

## Appendix

Examples of errors from real GTFS feeds are listed below.

### equal\_shape\_distance\_diff\_coordindates

Chicago Transit Authority (CTA) - Chicago, IL

The shape points with shape\_pt\_sequence 590002 and 590003 have the same *shape\_dist\_traveled* but different coordinates.

shapes.txt

shape_id	shape_pt_lat	shape_pt_lon	shape_pt_sequence	shape_dist_traveled
66201341	41.883199	-87.627225	1	0
66201341	41.883199	-87.627225	2	40
66201341	41.883199	-87.627078	3	80
...				
66201341	41.881930	-87.635635	101	4183
66201341	41.881930	-87.635890	102	4183
66201341	41.881984	-87.635887	103	4183
...				
66201341	41.846328	-87.744498	959	46919

### decreasing\_or\_equal\_stop\_time\_distance

Chicago Transit Authority (CTA) - Chicago, IL

The *shape\_dist\_traveled* for stops with *stop\_sequence* 3 and 4 are equal.

stop\_times.txt

trip_id	arrival_time	departure_time	stop_id	stop_sequence	shape_dist_traveled
81248587799	09:46:00	09:46:00	30175	1	0
81248587799	09:52:30	09:52:30	30297	2	21220
81248587799	09:54:00	09:54:00	30026	3	26031
81248587799	10:00:00	10:00:00	30027	4	26031
81248587799	10:02:00	10:02:00	30298	5	30842
81248587799	10:09:00	10:09:00	30176	6	52062

## trip\_distance\_exceeds\_shape\_distance

### Long Beach Transit - Long Beach, CA

The total distance covered by a trip exceeds the total distance covered by the shape by 28 meters.

#### stop\_times.txt

trip_id	arrival_time	departure_time	stop_id	stop_sequence	shape_dist_traveled
9955296	06:47:28	06:47:28	1618	58	12.4110
9955296	06:47:35	06:47:35	2520	59	12.4600
9955296	06:48:00	06:48:00	3052	60	12.6350

#### shapes.txt

shape_id	shape_pt_lat	shape_pt_lon	shape_pt_sequence	shape_dist_traveled
1120049	33.860203	-118.153236	590002	12.5370
1120049	33.860203	-118.152155	590003	12.5990
1120049	33.860204	-118.152019	590004	12.6070

## foreign\_key\_violation

### Mountain Transit - San Bernardino County, CA

The file `fare_leg_rules.txt` uses `from_area_id` and `to_area_id` to define fare rules that apply for traveling from one area to another. The fields `from_area_id` and `to_area_id` are foreign keys in `fare_leg_rules.txt`, that are referenced from `area_id` in `areas.txt`. In this example, '85' and '81' `area_id` are used but the file `areas.txt` does not contain them.

#### fare\_leg\_rules.txt

leg_group_id	fare_leg_name	network_id	from_area_id	to_area_id
Mountain	NaN	NaN	NaN	NaN
Mountain	NaN	RIMregional	85	85
Mountain	NaN	RIMregional	85	85
Mountain	NaN	RIMregional	85	81
Mountain	NaN	RIMregional	81	85

## invalid\_currency\_amount

### LaGoBus - Los Angeles County, CA

The fare amount is invalid according to the ISO 4217 standard as it specifies that the currency amount should have two decimal places for cities in the United States.

#### fare\_products.txt

fare_product_id	fare_product_name	amount	currency
oneway_topanga_cash_general	One-Way Fare General Public (Cash)	1.0	USD
oneway_topanga_cash_senior	One-Way Fare Seniors Age 60 and Over (Cash)	0.5	USD
oneway_topanga_cash_disabled	One-Way Fare Persons with Disabilities (Cash)	0.5	USD
oneway_topanga_sv_general	One-Way Fare General Public (TAP Card Stored Value)	1.0	USD
oneway_topanga_sv_senior	One-Way Fare Seniors Age 60 and Over (TAP Card Stored Value)	0.5	USD

## stop\_time\_timepoint\_without\_times

### Carroll Transit System - Carroll County, MD

The stops with `stop_sequence` 1 and 2 are marked as timepoints but do not have arrival or departure times.

#### stop\_times.txt

trip_id	arrival_time	departure_time	stop_id	stop_sequence	timepoint
4211559	16:23:00	16:23:00	1530405	0	1
4211559	NaN	NaN	1530406	1	1
4211559	NaN	NaN	1530404	2	1
...					
4211559	16:50:00	16:50:00	1530415	10	1

## duplicate\_key

### Fredericksburg Regional Transit - Fredericksburg, VA

These entries in `stop_times.txt` both reference the same `trip_id` (D6\_T1), but they also both have a `stop_sequence` value of 1. Both cannot be the first stop on the same trip.

#### stop\_times.txt

trip_id	arrival_time	departure_time	stop_id	stop_sequence	stop_headsign	shape_dist_traveled	timepoint
D6_T1	06:30:00	06:30:00	474	1	S Commuter Lot	0.0	1
D6_T1	08:15:00	08:15:00	523	1	S Commuter Lot	0.9	0

## block\_trips\_with\_overlapping\_stop\_times

### MARTA - Atlanta, GA

The trips with `trip_ids` '8920792' and '8890989' operate on the same day, and have the same `block_id` but the stop times (arrival and departure) overlap.

#### trips.txt

trip_id	route_id	service_id	block_id	shape_id
8920792	20845	25	1145063	114058
8890989	20845	25	1145063	114058

#### stop\_times.txt

Index	trip_id	arrival_time	departure_time	stop_id	stop_sequence
1343623	8890989	5:40:00	5:40:00	10266	1
1343624	8890989	5:41:16	5:41:16	10270	2
1344160	8920792	5:40:00	5:40:00	10266	1
1344161	8920792	5:41:16	5:41:16	10270	2

## missing\_required\_fields

### Pierce Transit - Pierce County, WA

Within `fare_attributes.txt`, the field `agency_id` is conditionally required if multiple agencies are present within the feed. Here, the `agency_id` is missing for rows 1,2 and 5.

#### `fare_attributes.txt`

fare_id	price	currency_type	payment_method	transfers	agency_id	transfer_duration	youth_price	senior_price
1	2.00	USD	0	0	NaN	NaN	1.00	1.00
31	2.75	USD	0	0	NaN	NaN	1.50	1.00
full	2.00	USD	0	NaN	3	7200.00	0.00	0.00
stexpress_full	3.25	USD	0	NaN	40	7200.00	0.00	0.00
32	3.75	USD	0	0	NaN	NaN	2.75	1.75

## fare\_transfer\_rule\_missing\_transfer\_count

### Norwalk Transit System (NTS) - Norwalk, CA

Within `fare_transfer_rules.txt`, the field `transfer_count` is conditionally required if `from_leg_group_id` and `to_leg_group_id` are same. Here, the `transfer_count` is missing for all rows.

from_leg_group_id	to_leg_group_id	fare_product_id	transfer_count	duration_limit	fare_transfer_type
Norwalk	Norwalk	NaN	NaN	NaN	0
Norwalk	Norwalk	NaN	NaN	NaN	0
Norwalk	Norwalk	NaN	NaN	NaN	0
Norwalk	Norwalk	NaN	NaN	NaN	0
Norwalk	Norwalk	NaN	NaN	NaN	0
Norwalk	Norwalk	NaN	NaN	NaN	0
Norwalk	Norwalk	EZpass_regular	NaN	NaN	0
Norwalk	Norwalk	EZpass_senior	NaN	NaN	0
Norwalk	Norwalk	EZpass_disabled	NaN	NaN	0