Another important variable is \*\*adherence\*\*, which compares the actual departure time to the scheduled time and is included in the ADHERENCE column. A negative adherence value means that a bus left a time point late and a positive adherence indicates that the bus left the time point early. Buses with adherence values beyond negative 6 are generally considered late and beyond positive 1 are considered early. However, there is some additional logic where the staff applies waivers to allow early departures, such as an express bus that has already picked up everyone at a park-and-ride lot and is only dropping people off at the remaining stops, and also allows for early timepoint records for all records where TRIP\_EDGE = 2 (end of trip), since it is not a problem if a bus ends its trip early as long as it didn't pass other timepoints early along the way. \*\*Note:\*\* When determining whether a bus is early or late, it is advised that you use the 'ADJUSTED\_EARLY\_COUNT', 'ADJUSTED\_LATE\_COUNT', and 'ADJUSTED\_ONTIME\_COUNT' columns in order to account for the adjustments.

Columns related to Adherence:

* ADHERENCE
* ADJUSTED\_EARLY\_COUNT
* ADJUSTED\_LATE\_COUNT
* ADJUSTED\_ONTIME\_COUNT
* SCHEDULED\_TIME: Scheduled stop time
* ACTUAL\_DEPARTURE\_TIME: actual departure time

\*\*Headway\*\* is the amount of time between a bus and the prior bus at the same stop. In the dataset, the amount of headway scheduled is contained in the SCHEDULED\_HDWY column and indicates the difference between the scheduled time for a particular stop and the scheduled time for the previous bus on that same stop.

\*\*Bunching\*\* occurs when there is shorter headway than scheduled, which would appear as a negative HDWY\_DEV value. \*\*Gapping\*\* is when there is more headway than scheduled and appears as a positive value in the HDWY\_DEV column. Note that you can calculate headway deviation percentage as HDWY\_DEV/SCHEDULED\_HDWY. The generally accepted range of headway deviation is 50% to 150% of the scheduled headway, so if scheduled headway is 10 minutes, a headway deviation of up to 5 minutes would be acceptable (but not ideal).

columns related to Headway:

* SCHEDULED\_HDWY
* ACTUAL\_HDWY
* HDWY\_DEV
* PREV\_SCHED\_STOP\_CANCELLED: flags whether the previous timepoint crossing was cancelled or waived. Useful for excluding records where the headway values are extremely high because the bus is just coming off a detour

80281 trips not on-time

9775 trips early, 59038 trips late

2. How does direction of travel, route, or location affect the headway and on-time performance?

* Direction of travel = ROUTE\_DIRECTION\_NAME
  + ~~remove: rows where scheduled headway = 0~~
  + ~~look at max hdwy dev perc of 24?? related to PREV\_SCHED\_STOP\_CANCELLED?~~
* route = ROUTE\_ABBR
  + divide by total # of trips/rows on that route?
* location =
  + BLOCK\_ABBR:
  + TIME\_POINT\_ABBR – match these with route?
  + ROUTE\_STOP\_SEQUENCE
  + LATITUDE/LONGITUDE – Ola said map these; be careful of main downtown hub

approach = filter by each of these factors to find aggregates values:

headway deviation pct

bunching

gapping

acceptable\_hdwy

'ADJUSTED\_ONTIME\_COUNT'

exploration of columns:

.value\_counts()

.isnull().value\_counts()

.describe() – if int or float

ontime answer from John Michael:

For on time performance, route 22 (Bordeaux) is the most "on time" route, location MCC5\_6 is the most "on time" location which is downtown around the wego central station, and going to downtown has a higher "on time" percentage than going from downtown.

headway answer from Emily:

* direction of travel: “from downtown” has a slightly higher headway deviation %; bunching and gapping are also higher from downtown. “to downtown” is on average more often within the acceptable headway deviation range.
* route:
  + route 50 (name) has the lowest headway deviation percentage, and route 3 (name) has the highest.
  + route 55 (name) has the most bunching on average and route 22 (name) has the lowest.
  + route 56 (name) has the most gapping on average and route 7 (name) has the lowest.
  + route 50 (name) has the most acceptable headway on average, and route 7 (name) has the lowest.
* location:
  + DCSCC (on route x) has the highest headway deviation percentage, and MLKS (on route x) has the lowest
  + MXWHARF (on route x) has the highest bunching
  + CV23 (on route x) has the highest gapping
  + 28&CHARL (on route x) has the most acceptable headway

Q3: How does time of day or day of week affect headway and on-time performance?

#converting columns to datetime (for #3)

wego['SCHEDULED\_TIME'] = pd.to\_datetime(wego['SCHEDULED\_TIME'])wego['ACTUAL\_ARRIVAL\_TIME'] = pd.to\_datetime(wego['ACTUAL\_ARRIVAL\_TIME'])wego['ACTUAL\_DEPARTURE\_TIME'] = pd.to\_datetime(wego['ACTUAL\_DEPARTURE\_TIME'])wego.info()

use scheduled time, actual departure time for hour

add late & early adherence to analysis throughout

group analysis fields together