets grows bigger. At the same time, it also brings extra development and maintenance costs. In this project, we started by manually performing the extract, transform, and load process, and then automated the pipeline by harnessing Tableau's ability to synchronize data from Google Sheets. Right now, most of the pipelines are fully automated under the control of a job scheduler.

2.3.2 Development Process

During the development of the COVID-19 dashboard, we've worked on several iterations to update the dashboard from manually process to automatic process. In the early stage of the development, many manual updates were used to display the data timely. Later, with the increasing variety of the data, semi-automatic and fully automatic process were applied to reduce the efforts needed for frequent maintenance.

In manual updates (Figure 4), there are two places where manual intervention is required. One is running the script for Extract and Transform (ET). The other is Transform and Load (TL), updating the Tableau Workbook's data source with the result. When multiple data sources need to be updated frequently, such an update would be cumbersome.

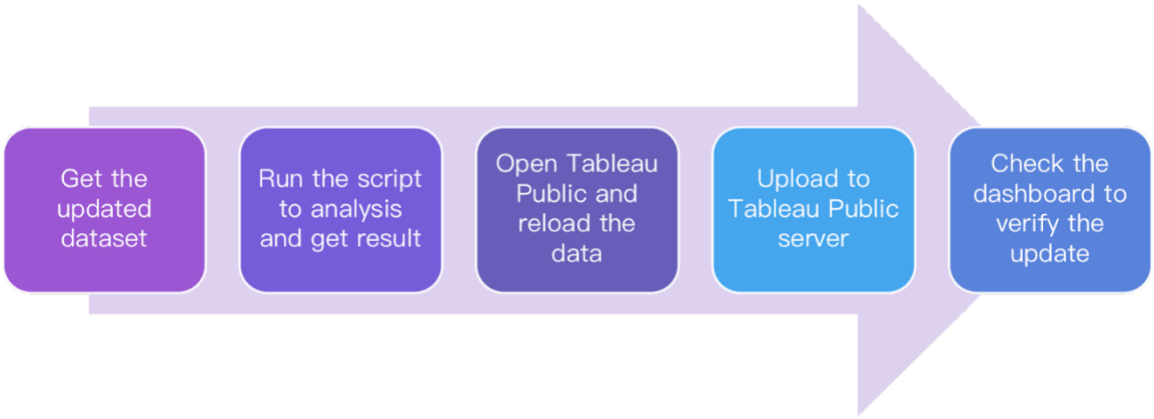


Figure 4: Manual Update Process

In semi-automation, Extract and Transform (ET) / Transform and Load (TL) are combined into one script (Figure 5). Since Tableau supports synchronization from Google Sheets, it is possible to automate the upload process. The transformed data are written to Google Sheet programmatically through Google Sheets API.

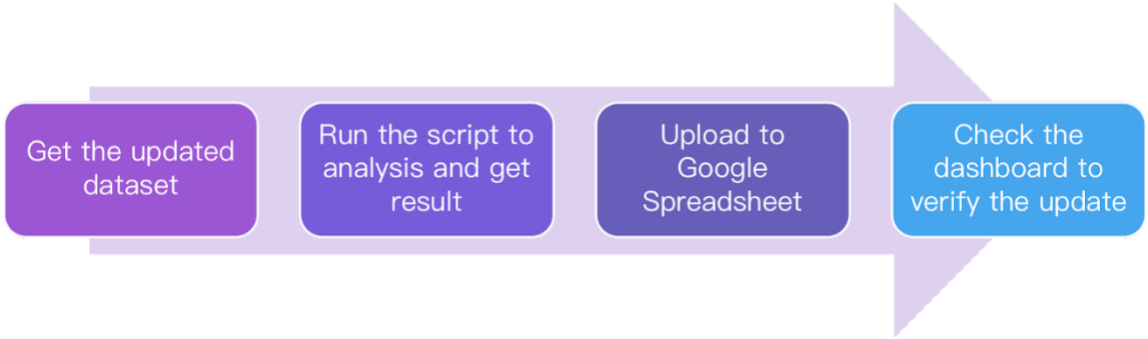


Figure 5: Semi-automation Update Process

In full automation, ETL is completely programmed. We use a job scheduler to execute the data processing script in a weekly routine (Figure 6). The script will fetch the latest data, transform them into the desired format, and write the result to Google sheet. Tableau dashboards that are connected to the sheets will then be updated.

The task scheduler runs under monitoring. From the admin interface, we can see if the update is successful and when it happened. If something goes wrong, we can quickly find the cause through logs and re-execute the ETL process in a few clicks.

All the weekly updated dashboards are fully automated. Ideally, they do not require any human intervention. However, since the data format and API may change when the open datasets are updated, inspections and maintenance are still required. Current practice the team is using is to perform a bi-weekly inspection. Moreover, some of the visualizations are updated in a semi-automation manner, as the input data is more complex and requires human intervention for advanced analysis. These data are updated less frequently (e.g., bi-weekly or monthly).

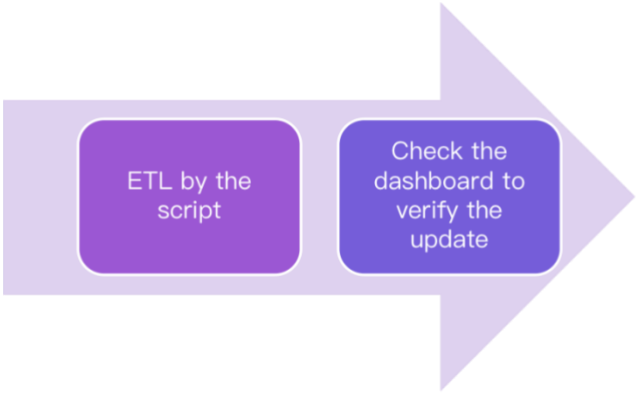


Figure 6: Full Automation Update Process

3. User Interface and Sub-boards

3.1 Homepage and User interface Design

The dashboard is implemented as a component of the C2SMART home page. The framework of the dashboard is written in JavaScript and CSS. Tableau Visualizations are embedded on the page Tableau Javascript API which allows dynamically loading and resizing visualizations. The design of the user interface of the COVID-19 Data Dashboard has iterated with three versions. Version 1.0 homepage used

a direct combination of multiple Tableau dashboards (Figure 7). Although version 1.0 is visually appealing, with the increase of the collected data, the loading speed of the homepage became slow.

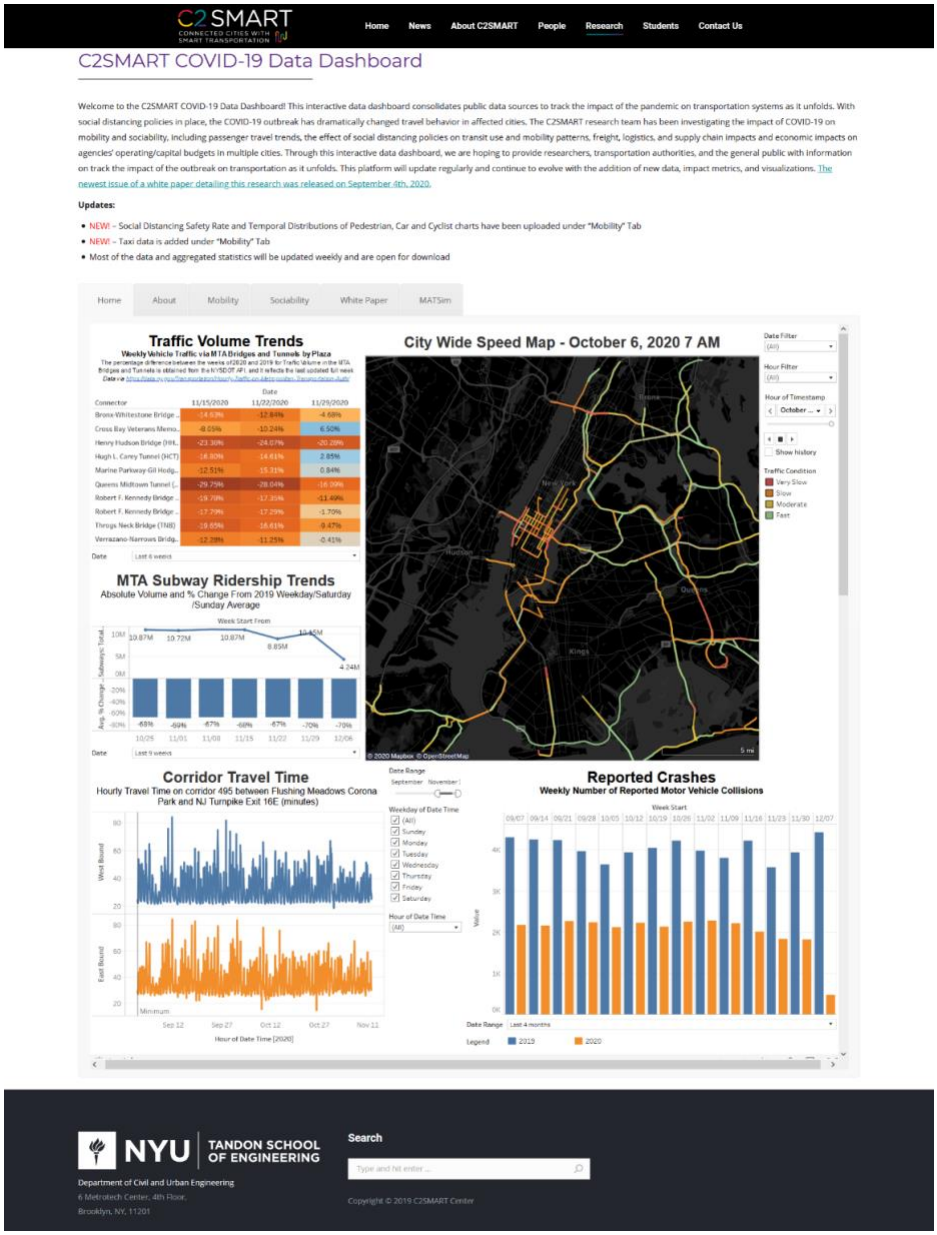


Figure 7 C2SMART COVID-19 Data Dashboard Homepage (Version 1.0).

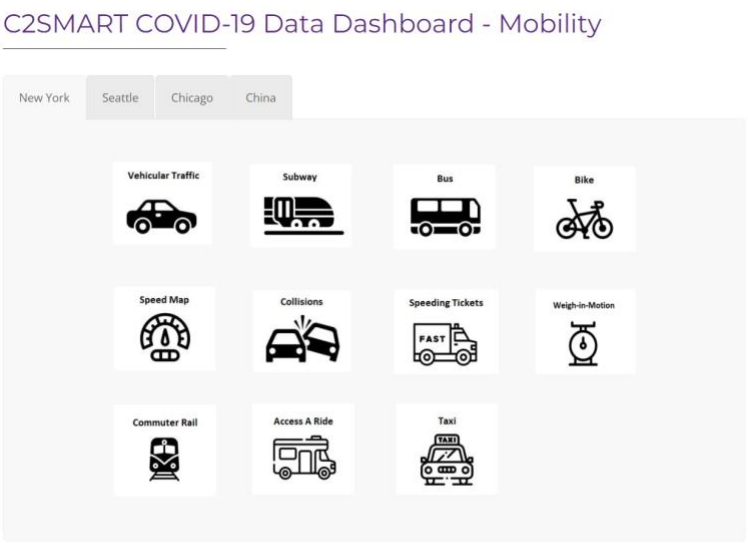


Figure 8 Lading Page of the Mobility Sub-board (Version 1.0 & 2.0).

Version 2.0 of the homepage removed the Tableau dashboards from the home page, thus improved the loading speed of the dashboard. However, since no up-to-date information was shown on the homepage, users need to click into each sub-board to see when the data was last updated (Figure 9). The archived version of the dashboard (version 2.0) can be viewed at https://archive.c2smart.engineering.nyu.edu/covid-19-dashboard/.

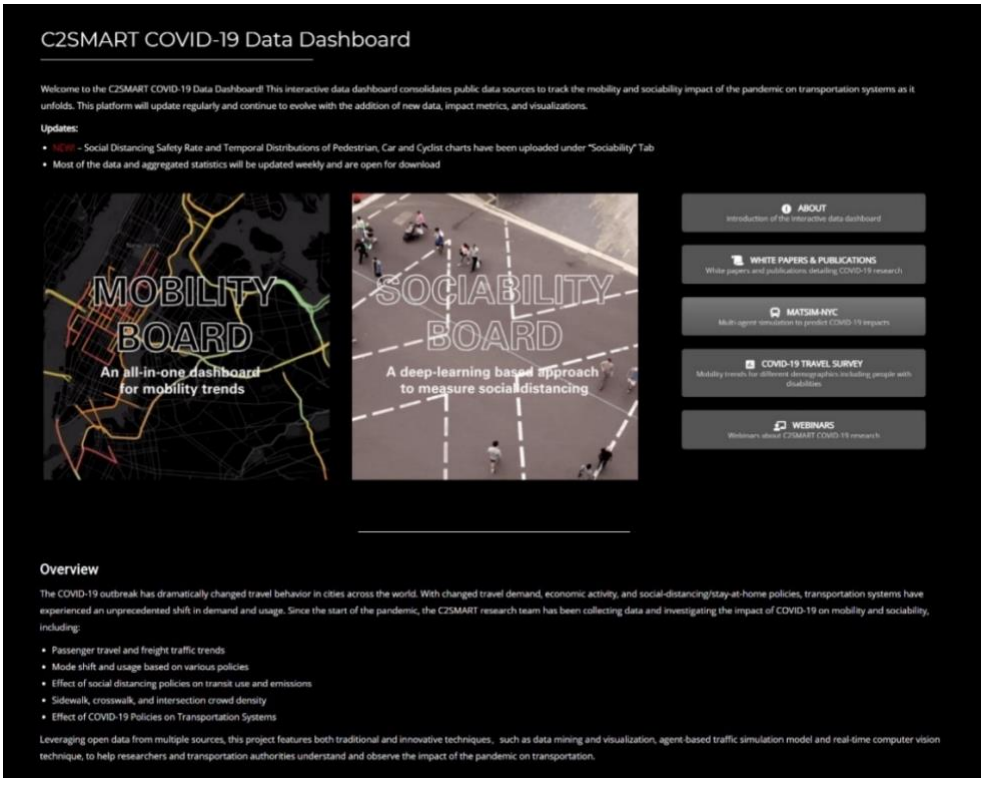


Figure 9 C2SMART COVID-19 Data Dashboard Homepage (Version 2.0).

Finally, the team updated the dashboard to version 3.0 with a new design of its homepage that combines certain up-to-date COVID statistics along with a new categorization of mobility-related data to avoid redundant navigations through the dashboard (Figure 10). The latest dashboard (version 3.0) can be visited at https://c2smart.engineering.nyu.edu/c2smart-data-dashboard/.

The dashboard contains two main sub-boards: mobility and sociability. The mobility board presents multi-data views such as weekly vehicular traffic or transit ridership while the sociability board presents average and maximum intersection crowd density and pedestrian social distancing compliance rates based on information extracted from 68 real-time traffic cameras. In addition, the dashboard also connects to C2SMART's MATSim agent-based simulation virtual testbed [6] that provides network performance and emission evaluation for reopening phases and C2SMART COVID-19 Travel Survey that collected responses nationwide about travel behavior changes.

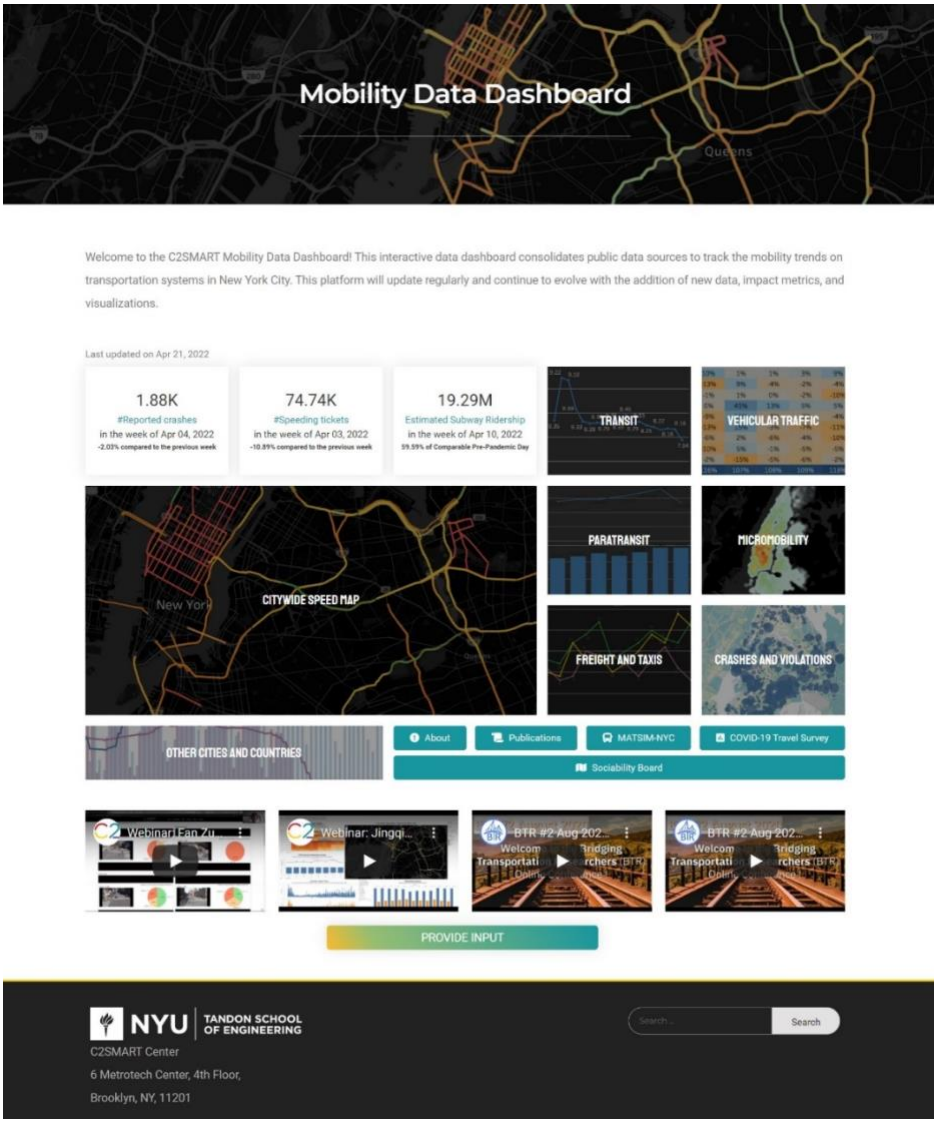


Figure 10 C2SMART COVID-19 Data Dashboard Homepage (Version 3.0).

In version 3.0, the front-end implemented a responsive design using the grid layout. Dashboards are grouped into different categories. Users can see snapshots of each category and select the topic of interest. In each category, they can switch between different visualizations by clicking on the tabs. Only one worksheet is rendered on the page at a time to guarantee performance.

Although Tableau also has the functionality to display multiple worksheets (single-view visualizations), by using dashboards (collections of worksheets) and stories (a sequence of worksheets or dashboards), which is used in the initial design of the dashboard (version 1.0). We decided not to use this function in version 2.0 and 3.0 and manually implement our dashboard based on the following reasons. First, displaying many worksheets on one page would slow down the data loading and rendering speed.

Second, since the data is from multiple sources and many of them are irrelevant, combining them into one tableau workbook brings unnecessary complexity to data management. Finally, the customized interface allows for better responsiveness and easier sharing of worksheets.

For users to read the latest data at first glimpse, we set up three panels: Reported Crashes, Speeding Tickets, and Estimated Subway Ridership, which show the most recent data aggregated by week and how it changed compared to the previous week or before the pandemic. These values are updated automatically as a by-product of the automated data pipeline. When performing weekly updates, there is a script that aggregates the data and stores the result in the database of the web page. This helps to accelerate data fetching and alleviate the burden of the storage engine for the original data. Users can investigate the corresponding visualizations by clicking on these panels.

3.2 Mobility Data Dashboard Data Collection and Visualizations

Multiple public data sources from NYC, Seattle, Chicago and various cities in China were consolidated into the mobility dashboard. Table 2 lists the collected data.

Table 2: Data Collection List

Data Category	Name/Data Source	Region
Vehicular Traffic	Weekly Vehicle Traffic via MTA Bridges and Tunnels, by Plaza	NYC
Vehicular Traffic	Overall Vehicle Traffic via MTA Bridge and Tunnels	NYC
Vehicular Traffic/Freight	Monthly Vehicular Traffic at PANYNJ tunnels and bridges (All Crossings)	NY/NJ
Transit	MTA Subway Ridership Trends	NYC
Transit	MTA Bus Ridership Trends	NYC
Transit	MTA Bus Speed Trends	NYC
Transit	MTA Long Island Rail Road Ridership Trends	NYC
Transit	MTA Metro North Railroad Ridership Trends	NYC
Paratransit	MTA Access-A-Ride Ridership Trends	NYC
Freight	Weigh-in-motion stations	NYC
Speed & Travel Time	NYCDOT real-time traffic speed map	NYC
Speed & Travel Time	Travel time (Virtual sensors)	NYC & NJ
Risk Indicators	NYPD crash reports	NYC
Risk Indicators	Camera violations – School zone speeding	NYC
Risk Indicators	311 service requests	NYC
Micromobility	Citibike ridership and spatial trends	NYC
Taxi & TNC	Yellow, green taxi and for hire vehicle trips	NYC

Social Distancing, Pedestrian Density	Real-time traffic cameras	NYC & Seattle
Speed & Travel Time	Travel time (Google API)	Seattle
Vehicular Traffic	Traffic volume (WSDOT detectors)	Seattle
Micromobility	SDOT bike/pedestrian counts	Seattle
Transit	Public transit demand	Seattle
Taxi & TNC	Taxi Ridership	Chicago
Transit	Subway Ridership	Six cities in China

MTA = Metropolitan Transit Authority, PANYNJ = Port Authority of New York and New Jersey, TNC = Transportation Network Company

Depending on the data, different performance metrics based on their temporal trend or spatial distribution are then generated and visualized. Figure 11 to Figure 17 illustrate some examples of the sub-boards. It is worth noting that these sub-boards allow users to perform interactive data analytics. For example, besides showing the overall daily traffic trend, temporal comparison of selected dates can be performed in the Seattle traffic volume trend sub-board (Figure 15).

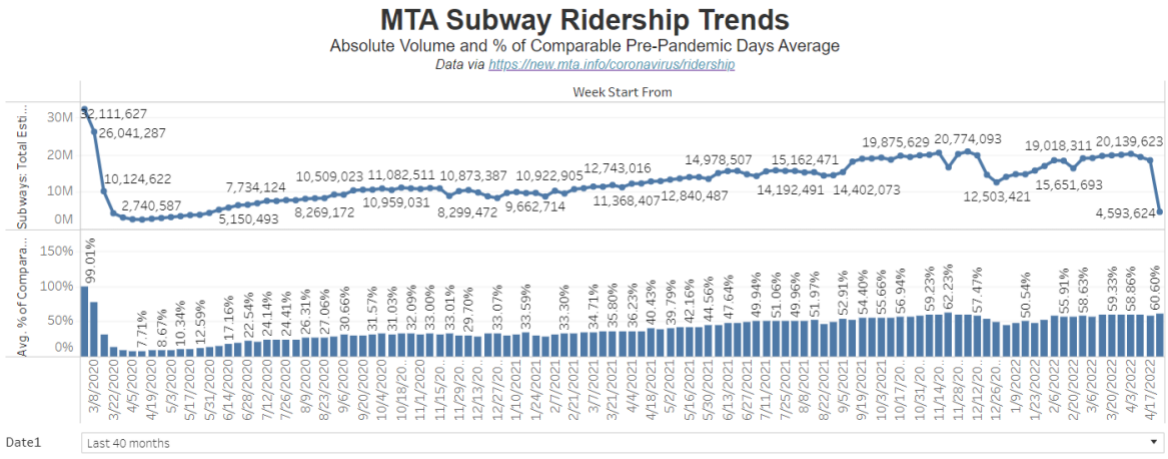


Figure 11 NYC Subway Ridership Sub-board.

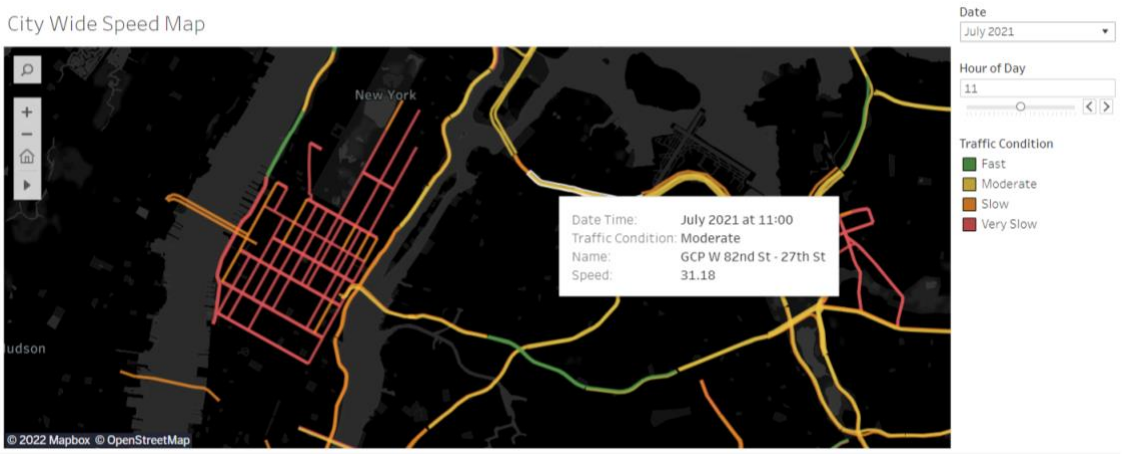


Figure 12 NYC City-wide Speed Map Sub-board.

Figure 13 NYC Speeding Ticket Sub-board (June 2020).

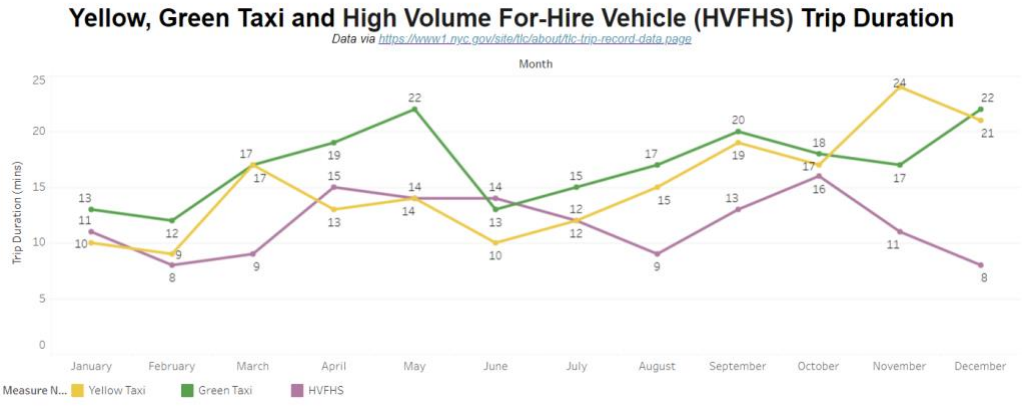


Figure 14 NYC Taxi Trip Duration Sub-board.

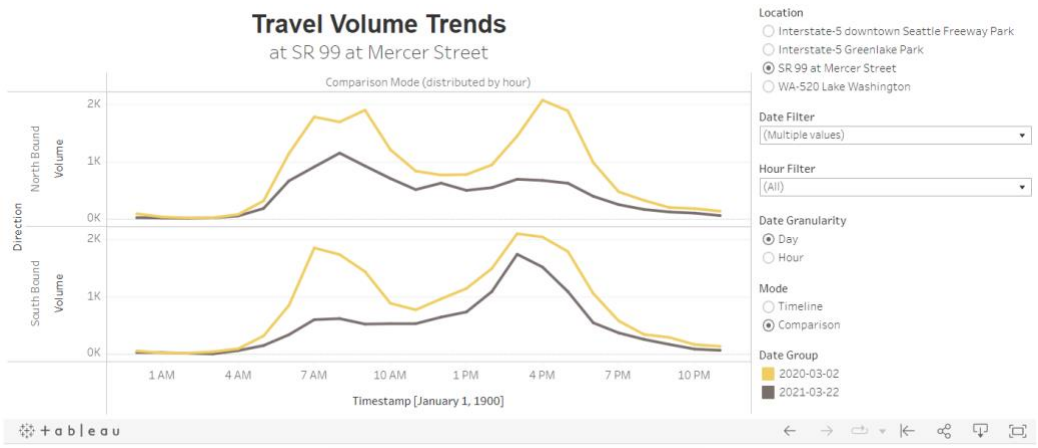


Figure 15 Seattle Traffic Volume Trends Sub-board (SR 99 at Mercer Street).

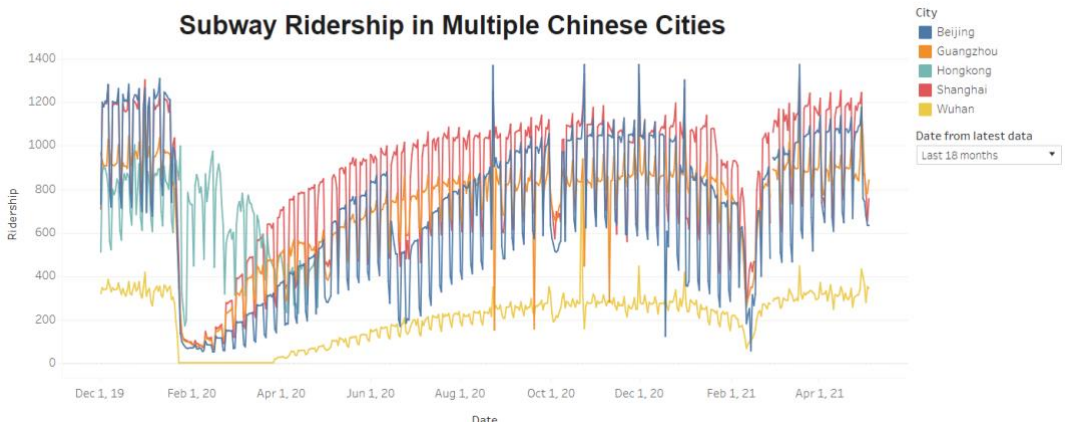


Figure 16 China Subway Sub-board.

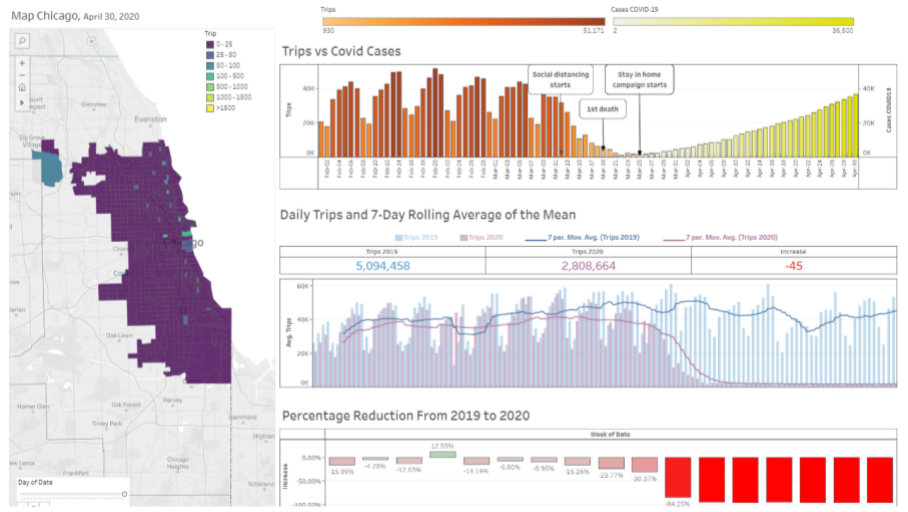


Figure 17 Chicago Sub-board.

3.3 Sociability Board Based on Computer Vision

This sociability board presents the findings from the deep-learning-based video detection system that is able to quantify pedestrian and vehicle densities and sociability indicators from publicly available real-time traffic cameras in NYC and Seattle. Perishable data was collected for 68 locations in NYC and 1 location in Seattle, including locations near hospitals, subway stations, and meal distribution centers. Besides temporospatial changes in pedestrian densities, we also quantify the safety rate of social distancing (i.e., % of people following social distancing guidance). Due to the numerous burdens of safely deploying any effective field data collection approaches during the pandemic, there are still few efforts to obtain precise information on social distancing patterns of pedestrians in urban environments.

A deep-learning approach with a pre-trained convolutional neural network and customized post-processing filters was designed to mine the massive amount of public video data captured in urban areas. As mentioned in chapter 3, the results quantify pedestrian density and microscopic pedestrian social distancing patterns using a generalized real-distance approximation method. Quantifying pedestrian density and social distancing adherence will provide decision-makers with a better understanding of the new norm of urban mobility amid the pandemic, and what patterns might hold for individual mobility post-pandemic or in the event of a future pandemic.

Compare with the current usage of traffic cameras (often used for monitoring and passive detection for incident management), the proposed approach leverages existing transportation infrastructures and its benefits include: 1) build on existing resources to reduce agency costs, 2) no additional cost and safety risk for field checks as it uses reference-free distance approximation, 3) enable real-time object detection for different classes (pedestrians, cars, buses, trucks and cyclists) to assist both system and local monitoring, 4) account for COVID-19 impacts such as risk indicator on social distancing and 5) can be easily adopted to other cities/states and expanded to other use cases such as traffic counts, incident and illegal parking detection, and lane occupancy monitoring. Figure 18 presents the user interface of the sociability board.

Figure 18 Sociability Board.

4. Measuring Social Distancing

4.1 Related works

Social distancing practices have been proved to be effective Nonpharmaceutical Interventions (NPIs) during the COVID-19 pandemic. For instance, De Oliveira, Silvano B., et al. [7] showed that the proportion of individuals staying home had a strong inverse correlation with the time-dependent reproduction number R(t) (rho<-0.7), which is a measure of disease transmissibility using aggregated mobile phone data. However, how to measure social distancing remains a challenge.

Instead of gathering data from conventional survey approach, recent advancements in information and communication technologies have made it possible to collect empirical data relevant to human behavior and contacts in real-world conditions. Solutions such as tracking WiFi and Bluetooth traces [8, 9], smartphone applications [10-12], and active radio frequency identifications devices (RFID) or other wireless sensors [13-16] have been lately employed to collect the close proximity data of human-to-human interactions [3]. Nevertheless, many of these applications are better suited for indoor localization and using them in dynamic outdoor environments may reduce their accuracy. Moreover, there are significant concerns about the scalability and privacy protection of such applications [9, 10].

Primary crowd detection technologies that can be used for both outdoor and indoor environments are summarized in

Table 3. Technologies such as LIDAR or Thermal sensors have a medium to high installation or equipment cost, and most of them usually integrate with existing systems, such as mobile phones and wearable devices (e.g., WiFi, Bluetooth, cellular, UWB, inertial sensors, visible light). A drawback of such approaches is that the device always needs to active some operational modes (e.g., Bluetooth mode), and the position accuracy can drop when the mobile phone is placed inside a pocket or bag [17]. Moreover, the temporal resolution (e.g., detection time) of these technologies is not sufficient for measuring many close contacts within a short duration [15, 18, 19] and its nature of deploying new hardware in the field or software on the phone makes it infeasible to quickly adapted to the current fast-evolving COVID-19 crisis. Therefore, very little precise data exist on COVID-19 related social distance behavior, especially continuous data combining environment facts and high temporal resolution.

Table 3 Crowd Detection/Positioning Technologies for Both Outdoor and Indoor Environments [3]

Technology	Range	Cost	Position Accuracy	Privacy
Wi-Fi	Outdoor: 100m Indoor: 50m	$	Medium (1m-5m)	Low
Bluetooth	Outdoor: 55-78m Indoor: 15-35m	$	Medium (1m-2m)	Low
Cellular	Short to Long	$	High (Less than 50cm)	Low to High
Ultrawideband (UWB)	Short to Medium	$	High (Less than 10cm)	High
Radar/LIDAR	100m	$-$$$	High	High
Seismic sensor	Short (~15m)	$$	High	High
Computer Vision	Depends on the cameras	$–$$	Low to High	Low
Infrared/Thermal sensor	IRP 1-10m, THC a few km	$–$$$	Medium	Low to high
Inertial sensors	Not applicable	$	Medium (Less than 1m)	High
Visible light	Short	$	High (1 cm – 20 cm)	High

Note: Synthesized based on References [17] [20] and [21]

Among the technologies discussed in

Table 3, computer vision offers a low-cost, appealing alternative for pedestrians detection [22] or safety assessment [23-25] as surveillance cameras have already installed in many cities. This became important as Lobe et al. [26] pointed out that although the pandemic provides a unique opportunity to study the crisis itself, social distancing mandates are restricting traditional in-person investigations of all kinds. For example, the Institutional Review Board (IRB) suspended all in-person human subjects research activities in response to the pandemic. The use of computer vision technology, in particular of object detection, provides a free-of-risk approach for data collection during the crisis. It can turn cameras into "smart"

sensors, which are capable of detecting crowd density and identifying whether people comply with social distancing requirements or not in real-time [17].

Two types of approaches that are commonly used in the computer vision domain include the following:

- **Region-based approach** (e.g., Fast-RCNN [27], Faster-RCNN [28], Mask RCNN [29], and RetinaNet [30]): This approach detect humans from images in two stages, including the region proposal and the procession according to the regions [31]. Although region-based approaches have a high detection accuracy, its applications may be limited due to its high complexity.
- **Unified-based approach:** This approach includes the You Only Look Once (YOLO) model [32], its improved versions [33-35], and Single Shot Multibox Detector (SSD). These approaches map the image pixels to bounding boxes and class probabilities to detect humans or objects and usually is faster than region-based methods [17].

Combining with distance approximation methods, these computer vision techniques can identify whether the number of gathering people complying with social distancing requirements or not, therefore, help in revealing human contact patterns and effectiveness of the social distancing policies [3].

Moreover, most relevant recent studies used aggregated mobility data with the assumption of homogeneous behavior within a city, state, or county [7, 36, 37]. Although such an approach is helpful for macro-level analyses, it is possible that different cities or regions may have non-synchronized social distancing patterns in open urban spaces. The computer vision-based approach leveraging existing data from cameras in cities is feasible to provide detailed spatiotemporal heterogeneity and dependence of social distancing patterns at a micro-level [3].

4.2 Reference-Free Video-to-Real Distance Approximation-Based Urban Social Distancing Analytics

For the sociability board, the team developed a novel video-to-real distance approximation-based method without the use of field measurement and homographic computing. By calculating the specific real-pixel distance rate for each person with a fixed real-height assumption and the pixel height obtained from the videos, the technology can approximate the distance between pedestrians without any facial features or other personally identifiable information. This generalized, reference-free distance approximation method allows the implementation of large-scale social distance calculations under different environment conditions (e.g., cameras with different angles). In the case of measuring social distancing, if the distance is less than a threshold of 6 feet, the event is flagged and recorded. **Figure 19** presents the overall workflow of the proposed methodology. The program is developed using Python and the primary machine learning modules are supported by open-source libraries, including TensorFlow [38], Keras [39], and ImageAI [40].

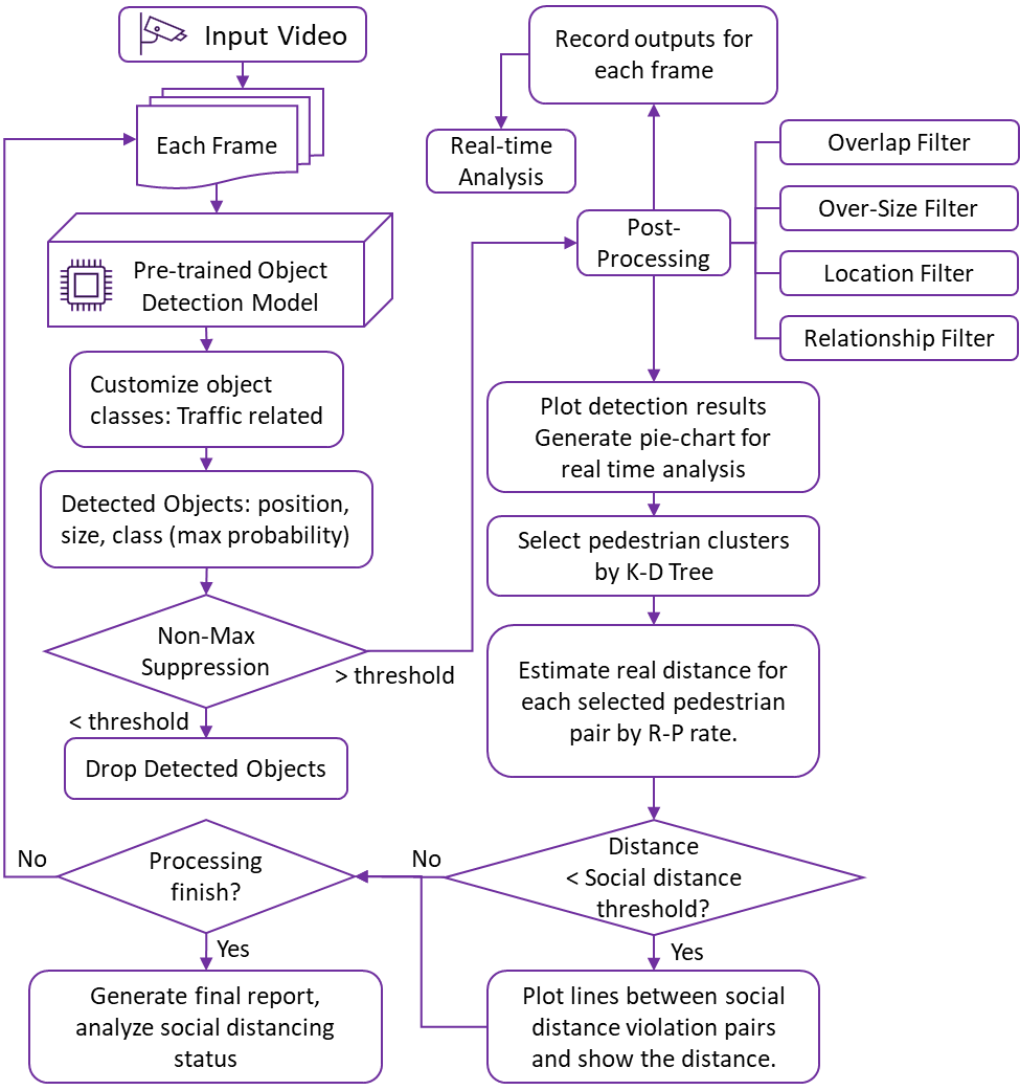


Figure 19 Proposed workflow of collecting and mining pedestrian social distance data from publicly available surveillance video data [3].

4.2.1 Pre-trained Object Detection Model

The object detection model detects objects (e.g., pedestrians) from each frame and finds boxes around the objects. The detector then returns a list of predicted potential classes of each object with probabilities and classifies the object into a relevant class type based on the highest probability. A pre-trained pedestrian detection model is selected for use in this study as pre-trained state-of-the-art object detection models are shown to have good generalization capability allowing efficient deployment to new environments, even with different video resolutions or camera angles [41, 42]. The Non-Max

Suppression method [43] with a fine-tuning threshold is used to control the output balance between false positives and false negatives by setting the minimum acceptable probability of identified classes.

4.2.2 Object Detection Model Selection and Evaluation

The performances of three well-known state-of-the-art object detection models, YOLOv3, RetinaNet, and Mask RCNN are compared with ground truth data in two selected intersection to specify the best model parameters for the subsequent multi-location analysis. Both RetinaNet and Mask RCNN used ResNet-101 as the network architecture. These models are all pre-trained using the COCO dataset [44]. The comparison is also used to validate the detection accuracy of the backbone models on the existing camera systems used in this study.

Each video contains four hours of surveillance data with a 30-second collection frequency. Precision-Recall (PR) curves were generated by tuning the non-max suppression threshold for the evaluation (Figure 20). A high area under the PR curve shows that the detector is returning accurate results (high precision) and a majority of all positive results (high recall). Based on computing costs and model performence, the YOLOv3 was selected for subsequent analysis in this study. The final detection accuracy using YOLOv3 for pedestrian is 89.18%, with a precision of 85.96% and recall of 84.32%.

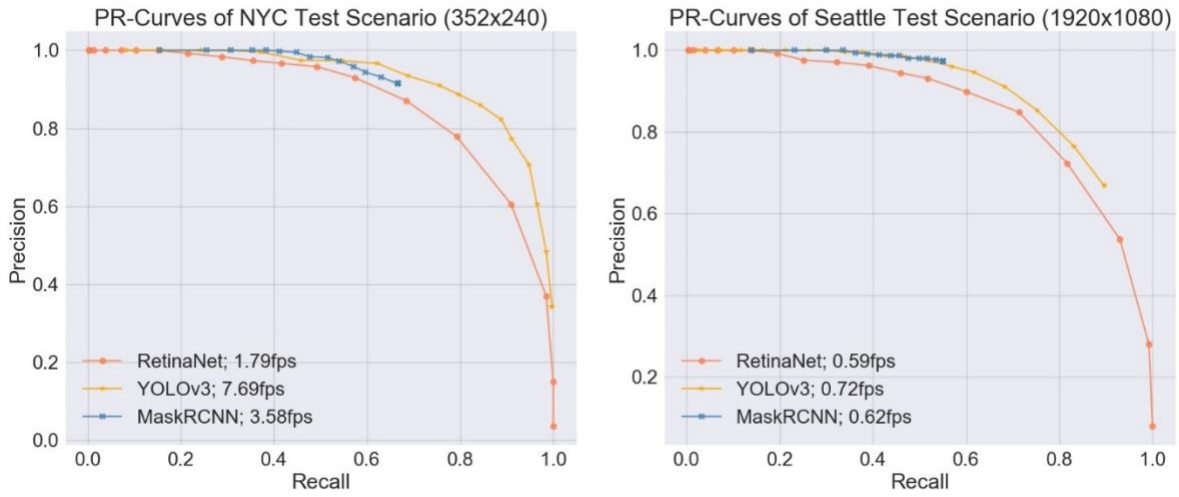


Figure 20 The performances of the three models tested at selected locations.

4.2.3 Post-Processing Filters

Four post-processing filters are developed for this study to improve detection accuracy and eliminate detection errors such as duplicate detections, oversized detections, or detection at impossible locations.

These post-processing filters are as follows:

- *Customize detection area*: This filter is used to remove any detected objects that appear in irrelevant or inaccessible areas (e.g., sky/building, parked lanes).
- *Intersection over Union (IoU) based filter to remove overlapped bounding boxes for the same object*: An Intersection over Union (IoU) based filter is developed to remove multiple bounding boxes generated for the same object, e.g., identifying a truck as both truck and a car. For areas of two bounding boxes B_1 and B_2, the IoU is calculated by:

$$\mathrm{IoU} = \frac{B_1 \cap B_2}{B_1 \cup B_2} \quad (1)$$

  This study uses an IoU = 0.8 (i.e., a bounding box with a lower-class probability is removed when the intersection over the union of two overlapped bounding boxes is larger than 80% of the union.)
- *A position filter to identify significant height difference at near hyper-plane*: A significant height difference of two pedestrians standing close to each other may affect due to the mechanism that the algorithm is built-on [3]. Therefore, a position filter is deployed to identify if two pedestrians are close in proximity. Given that all pedestrians are perpendicular to the same horizontal plane (the earth), the vertical position of proximate pedestrians should be similar in an image. The related objects are considered close if the vertical position difference between the bottom lines of detected bounding boxes is lower than a threshold. The level of vertical position difference is calculated as:

$$\mathrm{Diff}_{1,2} = \frac{\left|y^1_{\mathrm{bottom}} - y^2_{\mathrm{bottom}}\right|}{\min\left(h^1_p, h^2_p\right)} \quad (2)$$

  where the y^1_{bottom} and y^2_{bottom} are the vertical locations of the bottom line of bounding boxes and h^1_p and h^2_p are the estimated pixel-heights of each bounding box. If Diff is lower than 0.25, two detected pedestrians are considered close to one another, and the one with the higher pixel height will be assumed to have a real heigh equal to the pre-set height (5.74 feet in this study).
- *Remove incorrect size*: This filter removes any large bounding boxes that are larger than 75% of the size of the input image.

4.2.4 Real Distance Approximation

The main challenge of detecting social distancing patterns from surveillance videos is the accuracy of the measurement of the actual distance between pedestrians due to the perspective effect and a lack of references [3]. One common solution is to compute a homograph, which is a matrix represents the transformation between two planes, to morph the video frames from perspective view into a top-bottom view, then using preset objects with known measurement or existing objects with available measurement as references to compute the distances in the transformed frames [45]. Traditional

approaches to obtain reference distances include field visits, measurements from maps or civil infrastructure documents, and chessboard-like or point matrix calibration plates.

It may be challenging to get a reliable reference to calculate distances using camera feeds as cameras installed at different locations usually have different views of perspective and may be repositioned at different times, which demands a generalized method that can be used to measure distance from multiple cameras. With these complications, we propose a novel method to measure distance without the use of field measurement and homography computing.

Figure 21 illustrates the proposed real distance approximation method.

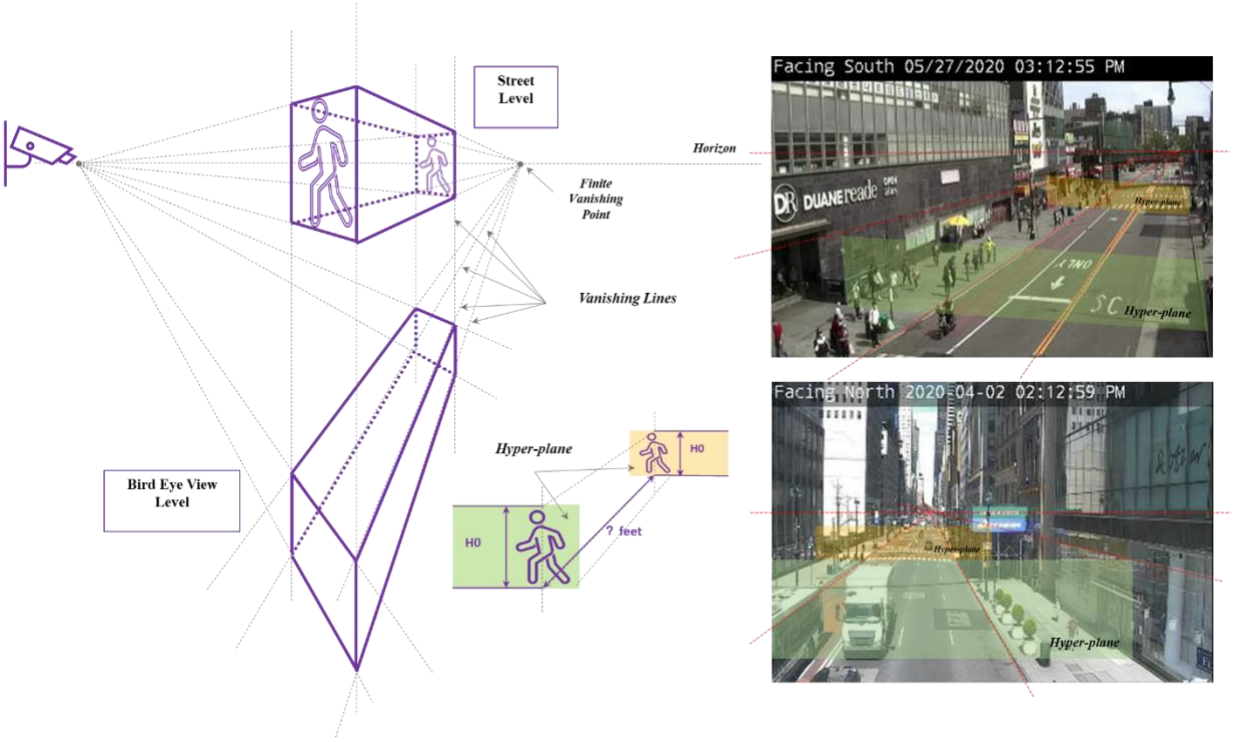


Figure 21 Proposed real distance approximation method.

The steps of this method are presented as follows:

- **Step 1:** Slice the image into several hyperplanes, which are perpendicular to the horizontal plane and the vanishing lines. Because of the perspective effect, the number of pixels in the image corresponding to a single real-world length can vary on different hyper-planes. In other words, each hyper-plane will have a specific real-pixel distance rate (RP-rate)—the "further" the hyper-plane, the larger the RP-rate.
- **Step 2:** It is assumed that each person in the image is perpendicular to the horizontal plane and has the same height h_r. Detection results may be affected by this height assumption, but because of the size and resolution of the videos, the impact of the height assumption is acceptable.

- **Step 3:** We identify the centroids of all the bounding boxes of all detected pedestrians and estimate the pixel-distance l between the centroids of bounding boxes for each pair. The pixel-height h_p^i of the bounding box is used as the pixel-height of the detected person i. For each pedestrian pair, it is always possible to find a "box" with four hyper-planes; two pedestrians belong to two of the hyper-planes, and the other two hyperplanes are perpendicular and intersect with the pedestrians (
- **Figure** 21 upper left). The generalized RP-rate r_{RP}^i for any person i can be represented as:

$$r_{RP}^i = \frac{h_r}{h_p^i} \tag{3}$$

- **Step 4:** We then slice the line between two centroids into small enough segments with a small pixel-distance Δp, and use each segment's RP-rate r_{RP} to calculate the real distance of that segment Δd:

$$\Delta d = \Delta p \cdot r_{RP} \tag{4}$$

Thus, the distance between person a and b can be calculated using the following formula:

$$D_{ab} = \int_a^b \Delta d = \int_a^b r_{RP}^i dp \tag{5}$$

Because the hyper-planes are perpendicular to vanishing lines, and the vanishing lines are straight lines, the transfer process of pixel-height h_p between two hyper-planes is correlated with pixel distance linearly, and the formula is shown below:

$$h_p = \frac{h_p^2 - h_p^1}{l} \cdot p + h_p^1 \tag{6}$$

Where h_p^1 and h_p^2 are the estimated pixel-heights of each person, l is the estimated pixel-distance between two persons, and p is the variable that indicates the pixel distance.

- **Step 5:** Finally, the real distance between two centroids can be calculated by combining equation (3), (5) and (6):

$$\begin{aligned} D &= \int_0^l r_{RP}^i dp = \int_0^l \frac{h_r}{h_p} dp = \int_0^l \frac{h_r}{\frac{h_p^2 - h_p^1}{l} \cdot p + h_p^1} dp \\ &= h_r \cdot \frac{log\left(\frac{h_p^2 - h_p^1}{l} p + h_p^1\right)}{\frac{h_p^2 - h_p^1}{l}} \Bigg|_{p=0}^{p=l} \\ &= h_r \cdot \frac{log(h_p^2)}{\frac{h_p^2 - h_p^1}{l}} - h_r \cdot \frac{log(h_p^1)}{\frac{h_p^2 - h_p^1}{l}} \\ &= h_r \cdot l \cdot \frac{log\left(\frac{h_p^2}{h_p^1}\right)}{h_p^2 - h_p^1} \end{aligned} \tag{7}$$

It should be noted that the transfer of the RP-rate is not linear. To simplify the computation, we assume that the RP-rate is transferred linear, so the equation (7) can be roughly simplified into:

$$D = \left(\frac{h_r}{h_p^1} + \frac{h_r}{h_p^2}\right) \cdot \frac{l}{2} \tag{8}$$

The K-D Tree algorithm [46], a space partitioning data structure for organizing points in a K-Dimensional space, is applied to quickly identify all pedestrians in close contact. The program will generate a blue line between the pair of pedestrians who are not following social distancing guidelines.

5. Mobility and Sociability Trends During COVID-19

5.1 New York City

5.1.1 Transit and Vehicular Traffic

The closing of essential businesses and stay-at-home policies caused an immediate direct impact on the usage of Metropolitan Transportation Authority (MTA) mass transit services, the largest public transit authority in the US, which typically carries over 11 million transit riders on an average weekday. Mobility board data collected through MTA shows steep declines in both transit ridership and vehicular traffic in the beginning of the pandemic. The decline rates reached up to 94% in peak transit ridership as of March 23rd, and up to 72% in vehicle traffic through MTA bridges and tunnels as of March 29th, compared to the same date/week in 2019 (Figure 22). Table 4 shows the comparison of transit ridership and vehicular traffic in 2020 and 2021 in NYC.

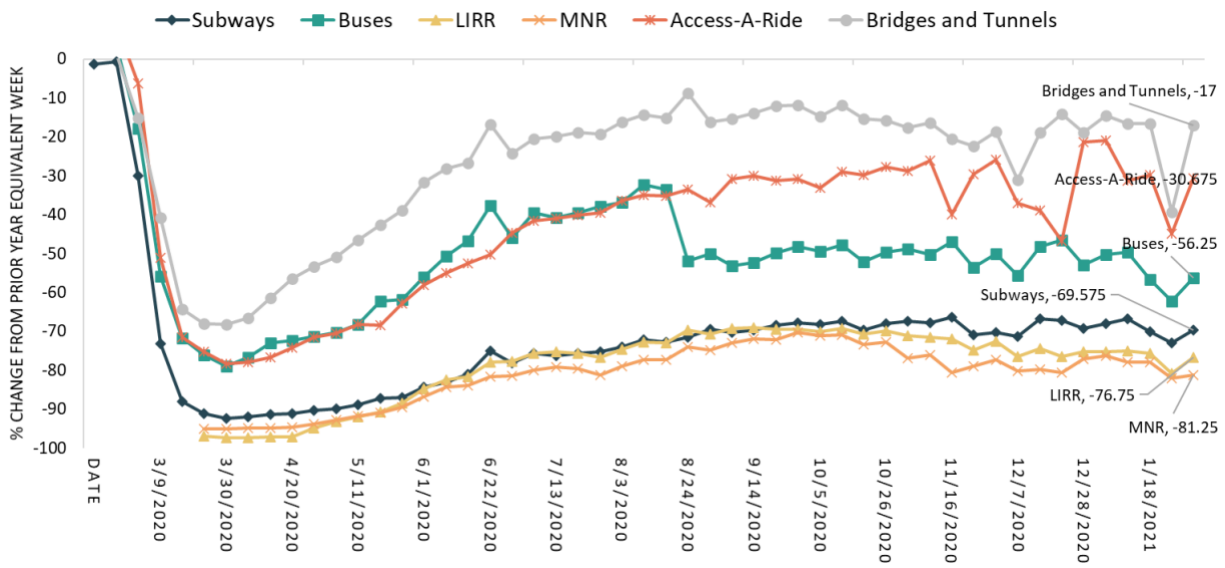


Figure 22 Transit Ridership Trends

Table 4 Comparison of Transit Ridership and Vehicular Traffic in 2020 and 2021 in NYC

Transit	Subway	Bus	Commuter Rail (LIRR)	Commuter Rail (MNR)	Access-a-ride
Worst week in 2020	-94%	-79%	-97%	-95%	-78%
Week of Jan 25, 2021	-70%	-56%	-76%	-78%	-30%
Vehicular Traffic	Via MTA bridges & Tunnels		Via PANYNJ crossings		Via BQE Queensbound
Worst week in 2020	-72%		-61% (-30% Truck)		-37% (-28% Truck)
Week of Jan 25, 2021	-17%		-11%* (-1% Truck)*		-4%** (+1% Truck)**

*Based on monthly data in January 2021. ** Based on weigh-in-motion data in November 2020.

5.1.2 Travel Time and Speed

Travel time on corridor 495 that connects Long Island and New Jersey (NJ) via Queens and Manhattan in NYC was analyzed in Figure 23. This sub-board uses virtual sensor data collected through crowdsourcing map applications [47]. A flatter travel time pattern of near free-flow speeds was observed following the stay-at-home orders, instead of typical spikes of commuter peaks (Figure 24). Increased traffic volume and travel time were observed in the week of May 4, 2020, indicating the start of the recovery period for NYC. Corridor 495 travel times in the first week of December 2020 were still about 17% lower (Eastbound) and 24% lower (Westbound) compared to pre-pandemic levels (Feb 2020).

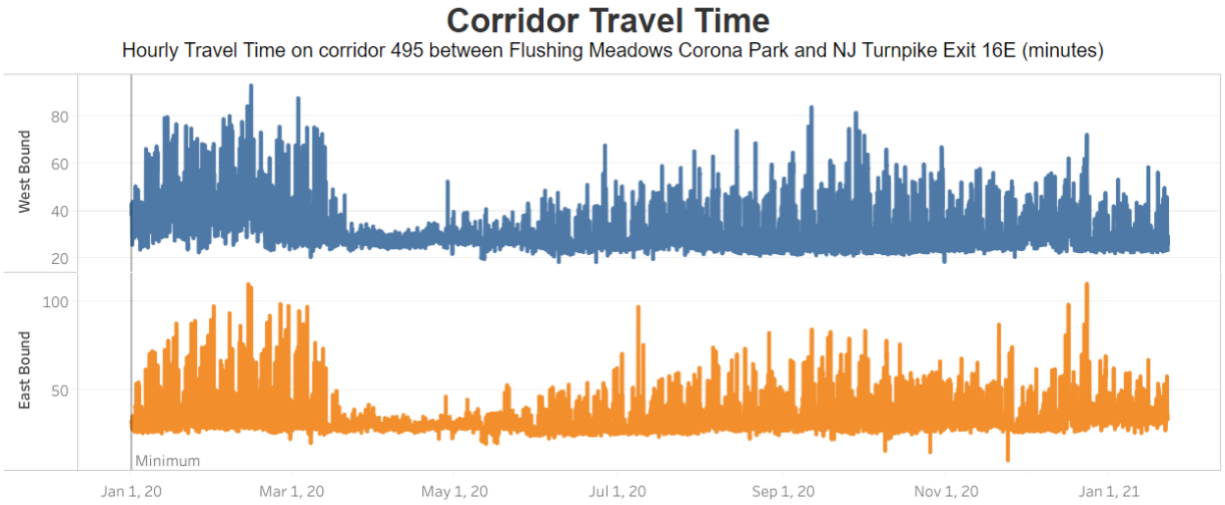


Figure 23 Travel time on 495 corridor (Flushing Meadows Corona Park to NJ Turnpike Exit 16E)

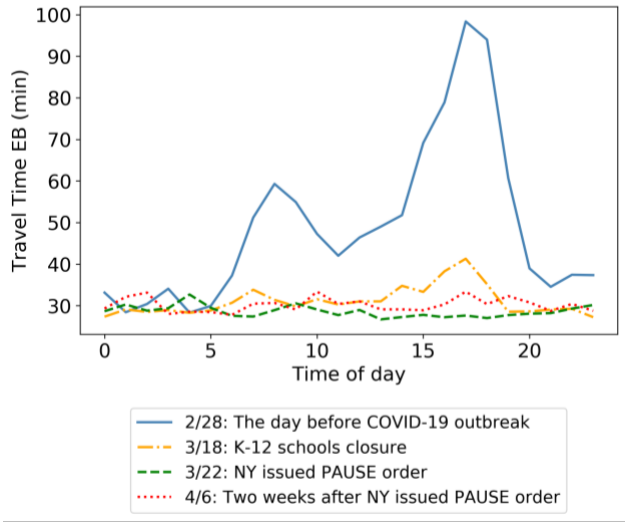


Figure 24 Hourly Travel Time Comparison on the 495 corridor

The speed maps in Figure 25 display the median travel speeds on major arterials/expressways of NYC as well as the Midtown Manhattan street network. With less traffic in the city during the outbreak, most roads and highways had higher speeds in the third week of April 2020, compared to the same week in February 2020. Average traffic speeds from 8AM to 6PM on north-south avenues between 34th and 57th Streets in Midtown Manhattan are shown in Figure 26. The average traffic speeds on avenues were up 108% in the third week of April, 2020, with a range from 22% to 253% on different avenues, compared to average traffic speed in the third week of February 2020. The average traffic speed on major crosstown streets, including 14th, 23rd, 34th, 42nd, and 57th Street, saw a 2-6 mph increase in the third week of April 2020 compared to the same week in February 2020.

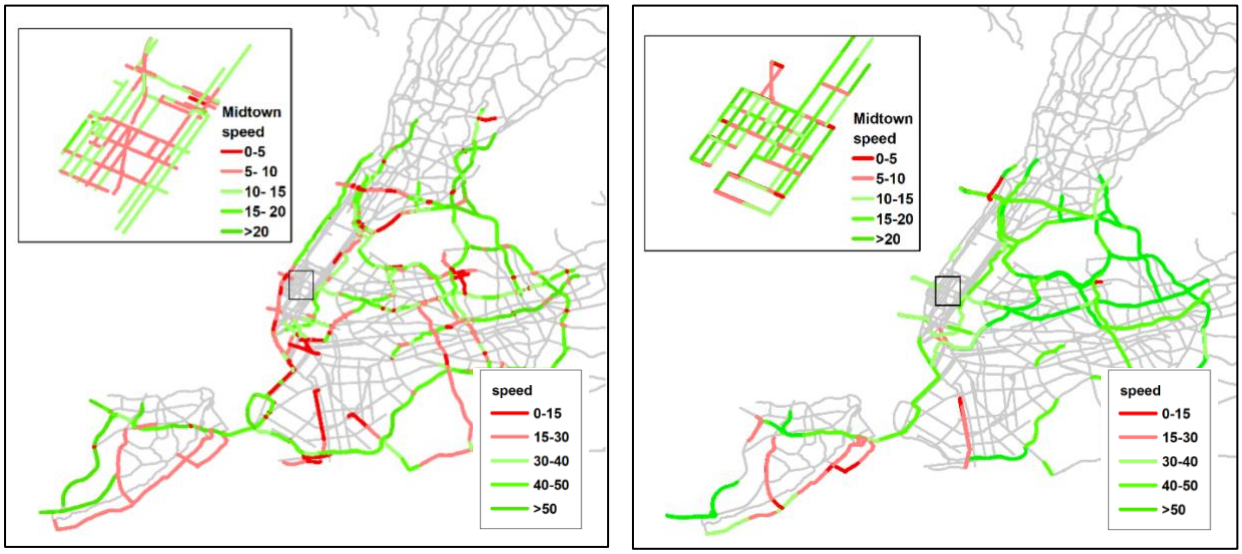


Figure 25 NYC Citywide average speeds (Citywide average speeds February 17-21, 8AM to 6PM (left), Citywide average speeds April 20-24, 8AM to 6PM (right))

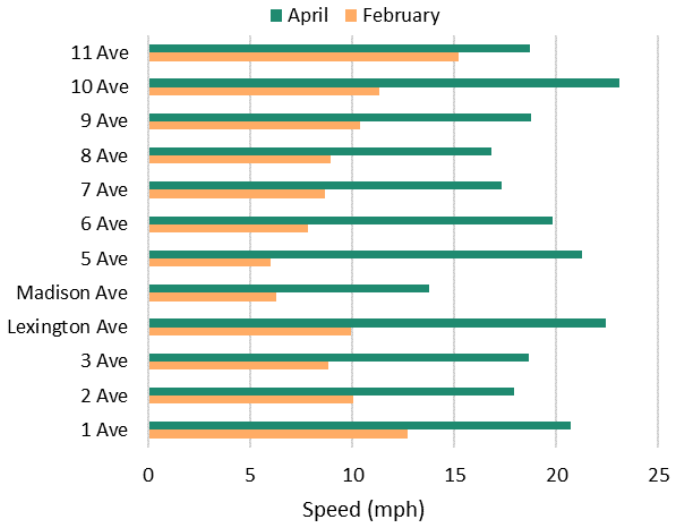


Figure 26 Midtown Avenue speeds, 8AM to 6PM

5.1.3 A Surge in Speeding

An increase in traffic speed and school zone speeding tickets was observed with some city streets operating at or near free-flow speed during COVID-19 crisis. For example, on April 15, 2020, during the traditional morning peak hour, 8.3% of the 145 local road segments where data is available, had an average traffic speed over the city speed limit of 25 mph. 28% of them were over 20mph, while only 3.4% had an average traffic speed over 20mph in February, before the pandemic. Based on the data collected from Open Parking and Camera Violations [48], the average number of daily school zone speeding tickets issued remains 70% higher in July 2020 and 62% higher in the first two weeks of August 2020 compared with pre-pandemic levels in 2020. Speeding violations are also likely to occur repetitively: 23% of ticketed vehicles received more than one speeding ticket in June, and these vehicles contributed to 44% of the total speeding tickets. Crashes remain low but the severity of crashes measured by fatalities is up. Unsafe speed was listed as the primary cause for half of the traffic fatalities in NYC in April (7 out of 14) and 42% of the fatalities in May (5 out of 12) according to NYPD's Motor Vehicle Collision reports [49]. Figure 27 presents the time series of weekly numbers of speeding tickets in 2020.

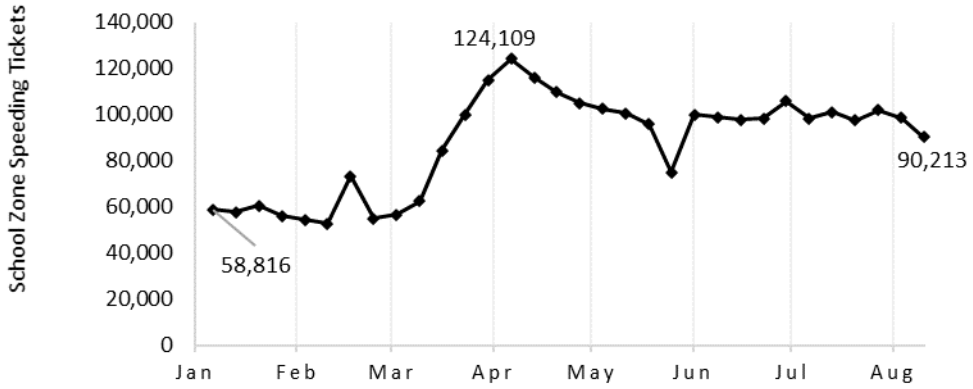


Figure 27 School Zone Speeding Ticket Trend in 2020 (weekly sum)

After examined unique plate IDs, speeding tickets per vehicle is summarized in Table 5. 14%, 29%, and 23% of the vehicles received more than one speeding ticket in February, April, and June 2020, respectively, and these vehicles contributed to 27%, 54%, 44% of the total speeding tickets in February, April, and June 2020, respectively.

Table 5 Speeding tickets per vehicle

	>1 tickets	>3 tickets	>5 tickets	>10 tickets
Feb	13.6%	0.8%	0.1%	0.0%
Apr	29.1%	5.6%	1.7%	0.2%
Jun	22.8%	3.2%	0.8%	0.1%

5.1.4 Freight Traffic

Weigh-in-motion (WIM) stations from C2SMART's testbed on the Brooklyn-Queens Expressway (BQE) showed changes in freight traffic for both Queens bound (QB) and Staten Island bound (SIB). A 37% reduction in overall traffic and 28% reduction in truck traffic were observed in the worst week in 2020 (Table 4). With the rebound of the traffic, as of August 16, WIM data indicated that average daily traffic (ADT)was down by only 2% for Queensbound (QB), and 4% for Staten Island-bound (SIB) traffic in the first two weeks of August, as compared with February 2020. Average daily truck traffic (ADTT) for SIB was down by 8% and is 5% higher QB in August compared to February. The average vehicle speed on BQE remained around the same level in the first weeks of August compared to June but is still 8% higher for SIB and 2% higher for QB, as compared to February.

5.1.5 Micromobility

During the first two months after the outbreak, bikeshare ridership remains down with the exception of the first 12 days of March. Ridership patterns did change, with 15% fewer Friday and Saturday trips and a 20% increase in average trip duration in March 2020 compared to the same month last year. Starting from April, cycling continued to increase. According to Citi Bike system data [50], on average, there were 67,422 rides in NYC per day in July 2020, a 9% increase compared to June 2020 and a 71% increase compared to February 2020 (Figure 28).

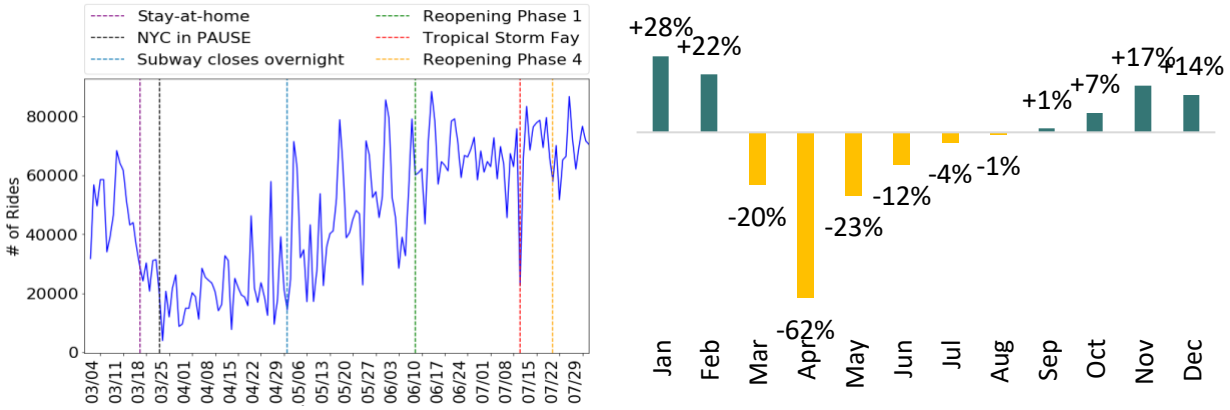


Figure 28 Citi Bike Ridership Trends (Citi Bike Daily Ridership (left), Citi Bike Monthly Ridership Change (2020 vs 2019, NYC only (right))

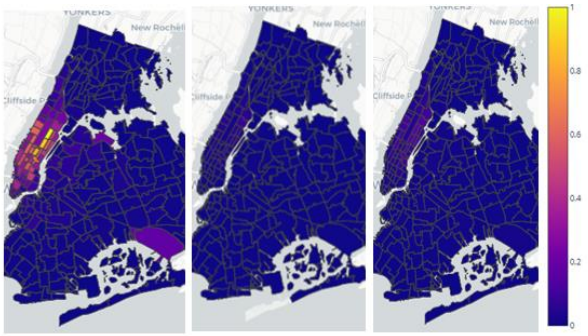
5.1.6 Taxi and For Hire Vehicles

Based on the NYC Taxi and Limousine Commission (TLC) data, it showed a sharp decrease in yellow taxi, green taxi, FHV, and high volume for-hire services (HVFHS) usage from the second week of March 2020 onwards. FHVs include Community Cars (aka Liveries), Black Cars, and Luxury Limousines, whereas HVFHS include records from Uber, Lyft, and Via. At the beginning of April 2020, trips in all TLC licensed vehicles dropped 84% compared to the same time last year. According to a report by TLC (8), in April 2020, 28,893 TLC drivers were on the road, compared with 122,076 TLC drivers in April 2019 (a 76% drop). The total number of monthly rides for yellow taxis dropped by 96%, 92% for green taxis, 79% for FHVs, and 76% for HVFHS in April compared to January 2020 (Table 6). The number of rides gradually rebounded in May and June but are still 60-90% lower than those in January 2020. In February, HVFHSs completed, on average, nearly 750,000 trips per day, Yellow Taxis completed 217,000 trips per day, and Green Taxis performed 14,000 trips per day. From March through April 2020, trips to parts of the city with hospitals increased as a percentage of all pick-ups, while trips to and from airports decreased (8).

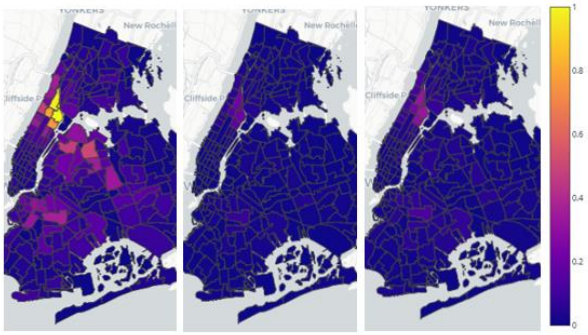
Table 6 Percentage changes in monthly ridership levels using January 2020 as baseline, Source: TLC (9)

Month	Yellow Taxi	Green Taxi	HVFHS	FHV
Feb	-2%	-11%	6%	-14%
Mar	-53%	-50%	-35%	-36%
Apr	-96%	-92%	-79%	-76%
May	-95%	-87%	-70%	
Jun	-91%	-86%	-63%	

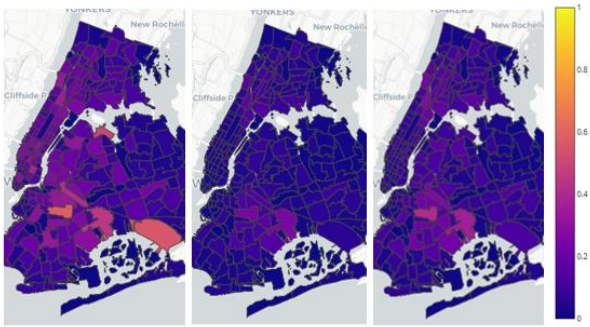
Figure 29 shows the spatial distribution in yellow taxi, green taxi, and HVFHS rides from March to June 2020. Green taxi rides ending in the upper Manhattan zones of Central Harlem North, Hamilton Heights, and Washington Heights North increased in June 2020. HVFHS had seen the most recovery in ridership levels since the start of the pandemic at JFK airport, Forest Hills, and Middle Village, Queens.



(a) Yellow Taxis (left-March, middle-April, right-June 2020)



(b) Green Taxis (left-March 2020, right-June 2020)



(c) HVFHS (left-March, middle-April, right-June 2020)

Figure 29 Spatial distributions of taxis and HVFHS

The average distance of a completed rides has increased during the pandemic (Figure 30). This, coupled with the drop in ridership, points at a trend where people are taking longer trips, but perhaps only when most necessary during the early stage of the pandemic. As ridership numbers increase, the average trip distance also become shorter. The average duration of a trip increased slightly during the pandemic and is gradually seen to be nearing pre-pandemic levels by June 2020 (Figure 30). The average number of passengers decreased since the start of the pandemic, especially for green taxis (from 3.4/vehicle in January 2020 to 1.4/vehicle in April 2020) - this might be because the passengers are suggested to not sit beside the driver in any taxis or FHVs, to adhere to social distancing norms.

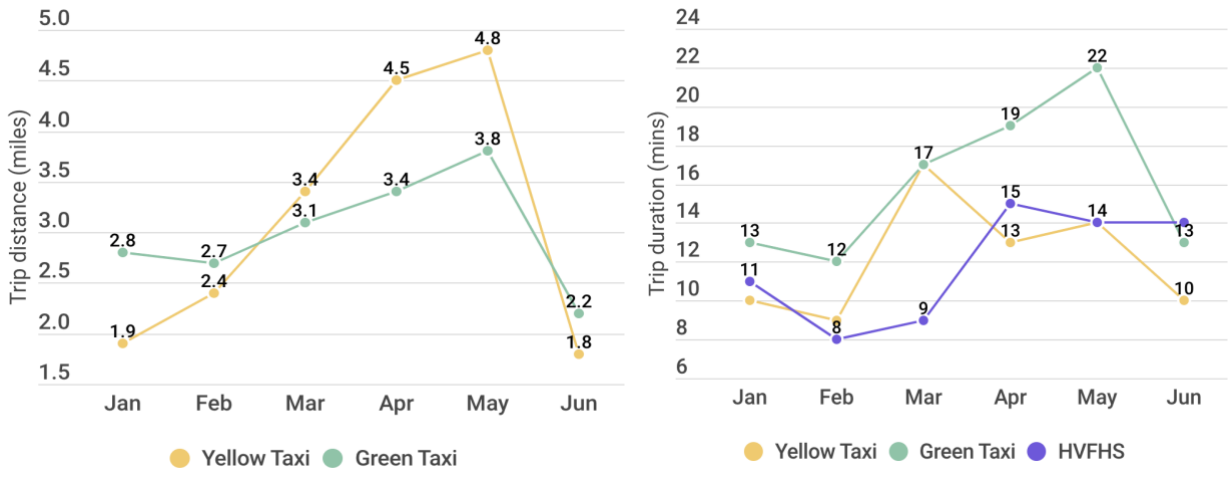


Figure 30 Yellow/Green Taxi Trip Distance and Trip Duration

5.2 Seattle

Seattle was the first major U.S. city to experience the drastic effects of the COVID-19 crisis. However, health concerns have abated quicker as compared to NYC, and data in 2020 suggested that Seattle may have begun returning to normal transportation conditions. Observing Seattle data may help to indicate what is to come in NYC.

5.2.1 Corridor Travel Time and Traffic Volume

Figure 31 illustrates the travel times on Interstate-5 (I-5) that goes through downtown Seattle area. The critical date analysis demonstrated that similar to NYC, the typical spikes of commute peaks disappeared since the release of stay-at-home order on March 23, 2020, indicating most people reacted immediately to the order [51]. Due to less traffic, the reliability of travel time became higher: the standard deviation of daytime travel time of the corridor reduced to 0.67 minutes in late April 2020 compared with 6.43 minutes in late February 2020.

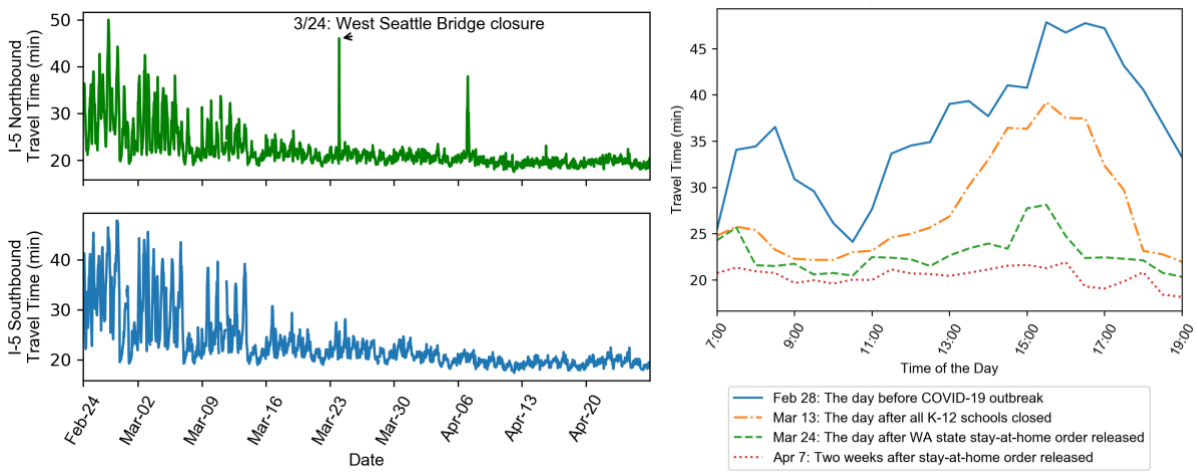


Figure 31 Hourly travel time on I-5 in Seattle (Hourly travel time in March and April (left), Critical day analysis for I-5 travel time (right)

Traffic volumes at three freeways locations (I-5 downtown Seattle, I-5 NE of Green Lake Park, and SR-520 toll bridge) showed a consistent drop in traffic volume in March, followed by a gradual increase in April 2020. Figure 32 presents an example of the daily traffic volume trends on I-5 in downtown Seattle. Using the same week of previous year as a baseline, traffic volume at the I-5 Downtown Seattle station dropped to its lowest level -46.91% in the week of March 30, 2020, and gradually increased to -41.95%, -35.50%, and -26.31% in the weeks of April 13, 27, and May 11, 2020, respectively. Similar trends have been observed for the other two stations.

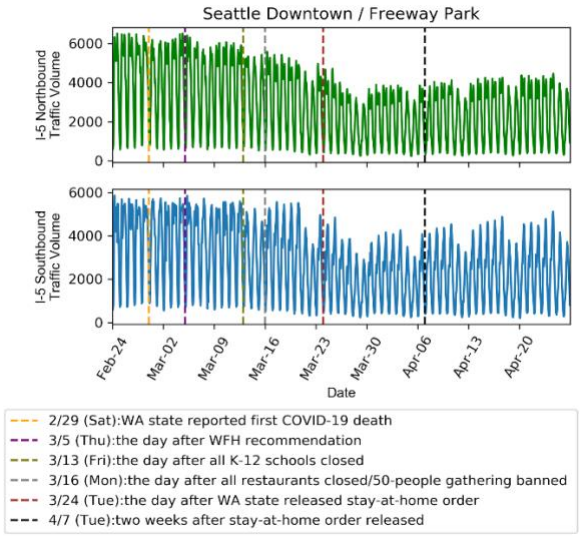


Figure 32 Traffic volume on I-5 (downtown Seattle)

5.2.2 Transit Ridership

According to Transit app data, public transit demand of Seattle began to decline on February 29, 2020, the same day when Washington state reported the first COVID-19 death [51]. The demand continuously decreased until March 26 (a 79% drop compared to a normal day), and demand has remained low since. The increase in traffic volumes while public transit ridership remains low indicates a potential shift in mode choice attitudes and preference for non-public modes of travel.

5.2.3 Micromobility

Figure 33 shows the total bike counts near Fremont Bridge between mid-February and late-April, 2020. During the first two weeks of March, 2020, the number of bikes gradually decreased, corresponding to work-from-home recommendation made on March 5, 2020 in Seattle [51]. Counts reached the lowest level at the end of March, which is consistent with the pattern change in travel time and traffic volume. Higher counts (+3,500) are observed on weekdays while fewer counts (~1,000) on weekends between mid-February to early-March. In April, bike counts grew to nearly 2,100, without a clear weekday/ weekend pattern, possibly due to an increase in the use of bikes for recreational purposes and as a result of the "Stay Healthy Streets closures" program, that allowed for safe social distancing building off of the 196 miles of Seattle's bike facilities and trails. Initial observations in 2020 indicate bike travel was up 299% from 2017 along Stay Healthy Streets in the Central District, which may contribute to the observed bike trend in Figure 33.

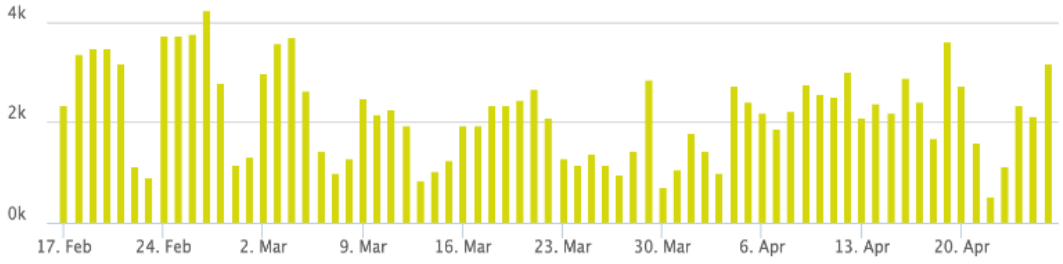


Figure 33 Seattle bike counts near Fremont Bridge (Figure source: Seattle DOT bike counters)

5.3 Pedestrian and Social Distancing

5.3.1 NYC311 Service Requests

The 311 service line provides a way for NYC residents to find information about services, make complaints, and report problems. Not following social distancing guidelines currently carries a penalty of $1,000. On March 28, 2020, a new category called “Social Distancing” was added as a subtype under the “Non-Emergency Police Matter” complaint in NYC311. A complaint could be made for violation of social distancing rules to report a business or location that is required to be closed, an essential business that is open but is not complying with necessary restrictions or overcrowding at a business or location.

The number of reports to 311 surged starting from the first day of stay-at-home, with most of the complaints being directed at stores or commercial buildings, followed by residential buildings, streets/sidewalks, and playgrounds. A steady increase of complaints in the following weeks made the “non-emergency police matter” category shoot up to the second spot among all complaint types reported to 311 (Table 7). There was a total of 17,470 such complaints made in under a month (March 28 to April 22, 2020), with 16,925 of them being about social distancing, reported from all five boroughs of New York City.

Table 7 The top 10 complaint types in NYC311 before and after stay-at-home order

	Feb 22 - Mar 21, 2020	**Mar 21 - Apr 22, 2020**
Rank	**Complaint Type**	**Complaint Type**
1	Noise- Residential	Noise- Residential
2	Heat/Hot Water	Non-Emergency Police Matter
3	Illegal Parking	Heat/Hot Water
4	Blocked Driveway	Consumer Complaint
5	Street Condition	Illegal Parking
6	Street Light Condition	Noise- Street/Sidewalk
7	Noise- Street/Sidewalk	Street Light Condition
8	General construction	Blocked Driveway
9	Consumer Complaint	Noise
10	Noise	Noise- Vehicle

5.3.2 Computer vision-based detection output

Using the aforementioned methodology in Chapter 4, the computer vision-based detection outputs all traffic-related objects (i.e., person, car, truck, bicycle, bus) as well as the total number of pedestrians that are in close contact in each frame. It builds a bounding box around the detected objects and assigns a class name and probability. The framework also highlights the pedestrians that are in close contact with blue lines in each frame. Figure 34 presents an example of the video processing output at one of the selected locations. Figure 34 also illustrates the number of detected objects in the video frame, including pedestrians, cars, and buses. Showcase of the detection output at various locations can be seen at https://c2smart.engineering.nyu.edu/covid-19-sociability-board/.

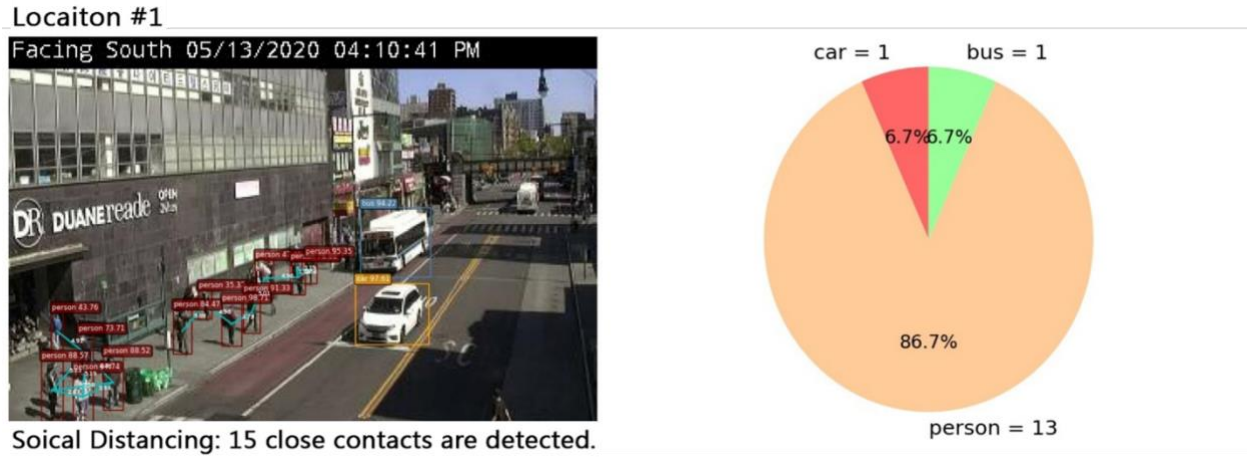


Figure 34 Example of detection output (including bounding box of identified objects; blue lines highlight the pedestrian pairs with a distance less than the threshold; and crowd density pie-chart.)

As a case study, we sampled 11 key camera locations that are spatially dispersed in the five boroughs of NYC and have different land use or sociodemographic characteristics (e.g., close to hospitals, meal distribution centers, or subway entrances) to evaluate the proposed method. Because the cameras are sometimes repositioned to view traffic from varying directions, representative weekdays with similar camera conditions are selected for these locations. The program is implemented in real-time and is run on an instance configured with Intel Core i7-7700HQ@2.8GHz CPU, 16GB RAM, and NVIDIA GeForce GTX 1060 GPU, Windows 10 64-bit operating system. The average running time is around 0.13 sec/frame/location without real-time visualization and 0.38 seconds/frame with real-time visualization.

The average pedestrian density and overall sociability metrics, considering various distances (e.g., 3-feet, 6-feet, and 12-feet) suggested by different agencies and experts [52-54] for the sampled locations, are summarized in Figure 35 and Table 8. Findings show that pedestrian density drops to the lowest point from mid-April to mid-May, where one pedestrian or less is captured in recorded frames for more than

half the time. The results show a positive skew, with the mean higher than the median, except on May 27, 2020, which displays a negative skew, and on Jun 18, 2020, which is normally distributed.

A gradual increase in pedestrian volumes is observed, starting from the end of May, reaching a peak in the middle of June. Accordingly, the ratio of people following social distancing guidelines (the social distancing rate) dropped slightly from April to May for all the suggested distances [3]. This rate slightly increases at the beginning of June 2020 and fluctuates around the new increased value.

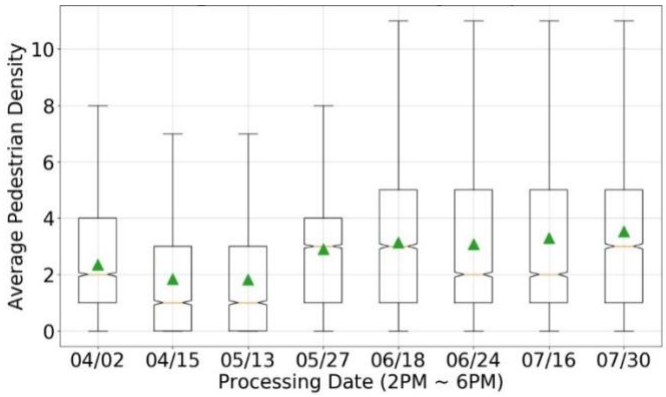


Figure 35 Box plot of the average pedestrian density/frame for different dates. The box shows the interquartile range (IQR), the 1st to the 3rd quantile of the data (Q1 and Q3). The whiskers extend 1.5 times the IQR from the Q1 and Q3. The green triangles are the mean values, and the "notches" indicate the 95% confidence interval of the median. We did not include the outliers in this plot, but Table 8 shows the maximum pedestrian densities for each date.

Table 8 COVID-19 Sociability Metrics, Selected Weekdays

	Apr 2	Apr 15	May 13	May 27	Jun 18	Jun 24
Average Peds Density (#/frame)	2.36	1.84	1.82	2.91	3.14	3.08
Maximum Peds Density (#/frame)	12	16	11	13	17	19
Social Distancing Adherence Rate (> 3 ft)	97.6%	96.3%	96.1%	95.2%	97.1%	97.1%
Social Distancing Adherence Rate (> 6 ft)	94.0%	91.8%	90.5%	88.7%	91.7%	91.3%
Social Distancing Adherence Rate (> 12 ft)	85.7%	83.7%	81.1%	75.8%	81.4%	80.3%

We compared the total number of pedestrians in close contact (distance<12, 6, and 3 feet) and the number of newly reported positive cases in NYC over time using the data generated by the proposed algorithm. In Figure 36, it is apparent that when the number of daily cases (annotations in the circle) decreases, more people go outside and get in closer contact with each other, with a lower number of people complying with social distancing guidelines [3]. There may be other underlying factors not

identified in this study that impact the decision for people to go outside in greater numbers and reduce observed distance between each other after April 15, 2020.

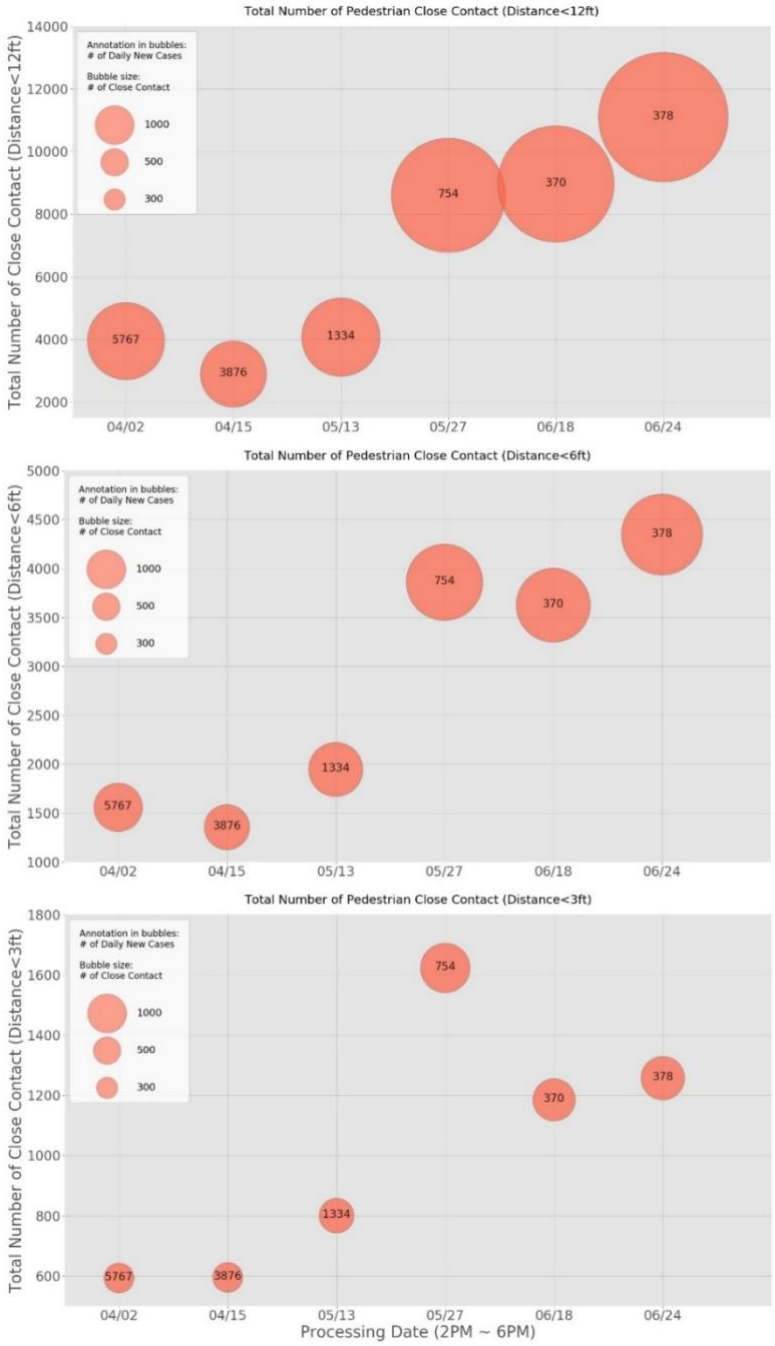


Figure 36 The total number of pedestrians in close contact (Distance<12ft) vs. the number of newly reported infection cases in NYC over time. The numbers in the circles are the daily new cases of COVID-19 in NYC for the given date, and the size of each circle indicates the number of pedestrians in close contact on that date.

Similar to NYC, the social distancing violation detection algorithm was implemented for a local street intersection (Broadway & E Pike St EW) in Seattle with a frame rate of 30 seconds. Results on May 18, 2020 show an average pedestrian density of 3.2 people/frame, a maximum pedestrian density of 12 people/frame, and a pedestrian social distancing compliance rate of 89%.

5.3.3 Heatmaps of clustered pedestrians

Through the analysis of heatmaps for each site, spatial patterns of social distancing have also been explored. This intends to highlight hotspots at which pedestrians are in close contact with each other. The heatmaps generated in Figure 37 use a 12-feet rule as a case study. Each pedestrian pair less than 12 feet apart is flagged and clustered to generate heatmaps for selected study locations and times. It is worth noting that instead of representing pedestrian densities, these heatmaps illustrate detection areas that capture high frequencies of close contact events. Generated heatmaps can serve as a visual cue to help understand and identify areas with a high risk of close contact between pedestrians.

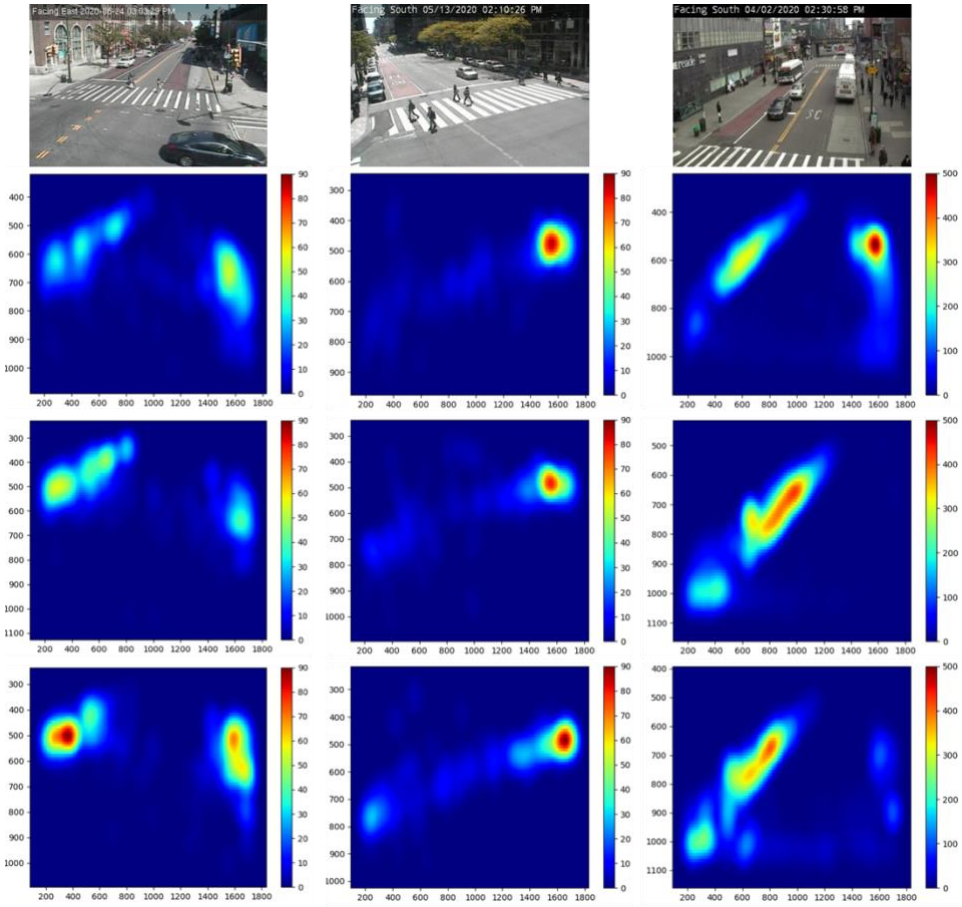


Figure 37 Heatmaps of clustered pedestrians (distance < 12 feet). For row two to four, from top to bottom: Apr 2, 2020, May 13, 2020, and Jun 24, 2020. Due to the camera movement, the third column's camera cannot capture the right-side sidewalk on May 13 and Jun 24, so the related heatmaps do not contain hot zones on the right-side.

5.3.4 24-hour Temporal Density Distribution

Based on the object detection outputs from the cameras, temporal density distribution profiles are constructed for one study location where pre-pandemic data is available to investigate potential temporal pattern changes. Figure 38 shows the 24-hour distribution of pedestrian densities at a selected site that had a typically high and consistent pedestrian density throughout the day before the outbreak of COVID-19. Compared to pre-pandemic levels, pedestrian density remains very low in April and May and gradually increases in June 2020 after the city's reopening. In addition, peak hours are shifted - afternoons (3:00 pm-5:00 pm) become the period with the highest pedestrian density for this location in May 2020, before the city's reopening. These temporal changes in pedestrian behavior may deserve more attention amid the COVID-19 pandemic, and appropriate response measures can be carried out (e.g., examining open street strategies, recommending staggered work hours to nearby companies) [3]. This temporal analysis assists in the investigation of the rebound of the pedestrian demand. The public should be continuously reminded of the potential risk of exposure in crowded environments, including in the open space of urban streets.

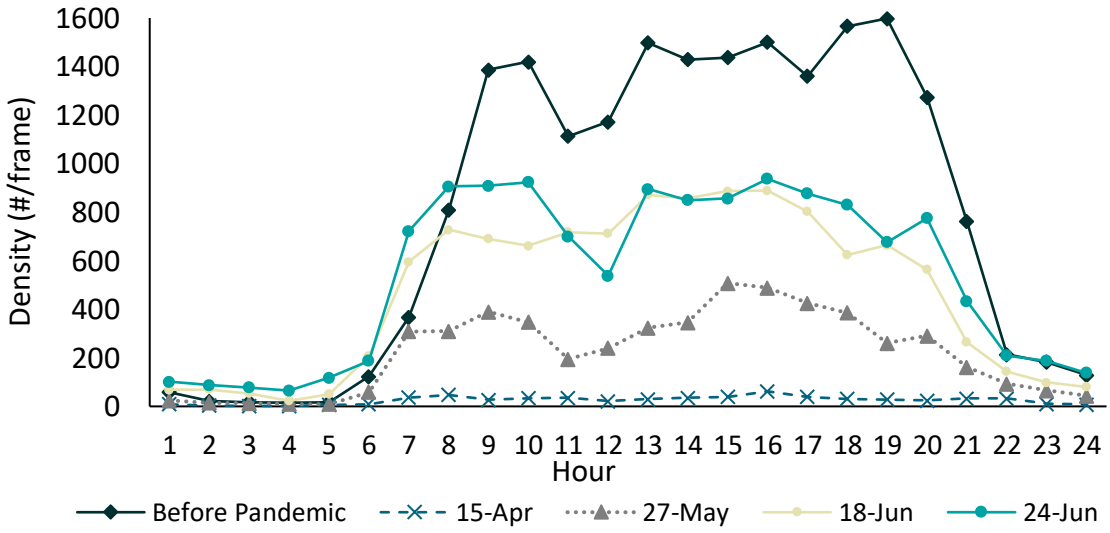


Figure 38 24-hour temporal distributions of pedestrians at one sampled location.

5.3.5 System Deployment Cost Estimation

Table 9 lists the estimated deployment costs for key components in deploying the video detection system based on 68 existing cameras in the NYCDOT system. It is assumed that the server will continue to collect the video feeds with an interval of every 30s, which results in a daily data size about 6GB. Although the current detection system uses a local server approach, the deployment costs based on cloud-based services are also estimated. The cost may vary depends on the number of cameras and data collection interval. Considering the requirement of the scalability, data availability, security, and performance, the Amazon cloud service (EC2 instance and Simple Storage Service (S3)) was used for estimating the cloud-based approach for server and storing the image data. The cloud-based service provides a platform which integrates access controls, auto backup service, data security and big data analytics. It also provides different pricing classes for users to choose based on their strategies. Two cost strategies are explored:

Plain File Storage

- Images are stored directly to the cloud-server.
- Users can easily view any existing images on the website with credentials.
- Users can set up queries to filter out the required images to process.
- The price includes storage pricing, request and data retrieval pricing, data transfer and transfer acceleration pricing, and data management features pricing.
- The total file amount can enormous because of the plain file storage method, which leads to a large number of requests for reading each image.

Zip File Storage

- Images are zipped based on time range or intersection before uploading to the cloud-server.
- Comparing to the Plain File Storage, the data request and retrieval price are decreased from million level to digit level, which reduces the cost.
- Users must download and unzip the file before processing.
- The price includes storage pricing, request and data retrieval pricing, data transfer and transfer acceleration pricing, and data management features pricing.

The price listed in Table 9 is estimated based on plain file storage strategy and a three-year contract with the Amazon cloud service. The price could be further reduced if zip file storage strategy is used. Moreover, the current system stores all the image files for validation purpose, once the program is validated, agencies have the flexibility of storing only the detection results without saving the image files, which can further reduce the storage requirements and cost.

Since the detection system is built on existing traffic camera system, no additional hardware maintenance costs are required. Cloud-based server maintenance fee is often included in the service and the maintenance cost for local server is minimal, so these costs are not quantified. The cost of running the program is not included as the detection program can be run on any regular desktop/laptop computer.

Table 9 Estimated system deployment cost (Annual cost based a 3-year estimation with 68 cameras; data collection interval is every 30s (daily data size: 6GB); plain file storage).

	Cost component	Cost
Local Server-based Services	One-time server acquisition	$600
	Network Attached Storage (NAS) device (up to 65TB) acquisition	$4390
	Annual Cost (Based on a 3-year estimate)	$1663 / year
Cloud-based Services – Option 1	EC2 server (t4g.micro, 2CPU, 1GB memory)	$67.56/year
	S3 One Zone-Infrequent Access storage	$394.2/year
	Data Transfer (180 GB/per month (Inbound) and 50 GB/per month (outbound))	$52.92/year
	Annual Cost (Based on a 3-year estimate)	$514.68/year
Cloud-based Services – Option 2	EC2 server (t4g.micro, 2CPU, 1GB memory)	$67.56/year
	S3 standard storage	$906.72/year
	Data Transfer (180 GB/per month (Inbound) and 50 GB/per month (outbound))	$52.92/year
	Annual Cost (Based on a 3-year estimate)	$1,027.20/year

*Cloud-based cost is estimated based on Amazon EC2 instance and S3 storage.

It worth noting that local server-based service can be cheaper in the long term, provided the data storage does not exceed the NAS device limit (e.g., $998/year for a 5-year estimate). Cloud-based services can be more expensive in the long term as more data needs to be stored on the cloud (e.g., $1631/year for a 5-year estimate if using S3 standard storage) and may need a local server to archive the data or other storage service such as S3 infrequent access or deep archive to reduce the costs.

5.4 Agent-based Simulation Virtual Testbed COVID-19 Travel Survey

The COVID-19 Data Dashboard also connects to C2SMART's MATSim agent-based simulation virtual testbed [6] that provides network performance and emission evaluation for reopening phases and C2SMART COVID-19 Travel Survey that collected responses nationwide about travel behavior changes.

5.4.1 MATSIM Agent-based Simulation Virtual Testbed

Well-calibrated agent-based simulation models built in MATSim [55] were used to study the impact of COVID-19 on the NYC transportation system as well as potential policies for the reopening of NYC. Two baseline models were developed and calibrated: 1) A Pre-COVID model that simulates typical travel behavior, and 2) a COVID model that represents travel behavior during the first month of the COVID-19 pandemic. C2SMART researchers developed the Pre-COVID model called MATSim-NYC, using a synthetic

population of more than 8 million New Yorkers and calibrated transit schedules [55]. This open-source, large-scale transportation model covers the entire NYC area and integrates data and modeling capabilities for new technologies and modes. The calibrated models are being used to study:

- Scenarios of multiple reopening phases
- Mode and system usage for various reopening policies and social distancing strategies
- Emissions and air quality impacts

Findings from the model indicate that because of changed preferences, a full reopening would perhaps only see as much as 73% of pre-COVID transit ridership and an increase in the number of car trips by as much as 142% of pre-pandemic levels, assuming mode preferences held during the crisis are maintained. Evidence from other cities further in the reopening process points to lingering mode choice preferences from the pandemic during reopening, however due to unique mode-choice in NYC, these numbers may represent an extreme case. The effect of capacity restrictions on public transit, like those applied in some cities, was also studied. Limiting transit capacity to 50% would potentially decrease transit ridership to as low as 64% while increasing car trips to as much as 143% of pre-pandemic levels by Phase 4 of NY's reopening plan. Details about the MATSim model can be found in [55].

5.4.2 C2SMART COVID-19 Travel Survey

A COVID-19 mobility survey was conducted as a parallel effort to explore the pandemic's impact on behavior across different demographics, with special attention paid to individuals with a disability, low-income households, and other disadvantaged or underrepresented groups in transportation research and planning. A total number of 2022 of responses were collected national wide. 1121 completed responses were received from residents from the five boroughs of New York City (NYC) with 45% respondents identified as living with a disability and 23% live in a low-income household (earning less than $50,000/year). At a time when certain mobility practices and challenges have become part of our new daily routines, the survey data will help understand what the real needs for disadvantaged population groups are, in the event of a pandemic. This, in turn, will aid the cities and transportation agencies in assessing a more equitable planning process to address their short-term and long-term trip needs. The following fact sheet presents a quick summary of the survey results. Please refer to [5] for more details.

5.5 White Papers

Over the course of the COVID-19 Data Dashboard's development, the team has published five white papers detailing this research. All white papers available for open access download. Please refer to Appendix A for more details.

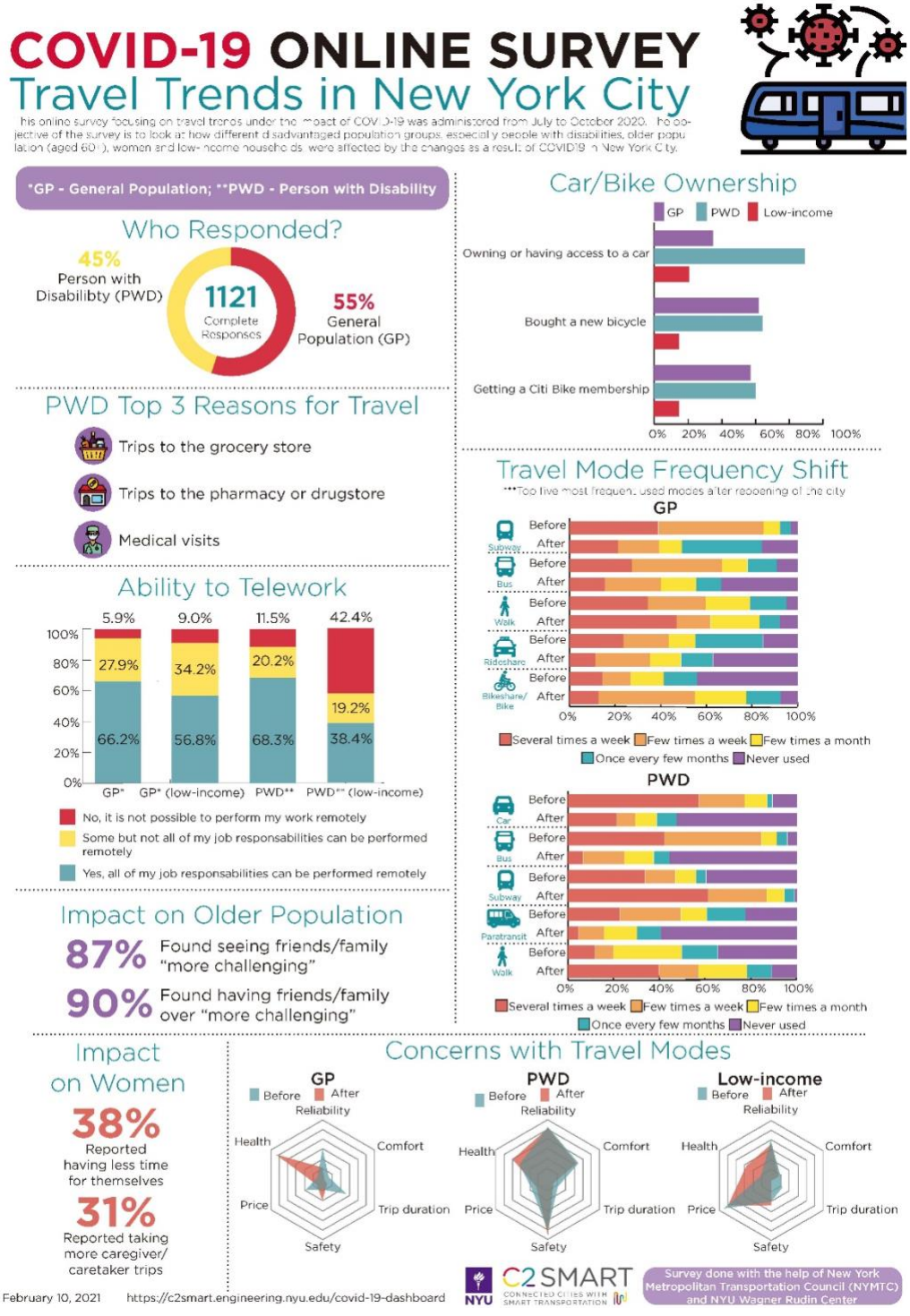


Figure 39. C2SMART COVID Mobility Survey Fact Sheet

6. Lessons Learned

6.1 Developing and deploying computer vision-based data collection applications

The following considerations could help when developing and deploying computer vision-based data collection applications, especially for pedestrian detection:

- Utilize existing transportation infrastructure and ITS devices (e.g., CCTV cameras) to minimize deployment cost when designing a computer vision-based data collection application.
- Develop an expandable framework to accommodate potential future needs of new features or add-ons to the current system.
- Engage with local partners to ensure relevant agencies and stakeholders can use the data and its findings to adapt policy interventions and outreach.
- Account for privacy when designing a computer vision-based application for pedestrian detection. Masks, blur filters, and reduced image resolution can be useful, simple solutions for protection of personal identities.
- Low resolution and frame rate cameras are sufficient to provide pedestrian density information with appropriate pre- and post-processing filters and existing computer vision techniques.
- Emerging technologies, such as AI-based automated detection, can be useful for short-term before-after evaluations since they can reduce the time and effort associated with manual data collection.
- Remote data collection and reference-free distance approximation techniques can improve flexibility, reduce labor costs, and minimize the dangers of fieldwork (e.g., during a pandemic) , as human investigators are not required.
- Detection accuracy largely depends on the resolution, sampling rate, and angle of the camera as well as the training image set. The detection model can be retrained using localized annotated images to improve performance.
- Cloud-based server and storage service is cheaper in the short-term (i.e., less than 5 years) when compared with local server and storage [56]. However, data archiving needs to be considered if cloud-based server and storage service is used for the long term.
- Local server storage is a good option for testing or piloting a video detection system. It is easy to set up but may suffer from single point failure (i.e., when the server is down, there is no incoming data, and data cannot be retrieved before it is restarted). To avoid this, a backup local server is needed, which may double the server acquisition cost.

6.2 Dashboard Performance

This study revealed the following lessons and suggestions on dashboard performance:

- Performance was a major issue we encountered. In the initial design, we planned to display multiple visualizations on the same page for comparison. However, this resulted in a long loading time. We recommend considering performance issues in advance when designing the interface and choosing visualization tools.
- The performance of web-based visualizations is affected by two main aspects: data transfer and rendering. Reduce the amount of data to make accelerate data transfer. Remove unnecessary fields and compress the data. For example, use a more efficient data format or perform lossy compression to reduce the granularity or precision of the data. When having control over data loading, there are some other optimization methods, such as transferring the data in chunks or using streaming data.
- In terms of rendering, there are three common ways to render shapes on a web page: SVG, Canvas API, and APIs for GPU like WebGL. They have different use cases and will affect the performance. SVG is for rendering high-quality vector graphics. It is suitable for rendering a small number of shapes (about 1000). As the number of shapes grows, the performance inevitably will drop, developers need to reduce the number of shaps or use other rendering methods. Canvas API has better performance than SVG when rendering thousands of shapes. But it still has a limit and would slow down on larger screens. WebGL and other APIs for GPU harness hardware optimization to support unlimited 2D and 3D rendering. This is the fastest way but is harder to implement. visualization tools may not support changing the way of rendering. Tableau, for example, uses Canvas and does not support GPU optimizations. If there is no other way but to build a customized visualization using Javascript libraries, we recommend using more high-level libraries to save time and ensure compatibility.

6.3 Dashboard Automation

Recommendations about automating data collection and display are provided as follows:

- Although we built an automated pipeline, human intervention is always required. The software updates may bring challenges to the automated pipeline and most importantly, we do not have control over the open data sets. The data format, API, and update time might be changed as time passed. Therefore, visualizations need to be checked regularly. It is also recommended to generate logs and use job schedulers in the backend to monitor the process so that when a visualization fails, maintainers can quickly detect the problem.

7. Conclusion and Future Work

An interactive data dashboard that serves as an all-in-one data fusion tool for COVID related transportation open data is introduced in this study. It uses data mining and deep learning techniques for data acquisition and processing and supports interactive data visualization and data analytics for quantifying multiple mobility and sociability metrics.

The current dashboard contains vehicular traffic, transit, micromobility, freight, taxi, and risk indicators, such as speeding ticket and crash records for NYC, as well as traffic data in Seattle, Chicago and six cities from China. The data sources are displayed in a user-friendly dashboard and are continuously updated giving researchers and interested members of the public access to dynamic information about the changing state of mobility during the pandemic and beyond. The crowd density and social distancing compliance for pedestrians based on data obtained from publicly available DOT cameras using state-of-the-art video processing techniques can provide useful insights into daily pedestrian and cyclist demands and their behavior in terms of maintaining social distance.

Two years after the outbreak of COVID-19 pandemic, data from both NYC and Seattle shows that while vehicular traffic volume is back to the pre-pandemic level, transit usage remains 40% lower, compared to same period in 2019. Adapting to the “new normal” and taking actions to address these enormous challenges in the aftermath of COVID-19 is crucial: providing safe and accessible micromobility network and open space for pedestrians, reducing traffic violence and inequality, balancing vehicular traffic and encouraging sustainable transportation alternatives of all ages and abilities, and maintaining the affordability and functionality of transit systems. Deeper consideration for vulnerability assessments and incremental improvements to make transportation systems more resilient and well-prepared are still needed.

As the world continues to adjust to the new reality in the aftermath of COVID-19, we will continue to collect data, including perishable mobility, safety, and behavior data, and will continue to monitor these trends. One of the future works is to integrate the mobility dashboard, as well as the data pipeline, to be part of the center’s data ecosystem (Figure 40). To collect and analyze large-scale data, we plan to use cloud-based products for integration, storage, and analysis. We can further upgrade the pipeline using Serverless frameworks. Currently, we run a server in the background to achieve automation that can be monitored. But since scripts are only run occasionally, usually no more than once a week, serverless would be a more scalable and cost-effective solution, especially when using cloud services. For most open data that only requires simple filtering and formatting, we can use Google Script with trigger, Google Spreadsheet, and Tableau Public. For other datasets that involve advanced analysis, we can use AWS Lambda, Google Spreadsheet, and Tableau Public.

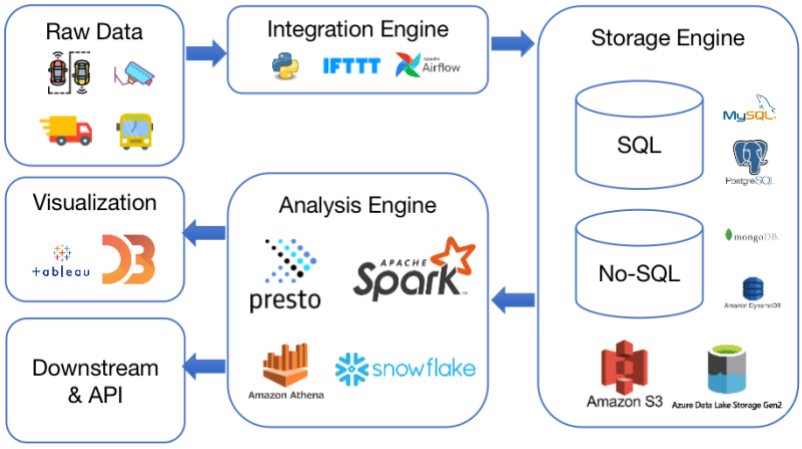


Figure 40 Data Ecosystem

Moreover, as an ongoing effort, this Artificial Intelligence (AI)-based object tracking approach will be extended to cover 100 camera locations to continuously evaluate the changes in crowd density and social distancing between pedestrians. It is also possible to translate these findings in near real-time and use them in the post-pandemic era, tracking density trends during special events where crowding may occur. These findings can also be used for predictive analysis forecasting future cycling rates.

Summary of Research Outputs and Tech Transfer

As of April 20, 2022, the COVID-19 Data Dashboard had over fifty thousand website views (52,876 views). The number of sub-board views are shown in Figure 41.

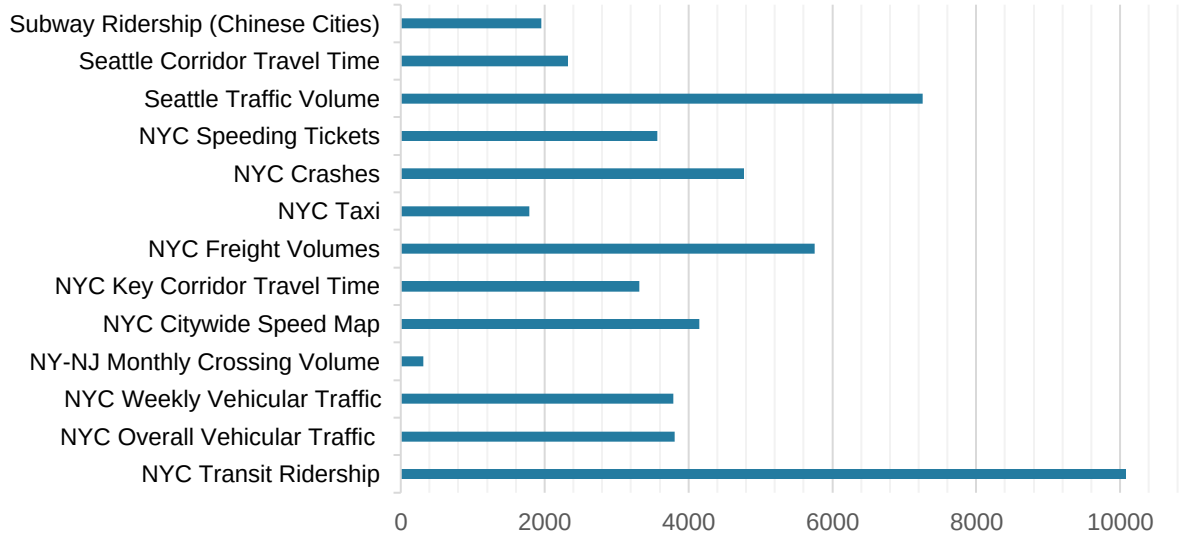


Figure 41 Sub-board Views

As an outcome of this research project, several research outputs were produced along with dissemination, including 5 journal and conference papers, 5 whitepapers, 2 videos, and 8 presentations. Moreover, this research was also highlighted and in the USDOT UTC newsletter and included in a case study for USDOT ITSJPO's ITS deployment database and the USDOT UTC video forum. The papers and whitepapers were cited by 146 researchers and the videos presenting the dashboard had 1,095 views, as of April 2022. Table 10 summarizes those results.

Table 10. Summary of research outputs

Output type	Description	Link/source
Dashboard	The developed COVID-19 Data Dashboard.	https://c2smart.engineering.nyu.edu/c2smart-data-dashboard/
Newsletter	University Transportation Center (UTC) Spotlight, UTC quarterly newsletter, Issue 3, July 2021	https://www.transportation.gov/utc/using-video-feeds-public-traffic-cameras-and-computer-vision-analyze-social-distancing-and
Case Study	ITSJPO ITS Development Database, Decision Support Resources	Submitted and approved, waiting for publication

Video	USDOT University Transportation Center Video Forums on Artificial Intelligence, March 18, 2022	https://vimeo.com/boozallencreativestudio/review/689399213/c3e226785d
Video	C2SMART COVID-19 Data Dashboard Showcase	https://youtu.be/9hTTc5qAVOU
Paper	Zuo, F., Wang, J., Gao, J., Ozbay, K., Ban, X., Shen. Y., Yang, H. and Iyer, S. (2020), An Interactive Data Visualization and Analytics Tool to Evaluate Mobility and Sociability Trends During COVID-19, UrbComp 2020: The 9th SIGKDD International Workshop on Urban Computing, San Diego, California, USA.	http://urban.cs.wpi.edu/urbcomp2020/file/05.pdf
Paper	Zuo, F., Gao, J., Kurkcu, A., Yang, H., Ozbay, O. and Ma, Q., Reference-free video-to-real distance approximation-based urban social distancing analytics amid COVID-19 pandemic, Journal of Transport & Health, Volume 21, 2021.	https://www.sciencedirect.com/science/article/abs/pii/S2214140521000268
Paper	Bian, Z., Zuo, F., Gao, J., Chen, Y., Venkata, S.S.C.P., Bernardes, S.D., Ozbay, K., Ban, X.J. and Wang, J., 2021. Time lag effects of COVID-19 policies on transportation systems: A comparative study of New York City and Seattle. Transportation Research Part A: Policy and Practice, 145, pp.269-283.	https://www.sciencedirect.com/science/article/abs/pii/S0965856421000276?via%3Dihub
Paper	Wang, D., He, B. Y., Gao, J., Chow, J. Y., Ozbay, K., & Iyer, S. (2021). Impact of COVID-19 Behavioral Inertia on Reopening Strategies for New York City Transit. International Journal of Transportation Science and Technology.	https://www.sciencedirect.com/science/article/pii/S2046043021000046
Paper	Wang, D., Tayarani, M., He, B. Y., Gao, J., Chow, J. Y., Gao, H. O., & Ozbay, K. (2021). Mobility in post-pandemic economic reopening under social distancing guidelines: Congestion,	https://www.sciencedirect.com/science/article/pii/S0965856421002299

	emissions, and contact exposure in public transit. Transportation Research Part A: Policy and Practice, 153, 151-170.	
White paper	Gao, J., Bernardes, S.D., Bian, Z., Ozbay, K., Iyer, S. 2020. Initial Impacts of COVID-19 on Transportation Systems: A Case Study of the U.S. Epicenter, the New York Metropolitan Area.	https://c2smart.engineering.nyu.edu/wp-content/uploads/2022/04/Whitepaper-Issue-1.pdf
White paper	Gao, J., Wang J., Bian, Z., Bernardes, S.D., Chen, Y., Bhattacharyya, A., Soorya Muruga Thambiran, S., Ozbay, K., Iyer, S., Ban, X. 2020. The Effects of the COVID-19 Pandemic on Transportation Systems in New York City and Seattle, USA.	https://c2smart.engineering.nyu.edu/wp-content/uploads/2022/04/Whitepaper-Issue-2.pdf
White paper	Wang, D., Zuo, F., Gao, J., He, Y., Bian, Z., Bernardes, S.D., Na, C., Wang, J., Petinos, J., Ozbay, K. and Chow, J.Y., 2020. Agent-based Simulation Model and Deep Learning Techniques to Evaluate and Predict Transportation Trends around COVID-19.	https://c2smart.engineering.nyu.edu/wp-content/uploads/2022/04/C2SMART-COVID-19-Whitepaper-Issue3-1.pdf
White paper	Bernardes, S.D., Bian, Z., Thambiran, S.S.M., Gao, J., Na, C., Zuo, F., Hudanich, N., Bhattacharyya, A., Ozbay, K., Iyer, S. and Chow, J.Y., 2020. NYC Recovery at a Glance: The Rise of Buses and Micromobility.	https://c2smart.engineering.nyu.edu/wp-content/uploads/2022/04/C2SMART-COVID-19-Whitepaper-Issue-4.pdf
White paper	Gao, J., Bhattacharyya, A., Wang, D., Hudanich, N., Sooryaa, S., Thambiran, M., Bernardes, S.D., Na, C., Zuo, F., Bian, Z. and Ozbay, K., 2020. Toward the" New Normal": A Surge in Speeding, New Volume Patterns, and Recent Trends in Taxis/For-Hire Vehicles.	https://c2smart.engineering.nyu.edu/wp-content/uploads/2022/04/C2SMART-COVID-19-Whitepaper-Issue5.pdf
Presentation	2020 C2SMART webinar. Jingqin Gao, Yubin Shen, Developing C2SMART's Interactive COVID-19 Data Dashboard	https://youtu.be/pnMxK0e7iqQ

Presentation	2020 C2SMART webinar. Fan Zuo, Public Traffic Cams and Computer Vision to Analyze Social Distancing Trends	https://youtu.be/EwVEuqSW5ig
Presentation	BTR #2 Aug 2020 - Track 1: COVID-19 Pandemic and Transportation. Joseph Chow, Impact of COVID-19 Behavioral Inertia on Reopening Strategies for NYC Transit.	https://youtu.be/vFN3fR8ZLjI?t=12596
Presentation	BTR #2 Aug 2020 - Track 1: COVID-19 Pandemic and Transportation. Kaan Ozbay, Remote Capturing of social distancing through advanced video analytics under COVID-19 conditions.	https://youtu.be/vFN3fR8ZLjI?t=20988
Presentation	ITS-NY Emerging Trends Webinar	https://static1.squarespace.com/static/5a998fe5a9e0286b82f94913/t/6179c8fa21c10f3ba883663b/1635371259705/Jannie_Ding_Presentation.pdf
Presentation	TRB Webinar: Visualizing Effects of COVID-19 on Transportation: A One-Year Retrospective	https://www.nationalacademies.org/event/03-08-2021/trb-webinar-visualizing-effects-of-covid-19-on-transportation-a-one-year-retrospective
Presentation	The Third Annual National Mobility Summit of US Department of Transportation University Transportation Centers, Apr 15, 2021	http://mobility21.cmu.edu/wp-content/uploads/2021/04/C2SMART_Slides_UTC_National-Mobility-Summit_20210415-r.pdf
Presentation	NYU Urban Research Day 2021	https://www.nyu.edu/content/dam/nyu/urbanInitiative/documents/2021%20URD%20Digital%20Booklet.pdf

References

[1] F. Zuo *et al.*, "An Interactive Data Visualization and Analytics Tool to Evaluate Mobility and Sociability Trends During COVID-19," *arXiv preprint arXiv:2006.14882,* 2020.

[2] N. M. Ferguson *et al.*, "Strategies for containing an emerging influenza pandemic in Southeast Asia," *Nature,* vol. 437, no. 7056, pp. 209-214, 2005.

[3] F. Zuo, J. Gao, A. Kurkcu, H. Yang, K. Ozbay, and Q. Ma, "Reference-free video-to-real distance approximation-based urban social distancing analytics amid COVID-19 pandemic," *Journal of Transport & Health,* vol. 21, no. 101032, 2021.

[4] P. Ryus *et al.*, "NCHRP Report 797: Guidebook on Pedestrian and Bicycle Volume Data Collection," 2014.

[5] A. Bhattacharyya, J. Gao, K. Ozbay, F. Zuo, and S. Iyer, "Understanding Transport Challenges to Help Advance Equitable Transportation for People with Disabilities and Low Income Households During Covid-19: A Survey-Based Approach in New York City," *Available at SSRN 3996139*.

[6] C2SMART. "Development and Tech Transfer of Multi-agent Virtual Simulation Testbed Ecosystem." c2smart.engineering.nyu.edu/tech-transfer-simulation-test-bed/ (accessed June, 2020).

[7] S. B. de Oliveira *et al.*, "Monitoring social distancing and SARS-CoV-2 transmission in Brazil using cell phone mobility data," *medRxiv,* 2020.

[8] M. Faggian, M. Urbani, and L. Zanotto, "Proximity: a recipe to break the outbreak," *arXiv preprint arXiv:2003.10222,* 2020.

[9] A. Berke, M. Bakker, P. Vepakomma, R. Raskar, K. Larson, and A. Pentland, "Assessing Disease Exposure Risk With Location Histories And Protecting Privacy: A Cryptographic Approach In Response To A Global Pandemic," *arXiv preprint arXiv:2003.14412,* 2020.

[10] H. Cho, D. Ippolito, and Y. W. Yu, "Contact Tracing Mobile Apps for COVID-19: Privacy Considerations and Related Trade-offs," *arXiv preprint arXiv:2003.11511,* 2020.

[11] T. L. Inn, "Smart City Technologies Take on COVID-19," *World Health,* 2020.

[12] B. Udugama *et al.*, "Diagnosing COVID-19: The Disease and Tools for Detection," *ACS nano,* 2020.

[13] S. Guo *et al.*, "Droplet-Transmitted Infection Risk Ranking Based on Close Proximity Interaction," *Frontiers in Neurorobotics,* vol. 13, p. 113, 2020.

[14] N. Zhang, B. Su, P.-T. Chan, T. Miao, P. Wang, and Y. Li, "Infection spread and high-resolution detection of close contact behaviors," *International Journal of Environmental Research and Public Health,* vol. 17, no. 4, p. 1445, 2020.

[15] P. Vanhems *et al.*, "Estimating potential infection transmission routes in hospital wards using wearable proximity sensors," *PloS one,* vol. 8, no. 9, 2013.

[16] A.Barrat, C.Cattuto, A.E.Tozzi, P.Vanhems, and N.Voirin, "Measuring contact patterns with wearable sensors: methods, data characteristics and applications to data-driven simulations of infectious diseases," *Clinical Microbiology and Infection,* vol. 20, no. 1, pp. 10-16, 2014.

[17] C. T. Nguyen *et al.*, "Enabling and Emerging Technologies for Social Distancing: A Comprehensive Survey," *arXiv preprint arXiv:2005.02816,* 2020.

[18] N. Zhang, J. W. Tang, and Y. Li, "Human behavior during close contact in a graduate student office," *Indoor Air,* vol. 29, no. 4, pp. 577-590, 2019.

[19] C. Cattuto, W. Van den Broeck, A. Barrat, V. Colizza, J.-F. Pinton, and A. Vespignani, "Dynamics of person-to-person interactions from distributed RFID sensor networks," *PloS one,* vol. 5, no. 7, 2010.

[20] M. Bernas, B. Płaczek, W. Korski, P. Loska, J. Smyła, and P. Szymała, "A survey and comparison of low-cost sensing technologies for road traffic monitoring," *Sensors,* vol. 18, no. 10, p. 3243, 2018.
[21] FHWA, "Traffic monitoring guide," ed: Federal Highway Administration of the US Dept. of Transportation Washington, DC, 2013.
[22] C. O. Manlises, J. M. Martinez, J. L. Belenzo, C. K. Perez, and M. K. T. A. Postrero, "Real-time integrated CCTV using face and pedestrian detection image processing algorithm for automatic traffic light transitions," in *2015 International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*, 2015: IEEE, pp. 1-4.
[23] K. Xie, K. Ozbay, H. Yang, and C. Li, "Mining automatically extracted vehicle trajectory data for proactive safety analytics," *Transportation research part C: emerging technologies,* vol. 106, pp. 61-72, 2019.
[24] K. Xie *et al.*, "Development of a comprehensive framework for video-based safety assessment," in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, 2016: IEEE, pp. 2638-2643.
[25] C. Li *et al.*, "Robust vehicle tracking for urban traffic videos at intersections," in *2016 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2016: IEEE, pp. 207-213.
[26] B. Lobe, D. Morgan, and K. A. Hoffman, "Qualitative Data Collection in an Era of Social Distancing," *International Journal of Qualitative Methods,* vol. 19, p. 1609406920937875, 2020.
[27] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440-1448.
[28] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91-99.
[29] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961-2969.
[30] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980-2988.
[31] Z.-Q. Zhao, P. Zheng, S.-t. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE transactions on neural networks and learning systems,* vol. 30, no. 11, pp. 3212-3232, 2019.
[32] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779-788.
[33] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7263-7271.
[34] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767,* 2018.
[35] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," *arXiv preprint arXiv:2004.10934,* 2020.
[36] S. Ghader, J. Zhao, M. Lee, W. Zhou, G. Zhao, and L. Zhang, "Observed mobility behavior data reveal social distancing inertia," *arXiv preprint arXiv:2004.14748,* 2020.
[37] S. Engle, J. Stromme, and A. Zhou, "Staying at home: mobility effects of covid-19," *Available at SSRN,* 2020.
[38] M. Abadi *et al.*, "Tensorflow: A system for large-scale machine learning," in *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)*, 2016, pp. 265-283.

[39] F. Chollet, "keras," ed, 2015.
[40] M. Olafenwa and J. Olafenwa, "ImageAI," 2018. [Online]. Available: https://github.com/OlafenwaMoses/ImageAI.
[41] J. Du, "Understanding of object detection based on CNN family and YOLO," in *Journal of Physics: Conference Series*, 2018, vol. 1004, no. 1: IOP Publishing, p. 012029.
[42] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, "Understanding deep learning requires rethinking generalization," *arXiv preprint arXiv:1611.03530,* 2016.
[43] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE transactions on pattern analysis and machine intelligence,* vol. 32, no. 9, pp. 1627-1645, 2009.
[44] T.-Y. Lin *et al.*, "Microsoft coco: Common objects in context," in *European conference on computer vision*, 2014: Springer, pp. 740-755.
[45] R. Szeliski, *Computer vision: algorithms and applications*. Springer Science & Business Media, 2010.
[46] J. L. Bentley, "Multidimensional binary search trees used for associative searching," *Communications of the ACM,* vol. 18, no. 9, pp. 509-517, 1975.
[47] E. F. Morgul *et al.*, "Virtual sensors: Web-based real-time data collection methodology for transportation operation performance analysis," *Transportation Research Record,* vol. 2442, no. 1, pp. 106-116, 2014.
[48] New York City Department of Finance. "Open Parking and Camera Violations, NYC Open Data Portal." https://data.cityofnewyork.us/City-Government/Open-Parking-and-Camera-Violations/nc67-uf89 (accessed July 6, 2021).
[49] NYPD. "Motor Vehicle Collisions." https://data.cityofnewyork.us/Public-Safety/Motor-Vehicle-Collisions-Crashes/h9gi-nx95 (accessed August, 2020).
[50] Citi Bike. "Citi Bike System Data." https://www.citibikenyc.com/system-data (accessed Sep, 2020).
[51] J. Gao *et al.*, "The Effects of the COVID-19 Pandemic on Transportation Systems in New York City and Seattle, USA, C2SMART COVID-19 Whitepapers," 2020.
[52] WHO. "Coronavirus disease (COVID-19) advice for the public." https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public (accessed.
[53] US CDC. "Social Distancing." https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/social-distancing.html (accessed.
[54] L. Bourouiba, "A sneeze," *New England Journal of Medicine,* vol. 375, no. 8, p. e15, 2016.
[55] D. Wang, B. Y. He, J. Gao, J. Y. Chow, K. Ozbay, and S. Iyer, "Impact of COVID-19 Behavioral Inertia on Reopening Strategies for New York City Transit," *arXiv preprint arXiv:2006.13368,* 2020.
[56] C2SMART. "Reference-Free Video-to-Real Distance Approximation-Based Pedestrian Detection System Amid COVID-19 Pandemic." https://www.itskrs.its.dot.gov/node/209597 (accessed August 12, 2021).

Appendix

White paper issue 1: Initial Impacts of COVID-19 on Transportation Systems: A Case Study of the U.S. Epicenter, the New York Metropolitan Area

https://www.dropbox.com/s/jajp7nz05q2t314/C2SMART%20COVID-19%20Whitepaper_draft_v10_final.docx?dl=0

White paper issue 2: The Effects of the COVID-19 Pandemic on Transportation Systems in New York City and Seattle, USA

https://www.dropbox.com/s/9qh1fvw8xga0t9u/C2SMART%20COVID-19%20Whitepaper_Issue2_v11_final.docx?dl=0

White paper issue 3: Agent-based Simulation Model and Deep Learning Techniques to Evaluate and Predict Transportation Trends around COVID-19

https://www.dropbox.com/s/f1x98dvw425vmb3/C2SMART%20COVID-19%20Whitepaper_Issue3_v16.docx?dl=0

White paper issue 4:NYC Recovery at a Glance – The Rise of Buses and Micromobility

https://www.dropbox.com/s/ge7cjjsibfodmkb/C2SMART%20COVID-19%20Whitepaper_Issue4_v12.docx?dl=0

White paper issue 5: Toward the “New Normal” – A Surge in Speeding, New Volume Patterns, and Recent Trends in Taxis/For-Hire Vehicles

https://www.dropbox.com/s/i5hfe9ejs1sac59/C2SMART%20COVID-19%20Whitepaper_Issue5_v8.docx?dl=0