

# 2023 Holiday Service Transit Traveler Information Research

Final Version completed on September 26, 2024 Prepared by Evan Siroky, Research Data Manager, Cal-ITP Contact info: evan.siroky@dot.ca.gov (916) 907-2133

# **Executive Summary**

The California Integrated Mobility Program within Caltrans conducted a thorough study on the representation of holiday service in public transit rider information. This investigation was prompted by numerous incidents of discrepancies in holiday service data found in General Transit Feed Specification files and reports from the public regarding incorrect information.

Caltrans analyzed the websites of 187 transit agencies across California websites and collected GTFS Schedule data. The study classified holiday service levels of "Regular Service", "Reduced Service" and "No Service." The analyses covered a timeline of potential holidays from Veteran's Day 2023 to Martin Luther King Junior Day in 2024.

Over 90% of agencies provide holiday service information in online materials. Significant variability existed in the service levels offered for each holiday and the presentation of information on websites. A super majority (74%) of agencies had at least one inconsistency between the holiday service levels stated on their websites and those reflected in their GTFS schedule data. While most transit agencies had few inconsistencies, some completely omitted or included outdated holiday service data in their GTFS data. The most common issue was the GTFS data having more service than the level stated on transit agency websites.

The potential impact of these discrepancies was considerable. On Thanksgiving Day 2023 alone, as many as 480,000 daily riders—representing 25% of statewide ridership – could have received conflicting information about holiday service.

As a result of our findings, we propose several strategies for enhancing the accuracy and delivery of rider information related to holiday services. Our recommendations focus on improvements that can be implemented at the Transit Agency level, through their vendors, and on an inter-agency or statewide basis.

# Motivation for Research

Reliable information on public transit holiday service is important for riders to obtain to have accurate information about public transit service to complete journeys on public transit. Travel patterns often diverge from the norm during holidays especially around Thanksgiving and Christmas when numerous people may travel away from familiar locations. Since holidays can have different service on days when riders may be expecting regular service, the publishing of incorrect information can result in riders planning impossible trips. This could result in delays for riders and distrust in public transit. This distrust may result on aggregate in lower public transit ridership.

Transit riders have options on how to access transit schedule data, including holiday schedules. Transit agencies typically alert the public about potential service outages or route changes in advance to manage rider expectations and travel plans. Service alerts can be shared on the specific transit agency's website, through transit app subscriber notifications, or on a specific transit station's media billboard.

Public transit agencies typically publish holiday information either on a website or brochure and within General Transit Feed Specification (GTFS) data. However, it is possible that some transit agencies may neglect to publish this information or that the information on the website may conflict with the information in the GTFS data. Additionally, some transit agencies may have published holiday information in the past, but not updated their data either on their website or in their GTFS data to reflect the current year's operations and dates.

According to the Federal Transit Administration (FTA) <u>National Transit Database 2023 Policy</u> <u>Manual</u>, transit agencies report holiday service under the day that most closely reflects the service. For example, agencies operating a Sunday schedule on Christmas Day, must indicate that this is an additional day of Sunday service (regardless of the day on which the holiday falls). Services that operate for a portion of any given day count as a day operated whereas service cancellations for an entire day despite the mode of travel does not count as a day operated. As a result, each transit agency defines their own holiday service classifications (service levels) based on FTA guidelines.

During some ad-hoc analyses during 2022, it was observed that some transit agency's websites did not have accurate information about holiday service on their website in different parts of California. Additionally, several complaints from members of the public to Cal-ITP regarding the quality of data about holiday service have been received.

Given this background, this research project idea was conceived to research how well transit agencies are providing information to their riders about holiday service. It was decided to investigate how well transit agencies are publishing rider information both on their websites and in GTFS Schedule data. It was expected that there should be an alignment between a transit agency's holiday website and the service provided in a transit agency's GTFS Schedule data. If there were inconsistencies, this was a likely indication that either the information on a transit agency's holiday website or in their GTFS Schedule data was inaccurate.

Given a high clustering of holidays occurring around the end of the year, this research examined the time between Veteran's Day (Observed) 2023 until Martin Luther King Jr Day 2024. The issue of data accuracy of holiday schedules extends beyond these holidays, but an assumption was made that these holidays would provide an effective sample for the research topic in general.

The holidays considered during these dates were as follows:

Holiday Name	Day of the Week	Date
Veteran's Day (Observed)	Friday	November 10, 2023
Veteran's Day	Saturday	November 11, 2023
Thanksgiving Day	Thursday	November 23, 2023
Day after Thanksgiving	Friday	November 24, 2023
Christmas Eve	Sunday	December 24, 2023
Christmas Day	Monday	December 25, 2023
New Year's Eve	Sunday	December 31, 2023
New Year's Day	Monday	January 1, 2024
Martin Luther King Jr Day	Monday	January 15, 2024

Table 1: Holidays Considered in Study Period

An analysis of the impact on riders was performed to determine the extent to which various problems of miscommunicating schedule information may have occurred for transit riders across California. And based on that analysis, we outline certain possible strategies for improving the delivery of rider information and institutionalizing ongoing checks for the correct communication of holiday schedules.

# Data Collection Methodology

This research involved the collection of three main types of data: 1) Information obtained through manual searches of transit agency websites regarding holiday service, 2) GTFS Schedule data downloaded from each transit agency, and 3) Ridership data sourced from the National Transit Database (NTD).

The study included all transit agencies in California that operate fixed-route services accessible to the public without reservations or specific rider eligibility requirements. On-demand and ridesharing services were excluded as they typically do not offer fixed-route with published schedules. At the time of this research, Caltrans identified 187 transit agencies that met these criteria.

Caltrans maintains a comprehensive database of all transit agencies operating transit services in California. This database includes details about the agencies, the services they provide, and the GTFS feeds they generate. For this project, additional data focused specifically on holiday service information and patterns were collected.

To ensure access to the most current GTFS data, Caltrans follows standard operating procedures. It downloads the latest GTFS Schedule data from nearly all transit agencies in California nightly. The data is stored in a data warehouse, organized into various database tables that can be queried with different analysis tools. Additionally, NTD data has been integrated into the warehouse, with efforts made to match it to the corresponding transit agencies whenever possible. However, some transit agencies are not represented in the NTD dataset.

# Collection of Data from Transit Agency Websites

Public transit agencies manage a public-facing website that contains operational and transit service information, potentially including information about holiday service. Caltrans staff visited each organization's website and located information about holiday service. There did not appear to be a standard pattern as to where organizations placed holiday schedules on their website.

Most organizations use either the Schedules or Routes header to share specific information, but Caltrans staff found holiday schedules stored in a variety of locations within an organization's website structure. We presume the holiday service information location (website or document) is based on organizational guidelines which may not align with making information easier for the public to access. For example, one organization may place the holiday service schedule on the home page whereas another organization the same information could be recorded on a subpage. Finding the holiday schedule was a matter of trial and error irrespective of an organization's size. In addition, the holiday schedule information can vary in format and detail depending on the organization.

Organizations choose to share their holiday schedule within a document or on their website (submenu or home page). If the holiday schedule is shared on a website, the public will typically experience the same formatting based on organization's website structure. If the holiday schedule is in a brochure, the public may be provided with limited or very detailed information depending on the organization. The holiday schedule document could also be a PDF or equivalent document that may require the public to download the file prior to accessing the schedule information.

## Timing of Data Collection from Websites

The first phase of the research began in October 2023 and involved gathering information about the transit agencies. Once the holiday schedule information was located, Caltrans staff stored data in an internal database. Caltrans staff updated their database with the holiday

schedule data collected, including the holiday service website, holiday service website status, and holiday service classifications. Also, Caltrans staff documented their findings and created supporting visual aids. In the first phase, we identified 16 transit agencies with Missing (12) or Old (4) holiday website statuses.

The second phase began in November 2023 to review transit agencies with the "Missing" or "Old" holiday website status to check if transit agencies have updated their holiday schedules. Using the holiday website address previously recorded in the Transit Database, holiday website status was updated to Current if the transit agency updated materials on their website to include the latest holiday schedule. Despite the holiday season being underway, only 1 of the transit agencies updated their outdated information.

## Format of Data Collected

When recording data of a transit agency's holiday website, we assigned values of "Current", "Missing", "Old", and "Off-Season". The meaning of each value is noted in the table below:

Holiday Website Status Value	Meaning of Value
Current	Any holiday website or schedule document dated after 2020
Missing	The holiday schedule could not be located either on the transit agency's website or in a document
Old	Any holiday website or schedule document dated before 2020
Off-season	An organization operates a seasonal service that is not operating during the holidays currently being tracked

#### Table 2: Holiday Website Status Values

We also collected the hyperlink to the transit agency's website or holiday schedule document location. In certain cases, we added notes to provide context about the holiday website, such as if the holiday schedule is visible to the public. This data is not published in this report but is available for use by Caltrans staff.

A set of categories were developed to record data about the level of service provided on each of the study holidays. Information was gathered on the holiday schedule details provided via the holiday website or a document/brochure that was downloaded from the transit agency's website. Any holiday not specifically identified on the transit agency's website or an online brochure was presumed to be operating regular service. The meaning of each value is noted in the table below:

#### Table 3: Holiday Schedule Values

Holiday Schedule Value	Meaning of Value
No service	The entire transit agency would not operate on the associated holiday
Regular service	The transit agency operates a normal weekday/weekend schedule, relative to that day of the week
Reduced service	The transit agency operates noted it operate on a "limited" or "reduced" schedule. Anything that was less than "Regular service", but more than "No service" was classified as "Reduced service".

During the study time period some holidays that rotate days each year occurred on weekends. This was the case for Veteran's Day, Christmas Eve and New Year's Eve. Some transit agencies noted that these were holidays, but some did not always note explicit dates that holidays occurred on. This led to ambiguity for holidays that change days of the week each year such as Veteran's Day, Christmas and New Year's Day. It was difficult to determine whether an agency would observe holidays on a different date if they fell on a weekend. Several holiday websites and online brochures were designed in a way to be agnostic to each year's nuances of what days of the week a holiday occurred on. In these instances, we assumed that unless explicitly noted, there were not any observed holidays.

# Analysis of Data from Transit Agency's Websites

Holiday schedules that are shared on the transit agency's website are categorized as being Current, Old, Missing, or Off-Season. As the chart below indicates, most transit agency's holiday schedules are dated after 2020 (document or website) so the website status is current. We also noticed that holiday schedules may not be published on a transit agency's website. In those instances, the holiday schedule is not made available to the public to view for unknown reasons. Lastly, some agencies provide seasonal services that operate for a limited time during the calendar year. For example, some transit agencies that operate services from Labor Day through Memorial Day weekend. As a result, these services have ended prior to the holidays included in this research project and are classified as off-season.



Figure 1: Holiday Schedule Status on Websites

We further analyzed those agencies with missing holiday service information using NTD data. We used the Vehicles Operated in Maximum Service (VOMS) data which represents the maximum number of revenue vehicles each transit agency operates at a single time. Per FTA guidelines, most transit agencies count VOMS annually during peak times on the busiest day they provide service because they have more than enough vehicles to operate a scheduled service. It was assumed that the amount of VOMS was a good proxy for determining the relative population in the surrounding area where the transit agency operates. For example, transit agencies in rural location typically do not operate many services since the population size is smaller (Manteca) than larger urban areas (San Francisco) with increased population size. The chart below shows the number of transit agencies in each VOMS range that had missing or old holiday information on their website.



Figure 2: Organizations with Missing or Old Holiday Schedules

The most common method that transit agencies use to publish the holiday schedule is to create a separate page under the main website header, such as "Routes or Schedules". The other method is for a transit agency to share the holiday schedule in an online document or brochure.



#### Figure 3: Holiday Transit Schedules by Location on Website

When the holiday schedule is on the website, most transit agencies list the holidays along with the affected routes. For example, Culver CityBus holiday schedule is as shown on the following page:

Home / Bus Schedules Bus Line Information & Schedules Line 1 - Washington Bhd Line 2 - Inglewood Bhd Line 3 - Crosstown Line 4 - Jefferson Bhd Line 5 - Braddock Bhd Line 6 - Seardheda Bhd		Culver CityBus Holiday Schedule Section on Website Page
Line 7 - Culver RHed Holiday Schedules Culver CityBus provides reduced bus service on Saturday, Sunday and Memorial Day, Independence Day, Labor Day, Thanksgiving Day and Ch are closed on holidays.	nolidays on Lines 1, 3 and 6. Weekend service is provided on New Year's Day, istmas Day. Lines 2, 5 and 7 DO NOT run on Saturday. Sunday or holidays. Offices	
Rider Tips When getting onto a Culver CityBus, please have the exact fare ready. Culver CityBus fareboxes accept dollar bills and coins; however, they D <u>Fare information</u> is posted on the farebox and on Culver CityBus scher High school students may be required to show their student ID as prov Senior citizens 62 and over, disabled, or Medicare recipients must sh	D NOT make change, and bus operators DO NOT carry change. ules. f of age. w their ID in order to pay a reduced fare.	
Real Time Route Map Welcome to Next CCBus ★ Favoriter Welcome to the Next CCBus © Stop Time Real-Time Map & Trip Planner	Log In Next CCBus is the real time passenger arrival information that tracks the location of your Culver CityBus and provides real time bus arrival predictions and rider alerts. CCB shows the information through the website, mobile app, call center, email, text messages, and solar LED signs at selected bus stop locations. Estimated stop times To ensure you don't miss the bus, please arrive at your stop at	
Subscribe  Track BulTaxt Log In   🛎   Terms   Contact us   Homepage	least to you don't miss the dus, prease all new all you stop at least to minutes before the estimated stop time. Please keep in mind that delays due to detours, traffic incidents, construction, weather or equipment probleme may affect the © take 0tytes 2023	

*Figure 4: Screenshot of Holiday Information on Culver City Bus Website* 

The other method is for a transit agency to share the holiday schedule in an online document or brochure. For example, San Diego Metropolitan Transit System Trolley holiday schedule is published within an online brochure that is accessible via the transit agency's website. The Trolley holiday service schedule is shown in the lower right corner on the last page as shown on the following page:



Figure 5: Screenshot of San Diego MTS Trolley Holiday Schedule in Brochure

Given what each transit agency described on their website, we did an analysis to see what service level each transit agency provided on each holiday. It was seen that there was a wide variety of service levels that each agency provided. A chart and table showing the service levels is shown on the following page.



Figure 6: Holiday Service Schedules Observed

Holiday	% with Regular Service	% with Reduced Service	% with No Service
Veteran's Day (Observed)	88.8%	6.7%	3.9%
Veteran's Day	62.9%	21.3%	15.2%
Thanksgiving Day	4.5%	24.2%	70.8%
Day after Thanksgiving Day	53.9%	30.9%	14.6%
Christmas Eve	66.3%	21.3%	11.8%
Christmas Day	5.1%	23.0%	71.3%
New Year's Eve	66.9%	19.7%	12.9%
New Year's Day	6.7%	26.4%	66.3%
MLK Day	52.8%	30.3%	16.3%

We also completed an analysis to determine what the number of unique holiday patterns were. To do this we created an algorithm that created a unique holiday pattern value. This value comprised of the combination of the service classification values for all 9 holidays in the study period. For example, if two agencies had regular service on each holiday except for Christmas Day, they would be given the same holiday pattern value. We counted the number of transit agencies that had the same holiday pattern values.

We found 69 different holiday pattern values. The most common variation accounted for almost 20% of all variations, but the commonality of other variations quickly dropped off. About 50% of transit agencies share their holiday pattern variation with 3 or less other transit agencies across the state. Additionally, 20% of transit agencies have their own unique variation of holiday service. The following chart shows these variations.



Figure 7: Holiday Service Types Across Agencies

The most common holiday pattern is shown in the table on the following page:

#### Table 5: Most Common Holiday Service Strategy

SERVICE STRATEGY	DAYS OBSERVED
Regular Service	Veteran's Day (Obs)
	Veteran's Day
	Day after Thanksgiving
	Christmas Eve
	New Year's Eve
	Martin Luther King Day
No Service	Thanksgiving Day
	Christmas Day
	New Year's Day

# Analysis of Consistency between Transit Agency's Websites and GTFS Data

The outcome of this section is to determine differences between the holiday information as represented in the GTFS data and what is posted on websites. We used our Caltrans data warehouse to compare the information on an agency's website, their normal number of trips on a reference day, and the number of trips they scheduled for a holiday in their GTFS.

## Data Cleaning and Preparation

To come up with an appropriate way to compare holiday service with GTFS data we did some data cleaning and preparation. Some agencies were not analyzed. These included agencies that had either off-season seasonal service or those where their holiday information could not be found on their websites including any online brochures. For the data about the holiday service from the transit agency websites, we frequently had to account for the holiday website data being in a year-agnostic format. We made some assumptions based on the stated service levels that a transit agency would apply for a holiday depending on which day of the week the holiday fell on. Additionally, we had to properly classify GTFS data to decide if there was a variance compared to a typical day of the week.

To translate what was present on a website, we had to make various assumptions of what this meant for a certain holiday if it ended up being on a Saturday or Sunday. We assumed that transit agencies would not have "observed" holidays unless explicitly noted on their website. We used the following translation table to derive what would be the expected holiday service level for a holiday. If the agency marked the date as "limited" we would assume that would mean reduced service beyond the typical service changes between weekdays and Saturday or Sunday service.

#### Table 6: Holiday Service Level Derived for Analysis

Holiday Service Level (as stated on website)	Typical Service Level for the agency on the day of week that the holiday occurred on.	Derived Holiday Service Level to use for Comparing to GTFS data
Regular	Regular	Regular
Regular	Reduced	Reduced
Regular	No Service	No Service
Reduced	Regular	Reduced
Reduced	Reduced	Reduced
Reduced	No Service	No Service
No Service	Regular	No Service
No Service	Reduced	No Service
No Service	No Service	No Service

Some examples of how this applied in certain situations are as follows:

- Some transit agencies would not have a holiday noted for the day after Thanksgiving.
   Since this is on a Friday which tended to have a weekday service, "Regular" weekday service was expected for this date.
- Some transit agencies would have a holiday for Veteran's Day which fell on a Saturday in 2023. Some of these transit agencies don't operate at all on weekends, so the expected service for those agencies on Veteran's Day was "No Service".
- Some of the agencies have Saturday service with less trips operating relative to weekday service. With the example of Veteran's Day which fell on a Saturday, if the transit agency said they would have reduced service, we expected that there would be even less than the typically observed Saturday service.
- Some of the agencies that do not have a holiday for Veteran's Day would therefore be expected to have roughly the same amount of service ("Regular") compared to the service observed on a typical Saturday.

In preparing to analyze the GTFS data, we picked some "reference" days. These days were when no holidays were assumed to be active and when a highly representative amount of total trips were present for the transit agency. In total we used 3 reference day types to properly account for usual variations in weekly schedules. The days used were as follows:

Reference Day Type	Date Used	
Weekday	Friday, December 15, 2023	
Saturday	Saturday, December 16, 2023	
Sunday	Sunday, December 17, 2023	

#### Table 7: Reference Days used in GTFS Analysis

We then compared the total number of trips observed on the reference day type to the day of each holiday to come up with a ratio of the number of trips.

$$Ratio = \frac{Number of trips on Holiday date}{Number of trips on Reference Day Type}$$

Using this ratio we then categorized the service observed in the GTFS data. We picked a somewhat arbitrary cutoff of 85% of reference day trips to change in classifying service from "Regular" to "Reduced" as noted in the following table:

Ratio of number of trips	Holiday Service Type Classification	
0.85+	Regular	
0.01 - 0.85	Reduced	
0-0.01	No Service	

Table 8: GTFS Classification Assigned based on Trip Ratio

For example, San Diego Metropolitan Transit System (SDMTS) has 7,541 trips scheduled in their GTFS data on their reference weekday. In their GTFS for Christmas Day, they listed 3,234 total trips (which also happens to be near their reference Sunday schedule of 3,422 trips). We ascribed a ratio of 3,234 / 7,541 which resulted in the value of 0.43, which falls within the "Reduced Service" range.

## Data Analysis Results

After classifying the resulting ratio of the number of trips, we were then able to make a comparison between the level of service as stated on the transit agency's website and the amount of service that was present in the GTFS data. In some instances, the transit agency's GTFS data showed more service than what was described on their holiday website. There were also instances when the GTFS data showed less service than what was described on a transit agency's holiday website.

There are 9 different possibilities of consistency of holiday service classification for a given holiday. The Agency's GTFS Service levels and the service level on their website are all classified into the 3 options of either "No Service", "Reduced Service", and "Regular Service". Accounting for all possible permutations of matching between the GTFS data and website data, there are 9 different combinations.

To continue with the San Diego MTS example, we found that their website noted that their transit agency had "Reduced Service" on Christmas Day. This matched the "Reduced Service" value of their GTFS data holiday service classification that was calculated upon analyzing their GTFS data as described in the previous section.

We did this analysis for each transit agency and for each holiday date and checked for whether each agency produced a matching amount of service. These can be visualized in a Confusion Matrix diagram as shown on the following page for Christmas Day:



#### Figure 8: Data Alignment Confusion Matrix for Christmas Day

The x-axis values are the holiday service classifications found on the transit agency websites. The y-axis shows the holiday service classifications that were derived from the GTFS data. The values in each box indicate the percent of transit agencies that had a certain permutation of the associated service level on their website and the associated service level in their GTFS data.

Using the values from within the confusion matrix, the data was classified into categories of information alignment. The alignment was achieved between the service level classification described for the holiday on the holiday website to the service level classification derived from GTFS data. The simplified patterns can generally be described as "GTFS matches website", "GTFS greater than website" and "GTFS less than website". The meaning of each of these values is described in the table on the following page:

#### Table 9: Classification of Confusion Matrix Cells

Alignment Value	<b>Confusion Matrix Color</b>	Meaning
GTFS matches	Blue	Alignment was achieved between
website		the information on the transit
		agency's holiday website and the
		service levels observed in GTFS data.
GTFS greater than	Green	The holiday service classification
website		derived from GTFS data was more
		than the holiday service
		classification stated on the transit
		agency's website.
GTFS less than	Orange	The holiday service classification
website		derived from GTFS data was less
		than the holiday service
		classification stated on the transit
		agency's website.

For this example of Christmas Day, a super-majority of transit agencies achieved the "GTFS matches website" pattern. However, there were several transit agencies that had either "GTFS greater than website" or "GTFS less than website" patterns. A significantly higher number of agencies with mismatches fell within the "GTFS greater than website" pattern. The most common problem were instances where the GTFS data indicated regular service was occurring on days with no service.

This analysis was repeated for all holidays and transit agencies. The following charts and tables show the results for each holiday. Thanksgiving and Christmas Day had the highest levels of mismatching patterns between the GTFS data and websites.



Figure 9: Comparison of GTFS and Holiday Service Levels

Table 10: Percent of Agencies	by Classification	Result by Holiday
-------------------------------	-------------------	-------------------

	Day of Week	GTFS Matches Website	GTFS Greater than Website	GTFS Less than Website
Veteran's Day	Friday	85 %	4	11
(Observed)				
Veteran's Day	Saturday	73 %	22	4
Thanksgiving Day	Thursday	72 %	27	1
Day after	Friday	78 %	17	6
Thanksgiving h				
Christmas Eve	Sunday	89 %	6	5
Christmas Day	Monday	71 %	27	3
New Year's Eve	Sunday	86 %	8	6
New Year's Day	Monday	70 %	24	6
MLK Day	Monday	79 %	11	10

Comparison of GTFS / Holiday Service Levels

As can be seen in the results, at a statewide level, the most common classification between the holiday website and GTFS data did seem to correctly account for their holiday service in their GTFS Schedule data. The most common mismatch problem was that more service was being shown in the GTFS data than was stated to occur as noted on holiday service websites. This problem seemed to be more pronounced on the holidays occurring on weekdays and among those where more agencies noted on the holiday website that they would have less than regular service on these holidays.

Additionally, we examined the extent of these discrepancies to assess how consistently each agency aligns its holiday schedules across platforms. We then calculated the number of instances where a transit agency's holiday schedule match was classified either "GTFS Greater than Website" or "GTFS Less than Website". Following this, we determined the percentage of agencies that had varying numbers of holidays with mismatching information.



Frequency of Holiday Schedule Mismatches Between Website and GTFS

Figure 10: Frequency of Holiday Schedule Mismatches

Table 11: Number and Percent of Agencies by the Number of Instances of Information
Mismatches Between Website and GTFS Data

Number of mismatches	Number of Agencies	Percent of Agencies
0	41	26.1%
1	37	23.6%
2	24	15.3%

3	23	14.6%
4	12	7.6%
5	15	9.6%
6	3	1.9%
7	1	0.6%
8	0	0.0%
9	1	0.6%

The results show that nearly 74% of transit agencies had at least one holiday where there was a mismatch. Of those that had at least one mismatch, the distribution of the number of mismatches was concentrated among a smaller number of mismatches. This indicates that transit agencies are having at least some level of mismatch between their website and their GTFS data. Very few transit agencies are mismatching on a supermajority or more of the holidays we analyzed. Some of this mismatch could be due to our selection of the percent of service that we use as cutoffs to indicate regular or reduced service. Overall, this indicates that the distribution of problems isn't always isolated to a single or even select group of agencies.

# Analysis of Impact

Discrepancies between the information presented on a transit agencies website and the data found in GTFS can lead to conflicting information for riders. If passengers rely on only one source of information regarding holiday services, they may inadvertently plan trips based on incorrect data. For instance, If the GTFS data indicates more service than what is operating, riders may expect certain trips to be available when they are not. Conversely, if the GTFS data underrepresents available services rider may miss out on opportunities to utilize transit options they were unaware of.

Additionally, users of GTFS Schedule data who do not rely on it for trip planning can encounter issues leading to inaccurate calculations. For example, researchers analyzing transit agency service hours might overestimate the total service hours if the GTFS Schedule data reflects more service than what is being provided.

At an aggregate level, we matched each transit agency with their annual ridership data from the National Transit Database (NTD). It is important to note, that some transit agencies do not report ridership data to the NTD, so our analysis only includes those that do. We assumed that the average annual weekday ridership could serve as a proxy for the potential number of riders affected by inaccuracies in holiday service data. Subsequently, we examined how many riders were impacted by discrepancies between a transit agency's website and the GTFS data.

Regarding website accuracy, we found that transit agencies with outdated information tended to be those with lower ridership. However, several agencies had no holiday information

available indicating a "missing" status value for their website. As a result, riders served by these agencies face uncertainty about whether their transit services will operate on certain holidays.

	Estimated Ridership	Percent of Total	
		Ridership	
Current	1,881,743	98.7	′%
Off-Season	67		0
Old	571	^	~0
Missing	23,468	1.2	!%

Table 12: Website Status Estimated Impact on Average Annual Weekday Ridership by Holiday

When analyzing the consistency between a transit agency's holiday website and their GTFS data, we only included agencies where holiday information on the website was available. We used the results from the previous analyses to classify each agency based on whether their GTFS data matched, exceeded, or fell short of the service levels indicated on their holiday website. Once classified we tallied the ridership for each transit agency under the appropriate category.

During this analysis, we encountered an outlier situation that required special handling. The transit agency with the highest ridership in the state (Los Angeles Metro) was flagged for a mismatch between its holiday website and GTFS data. For New Year's Day, 2024, Los Angeles Metro websites indicated that it would run "Reduced Service" on this day. However, the number of trips observed on that date was 88% of the typical reference day's trips, slightly above our threshold of 85+% for "Regular Service". This mismatch was isolated to Metro rail feed, as Los Angeles Metro has two GTFS schedules one for bus and one for rail services. To account for this, we split up the NTD ridership proportion to the ridership split between rail and bus services, based data obtained from LA Metro's website.



Figure 11: Agency Ridership by Holiday Information Alignment

	Day of Week	GTFS Matches Website	GTFS Greater than Website	GTFS Less than Website
Veteran's Day	Friday	1,575,428	3,226	302,287
(Observed)				
Veteran's Day	Saturday	1,502,673	353,444	24,823
Thanksgiving Day	Thursday	1,401,726	476,905	2,310
Day after	Friday	1,518,371	344,439	18,131
Thanksgiving h				
Christmas Eve	Sunday	1,759,105	39,915	81,921
Christmas Day	Monday	1,734,335	92,032	54,574
New Year's Eve	Sunday	1,601,445	93,002	186,494

Table 13: Average Annual	Weekday Ridershi	ip of Agencies b	y Classification	Result by	Holiday
		1	/ /	/	/

New Year's Day	Monday	1,425,577	212,974	242,391
MLK Day	Monday	1,832,362	25,210	23,369

As can be seen, the impact to the number of riders was of different proportions than when comparing the percent of agencies. While there were still a large majority of ridership that appeared to be presented with consistent information, there were certain holidays that did have a noticeable amount of riders affected by mismatching information. Again, there did appear to be more mismatches on holidays that fell on weekdays. The variability in ridership on each holiday can be explained by an earlier analysis that showed that numerous agencies had at least 1 mismatch. Therefore, the ridership impacted on each holiday is likely happening at different transit agencies on different holidays.

# Possible Sources of Data Problems

As noted in the previous sections, several transit agencies were found that provided potentially incorrect holiday service in their GTFS data. Caltrans conducted an analysis into why these problems were occurring. Caltrans analyzed the general structure of the GTFS data in the GTFS Schedule feeds that had a mismatch to what was stated on the associated holiday service website. Based on these findings, we then reached out to the transit agencies to try to ascertain what business processes might have led to these problems occurring.

## Analysis of GTFS Data Structure

Caltrans analyzed 24 of the GTFS feeds at agencies that seemed to have multiple holidays with incorrect information. When analyzing the structure of the GTFS data among agencies that had a mismatch between their holiday website and GTFS data, we classified the reason for problems occurring into generalizations of how the data appeared to be produced. A table showing the percentage occurrence of these generalization is shown below.

Problem Type	Percent of Problems of This Type
Omitted Data	48%
Outdated Data	36%
Other Problem	16%

Table 14: GTFS Problems Resulting in Mismatches

The most common problem observed was that the necessary data to flag holiday information was simply omitted. This most frequently occurs in the GTFS data when the GTFS data includes the *calendars.txt* file, but either completely lacks the *calendar\_dates.txt* file or includes the *calendar\_dates.txt* file without any records. The second most common problem was that the *calendar\_dates.txt* file was included but with outdated data. This indicates that the GTFS feed was kept up to date in only the *calendars.txt* file that described regular service, but the most

current holiday service information was not added. These two items together accounted for 84% of the problems observed.

There were two problems classified as "other" and they were as follows:

- The website had incorrect information and the GTFS was correct.
- The GTFS appeared to be too aggressive with holiday service. It showed the agency had "No Service" when the holiday website indicated it should have shown "Reduced Service."

# Outreach to Transit Agencies

Caltrans staff reached out directly to several transit agencies to inquire about potential reasons for these problems existing. In these emails we notified them of ongoing problems with their holiday service that extended beyond the analysis period. In the email conversations we also asked the transit agencies what difficulties they encountered with making the GTFS data consistent with what was described on their holiday websites.

We sent out emails to a total of 24 agencies. We received a response from 18 agencies. Nearly all agencies we received a response from said that they would investigate the matter, however we only received a detailed response into what may have caused the data problem from 11 agencies. Given this feedback we used our professional judgement to make a hypothesis upon what the business process problem was that caused the data problem. The resulting business process problems were varied as shown in the chart below:



Figure 12: Business Process Problems Resulting in Information Misalignment

The problems identified were varied in type and were sometimes attributable to problems attributable to a transit agency, a vendor or the relationship between a transit agency and their vendor.

In the instances of the "Vendor Data Entry Problem" or "Agency Data Entry Problem", either the vendor or the agency responsible for GTFS upkeep knew how to update GTFS data but had not done so.

In several instances, transit agencies sometimes had multiple vendors producing GTFS data. One agency was transitioning between GTFS production processes and Caltrans was unable to confirm which GTFS feed was the correct one to use. The other two instances were more complex. Caltrans learned that one transit agency was unaware that they had an additional GTFS Schedule feed that was provided by a vendor under a subcontract from a prime contractor that provided managed transportation services for the agency. Caltrans was therefore using a GTFS Schedule feed that perhaps wasn't intended to be the official GTFS Schedule feed for the transit agency.

In one instance the transit agency had incorrect information about holiday schedules on its website which it promptly corrected. In another instance the transit agency was unable to complete a procurement for a vendor that could update their GTFS data.

The final kind of problem we learned about were instances where the transit agency or vendor had software that seemed to be capable of outputting data about holiday information. However, there may have been some confusion or lack of necessary training of agency staff such that they were unable to properly utilize these features. This was more common in agencies where they operated reduced service on holidays.

In a few instances we also learned that transit agencies resorted to using other means of communication besides GTFS Schedule data to communicate holiday service. This included the use of special news bulletins on the transit agency website or the use of GTFS Realtime Service Alerts stating that some service was not operating on a regular schedule.

# Possible Next Steps

The results of this research provided numerous different insights into holiday service and its representation of traveler information both on transit agency websites and in GTFS data. This section seeks to propose some possible next steps given all these findings. These next steps would need varying levels of coordination and support from a variety of stakeholders. Additionally, some items would require the production and collection of more data. The next steps are broken out into various sections discussing possible organizational changes and technical changes.

# Transit Agency Actions

As transit agencies are the ones that determine how much service they run, they are typically in control of how they chose to run or not run service on holidays. Transit agencies should strive to pick a holiday service pattern that is not too complex. When introducing new variations on regular service, transit agencies are on the one hand reducing expenditures from a reduction in service hours but are creating complexity that necessitates more work and potential for increased overhead costs. Transit agencies must properly account for these deviations from regular service in a variety of ways that require additional internal communication and coordination.

With additional increases in the variation of service on holidays, riders will have some level of difficulty in keeping track of these variations. It is critical for transit agencies to publish consistent information about their holiday service so that riders can correctly determine what service is available. Transit agencies should at a minimum always keep their own publications such as rider guides and their website up to date with the latest holiday information.

Transit agencies must also take steps to ensure that their GTFS data is kept up to date with the current holiday information. If the transit agency manages their own GTFS data, they should ensure that adding information about holidays is part of their standard operating procedures. Transit agencies should be careful to select software vendors that make editing holiday information as easy as possible. If a vendor manages a transit agency's GTFS data, the transit agency should hold their vendor accountable through their contract with the vendor. The contract should have specific clauses to ensure their vendor is held responsible for adding data about holiday service. The agency should regularly check that the vendor performed these updates.

## Vendor Actions

Transit agencies typically procure software to assist them with scheduling of their transit system. Vendors should ensure that their software is as easy to use as possible for when a transit agency needs to create a deviation from their regular schedule. Vendors should provide training to transit agencies on how to properly account for holidays using their software.

For vendors that are responsible for directly creating GTFS data, they should have standard operating procedures that ensure that they always enter in the current holiday information in the GTFS data.

## Inter-agency, Statewide Coordination

As was seen in the analysis of holiday service patterns, there is a large variety in the level of service for each holiday that each transit agency decides to run. This suggests that riders wanting to complete inter-agency journeys may encounter some legs of their journey that are associated with transit agencies that may have reduced or no service. Statewide coordination

could be useful to ensure that transit agencies or at least certain routes of statewide importance are closely coordinated to ensure proper connectivity.

## Representation of holiday service on websites

In our data collection activities, we observed a wide variety of how and where information was presented about holiday service. There was a large variety in how transit agencies communicated and presented holiday service on their websites. This large variety of formats of information could make it difficult for riders to determine how each transit agency operates. It could end up being useful to come up with some standardized guidelines or requirements for showing some basic information about the transit agency. This could help riders quickly determine how each transit agency approaches common operational characteristics.

When it comes to holiday service, it could be possible for agencies to produce a standardized graphic showing each holiday and whether they offer regular, reduced or no service. Perhaps the reduced service could also have a percentage of regular service shown as well. Additional aspects of how the transit agency operates with respect to its paratransit or other forms of service could also be included in this standardized graphic.

## Representation of holiday service in GTFS data

Some transit agencies and vendors seemed to prefer other GTFS solutions to communicate to riders about holiday service in-lieu of adding this data to their GTFS Schedule data. We heard of some transit agencies simply posting a GTFS Realtime Service Alert noting that holiday service would result in less or no service for certain routes. While being better than not communicating anything at all, this practice potentially goes against what is outlined in the GTFS Schedule service in the GTFS data as explicitly reduced or cancelled.

The documentation for the GTFS Schedule standard does mention that "the published GTFS dataset should be valid for at least the next 7 days". However, it does not say anything explicitly about how holiday service should or should not be accounted for. It might be a good idea to add something to the GTFS Schedule documentation that explicitly states that holidays should be accounted for.

## Institutionalization of Caltrans Data Quality Checks

At Caltrans, it could be possible to institutionalize the ongoing checking of holiday service with the transit agency website data and GTFS data that transit agencies publish. Caltrans has already collected data about holiday service for all transit agencies as part of this research project, so it has a decent amount of data that could be used in future checks. Some of the analysis work could be reran later to discover mismatches in additional holiday seasons. Caltrans currently does not have data about holidays outside of the research period. An extra effort would be needed to collect holiday information about all holidays throughout the year and to refresh the data for future years. If Caltrans were to institutionalize ongoing checks of Holiday Service, it would have to determine how often it should do these checks. The checks would likely have to be coordinated to ensure timely analysis and communication of any issues observed to the transit agencies. Since some holidays have a higher number of agencies reducing their schedule, it would likely have a larger impact to focus on some specific holidays instead of having monthly or even quarterly checks. In particular, the sequence of holidays of Thanksgiving Day, Christmas Day and New Year's Day all seemed to have many agencies reducing service and all occur within a relatively close amount of time to each other. A once-a-year check for these holidays could also help bring awareness of this issue to transit agencies. With a regular reminder, transit agencies may begin to ensure good data quality for holidays at other times of the year.

# Additional Research Ideas

While this project looked at specifically holiday service, there are many other deviations from regular service that a transit agency may include as part of their service. A very common deviation is that of detours. Detours can sometimes be very temporary in affecting just a single or a few trips. However, some detours can be extensive and known in advance such as detours related to construction activity. Further research into how to best publish detour information would be an interesting topic to pursue.

# Research Acknowledgements

This research was done by the following people:

- Gillian Gillett, Chief of Division of Data & Digital Services at Caltrans
- Evan Siroky, Research Data Manager of the Transit Data Quality team within the Division of Data & Digital Services at Caltrans
- Samantha Lewis, Research Data Analyst II of the Transit Data Quality team within the Division of Data & Digital Services at Caltrans
- Vivek Bansal, Research Data Analyst II of the Transit Data Quality team within the Division of Data & Digital Services at Caltrans

Gillian Gillette approved this research. Evan Siroky directed this research and wrote some of the background for this research paper. Samantha Lewis collected data from transit agencies websites and contributed to this research paper. Vivek Bansal completed various analyses of GTFS data, NTD data and how well it matched the data present on transit agency websites.