



Understanding COVID-19 Effects on Mobility: A Community-Engaged Approach

Arun Sharma, Majid Farhadloo, Yan Li, Jayant Gupta, Aditya Kulkarni, and Shashi Shekhar
Department of Computer Science, University of Minnesota-Twin Cities, Minneapolis, Minnesota

Correspondence: Arun Sharma (sharm485@umn.edu)

Abstract. Given aggregated mobile device data, the goal is to understand the impact of COVID-19 policy interventions on mobility. This problem is vital due to important societal use cases, such as safely reopening the economy. Challenges include understanding and interpreting questions of interest to policymakers, cross-jurisdictional variability in choice and time of interventions, the large data volume, and unknown sampling bias. The related work has explored the COVID-19 impact on travel distance, time spent at home, and the number of visitors at different points of interest. However, many policymakers are interested in long-duration visits to high-risk business categories and understanding the spatial selection bias to interpret summary reports. We provide an Entity Relationship diagram, system architecture, and implementation to support queries on long-duration visits in addition to fine resolution device count maps to understand spatial bias. We closely collaborated with policymakers to derive the system requirements and evaluate the system components, the summary reports, and visualizations.

Keywords. Information systems applications, Data mining, Decision support systems, Data analytics

1 Introduction

The coronavirus disease 2019 (COVID-19) has impacted public health with hundreds of thousands of mortalities and millions of confirmed cases. COVID-19 policy interventions significantly changed people's mobility patterns in many places such as urban trips and traffic congestion.

Given aggregated privacy-protected mobile device data, we aim to understand the impact of COVID-19 policy interventions on mobility. In collaboration with policymakers in Minnesota since Spring 2020 Shekhar (2020); Li (2020); Sharma et al. (2021), we have explored issues such as: Where are the potential hotspots of the hangouts (e.g., long duration visits)? How are these hotspots evolving? Is the public complying with policy interventions? Providing answers to these questions is important for not only for

addressing the issue of safely reopening the economy but also devising new ways to reduce the spread of coronavirus disease. Hence, we design a community-engaged tool via close collaboration with end-users such as policymakers, healthcare, and transportation analysts.

The challenges of this problem are three-fold. First, we are required to work closely with end-users in order to understand their policy and socioeconomic related questions (e.g., closing and reopening of businesses, etc.) The second challenge concerns handling cross-jurisdictional variability given the choice and timing of interventions within a given geographic region. Policy intervention differs within sub-regions of a given geographic area. Hence, it is challenging to create a learning model that encompasses all spatial variability in a fine-geographic space (i.e., census tracts or block groups) thereby increasing geographic complexity. The third challenge is handling large data volume which involve several points of interest throughout the US. Other challenges are related to addressing data quality questions (e.g., unknown sampling bias, location privacy) posed by the end-users.

Google Datasets (2020); 'Reports' (2020), StreetLight "StreetLight" (2020) and other sources provide rich mobility location data for understanding changes in human mobility over time. They provide aggregated mobility data in the form of reports 'Reports' (2020) summarizing mobile device data. However, such reports are limited to the frequency of visits to certain points of interest (POIs) aggregated by certain geographic areas (e.g., cities, states, and countries) and do not consider visit duration. Hence, these reports do not separate short and long-duration visits, which are of interest to local and state governments towards reducing disease spread. Other platforms include early work based on an interactive dashboards for human mobility trends Desjardins et al. (2020); Dong et al. (2020); Gao et al. (2020); Samet et al. (2020), which includes Geographic Information System (GIS) visualization at the county-level computing on the fly statistical measures. However, none of the application dashboards actively engage end-users to address complex queries and generate custom reports at finer geographic complexity.

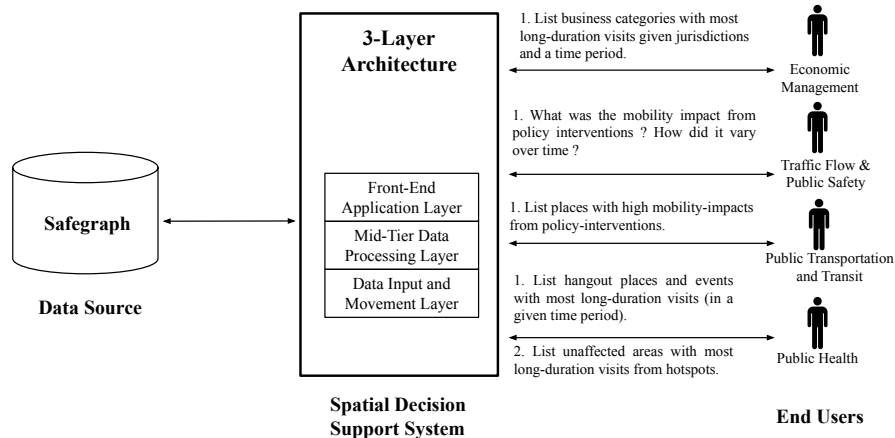


Figure 1. Context Diagram with end-users and data sources

Contributions: Our main contributions are as follows:

- We describe a community-engaged COVID-19 decision support system, based on close collaboration with end-users, who provided the system requirements such as the queries of interest (e.g., business categories with many long-duration visits, sampling bias, data privacy safeguards, etc.)
- We present the Entity-Relationship diagram to improve understanding of the aggregated mobile device data and facilitates a richer set of queries such as those related to long duration visits.
- We report the user feedback on summary reports and visualizations generated by our system leveraging the Entity-Relationship diagram.

Scope: Our aim is to understand spatiotemporal patterns of mobility affected by COVID-19 in Minnesota using aggregated privacy-protected mobile device data. COVID cases correlation and detailed assessment of data quality beyond simple sanity checks and the characterizing relationship between disease spread and mobile-phone data fall outside the scope of this paper.

Outline: The rest of the paper is organized as follows: Section 2 describes application domains and policy context. Section 3 uncovers the overall system architecture with a brief description of each layer. Section 4 details the proposed entity-relationship diagram and identified in response to the questions from end-users. Section 5 describes the validation process used to verify the design of a new schema via a case study. Section 6 discusses user feedback and data quality issues along with data and software availability. Section 7 gives a broad overview of related work. Lastly, Section 8 concludes this work and lists plans for future research.

2 Application Domain and Policy Context

In March 2020, we were invited by policy analysts and the Center for Transportation Studies at the University of Minnesota and policy analysts to investigate a number of

questions related to the impact of the pandemic and state-mandated interventions on mobility. In collaboration, we engaged with this audience through a sequence of online meetings, interviews, and email exchanges to better understand their information needs and understand the requirements for our decision-making system. The four main categories of end-users participating in these exchanges were researchers and policymakers in Public Health, Economic Management, Traffic Flow and Public Safety, and Public Transportation and Transit summarized in Figure 1.

Public Health: Public health policymakers and researchers were interested in calibrating mobility-sensitive parameters in disease transmission dynamics models (e.g., SEIR model Bjørnstad et al. (2020)). The mobility information of interest to this group included the size of gatherings and the number of long-duration visits at vulnerable places and events (e.g., indoor hangouts, high-density outdoor gatherings, such as super spreader events) since these may help estimate the number of contacts. In addition, they were interested in travel between disease hotspots and other areas to estimate the probability of future spread to new geographic locations.

Economic Management: Economic management policymakers and analysts were interested in understanding the impact of the policy interventions as well as how to safely reopen the economy. They asked for reports on the number of visits and the number of long-duration visits for different business categories (e.g., bars, full-service restaurants, and limited-service restaurants). They also requested an aggregate reports on trends in average distance traveled and average time spent at home in order to assess compliance such as Minnesota Stay-at-Home orders and phased reopening Plan (2020) of the economy.

Traffic Flow and Public Safety: The traffic flow and public safety community were interested in understanding the current and future impact of the pandemic on travel demand (e.g., vehicle miles traveled for commuting and delivery of goods and services, etc.), quality of road service (e.g., congestion, average speed), safety (e.g., traffic accidents) and the environmental impacts (e.g., emis-

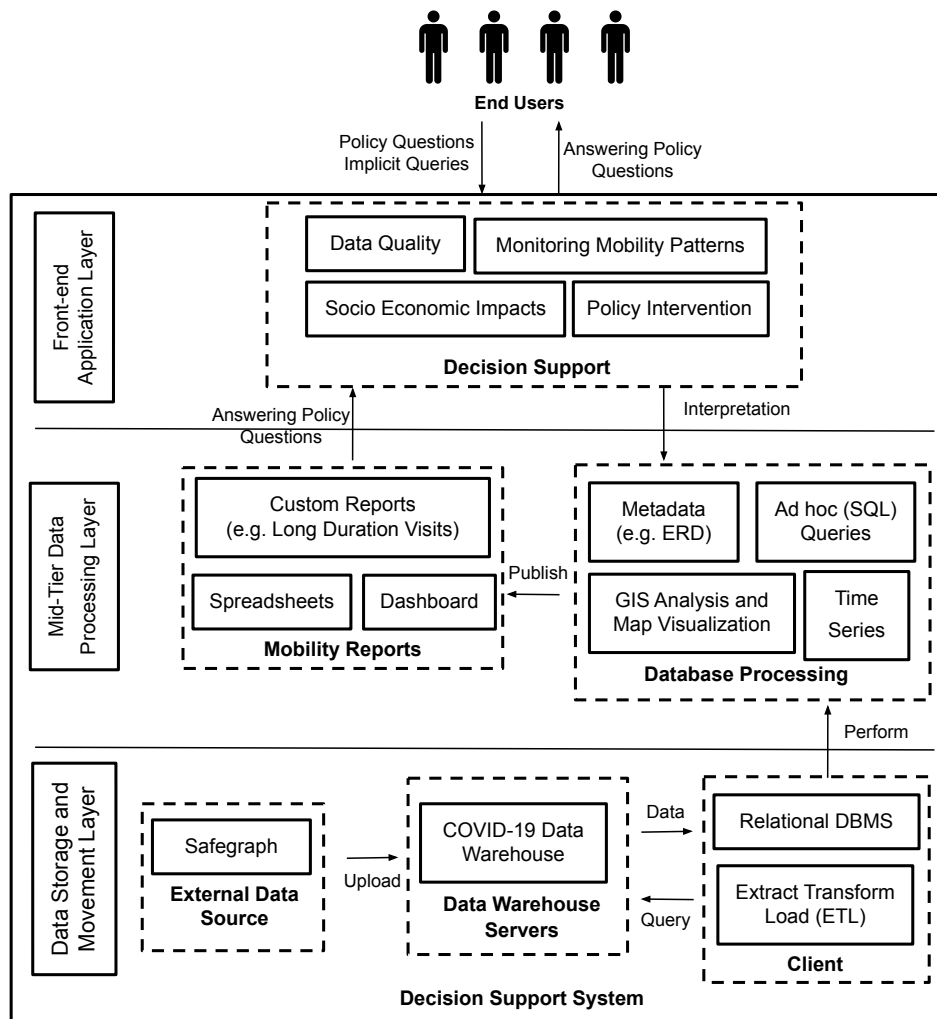


Figure 2. 3-Layer Architecture of the Community-Engaged Decision Support System

sions). They also asked questions on the quality of data in terms of sampling bias, geographic coverage, sampling frequency, and comparison with the ground truth dataset (e.g., travel surveys and loop detectors). Furthermore, they also requested a deep engagement by inviting us to present our findings at technical conferences, public webinars, and legislative hearings for the State House of Representatives, Transportation, Finance, and Policy Committees.

Public Transportation and Transit: Analysts from transit were interested in most frequently visited Points of Interest (e.g., Minneapolis-St.Paul Airport) during the pandemic phase. This may help to plan new major bus routes surrounded by such POIs and their usage during the stay-at-home order or reopening phase. In addition, they were also interested in understanding mobility patterns based on finer temporal granularity (e.g., commute hours) for certain days in the week (e.g., Weekdays vs Weekends) and different calendar events.

3 Community-Engaged Decision Support System

This section provides a brief overview of the proposed decision support system developed in consultation with our end-users. The system generates weekly reports based on policy intervention questions posed by the end-users during the course of the pandemic. The end-users from interdisciplinary fields can interact with our decision support system and pose questions of interest (e.g., long duration visit queries). The system outputs a customized report along with supplemented details for answering data quality concerns (e.g., sampling bias, differential privacy). The system has three layers: a Data Storage and Movement Layer, a Mid-Tier Data Processing Layer, and a Front-end Application Layer. Details are as follows:

Data Storage and Movement Layer: The first layer consists of three components: External Data Sources, Data Warehouse Servers, and the Client. The primary goal of this layer is gathering mobility data from remote geospatial servers to the client and mid tier Data Processing Layer. The Data Source Layer contains source files such

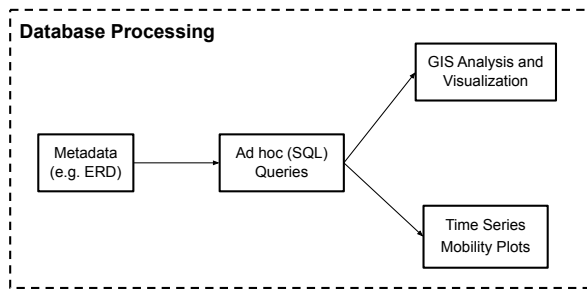


Figure 3. Database Processing Component.

as Safegraph that are periodically fetched by the Data Warehouse Servers and saved in a COVID-19 Data Warehouse. The COVID-19 Data Warehouse is timely maintained by USpatial Uspatial (2020), which periodically uploads comma-separated variables files to their PostgreSQL servers across different tables (e.g., weekly patterns and social distancing). These records are fetched by the client using secured VPN performing Extract Transform Load (ETL) operations and saved in a relational database table. The data is then prepared for more complex analysis based on policy intervention questions, as discussed in the Mid-Tier Data Processing Layer.

Mid Tier Data Processing Layer: The main objective of Mid-tier data processing is to perform desired spatial and spatiotemporal mobility analytics based on a given pre-processed dataset from the client. The layer has two components, Database Processing, and Mobility Reports. Within Database Processing component we first check if policy intervention questions can be answered via an Entity Relationship Diagram. If yes, then the questions can be further formalized via SQL queries. These queries may require additional I/O operations from the data warehouse servers in the Data Storage Layer. After formalizing and executing ad-hoc SQL queries, we perform more detailed mobility patterns via time-series plots and other geographic maps and visualizations (shown in Figure 3). Based on the policy questions, such queries further involve join operations with different tables in Safegraph for providing Time Series Mobility Analysis. These findings are then documented in the form of spreadsheets and presentations to communicate with policymakers in the interpretation of mobility patterns.

Front End Application Layer: The primary objective of this layer is to provide decision support and interpretation of the results published in the form of reports the end-users. In addition, end-users can pose questions and queries (e.g., long-duration visits) and evaluate policy decisions based on human mobility patterns. These reports are further verified for accordance with the policy intervention calendar Plan (2020) of Minnesota.

4 Entity Relationship Diagram for Weekly Pattern Data on Safegraph Dataset

‘SafeGraph’ (2020) is a mobility data vendor company that provides anonymized aggregated location data from mobile devices. Even though Safegraph provides invaluable data, its denormalized (a single flat-table) schema makes it hard to discern the semantic richness of the data and support ad-hoc queries such as those related to long-duration visits. Figure 4 shows the denormalized schema of the "weekly pattern" table from the SafeGraph dataset. With this format, it is hard to answer queries that distinguish between frequently visited places (e.g., grocery stores, fast-food restaurants) and hangout places with long-duration visits (e.g., bars and full-service restaurants), a critical requirement for this study. For example, even though the number of visits to limited-service restaurants (e.g., fast food restaurants) is much greater visits the number of visits to alcoholic bars, bars are found to be more important for COVID-19Walsh (2020).

To overcome these limitations, we designed a conceptual schema (Entity-Relationship diagram) that satisfies the 3rd normal form rule to support a richer set of ad-hoc queries for the currently available weekly pattern data from the Safe-Graph dataset. We initially asked data engineers at SafeGraph whether a conceptual data model (e.g., entity relationship diagram) existed on the current design of the database. Despite relatively complicated and inflexible design, no effort had previously been made towards building a conceptual data model. We also carried out an extensive search of recently published works by organizations and academic researchers involved in the Safe-Graph COVID-19 Data Consortium ‘SafeGraph’ (2020) but could not find any conceptual data model. Defining data conceptually improves understanding of the semantics of the data and enhances the interpretability of the data to a more general audience. It also facilitates representation of a richer set of queries by defining relational tables in 3rd normal form. Therefore, we designed an Entity-Relationship (ER) diagram that satisfies 3rd normal forms on the weekly pattern data (Figure 5).

4.1 Proposed Schema on Weekly Patterns Dataset

In the early stage of the Entity Relationship (ER) diagram design, we encountered many challenges in specifying the major entities. In current dataset, the mobile device data is currently aggregated at a Census Block Group level to address privacy concerns and data protection. Furthermore, a data redundancy issue arises since various data attributes are replicated across several key-value pairs that make it hard to distinguish major entities.

To address these issues, we have incorporated specific design decision steps that enable us to model ad-hoc queries (e.g., queries related to hangout places with many concurrent long-duration visits) in as much detail as possible to understand people’s mobility affected by COVID-19. First, we consolidated the overall structure of

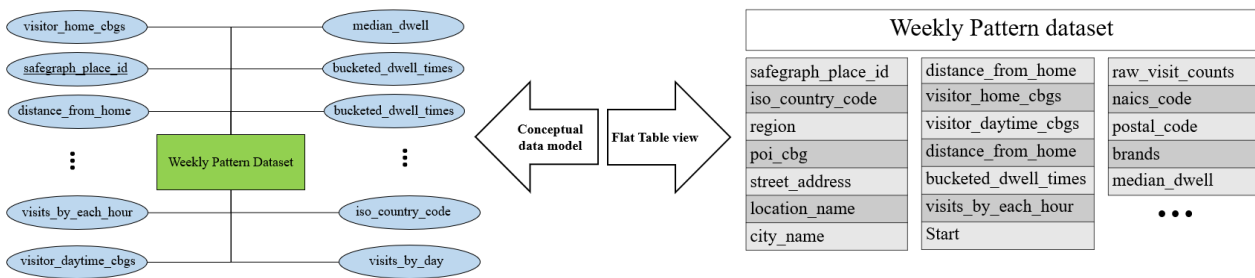


Figure 4. Denormalized Schema of Weekly Patterns Data (Best in Color)

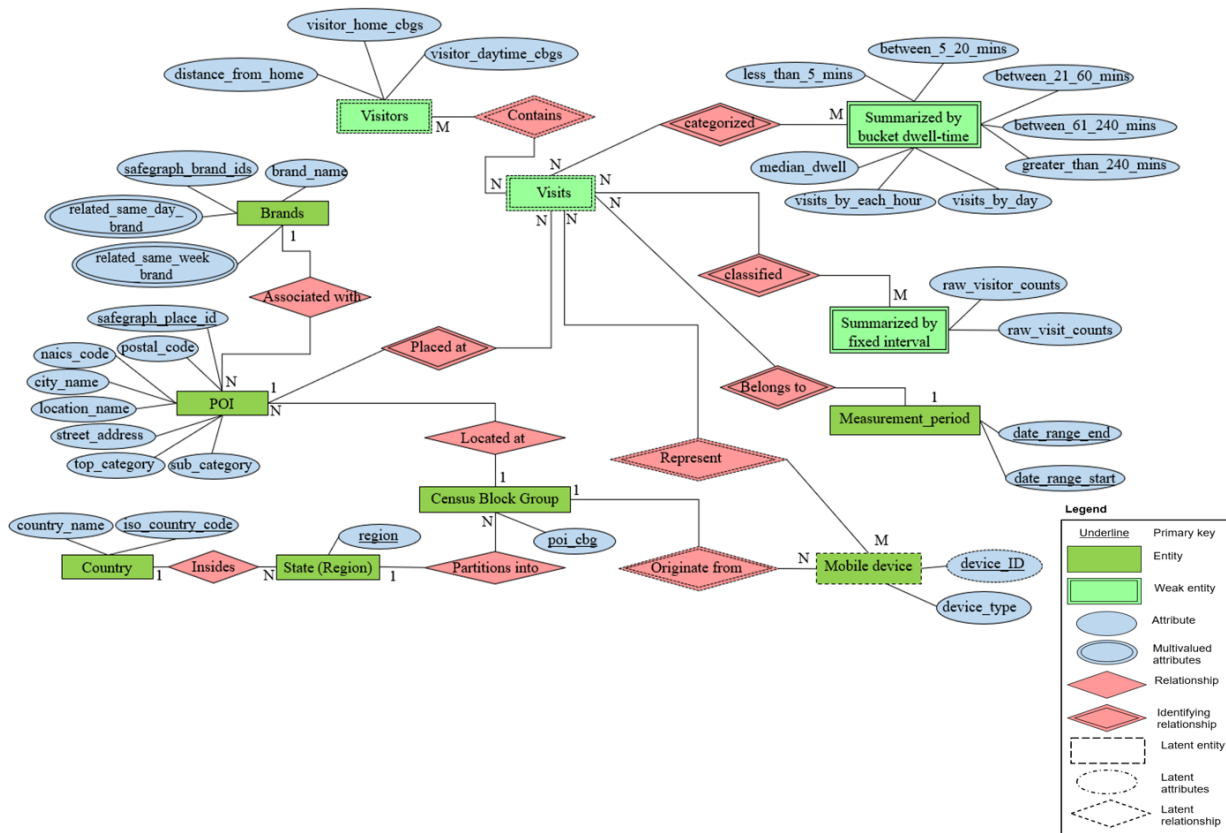


Figure 5. An Entity-Relationship diagram of the weekly pattern data from the SafeGraph dataset (Best in Color).

the ER diagram by adding latent entities, such as “Mobile device”, “Visitors”, and “Visits” (green dashed green boxes in Figure 5), which substantially improve our understanding of the denormalized schema. Due to privacy concerns as well as data protection, relevant information on latent entities are not currently available. However, having latent entities allows us to correctly identify entities from the available dataset and subsequently assign necessary existing or derived data attribute columns to each entity. We further improved the design of the ER diagram by summarizing the frequency of visits into bucket dwell times and fixed interval entities. For example, we model queries related to hangout places using bucket dwell

time data attributes, including “between_21_60_mins”, “between_61_240_mins”, and “greater_than_240_mins” from the “Summarized by bucket dwell-time” entity, which helps us address concerns from end-users regarding safely reopening of the economy. The complete list of entities for designing the ER diagram is shown in Table 2. Table 3 relationships (reddish diamonds in Figure 5) associate entities to one or more other entities. For example, a state is “inside” a country, each state “partitions into” many Census Block Groups (CBGs), and many Points Of Interests (POIs) are “located at” each CBG. Next, we incorporate a key design improvement into the ER diagram by adding latent identifying relationships (double dashed

diamonds in Figure 5) and derived identifying relationships (double diamonds in Figure 5). Similar to identifying entities, having latent relationships in our ER diagram design enables us to go further in specifying the correct entities and data attribute columns for each entity. For example, mobile devices “represent” visits that “took place at” a POI, which “belongs to” various times and days of the week. The number of visits can be further “categorized/classified” as a fixed interval or bucket dwell time (e.g., visit duration longer than 20 minutes). Lastly, cardinality constraints (the numbers in Figure 5) determine the least and greatest number of occurrences of an entity that could be related to a single occurrence of another entity. Cardinality constraints (e.g., one-to-many, many-to-many) represent business rules. For instance, we were able to create a hangout query place by modeling the frequency of visits that represent “many” mobile devices, which “originate from” a CBG (e.g., home, workplace, etc.) and travel to another CBG. “Many” POIs (e.g., grocery store, full-service restaurants, bars) are also “located” at “one” CBG that is further “associated” with “many” brands (e.g., Walmart, McDonald’s, Sally’s). Table 3 shows the complete list of the relationships between pairs of entities.

4.2 An examination of the Entity Relationship Diagram

A benefit of having a conceptual point of view Entity Relationship Diagram (ERD) is that it enables us to introduce a richer set of queries and subsequently assess which if any can be answered using the current dataset. Accordingly, we devised a set of 8 queries, 4 of which could be answered with the currently available dataset and 4 which require additional information or a supplementary dataset, as shown in Table 1.

The first four queries could be answered using the existing dataset since the required data, i.e., spatial extent of each census block group, the number of visits, Points Of Interests (POIs), and visit duration are available. On the other hand, the last four queries may provide useful and interesting patterns that can help policymakers; however, due to a lack of information (e.g., confirmed COVID-19 cases and unemployment rates) such queries could not be executed. To investigating these types of queries requires integrating complementary datasets.

5 Validation

We validated the proposed decision support system via a case study using a real world human mobility dataset for Minnesota. Section 5.1 describes SafeGraph COVID-19 Data Consortium ‘SafeGraph’ (2020) data, including Weekly Patterns and Social Distancing Dataset. In addition, we briefly talk about the Minnesota calendar of policy interventions. In Section 5.2, we investigate weekly visit trends via a case-study over top business categories with long-duration visits greater than 20 minutes. We further explore frequently of visits in various categories at

different bucket dwell times, which leads us to find new interesting spatiotemporal patterns. The overall validation framework is shown in Figure 6.

5.1 Dataset Description and Case Study in Minnesota

The dataset description and Case Study in Minnesota as described in Figure 6 and MN Policy Calendar (shown in Table 4) are as follows:

SafeGraph: The mobility data in this work was supported by COVID-19 Response SafeGraph Data Products ‘SafeGraph’ (2020) providing various datasets (e.g., social distancing, weekly patterns, monthly patterns, etc.) to academic researchers, non-profit organizations, and governments to study the primary and secondary effects of the novel coronavirus. The raw data was generated by using a panel of positioning system pings from anonymous mobile devices. We used weekly pattern data to monitor mobility across the most frequently visited business categories. The current weekly pattern data is given in a single flat table format (see Figure 4) that consists of many data attribute columns with a short description of each attribute. A core place table provides information such as a North American Industry Classification System (NAICS) Census (2021) categorizing business categories for Points Of Interests, which used in conjunction with Weekly Patterns. Social Distancing Metrics provide aggregated statistics of distance traveled within each Census Block Group.

A Case Study in Minnesota: As the case study in this work was conducted in Minnesota, we describe the summary statistics of SafeGraph in this context. The SafeGraph data for Minnesota was derived from 294,014 individual mobile devices, which is roughly 5% of the total state population. Further, it covers 73,548 points of interest (POIs) across 261 different business categories (e.g., full-service restaurants, limited-service restaurants, gasoline stations with convenience stores). Due to privacy concerns and data protection, SafeGraph aggregates individual-level anonymous data points to the census block group (CBG). Its spatial coverage in Minnesota is 4107 out of 4111 Bureau (2020) CBGs. This data was used for both specific and generic query types. For validating our community-based decision support system, we provide mobility analysis for specific queries related to long visits duration.

MN Policy Intervention Calendar: After figuring out which types of mobility questions were answerable with the available data, we then analyzed mobility patterns based on MN Policy Intervention Calendar, as shown in Table 4. These interventions were imposed by the state government in order to control and minimize the spread of COVID-19. The policy intervention calendar starts on March 17, 2020, with the University of Minnesota (U of M) School Closing ordered just after Spring Break (i.e., March 9, 2020). This order was followed by the Minnesota Stay-at-Home order issued on March 27, 2020. After holding a Stay-at-Home order for a month, a small reopening

Table 1. Query evaluation based on the proposed Entity Relationship Diagram

End-user	Query	Evaluation	Implicit with current schema	Explicit with current schema	Implicit with ERD	Explicit with ERD
Economic Management Public Transportation and Transit	1. What are the business categories located at a given Census Block Group and what is the distribution of visits in a given Census Block Group?	✓	✓	x	x	✓
Economic Management	2. Which business category (e.g., grocery stores, full-service restaurants, limited-service restaurants, etc.) has the highest number of visits?	✓	✓	x	x	✓
Economic Management Public Health	3. List the top ten bars in the state of Minnesota by long visit duration. A long visit duration is a visit that lasts longer than 20 minutes.	✓	✓	x	x	✓
Economic Management	4. Which business category has the least impact from COVID-19 pandemic and related policy interventions since the stay-at-home order took effect in early April?	✓	x	x	✓	x
Public Transportation and Transit	5. What is the difference in median distance traveled between commuters (e.g., people who go shopping) and delivery vehicles (e.g., USPS, DoorDash)?	x	[rgb]0.396,0.396,0.396	[rgb]0.396,0.396,0.396	[rgb]0.396,0.396,0.396	[rgb]0.396,0.396,0.396
Public Health	6. Does Census Block Group have the highest number of confirmed COVID-19 cases?	x	[rgb]0.396,0.396,0.396	[rgb]0.396,0.396,0.396	[rgb]0.396,0.396,0.396	[rgb]0.396,0.396,0.396
Traffic Flow Public Safety	7. How many smartphones reported from the I-35W bridge during a given protest event?	x	[rgb]0.396,0.396,0.396	[rgb]0.396,0.396,0.396	[rgb]0.396,0.396,0.396	[rgb]0.396,0.396,0.396
Economic Management	8. What is the difference in the unemployment rate based on gender, and which business category, and what brand contributes the most to the unemployment rate?	x	[rgb]0.396,0.396,0.396	[rgb]0.396,0.396,0.396	[rgb]0.396,0.396,0.396	[rgb]0.396,0.396,0.396

Table 2. Entity name and description in the weekly pattern data from the SafeGraph dataset.

Entity Name	Description
Country	The 2 letter ISO 3166-1 alpha-2 country code
State	The state or territory
Census Block Group	The census block group the POI is located within
Points of Interest (POI)	The business categories that are located within the census block group
Brands	Brands that are associated with the POI
Measurement period	The start and end time of the measurement period
Mobile device	The mobile devices that are located within a census block group
Visits	The mobile devices that visit a POI at a certain time
Visitors	The summary statistics on visitors for certain POIs in the given time period
Summarized by bucket dwell-time	The summary statistics on visitors for certain POIs based on bucket dwell time (shorter, average, or longer) visits
Summarized by fixed interval	The summary statistics on the visitors for certain POIs in a fixed interval

Table 3. Relationship descriptions in weekly pattern dataset between pairs of entities.

Entity 1	Relationship	Entity 2	Description
State	Inside	Country	A state or territory is located in a country
State	Partition into	Census Block Group	A state is spatially partitioned into multiple census block groups
POI	Located at	Census Block Group	Multiple POIs (i.e., business categories) are located in a census block group
Brands	Associated with	POI	One brand is associated with many POI
Mobile device	Originate from	Census Block Group	Multiple mobile devices originate from a census block group (i.e., visitors' home)
Visits	Placed at	POI	Many visits are placed at a POI.
Visits	Categorized	Summarized by bucket dwell-time	Many visits are categorized by many bucket dwell-times (e.g., how long a visit lasts)
Visits	Classified	Summarized by fixed interval	Many visits are categorized by many fixed intervals.
Visits	Belongs to	Measurement period	Visits to a POI from a mobile device belong to certain time and days of the week.
Visits	Contains	Visitors	Related information (e.g., distance from home) from visitors are contained within visits.
Mobile device	Represent	Visits	Many Multiple mobile devices represent visits are happened in a POI at given time.

Table 4. MN COVID-19 policy Interventions Calendar Plan (2020)

Dates	Social setting
Mar 9, 2020	University of Minnesota Spring break
Mar 17, 2020	University of Minnesota school closing
Mar 27, 2020	MN stay-at-home
May 18, 2020	MN reopening Phase 1
June 1, 2020	MN reopening Phase 2
June 10, 2020	MN reopening Phase 3
November 16, 2020	MN shutdown order for Bars and Restaurants
January 11, 2021	MN reopening order for Bars and Restaurants
May 27, 2021	No limits on size and no social distancing requirements.

phase was ordered with gatherings of 10 or less. Subsequent re-openings, namely, Phase 2 and Phase 3, were issued on June 1 and June 10, 2020, respectively, along with gradual increasing of gathering capacity for essential services. Due to an increase in COVID incidents, the state government issued another shutdown for certain business categories in mid-November which later reopened in early January. Further restrictions related to capacity and social distancing were lifted on May 27, 2021.

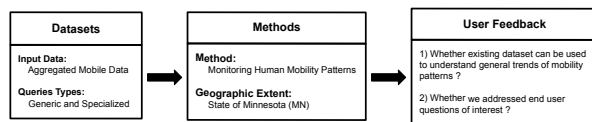


Figure 6. Validation Framework

5.2 Mapping and analyzing human mobility changes

To validate our schema with end-user queries, we provide a mobility analysis based on number of long-duration visits (i.e. visits with durations greater than 20 minutes) aggregated by weeks for about one year (i.e. from March 1, 2020 to June 28, 2021).

We first created summary reports of the most frequently visited business categories (as shown in Figure 7) based on long-duration visits. This led us to distinguish anomalous behaviors and mobility impacts in different business categories. For example, elementary and secondary schools were the most frequented business categories in early March 2020. But starting at 2020 spring break, and shortly after the closure of the schools, we observed an enormous drop in that category, and it has stayed low over the course of the pandemic. On the other hand, full-service restaurants, malls, and natural parks had an extensive initial drop; however, subsequently they illustrate positive trends, and almost recover to their original levels even before the start of the first phase of the opening. Next, we evaluated them based on the duration of people's visits.

Using the ERD, we quantified interactions based on bucket dwell times (e.g., visits longer than 20 minutes). We produced a summary of hangout place reports that identified POIs with long-duration visits. Classifying the frequency of visits based on visit duration allowed us to identify business categories (e.g., full-service restaurants,

bars), which despite having relatively fewer visits in comparison with general categories (e.g., grocery stores), included more hotspots of disease Nelson (2020); Mayerle (2020).

Figures 8, 9 and 10 show the frequency of visits to different business categories, including full-service restaurants, limited-service restaurants, and bars, based on various bucket dwell times ranging from shorter visits to longer visits (e.g., 61-240 minutes). We observed a rapid increase to bars and full-service restaurants that emerged shortly by the first phase of reopening (Figures 8 and 10), whereas the same is not true in case of Limited-Service Restaurants (Figure 9).

We also compared these trends with the outbreak reports by the Minnesota Department of Health (MDH) of Health (2021) that specifically name bars and restaurants deemed as venues with significant impacts on COVID spread Emily Haavik (2020) Johnson (2020) that were linked to COVID-19 cases during different months. According to the MDH Turtinen (2021), an outbreak location is a restaurant or bar with at least seven unrelated COVID cases from seven different households that only visited one restaurant or bar establishment during that month. To find patterns that exclusively apply to outbreak locations, we included non-outbreak locations to act as a control group which was selected to keep multiple external variables between the two groups constant. For each matched pair of an outbreak and non-outbreak location, we picked places within proximity of each other to maintain consistent local mobility patterns for both locations. Then, we ensured that locations were both part of the same business category either both full-service restaurants or both bars. Lastly, we checked that both locations had a similar number of weekly visits to their locations before the most recent COVID-19 lockdown: either visits before the March or November shutdown of indoor dining, depending on a specific month.

We analyzed the trend of long-duration visits to 15 locations linked to COVID outbreaks in June 2020 and 15 non-outbreak locations. As shown in Figure 11, we first notice an increase in long-duration visits for both outbreak and non-outbreak groups after the reopening period, which is expected. However, by June 29, 2020, the outbreak group's long-duration visits reached pre-COVID levels of visits (first week of March), whereas the non-outbreak group only reached 50%. That suggests that large increases in long-duration visits are associated with outbreaks. However, such observations needs to be further validated by disease transmission model (e.g., SIR/SEIR Martcheva (2015)) simulations which consider indoor information, such as mask use and spatial distance between individuals (e.g., where they are sitting, etc.).

6 Closing the Loop: User Feedback

To assess how well our decision-making system met the needs of the users, we sought feedback from our collaborators in transportation, economic management and public health. We begin by discussing their concerns with data

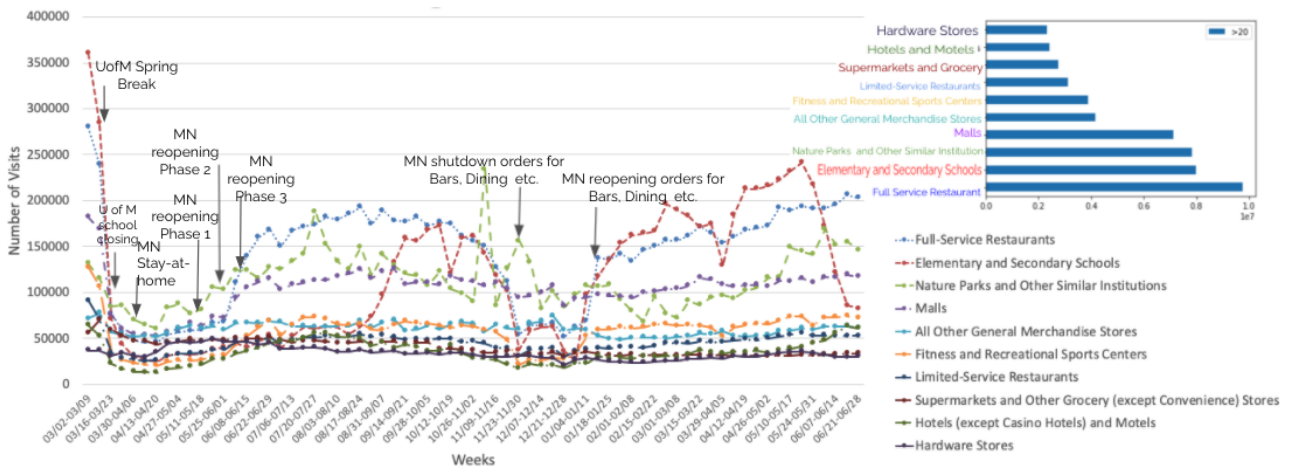
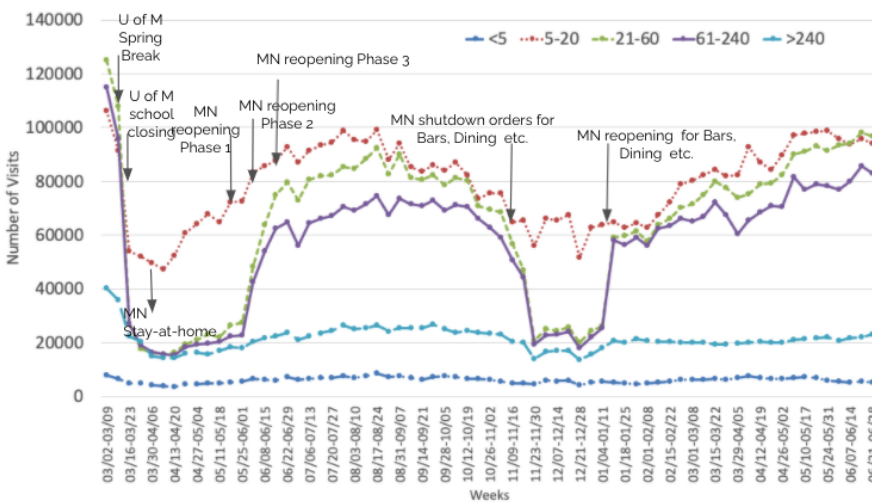
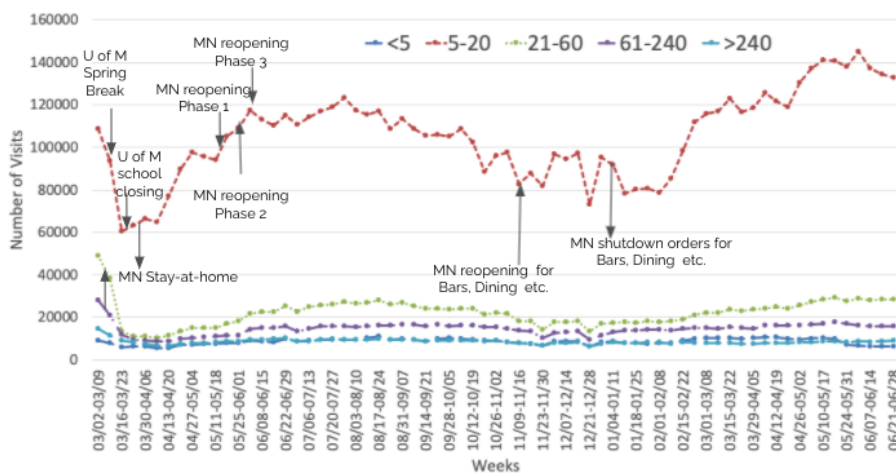


Figure 7. Most frequent long-duration visited business categories in Minnesota (March 2 ,2020 - June 28, 2021) (Best in Color).



Dwell Time	Visits
<5	416949
5-20	5446795
21-60	4416678
61-240	3867499
>240	1467323

Figure 8. Analysis of full-service restaurants visits based on the bucket dwell times (Best in Color).

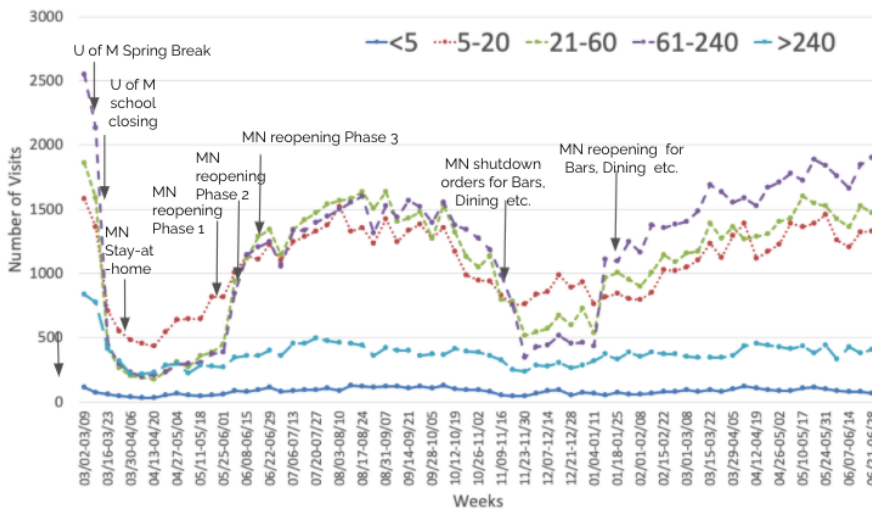


Dwell Time	Visits
<5	592995
5-20	7262497
21-60	1522801
61-240	1000987
>240	590133

Figure 9. Analysis of Limited Service Restaurants based on the bucket dwell times (Best in Color).

quality, then we summarize other feedback such as the value derived and the need for additional information to

interpret reports. This section provides a summary of the feedback.



Dwell Time	Visits
<5	5646
5-20	73095
21-60	74254
61-240	81226
>240	25724

Figure 10. Analysis of bar visits based on bucket dwell times (Best in Color).

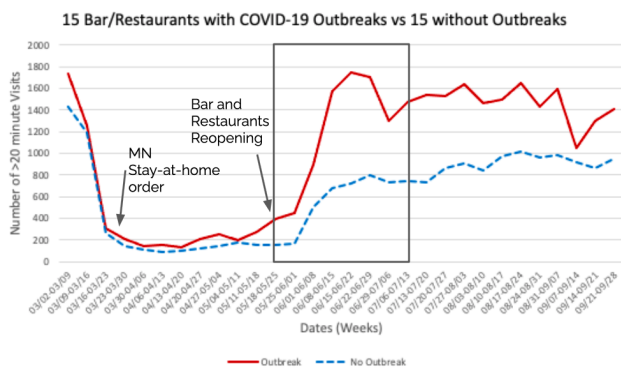


Figure 11. Long-duration visits for outbreak and non-outbreak groups (Best in Color).

6.1 Data Quality

In Spring 2020, many end-users expressed concerns about Safegraph data such as the following:

- How does the dataset protect the individual’s privacy?
- What is the sampling rate?
- Is the spatial and temporal granularity reasonable for the sample size ?
- How does the mobile device data compare with well known datasets in our domain ?

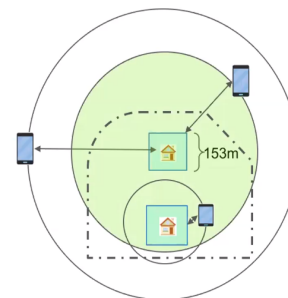
Next, we share our findings in context of above questions.

6.1.1 How does the dataset protect the privacy of individuals ?

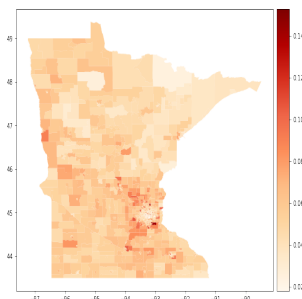
The mobile device data aggregates to census block groups (CBGs) because a geo-referencing for a home’s location includes a 153m × 153m buffer around the home, as shown in Figure 12a. However, this geocoding referencing may be considered as diluted, as it underestimates the effect of mobile people such as college students. To apply additional privacy and data protection, the median in some

data attributes (e.g., distance_traveled_from_home, median_home_dwell_time) are reported.

For example, as shown in Figure 12a, three mobile devices are within the range of a census block group, and the median range from those three devices is reported as the traveled distance from home. It is worth mentioning that if a census block group has fewer devices, that census block group is suppressed. Lastly, similar to US Census 2020 data, noise has been applied to safeguard data for protecting individual information.



(a) Reporting median range for a Census Block Group.



(b) A map of county-level sampling rate in Minnesota.

Figure 12. Data Quality assessment based on (a) data privacy and (b) sampling bias (Best in Color).

6.1.2 What is the sampling rate? How does it vary over the counties? Does vendor report any other sample bias (e.g., demographic)?

The visualization in Figure 12b provides a sampling rate based on the ratio between sampled devices and population in each census tract in Minnesota. As can be seen, the sampling rate ranges from 1% to 15% and is unevenly distributed. This helps us to understand the sampling bias based on population distribution i.e., around suburban regions near the Minneapolis Downtown region. However, our sampling bias discussion is limited to population distribution due to the paucity of demographic information in the given Safegraph dataset. Sampling bias issues in Safegraph data were also raised in literature Coston et al. (2021) such as lack of demographic interpretation. Because of the low sampling rate and uneven distribution, the data needs to be aggregated to a certain spatial resolution to avoid conveying misleading information. The original dataset groups the sampled mobile devices by the census block groups where their “home” are. However, there are census block groups with very few or no sampled devices, which likely does not represent the population.

6.1.3 Is the spatial and temporal granularity reasonable for the sample size ?

The mobile device data is spatially aggregated at a variety of levels starting from the census block group level up to the state level for all fifty states. In terms of temporal granularity, each mobile device is tracked at intervals ranging from monthly to hourly scale. However, the significance of a particular granularity varies from one disciplinary field to another. For instance, hourly monitoring of infection provide more value to the public health domain. Further, US Census data provide privacy for protecting individual information result in low accuracy as mentioned in Section 6.1.1. Such spatial aggregations is the most crucial factor affecting the data quality.

6.1.4 How does the mobile device data compare with well known datasets in our domain ?

Due to the data quality issues within the existing dataset, we needed to know whether the trends found in the mobile device dataset could be validated by ground truth, such as loop detector data. To validate, we visualized the vehicle miles traveled using loop detector data from the Minnesota Metropolitan Council Declines (2020). The dataset consists of observed vehicle miles traveled and predicted vehicle miles traveled. Predicted values are calculated based on a generalized additive model Hastie and Tibshirani (1990) using historical data from previous years.

Figure 13a shows the trends for observed and predicted miles traveled values over the entire freeway network in 2020. As can be seen, the network traffic flow dramatically dropped starting the third week of March compared to typical traffic with no outbreak or stay-at-home order in place. Figure 13b shows the relative trends in mobility for

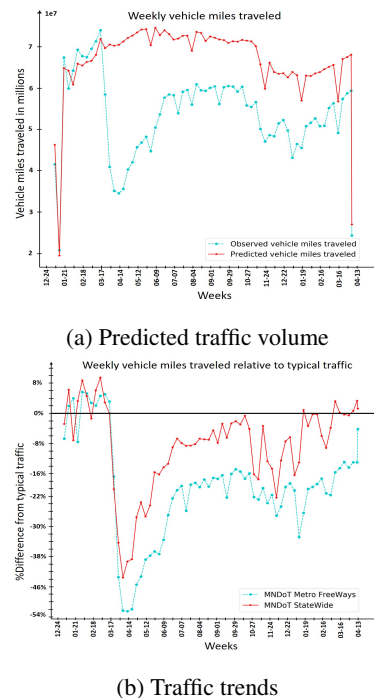


Figure 13. Decrease in traffic network across the COVID-19 timeline (Best in Color).

Minnesota in 2020, where the horizontal line at zero is the baseline, and the further below that line, the more traffic has decreased. As can be seen, the lowest point of network traffic for both the metro area freeway and statewide sensors happened just as the “stay-at-home” order took effect in late March.

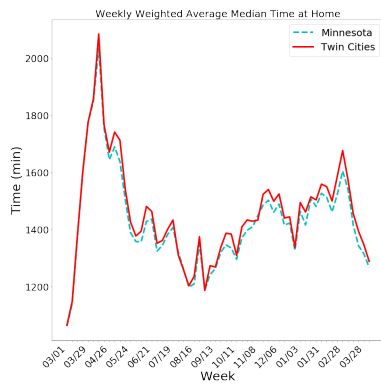
Similarly, we created mobility reports on two key data attributes, “time spent at home” and “distance travel from home” using the social distancing dataset. As shown in Figure 14, we observed similar trends to the loop detectors dataset for both metropolitan areas as well as for the state of Minnesota. Given these comparably equivalent trends and despite the existing data quality issues, the mobile device dataset may be used to identify novel spatiotemporal patterns and further evaluate and understand the COVID-19 impacts on mobility at a fine geographic resolution.

6.2 Information Value and Report Quality

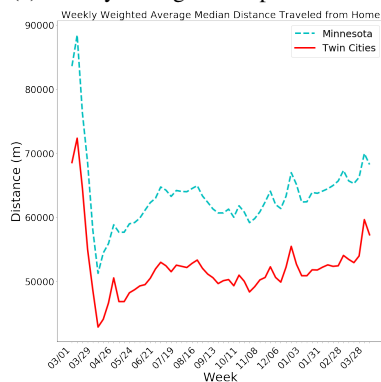
6.2.1 Information Value :

We sought feedback from users in our partner communities regarding the value of our reports to their decision making analysis. The feedback was obtained through multiple meetings and interviews with user groups, namely, Traffic Flow and Public Safety, Public Transportation and Transit, Economic Management, and Public Health.

Traffic Flow and Public Safety: Transportation analysts appreciated the capabilities of the mobile device spatiotemporal data based on geographic extent and sampling frequency compared with Loop Detectors and Travel Surveys respectively. The use of spatial big data in transporta-



(a) Weekly average time spent at home



(b) Weekly average distance travel from home

Figure 14. (a) Weekly average time spent at home, (b) Weekly average distance travel from home (Best in Color).

tion research Shekhar et al. (2012) is promising due to both high frequency and high geographic coverage (Figure 15).

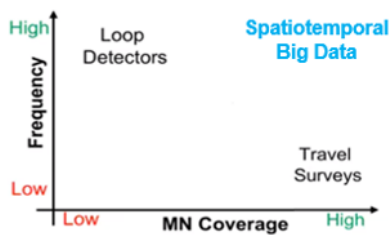


Figure 15. Frequency and geographic coverage of traditional datasets and spatial big data (Best in Color).

Travel surveys are limited in number, and while loop detectors cover urban highways, they cannot capture public squares, avenues, parkways, etc. Further, analysts also acknowledged the potential of Safegraph to handle large mobile device data at certain geographic levels (e.g. census tracts, census block groups) irrespective of data quality issues. In addition, we also tried to address specific queries of interest. For instance, the transportation analysts were interested in separating vehicle miles traveled from commuting from vehicle miles traveled for delivery goods. However, we were unable to provide the results since it

was not possible to slice vehicle miles traveled by mode of transportation (e.g. cars vs delivery trucks).

Public Transportation and Transit: Metro transit analysts acknowledged they found the weekly reports helpful for tracking the types of trips that might be recurring, and perhaps eventually returning to transit. They were interested in certain Points of Interest such as "Fourpost" (as shown in Figure 16) since the number of visits there increased from 06/30/2020 but eventually went down to zero after 11/16/2020. According to a news report Norfleet (2020), "Fourpost" was closed early 2020 before the stay-at-home order and offered physical space to other retailers, explaining some activity related to long or short duration visits after 06/30/2020. Hence, data quality issues still persist since the Safegraph dataset may not have been updating or removing old POIs frequently enough.

Economic Management: Economic management colleagues acknowledged the value of our human mobility patterns based on long visit duration with COVID-19 policy Interventions Calendar. They encouraged us to share weekly reports based on mobility traffic for certain places, businesses and business categories. We also conducted regular meetings to discuss questions posed by policymakers and analysts for certain business categories. However, the most significant patterns of interest to policymakers remained Bars, Limited and Full-time Restaurants.

Public Health: When discussing Safegraph data, public health researchers posed initial questions regarding data quality such as sampling bias and data transparency. However, such issues are not addressed in Safegraph data which presents a barrier to perform calibration in disease transmission models (SEIR). Nevertheless, they were also interested in identifying long duration visits for estimating number of contacts with hourly granularity.

6.2.2 Report Quality:

We published periodic reports to Economic Management policymakers and analysts, and they found the report quality acceptable. In addition, they requested supplement at materials such as summary data in tabular format to supplement the trend via visualizations.

Besides information value and report quality, we also provide a brief justification on how the proposed community-engaged decision support system addresses problems beyond the limited questions of specific community stakeholders presented in Figure 1. We further analyzed **compliance behavior towards stay at home (SAH) order** generated by the proposed system. The state government issued Stay at Home orders from 27th March 2020 - 18th May 2020, which included the closing of bars and restaurants of Health (2020). Figure 17a shows a significant drop in average distance traveled from home until April 2020, after which we see an increase in mobility activity. This accords with the trend we see in Figure 17b, which shows an increase in Average Time at Home until 10th April, followed by a decrease in later weeks. In addition, the stay-at-home order was followed by an extended

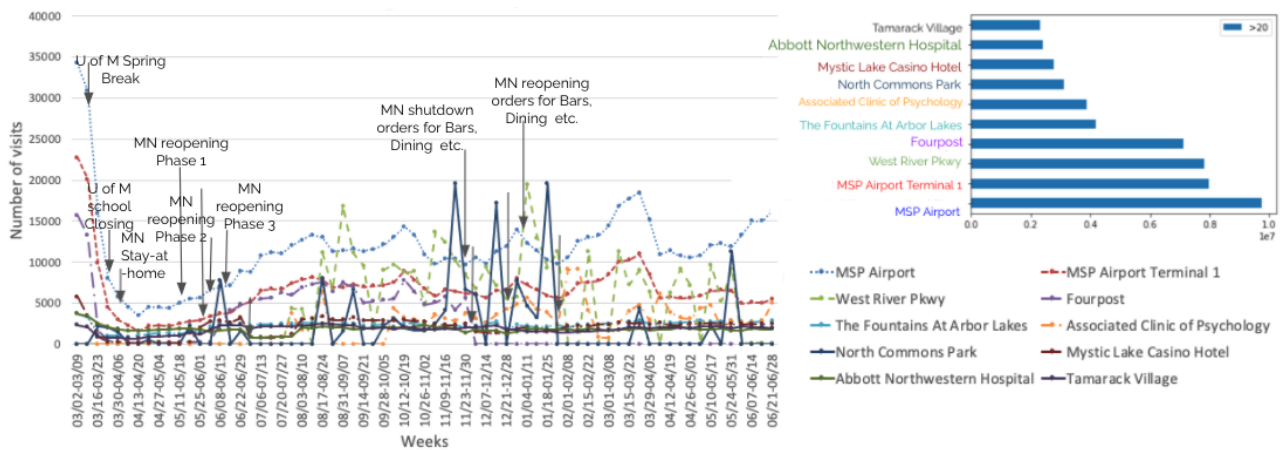


Figure 16. Most frequent long-duration visited Points of Interest in Minnesota (Since March 2, 2020 - June 28, 2021) (Best in Color).

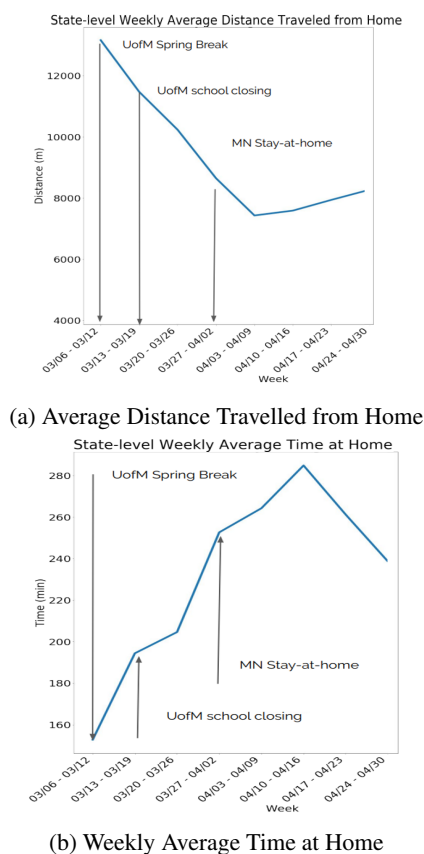


Figure 17. Mobility assessment in MN State between 03/02/2020 - 04/30/2020

stay-at-home order of Health (2020). This order called for the gradual reopening of businesses after June 10, when full-service restaurants could operate at a limited indoor capacity. The result was an increase in foot traffic, as shown in Figure 7. To monitor adherence to the extended stay-at-home on order hourly or daily basis, information on the number of long (i.e., greater than 20 mins) duration visits to certain businesses (e.g., full-service restaurants) is necessary (as shown in Figure 8-10).

6.3 Data and Software Availability

Safegraph was purchased and can be remotely accessed. The compiled dataset cannot be redistributed due to licensing restrictions. Figures 7-14 and Figures 16-17 can be reproduced via code in Github repository Sharma (2021a). More details can also be found in Sharma (2021b).

7 Related Work

A web-based mapping platform Dong et al. (2020) by Johns Hopkins University was the earliest work in response to tracking the COVID-19 outbreak in real-time. Following the same line of work, other surveillance applications and web-based platforms Gao et al. (2020); Samet et al. (2020); Desjardins et al. (2020); Gao et al. (2020) were developed to visualize and provide real-time quantification data, such as mobility changes and emerging hotspots. For instance, CoronaViz Samet et al. (2020), visualizes COVID-19 spread across the globe using animation, which allows users to change the spatial region and time span interval. Desjardins et al. (2020) uses space-time scan statistics Kulldorff et al. (2005) to detect emerging hotspots and calculates the relative risk score at the county level on a US map. While all of these web-based map visualizations are useful for understanding the COVID-19 spread at a low spatial scale (i.e., county), the main component of engaging end-users (e.g., policymakers) and closing the loop by delivering customized reports is missing. Customized reports can improve decision-making and allocation of resources.

Many studies Park et al. (2020); Aleta et al. (2020); Kretzschmar et al. (2020); Mokbel et al. (2020) have demonstrated the promise of contact tracing for safe reopening of the economy and business following state and federal policy interventions in response to the COVID-19 crisis. In Huang et al. (2020); Kang et al. (2020) discusses the change in human mobility based on transportation-related behaviors (e.g., modes of transportation, etc.) and their spatial interactions patterns at different scales (e.g.,

census tracts, county, and state level). In Coston et al. (2021), the authors link smart-phone based data (SafeGraph) to high-fidelity ground truth administrative data and further audits mobility data for bias in the absence of demographic information and ground truth labels which further illuminates demographic disparities and how such disparities distort policy decisions.

8 Conclusion and Future Work

In this work, we proposed a community-engaged COVID-19 decision support system which addresses queries (e.g., long-duration visits) posed by end-users. The architecture provides custom reports related to user specific queries related to actual decision making questions which are not addressed by state of the art dashboards. To address such queries, we designed an Entity-Relationship diagram on weekly pattern data from the SafeGraph dataset which further enhance our understanding of the Safegraph data. Finally, we validate the proposed decision support system on a real-world mobility data with a case study and briefly summarize end-user feedback along with a discussion of data quality.

Future Work: In future work, we are interested in investigating part-time and full-time employment trends and modeling the mobility impacts with respect to those attributes. In addition, we will investigate data quality by further characterizing the relationship between disease spread and mobile-phone data. We will also address the effect of 3rd normal schemas on data storage cost, processing time, choice of indexing and query processing strategies. We will also integrate a real-time data stream in the proposed system. In addition, we will compare Safegraph data trends with other data sources (e.g., Foursquare) to assess data quality issues. Finally, we will explore a new spatial data mining technique that can automate routine tasks and further consolidate it with an interest measure that can distinguish hangout places and emerging hotspots.

Acknowledgement

This material is based upon work supported by the National Science Foundation under Grants No. 2040459 and 1737633. We would also like to thank our collaborators from the data provider (SafeGraph) to MIT media lab (Dr. Esteban Moro Egido and Dr. Alex Pentland), Metropolitan council (Dr. Ashley Asmus), State of Minnesota (Mn-DOT, MMB, and MN DHS), and the University of Minnesota (Center for Transportation Studies (Laurie McGinnis), School of Public Health (Dr. Eva Enns, Dr. Shalini Kulasingam, and Dr. Kelly Searle), U-Spatial (Len Kne, Adam Null), Metro Transit (Dr. Eric Lind) and Dept. of Industrial and Systems Eng. (Dr. Ankur Mani) for their insightful comments and constructive guidelines. We also want to thank Kim Koffolt and the spatial computing research group for their helpful comments and refinements.

References

- Alberto Aleta, David Martin-Corral, Ana Pastore y Piontti, Marco Ajelli, Maria Litvinova, Matteo Chinazzi, Natalie E Dean, M Elizabeth Halloran, Ira M Longini Jr, Stefano Merler, et al. 2020. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. *Nature Human Behaviour*, Nature Publishing Group 4, 9 (2020), 964–971.
- Ottar N Bjørnstad, Katriona Shea, Martin Krzywinski, and Naomi S Altman. 2020. The SEIRS model for infectious disease dynamics. *Nature methods*, Nature Publishing Group 17, 6 (2020), 557–558.
- “U.S. Census Bureau”. 2020. United State Census Bureau Quick Facts, Minnesota Basic Information. *US Census Bureau* (2020). <https://www.census.gov/geographies/reference-files/2010/geo/state-local-geo-guides-2010/minnesota.html>
- US Census. May 21, 2021. North American Industry Classification System. *US Census* (May 21, 2021). <https://www.census.gov/naics/>
- Amanda Coston, Neel Guha, Derek Ouyang, Lisa Lu, Alexandra Chouldechova, and Daniel E Ho. 2021. Leveraging Administrative Data for Bias Audits: Assessing Disparate Coverage with Mobility Data for COVID-19 Policy. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. 173–184.
- Datasets. 2020. Google COVID-19 public datasets. <https://cloud.google.com/blog/products/data-analytics/publicly-available-covid-19-data-for-analytics>
- COVID-19 Outbreak – Metro Area Travel Declines. Sep 16, 2020. *Minnesota Metropolitan Council* (Sep 16, 2020). <https://metrotransitm.n.shinyapps.io/covid-traffic-trends/>
- MR Desjardins, Alexander Hohl, and EM Delmelle. 2020. Rapid surveillance of COVID-19 in the United States using a prospective space-time scan statistic: Detecting and evaluating emerging clusters. *Applied Geography*, Elsevier 118, 102202.
- Ensheng Dong, Hongru Du, and Lauren Gardner. 2020. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet infectious diseases*, Elsevier 20, 5 (2020), 533–534.
- John Croman Emily Haavik. June 5, 2020. Walz Announces Phase 3 for Reopening Minnesota. available at: <https://tinyurl.com/yckm2krj>.
- Song Gao, Jinmeng Rao, Yuhao Kang, Yunlei Liang, and Jake Kruse. 2020. Mapping county-level mobility pattern changes in the United States in response to COVID-19. *SIGSPATIAL Special*, ACM New York, NY, USA 12, 1 (2020), 16–26.
- Trevor J Hastie and Robert J Tibshirani. 1990. Generalized additive models. *CRC press* 43 (1990).
- Jizhou Huang, Haifeng Wang, Miao Fan, An Zhuo, Yibo Sun, and Ying Li. 2020. Understanding the impact of the COVID-19 pandemic on transportation-related behaviors with human mobility data. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 3443–3450.
- Carla K. Johnson. 2 July 2020. Closing Bars to Stop Coronavirus Spread Is Backed by Science. *The As-*

- sociated Press (2 July 2020). apnews.com/article/a288340b3bd3fbc62e564b3d0adfaa2e
- Yuhao Kang, Song Gao, Yunlei Liang, Mingxiao Li, Jinqiang Rao, and Jake Kruse. 2020. Multiscale dynamic human mobility flow dataset in the US during the COVID-19 epidemic. *Scientific data* 7, 1 (2020), 1–13.
- Mirjam E Kretzschmar, Ganna Rozhnova, Martin CJ Bootsma, Michiel van Boven, Janneke HHM van de Wiggert, and Marc JM Bonten. 2020. Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study. *The Lancet Public Health, Elsevier* 5, 8 (2020), e452–e459.
- Martin Kulldorff, Richard Heffernan, Jessica Hartman, Renato Assunção, and Farzad Mostashari. 2005. A space–time permutation scan statistic for disease outbreak detection. *Plos med, Public Library of Science* 2, 3 (2005), e59.
- Yan Li. August 19, 2020. Spatial Big Data Analysis to Understand COVID-19 Impacts on MN. *Virtual Snowbird Conference, CRA Virtual Conference 2020* (August 19, 2020). <https://cra.org/highlights-from-cra-virtual-conference-2020/>
- Maia Martcheva. 2015. An introduction to mathematical epidemiology. *Springer* 61 (2015).
- Jennifer Mayerle. July 29, 2020. Mayor Jacob Frey Announces Bar Areas Must Close Again Due To COVID-19 Spread. *CBS Minnesota* (July 29, 2020). <https://tinyurl.com/2p934zb5>
- Mohamed Mokbel, Sofiane Abbar, and Rade Stanojevic. 2020. Contact tracing: beyond the apps. *SIGSPATIAL Special, ACM New York, NY, USA* 12, 2 (2020), 15–24.
- Joe Nelson. July 30, 2020. List of 19 bars in Minnesota linked to COVID-19 outbreaks. *Bring Me the MN News* (July 30, 2020). <https://bringmethenews.com/minnesota-news/list-of-19-bars-in-minnesota-linked-to-covid-19-outbreaks>
- Nicole Norfleet. September 29, 2020. Mall of America gives free space to small retailers impacted by pandemic. *Star Tribune* (September 29, 2020). <https://tinyurl.com/dknhsvnu>
- Minnesota Dept. of Health. 19 Feb, 2021. COVID-19 Outbreak Summary for Business Categories. *Minnesota Dept. of Health* (19 Feb, 2021). www.health.state.mn.us/diseases/coronavirus/stats/index.html#warchive1
- Minnesota Department of Health. 2020. Stay at Home Order. <https://mn.gov/covid19/for-minnesotans/stay-home-order/index.jsp>
- Young Joon Park, Young June Choe, Ok Park, Shin Young Park, Young-Man Kim, Jieun Kim, Sanghui Kweon, Yeonhee Woo, Jin Gwack, Seong Sun Kim, et al. 2020. Contact tracing during coronavirus disease outbreak, South Korea, 2020. *Emerging infectious diseases* 26, 10 (2020), 2465–2468.
- Minnesota’s Stay Safe Plan. 2020. <https://mn.gov/covid19/for-minnesotans/stay-safe-mn/stay-safe-plan.jsp>
- ‘Reports’. 2020. COVID-19 Community Mobility Reports. <https://cloud.google.com/blog/products/data-analytics/publicly-available-covid-19-data-for-analytics>
- ‘SafeGraph’. 2020. The Source of Truth for POI Data & Business Listings, Safegraph Consortium. <https://www.safegraph.com/blog/stopping-covid-19-with-new-social-distancing-dataset>
- Hanan Samet, Yunheng Han, John Kastner, and Hong Wei. 2020. Using Animation to Visualize Spatio-Temporal Varying COVID-19 Data. In *Proceedings of the 1st ACM SIGSPATIAL International Workshop on Modeling and Understanding the Spread of COVID-19, ACM*. 53–62.
- Arun Sharma. 2021a. <https://anonymous.4open.science/r/COVID-19-4014/README.md>
- Arun Sharma. 2021b. http://www.spatial.cs.umn.edu/Project/covid_19.html
- Arun Sharma, Yan Li, and Shashi Shekhar. May 21, 2021. SafeGraph Data Analysis to Assess COVID-19 Effects on MN Mobility. *Impacts of the COVID-19 Pandemic on Minnesota’s Traffic and Transit Networks, Center for Transportation Studies, University of Minnesota* (May 21, 2021). http://www.spatial.cs.umn.edu/Project/covid_19/Appendix_Weekly_Mobility_Patterns_05212021.pdf
- Shashi Shekhar. June 23, 2020. Spatial Big Data Analysis to Understand COVID-19 Effects on MN. *Impacts of the COVID-19 Pandemic on Minnesota’s Traffic and Transit Networks, Center for Transportation Studies, University of Minnesota* (June 23, 2020). http://www.spatial.cs.umn.edu/Project/covid_19.html
- Shashi Shekhar, Viswanath Gunturi, Michael R Evans, and KwangSoo Yang. 2012. Spatial big-data challenges intersecting mobility and cloud computing. In *Proceedings of the Eleventh ACM International Workshop on Data Engineering for Wireless and Mobile Access*. 1–6.
- “StreetLight”. 2020. STREETLIGHT DATA, Big Data for Mobility. <https://www.streetlightdata.com/>
- Melissa Turtinen. 23 Feb. 2021. 39 Restaurants, Bars Linked to COVID-19 Outbreaks so Far This Year. *MN Dept. of Health Report, MN Dept. of Health* (23 Feb. 2021). bringmethenews.com/minnesota-news/39-restaurants-bars-linked-to-covid-19-outbreaks-this-year.
- Uspatial. 2020. <https://research.umn.edu/units/uspacial>
- Paul Walsh. 20 July, 2020. 14 Minnesota bars and restaurants warned for breaking coronavirus rules. *Star Tribune* (20 July, 2020). <https://www.startribune.com/state-warns-14-bars-and-restaurants-over-covid-violations/571835461/>