

# Metro\_GO\_Transit

April 23, 2025

```
[1]: import os
import pandas as pd

# Path to the folder where GTFS files are extracted
folder_path = r"C:\Users\arham\Downloads\GO-GTFS"

# Dictionary to store all dataframes
gtfs_data = {}

# Load each .txt file into a separate DataFrame
for filename in os.listdir(folder_path):
    if filename.endswith('.txt'):
        file_path = os.path.join(folder_path, filename)
        df_name = filename.replace('.txt', '') # e.g., stops.txt -> stops
        gtfs_data[df_name] = pd.read_csv(file_path)

# Example: Accessing the stops dataframe
stops_df = gtfs_data['stops']
print(stops_df.head())
```

C:\Users\arham\AppData\Local\Temp\ipykernel\_21612\2421658755.py:15:

DtypeWarning: Columns (0) have mixed types. Specify dtype option on import or set low\_memory=False.

```
gtfs_data[df_name] = pd.read_csv(file_path)
```

C:\Users\arham\AppData\Local\Temp\ipykernel\_21612\2421658755.py:15:

DtypeWarning: Columns (7) have mixed types. Specify dtype option on import or set low\_memory=False.

```
gtfs_data[df_name] = pd.read_csv(file_path)
```

	stop_id	stop_name	stop_lat	stop_lon	zone_id	\
0	WR	West Harbour GO	43.266775	-79.866222	18	
1	WH	Whitby GO	43.864840	-78.938180	93	
2	WE	Weston GO	43.700220	-79.514671	4	
3	UN	Union Station	43.645195	-79.380600	2	
4	UI	Unionville GO	43.851689	-79.314332	71	

  

	stop_url	location_type	\
0	https://www.gotransit.com/find-a-station-or-st...	0	
1	https://www.gotransit.com/find-a-station-or-st...	0	

```

2 https://www.gotransit.com/find-a-station-or-st... 0
3 https://www.gotransit.com/find-a-station-or-st... 0
4 https://www.gotransit.com/find-a-station-or-st... 0

```

```

parent_station wheelchair_boarding stop_code
0           NaN                1         NaN
1           NaN                1         NaN
2           NaN                1         NaN
3           NaN                1         NaN
4           NaN                1         NaN

```

```

[3]: # Load the data
stop_times = gtfs_data['stop_times']
stops = gtfs_data['stops']
trips = gtfs_data['trips']
routes = gtfs_data['routes']

```

## 1 Top 10 Stops by Scheduled Arrivals

```

[39]: # Merge stop names into stop_times
stop_times_named = pd.merge(stop_times, stops[['stop_id', 'stop_name', 'stop_lat', 'stop_lon']], on='stop_id', how='left')

# Group by stop_name (not stop_id) and sum counts
top_stops = (
    stop_times_named.groupby('stop_name')
    .agg({
        'stop_id': 'count',          # This becomes num_times_served
        'stop_lat': 'mean',
        'stop_lon': 'mean'
    })
    .reset_index()
    .rename(columns={'stop_id': 'num_times_served'})
    .sort_values(by='num_times_served', ascending=False)
    .head(10)
)
top_stops

```

```

[39]:
stop_name  num_times_served  stop_lat \
507      Square One          36904  43.594611
261      Hwy 407 Bus Terminal  26176  43.783063
61       Bramalea GO Bus      25363  43.702269
562      Union Station        24324  43.645195
696      Yorkdale Bus Terminal 19754  43.724937
154      Durham College Oshawa GO 17608  43.870943
232      Hamilton GO Centre Bus  17238  43.252945
552  Trafalgar Rd. @ Hwy. 407 Park & Ride 15629  43.506545

```

485	Renforth Transitway Station	15358	43.663035
140	Dixie Transitway Station	15358	43.630030

```

stop_lon
507 -79.647546
261 -79.524059
61 -79.690105
562 -79.380600
696 -79.448875
154 -78.885477
232 -79.869586
552 -79.745813
485 -79.591200
140 -79.614657

```

```
[41]: top_stops.to_csv("top_10_stops_grouped_by_name.csv", index=False)
```

## 2 Top 30 Most Frequent Stop Pairs

```
[17]: stop_times = stop_times.sort_values(by = ['trip_id', 'stop_sequence'])
stop_times['next_stop_id'] = stop_times.groupby('trip_id')['stop_id'].shift(-1)
stop_pairs = stop_times.dropna(subset=['next_stop_id'])
stop_pairs['next_stop_id'] = stop_pairs['next_stop_id'].astype(str)
```

C:\Users\arham\AppData\Local\Temp\ipykernel\_21612\957052164.py:4:

SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.

Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
stop_pairs['next_stop_id'] = stop_pairs['next_stop_id'].astype(str)
```

```
[19]: pair_counts = stop_pairs.groupby(['stop_id', 'next_stop_id']).size().
      ↪reset_index(name='count')
pair_counts = pd.merge(pair_counts, stops[['stop_id', 'stop_name']],
      ↪on='stop_id', how='left').rename(columns={'stop_name': 'stop_name_from'})
pair_counts = pd.merge(pair_counts, stops[['stop_id', 'stop_name']],
      ↪left_on='next_stop_id', right_on='stop_id', how='left').
      ↪rename(columns={'stop_name': 'stop_name_to'})
pair_counts['stop_pair'] = pair_counts['stop_name_from'] + " → " +
      ↪pair_counts['stop_name_to']
top_30_pairs = pair_counts[['stop_pair', 'count']].sort_values(by='count',
      ↪ascending=False).head(30)
top_30_pairs.to_csv("top_30_stop_pairs.csv", index=False)
```

```
[43]: top_30_pairs
```

```
[43]:
```

	stop_pair	count
997	Yonge St. @ Hwy. 407 → Richmond Hill Centre	12066
222	Richmond Hill Centre → Yonge St. @ Garden Ave.	12066
743	Main St. W. @ Haddon Ave. S. → McMaster Univer...	9341
911	Hwy 407 Bus Terminal → Bramalea GO Bus	8820
400	Main St. W. @ Macklin St. S. → Main St. W. @ D...	8807
281	Main St. W. @ Longwood Rd. S. → Main St. W. @ ...	8807
405	Main St. W. @ Poulette St. → Main St. W. @ Pea...	8807
398	Main St. W. @ Caroline St. S. → Main St. W. @ ...	8807
399	Main St. W. @ Dundurn St. S. → Main St. W. @ P...	8807
404	Main St. W. @ Ray St. S. → Main St. W. @ Carol...	8807
406	Main St. W. @ Summers Ln. (Hamilton City Hall)...	8807
403	Main St. W. @ Pearl St. S. → Main St. W. @ Ray...	8807
401	Main St. W. @ Paisley Ave. S. → Main St. W. @ ...	8807
10	Yorkdale Bus Terminal → Hwy. 401 @ Keele St.	8455
393	King St. W. @ Caroline St. N. → King St. W. @ ...	8431
83	Hamilton GO Centre Bus → King St. E. @ Hughson...	8431
988	King St. W. @ Queen St. N. → King St. W. @ Pea...	8431
396	King St. W. @ Strathcona Ave. N. → King St. W...	8431
394	King St. E. @ Hughson St. N. → King St. W. @ S...	8431
395	King St. W. @ Pearl St. N. → King St. W. @ Str...	8431
397	King St. W. @ Summers Ln. (Hamilton Place) → K...	8431
20	Keele St. @ Hwy. 401 → Yorkdale Bus Terminal	8234
1042	Yonge St. @ Poyntz Ave. → Yonge St. @ Florence...	7778
899	Renforth Transitway Station → Dixie Transitway...	7708
897	Dixie Transitway Station → Square One	7708
63	Square One → Dixie Transitway Station	7650
898	Dixie Transitway Station → Renforth Transitway...	7650
71	Square One → Erin Mills Transitway Station	7222
760	Erin Mills Transitway Station → Square One	7133
1071	Bramalea GO Bus → Hwy 407 Bus Terminal	6883

### 3 Hourly Stops with Day Routes

```
[23]: trips['trip_date'] = pd.to_datetime(trips['trip_id'].str[:8], errors='coerce',
    ↪format='%Y%m%d')
trips['day_of_week'] = trips['trip_date'].dt.day_name()
trips = pd.merge(trips, routes[['route_id', 'route_long_name']], on='route_id',
    ↪how='left')
trips
```

```
[23]:
```

	route_id	service_id	trip_id	trip_headsign \
0	04250625-47	20250625	20250625-47-47514	47G - Erin Mills
1	04250625-18	20250625	20250625-18-18901	18 - Hamilton GO

2	04250625-56	20250625	20250625-56-56720	56 - Oakville GO
3	04250625-41	20250625	20250625-41-41502	41B - Centennial College
4	04250625-41	20250625	20250625-41-41731	41 - Hamilton GO
...	...	...	...	...
113016	04250625-32	20250509	20250509-32-32461	32 - Trinity Common
113017	04250625-41	20250509	20250509-41-41373	41C - Square One
113018	04250625-32	20250509	20250509-32-32280	32 - Finch Terminal
113019	04250625-32	20250509	20250509-32-32200	32 - Finch Terminal
113020	04250625-33	20250509	20250509-33-33621	33B - Mount Pleasant GO

	trip_short_name	direction_id	block_id	shape_id	\
0	NaN	0	MDDS1142	388876	
1	NaN	1	MDDS1141	388773	
2	NaN	1	MDDS1141	388893	
3	NaN	0	MDDS1141	388870	
4	NaN	1	MDDS1140	388871	
...	...	...	...	...	
113016	NaN	1	MDDS3210	388840	
113017	NaN	1	MDDS3210	388868	
113018	NaN	0	MDDS3210	388842	
113019	NaN	0	MDDS3210	388842	
113020	NaN	1	MDDS3209	388853	

	wheelchair_accessible	bikes_allowed	route_variant	trip_date	\
0	1	1	47G	2025-06-25	
1	1	1	18	2025-06-25	
2	1	1	56	2025-06-25	
3	1	1	41B	2025-06-25	
4	1	1	41	2025-06-25	
...	...	...	...	...	
113016	1	1	32	2025-05-09	
113017	1	1	41C	2025-05-09	
113018	1	1	32	2025-05-09	
113019	1	1	32	2025-05-09	
113020	1	1	33B	2025-05-09	

	day_of_week	route_long_name_x	\
0	Wednesday	Hamilton / Hwy 407 Terminal	
1	Wednesday	Lakeshore West	
2	Wednesday	Oshawa / Oakville	
3	Wednesday	Hamilton / Pickering	
4	Wednesday	Hamilton / Pickering	
...	...	...	
113016	Friday	Brampton Trinity Common / North York	
113017	Friday	Hamilton / Pickering	
113018	Friday	Brampton Trinity Common / North York	
113019	Friday	Brampton Trinity Common / North York	

113020            Friday                            Guelph / North York

```

                                route_long_name_y
0          Hamilton / Hwy 407 Terminal
1                                Lakeshore West
2                                Oshawa / Oakville
3          Hamilton / Pickering
4          Hamilton / Pickering
...
113016  Brampton Trinity Common / North York
113017          Hamilton / Pickering
113018  Brampton Trinity Common / North York
113019  Brampton Trinity Common / North York
113020          Guelph / North York

```

[113021 rows x 15 columns]

```
[27]: stop_times_merged = pd.merge(stop_times, trips[['trip_id', 'trip_date',
↪ 'day_of_week', 'route_long_name_x']], on='trip_id', how='left')
```

```
[29]: stop_times_merged['hour'] = stop_times_merged['arrival_time'].str.split(':').
↪ str[0].astype(int)
```

```
[31]: stop_times_merged
```

```
[31]:
      trip_id arrival_time departure_time stop_id stop_sequence \
0      20250416-11-11061      06:11:00      06:11:00      00180          1
1      20250416-11-11061      06:19:00      06:19:00      01010          2
2      20250416-11-11061      06:19:00      06:19:00      00732          3
3      20250416-11-11061      06:20:00      06:20:00      01009          4
4      20250416-11-11061      06:21:00      06:21:00      01008          5
...
1923875  20250627-ST-7909      08:10:00      08:10:00      UI          9
1923876  20250627-ST-7909      08:16:00      08:16:00      MK         12
1923877  20250627-ST-7909      08:23:00      08:23:00      AG         13
1923878  20250627-ST-7909      08:31:00      08:31:00      KE         16
1923879  20250627-ST-7909      08:50:00      08:50:00      UN         28

```

```

      pickup_type drop_off_type \
0              0              0
1              0              0
2              0              0
3              0              0
4              0              0
...
1923875          0              0
1923876          0              0

```

1923877	0	0
1923878	0	0
1923879	0	0

		stop_headsign	next_stop_id	trip_date	\
0		NaN	01010	2025-04-16	
1		NaN	00732	2025-04-16	
2		NaN	01009	2025-04-16	
3		NaN	01008	2025-04-16	
4		NaN	01013	2025-04-16	
...		...	...	...	
1923875	Unionville GO 08:10 - Union Station 08:50			MK	2025-06-27
1923876	Milliken GO 08:16 - Union Station 08:50			AG	2025-06-27
1923877	Agincourt GO 08:23 - Union Station 08:50			KE	2025-06-27
1923878	Kennedy GO 08:31 - Union Station 08:50			UN	2025-06-27
1923879		NaN		NaN	2025-06-27

	day_of_week	route_long_name_x	hour
0	Wednesday	Brock University	6
1	Wednesday	Brock University	6
2	Wednesday	Brock University	6
3	Wednesday	Brock University	6
4	Wednesday	Brock University	6
...	...	...	...
1923875	Friday	Stouffville	8
1923876	Friday	Stouffville	8
1923877	Friday	Stouffville	8
1923878	Friday	Stouffville	8
1923879	Friday	Stouffville	8

[1923880 rows x 13 columns]

```
[33]: stop_times_merged[['hour', 'day_of_week', 'route_long_name_x']].
      ↪to_csv("hourly_stops_with_day_route.csv", index=False)
```