1. TSRTC Hyderabad all depots bus schedules data analysis report
2. By: Nikhil VJ
3. Date: 8.Aug.2018
4. Status: v2
5. Initial notes:

1. File received from TSRTC having data for all depots in Hyderabad: TSRTC-alldepots-ORIG.csv

2. Download all related files here: <http://nikhilvj.co.in/files/tsrtc/>

3. Rudimentary GTFS feed for the full network is **ready** and available under “gtfs1” folder in above link.

4. Analysis sheets are stored together in “TSRTC-analysis.xlsx” file and the same is shared [on google drive here](https://docs.google.com/spreadsheets/d/13ySs3HbyPJ1tRNaeeBNpPuueSkeRFTgzDxSf7trh4bo/edit?usp=sharing). (I might create some charts there).

5. Programs were done in python jupyter notebooks which allow for exploratory programming and documentation, they are up in .ipynb format, on first link shared above.

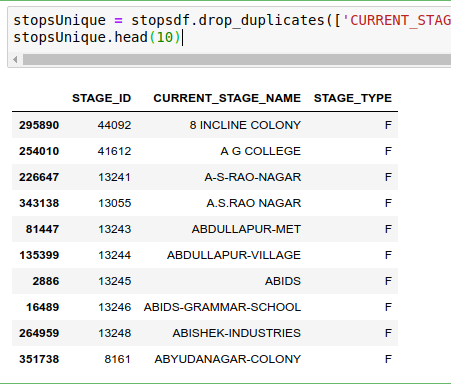
6. This doc is split into 3 broad sections: Initial/GTFS, Analysis, Anomalies.

# 1. Initial processing, GTFS making

## 1.1. Stops

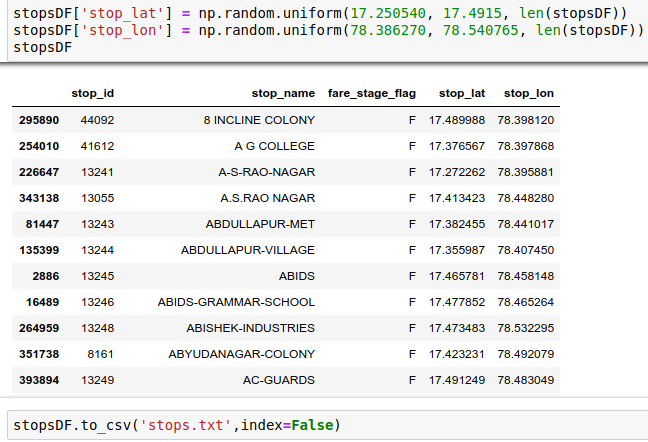
Sound some stray spaces/tabs in some stop entries (CURRENT\_STAGE\_NAME), shared it in anomalies section. Fixed them.

Filtered original csv to get unique listing of stops, 1737 total.



Assigned randomized lat-longs.

Then created stops.txt from that. Preserved the stage\_type column as I expect it may be needed further along.



## 1.2. Routes

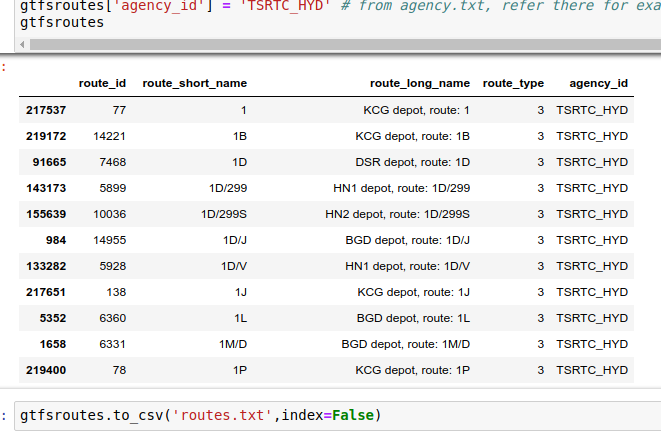
News: ROUTE\_NAME from KCG file is replaced now by a ROUTE\_NUMBER column, and it has a much better representation of actual route titles. (see a quick list of all routes in appendix section)

Filtered main csv to give unique entries by route\_id, route\_number and depot\_code. Total 833.

There were anomalies, see anomalies section below for details.

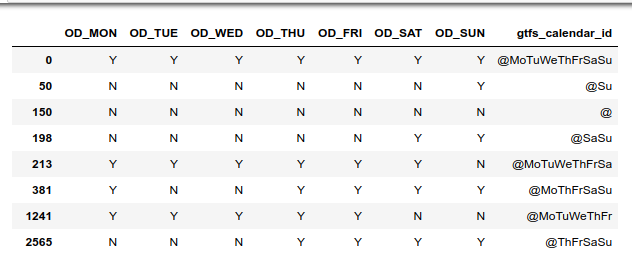
Some route names have multiple route\_id’s attached to them, so when counting by route\_number column, there are 695 routes.

Created gtfs routes.txt. 833 routes total.

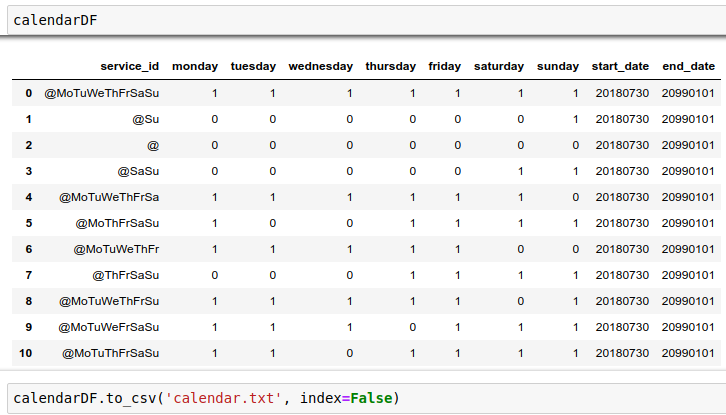


## 1.3. Calendar

Generated a GTFS id for calendar (also called service\_id, don’t confuse it with TSRTC’s) for each trip by processing the OD\_MON through OD\_SUN columns. Got 29 calendar combinations total.



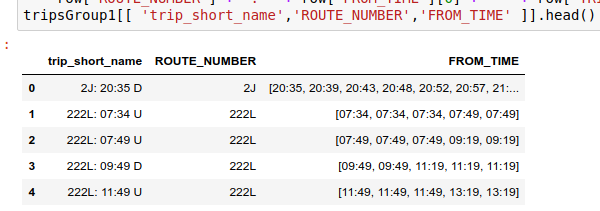
After generating the id’s, generated the gtfs calendar.txt file.



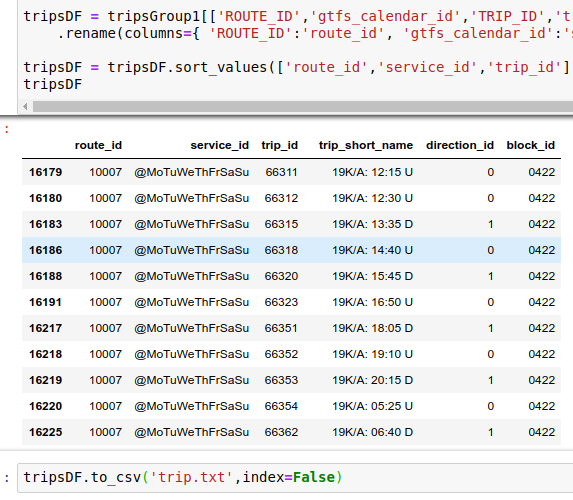
## 1.4. Trips

Grouped TSRTC file by trip\_id. Trip\_short\_name field derived from route number, first stop’s time and direction.

~~35657~~ 35550 trips total.



Created trips.txt :

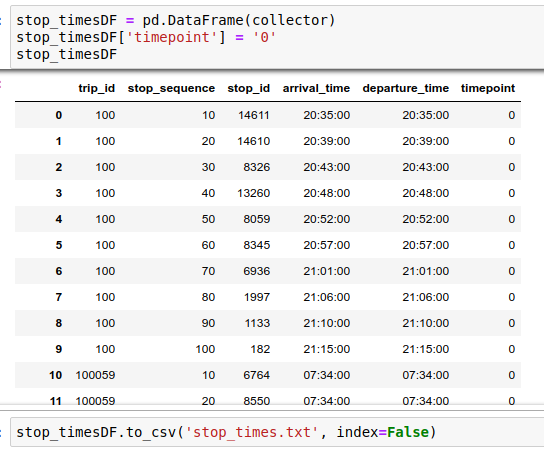


Used same trip\_id as TSRTC’s TRIP\_ID as it was unique dataset.

Used SERVICE\_ID from the data as gtfs’s block\_id field.

## 1.5. Stop\_times

* Created.. 435,859 rows total.
* Have kept the sequence numbers as 10,20... instead of 1,2.. so that intermediate stops between stages can be inserted later on. The GTFS spec allows to use any numbers for sequence as long as they’re in ascending order.
* Arrival and departure timings are kept the same, based on FROM\_TIME row.
* One change in values : in trips that cross midnight, 24hrs are added as per gtfs specs. (ex: 00:45 becomes 24:45) . For this, a new column FROM\_TIME\_NEW was created.



## 1.6. Validation of rudimentary GTFS feed

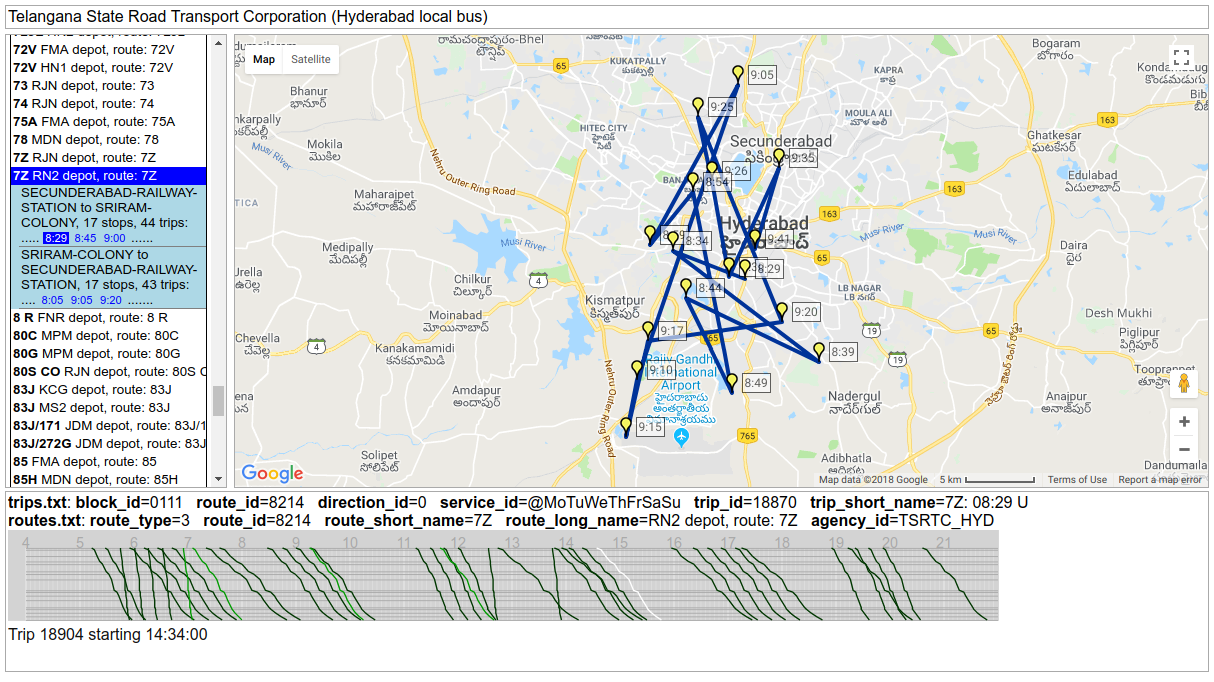
* Find full Validation report here: http://nikhilvj.co.in/files/tsrtc/gtfs1/validator-hydgtfs1.html
* No errors, many warnings.
* Created using same program as the one behind [http://gtfsfeedvalidator.transitscreen.com](http://gtfsfeedvalidator.transitscreen.com/) but with longer listing for each type of error.

## 1.7. Visualization of rudimentary GTFS feed

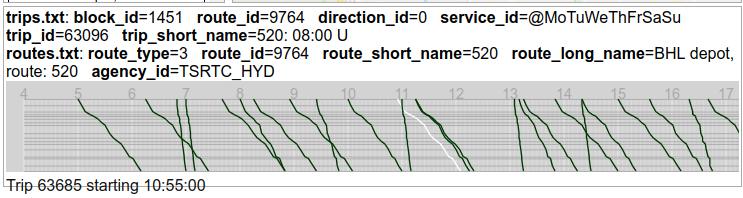
Using [ScheduleViewer](https://github.com/google/transitfeed/wiki/ScheduleViewer) by google transit

Instructions for running on windows:

* Go to <https://github.com/google/transitfeed/releases>
* Look for windows-binary (should be under the 1.2.15 one)
* Download and unzip
* Place the gtfs feed zip file in this unzipped folder, next to schedule\_viewer.exe
* Double-click schedule-viewer.exe
* In the black box, type in the zip filename (including .zip extension) and press Enter
* Wait.. after many messages the black box will tell to open a URL
* Open [http://localhost:8765](http://localhost:8765/) in your browser.



See [full page screenshot here](https://i.imgur.com/4zrR5CF.png).



The trips timeline at bottom shows up trips that are criss-crossing, having too short times etc.

# 2. Analysis

## 2.1. Depot\_code

['KCG', 'MI2', 'JDM', 'HN1', 'HN2', 'KGD', 'CNT', 'FMA', 'UPL',

'BKP', 'RN2', 'DSR', 'MPM', 'MS2', 'RJN', 'FNR', 'MS1', 'MHM',

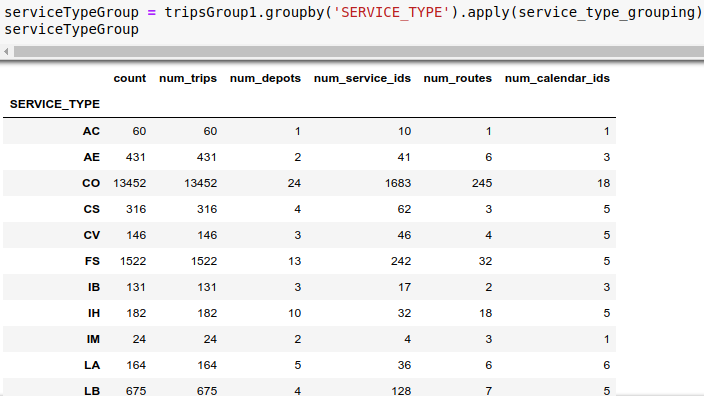
'BGD', 'MDN', 'BHL', 'KPL', 'HPT', 'MI1', 'CCL', 'HCU', 'RN1']

27 Depots

Have to get the full names of these depots. See appendix .

## 2.2. Service\_type

Frequency distribution of service\_types



See sheet “**serviceType\_grouping**” in analysis excel.

Service\_type column has many types of services that need deciphering. See Appendix

## 2.3. Routes

### 2.3.1. Naming of routes

There are routes with are number-only, mixed (1B, 1M/D etc), and word (like GUNTUR, CTPL). Breakdown of numbers:

alphanumeric

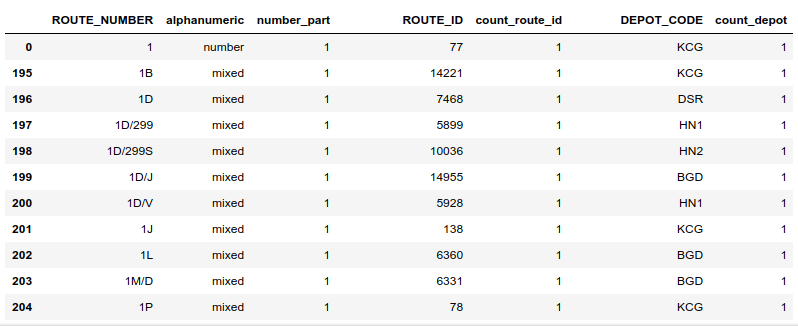
mixed 528

number 121

word 46

### 2.3.2. Listing

Made a full listing of route\_number and corresponding route\_id and depot\_code. Where repeats happen, they are concatenated.



See “**Route\_id\_depot\_listing**” in analysis excel

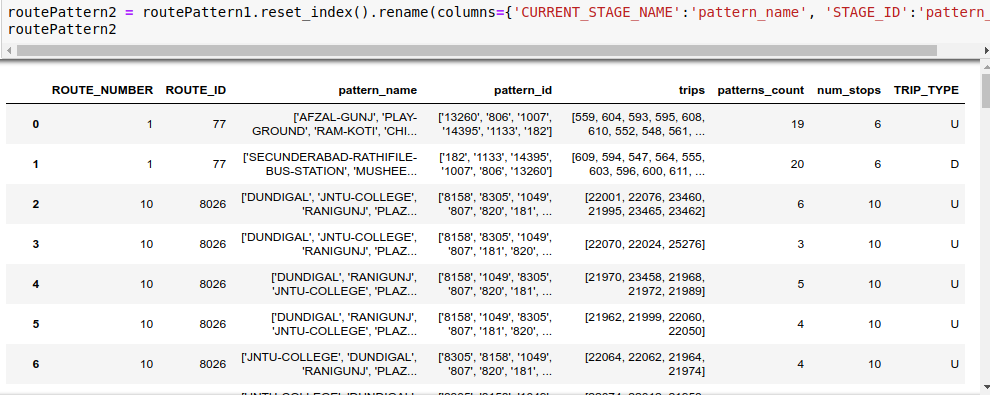
### 2.3.3. Routes-services mapping

We can compare this with MTD141 data.

See “**Route\_service\_mapping**” sheet in analysis excel.

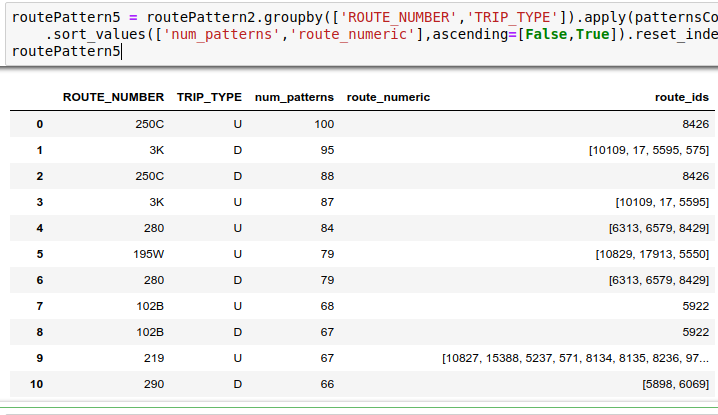
### 2.3.4. Route-patterns mapping

Pattern: a sequence of stops covered in a trip.



When grouping the whole dataset by patterns, it is found that there is one pair of up-down patterns that are repeating under both routes 1 and 1P.

Pattern Counts:



Route “250C” has highest number of varying patterns: 100 Up, 88 Down.

Out of 695, there are 384 routes that have just one Up pattern and one Down pattern. MPM depot has max, 36 such "proper" routes.

See “**Route\_patterns**” and “**Route\_pattern\_counts**” sheets in analysis excel.

## 2.4. Trips

### 2.4.1. Lengths of trips

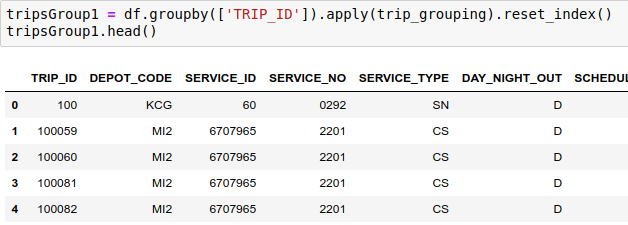
Frequency distribution made of how many trips or routes are 4 stops long, how many have 20 stops, etc.

## 

See sheet “**TripLengths\_grouping**” in analysis excel.

### 2.4.2. Tripwise full listing

Grouped the csv data by Trip\_id, to get each trip in its own row.



See “**grouped-tripwise.csv**” file. Saved this separately as it was large.

# 3. Issues, Anomalies

Suggestion: We could relay these back to TSRTC and they might make changes at source and give us fresh data.

## 3.1. Stop Anomalies

### 3.1.1. Leading or trailing spaces

Some stops have spaces or tabs in their names. ‘Trimmed’ them.

['S.R.NAGAR\t', 'EENADU\t', 'MUNICIPAL OFFICE\t', 'ECIL\t',

'AGRICULTURAL COLLEGE \t']

### 3.1.1. Repeating stops

Cases of different stop id’s, same name

## 

See “Stops\_Repeating” sheet in analysis excel.

## 3.2. Route Anomalies

### 3.2.1. Blanks

Route\_id 16390 has blank route\_number value.

Have set it to ‘BLANK’ for now. Need to find out which route it originally is.

It has 37 trips total and is under CCL depot

Here is its stops sequence:

['CHENGICHERLA-DEPOT',

'CHENGICHERLA',

'RTC-COLONY-CHENEGICHERLA',

'MEDIPALLY',

'PEERAZADI-GUDA',

'UPPAL-BUS-STAND',

'MODERN-BAKERY',

'RAMANTHAPUR-PUBLIC-SCHOOL',

'AMBERPET-GANDHI-STATUE',

'KACHIGUDA',

'KOTI']

### 3.2.2. Leading or trailing spaces

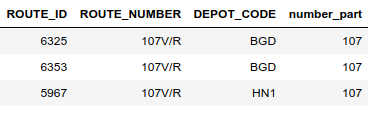
ROUTE\_ID 15296 had a leading space in its ROUTE\_NUMBER : ‘ 212/568’

Have trimmed it.

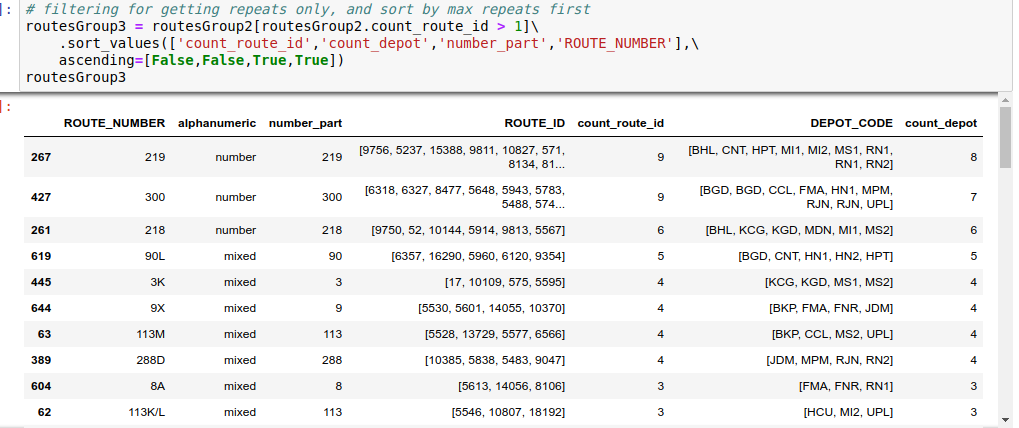
### 3.2.3. Repeating routes

Route\_number values are repeating in 97 cases.

Example:



Listing made of repeating route\_number values, with lists of route\_ids and depot\_codes.

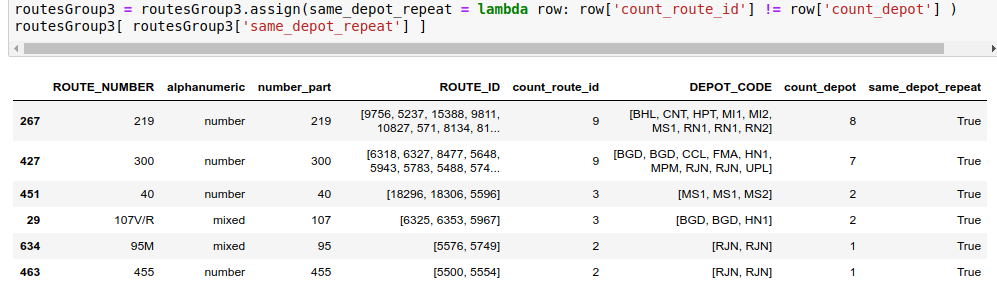


When counting by route\_number, there are 695 unique names. (vs 833 unique id’s)

See “**Routes\_repeating**” sheet in analysis excel.

### 3.2.4. Routes repeating in same depot

In a few cases, under the same depot there are instances of two route\_id’s and same route\_number title.



See “**Routes\_same\_depot**” in analysis excel.

### 3.2.5. Routes having only one direction

36 routes have only Up direction:

['22R' '455' '78' '95M' '120K' '14X' '16X' '188/251' '188H' '19F/A' '19KA'

'251J SU' '251M SU' '251P' '251SS' '2U' '449' '44EX' '450' '453' '454'

'532 HK' '73' '74' '8AU' '92' '92A' '92S' '94/95' '94K' '94R' '94U' '95N'

'95P' '95R' 'MP-NZB']

3 routes have only Down direction:

['288N/B' '90K' 'DVRM']

28 of these are under RJN depot.

Depot-wise counts of one-direction-only routes:

[RJN] 28

[CNT] 3

[MDN] 2

[FNR] 2

[MPM] 2

[KCG] 1

[HPT] 1

See “**Route\_pattern\_counts**” in analysis excel.

## 3.3. Timings/trips Anomalies

### 3.3.1. Blanks

5 trips under 2 routes had all timings cells as blank.

Route number 94U : trip\_id 82427

Route number 567: trip\_id: 96745, 96746, 96747, 96748

They were dropped from the data to process for now, as without timings info these cannot make valid gtfs.

### 3.3.2. Trips with only two stops

793 trips found having only two stops / stages.

They come under the following route numbers:

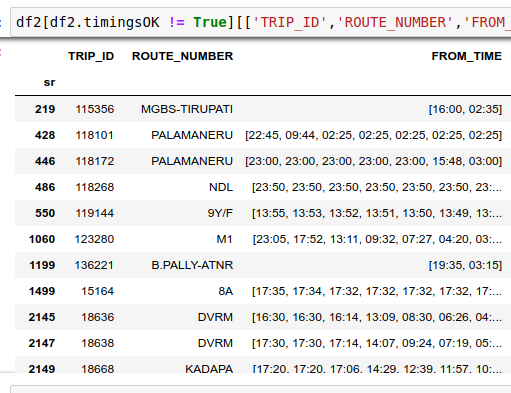
['MGBS-TIRUPATI', '1Z', '45F', '9F', 'B.PALLY-ATNR', '1B', '3H/N',

'190R', '113YK', '45RF', '6G', 'VJA', '158F/VSN', '10AS', '10F',

'KP', '19S/F', '19K/A', '158YF']

### 3.3.3. Odd timings

**105** trips were found to have timings that were descending or irregular. See screenshot below.



It is difficult to programmatically correct them. Some are simple cases of reversals, but in cases of trip crossing midnight, it is difficult to sift them out from trips that are simply problematic. So, for now, these 105 trips are also dropped out from the processed trip-wise data.

There are 60 routes under which they come, route VJA-ATNR has max odd trips at 8. Majority of the 60 routes have only 1 or 2 odd trips.

# 4. Appendix

## 4.1. All Route\_number values

Listing of routes from their ROUTE\_NUMBER values in the TSRTC file.

Note: subsequently, a blank one has been renamed to BLANK, and one having a leading space has been trimmed.

[, 212/568, 1, 10, 100, 100D, 100R, 101K/D, 102, 102/253L, 102A/G, 102B, 102B/218L, 102C, 102K, 102M, 102P, 102W, 103, 104A, 104A/216, 104M/127K, 104R, 104R/127K, 104R/218L, 104S, 105, 107J/21, 107JD, 107JS, 107JS-XC, 107V/R, 107VR, 107VR-ME, 107VR-SN, 10A/S, 10AS, 10BK, 10F, 10F/V, 10H, 10H/16A, 10H/25I, 10H/L, 10HA, 10K, 10K/L, 10K/P, 10KB, 10KJ, 10KM, 10L, 10L/18, 10Y/F, 10YF, 10YH, 113A, 113E/L, 113F, 113I/KP, 113I/M, 113I/V, 113K, 113K/225L, 113K/L, 113M, 113M/216, 113M/288, 113M/W, 113MG, 113Y, 113Y/K, 113YK, 115, 115C, 115M, 116, 116N, 116P, 117, 117L, 118D, 118G, 118H, 118P, 118W, 11W, 120, 120K, 120M, 123, 126/300, 126M, 127K, 127K/V, 127S, 127k, 136H, 136N, 142N, 147, 14X, 156, 156/118, 156/126, 156/216G, 156/288, 156/299, 156B/118, 156L, 156V, 158/277D, 158F, 158F/VSN, 158FV, 158J, 158J/L, 158JL, 158L, 158S, 158V/J, 158YF, 15H, 15H/242RG, 15IG, 16A, 16A-XC, 16A/219, 16A/226, 16A/49M, 16AK, 16AS, 16C, 16D, 16D-XC, 16R, 16V, 16X, 171, 171D, 171M, 171R, 171SM, 178F, 178G, 178K, 178N, 17D, 17H, 17H/10W, 17H/219, 17HN, 17HN/90LV, 17JJ, 17S, 17SP, 18/10K, 18/219, 185J, 186, 187, 187/224JN, 187B/C, 187D/V, 187K/V, 187P, 188/251, 188H, 189M, 189M/A, 189MQ, 18B, 18C, 18C/10J, 18C/250S, 18N, 18R, 18V, 190R, 193M, 193S, 194 H, 194B, 194V, 195W, 198, 199W, 19F, 19F/A, 19K, 19K/A, 19KA, 19KJ, 19M, 19S/F, 19V, 19YF, 1B, 1D, 1D/299, 1D/299S, 1D/J, 1D/V, 1J, 1L, 1M/D, 1P, 1P/25S, 1V, 1Z, 201, 201 G/I, 201G, 201K, 201M, 202T, 203A, 203A/M, 203A/R, 203K, 204, 204K, 204P, 204U, 205, 205A, 205B, 205M, 205T, 206, 208, 208C, 20P, 20P-SN, 20X/251, 21, 211A, 211B, 211C, 211D, 211DY, 211E, 211K, 211M, 211T, 211T/D, 211U, 212, 212/568, 212/570, 212/702, 216, 216K/L, 216L/W, 216LG, 216MG, 216T, 216W, 217, 217/300, 217A, 217D, 217M, 218, 218C, 218D, 218D/L, 218L, 218L/V, 219, 219-SU, 219/229, 21B, 21W, 220J, 220K, 220M, 220P, 220T, 220V, 222, 222A, 222L, 224/205F, 224G, 224HS, 224JW, 224K, 224X, 225, 225L, 226, 226L, 226M, 22D, 22K, 22M, 22R, 22T, 230A, 230D, 230P, 230V, 231K, 23B, 23GF-PC, 241T, 242, 242A, 242B, 242G, 242PG, 242RG, 245A, 245J, 24B, 24B-PC, 24B-PN, 24E, 24N, 24N/25S, 24S, 24S/219L, 24S/281E, 250/281, 250C, 251, 251AK, 251B, 251J SU, 251K, 251L, 251M SU, 251P, 251SS, 252, 252A, 252M, 252NP, 252S, 253G, 253H, 253K, 253KT, 253L, 253M, 253S, 253T, 254K, 254M, 25A, 25A/J, 25B, 25J, 25M, 25S, 26, 26G, 26M, 26N, 26S, 272, 272G, 277N, 279, 279-PA, 279R, 280, 280/564M, 280B, 280E, 280K, 280N, 280R, 280S, 280ST, 280T, 281, 283C, 283D, 283I, 283R, 283RG, 283S, 284P, 285 PS, 287N, 288, 288A, 288B, 288C, 288D, 288M, 288N/B, 288P, 288R, 288S, 288X, 288Y, 288Z, 289M, 290, 290/463, 290AF, 290F, 290K, 290K/P, 290U, 291, 293S, 299, 299H, 299M, 299P, 299S, 29B/272G, 29B/272I, 29B/272J, 29CD, 29R, 29S, 2C, 2J, 2U, 2Z, 3, 30, 30/280, 30/290U, 300, 300/126, 300/126M, 300/216, 30C, 30KPK, 30S, 316, 37, 37D, 38A, 38EX, 38S/47V, 38X, 3D, 3H, 3H/N, 3HN, 3K, 3K/252, 3KL, 3KN, 3M, 3Y, 40, 443N, 444, 445, 449, 44EX, 450, 451, 451M, 452, 453, 454, 455, 455M, 458, 45F, 45MF, 45RF, 460, 462, 463, 464, 47/224, 471, 471B, 471C, 472, 473D, 474, 475, 477, 478, 479, 47L, 47Y, 47YD, 47Z/C, 488, 49, 490, 490S, 491, 493, 494, 495G, 496, 497, 498KJ, 498VJ, 499, 49M, 49M/250, 49M/250C, 49M/251, 49M/T, 502, 503, 505, 50B, 51, 511B, 511D, 515, 516, 517, 518, 519, 520, 523K, 525, 526N/G, 527, 528, 530, 532, 532 HK, 535, 536, 537, 538, 540C, 540S, 543, 544, 545, 546, 567, 568, 571, 572, 572M, 573, 577, 578, 579, 580, 592, 593, 5G, 5GB, 5K, 5K/16A, 5K/16C, 5K/92, 5M-SN, 5MD, 5R, 5RW, 63M, 63S, 65, 65G, 65M, 65S, 65S/120, 66G, 67L, 67L/S, 67U, 6G, 6H/288, 6HN, 6LG, 6NG, 70, 70A, 71, 71A, 72H, 72J, 72JL, 72V, 73, 74, 75A, 78, 7Z, 8 R, 80C, 80G, 80S CO, 83J, 83J/171, 83J/272G, 85, 85H, 85Q, 85R, 85S, 85V, 85W, 85WS, 85j, 8A, 8A/16A, 8A/K, 8A/U, 8AU, 8C, 8J/M, 8JM, 8M, 8R, 90/251, 90D, 90D/47Y, 90K, 90K/253M, 90L, 90L/229, 90L/251, 90L/290F, 90N/U, 90R, 92, 92A, 92S, 94/95, 94C/70, 94K, 94R, 94U, 95K, 95M, 95N, 95P, 95R, 9E, 9F, 9K, 9K/272G, 9M, 9Q, 9X, 9X/171C, 9X/230X, 9Y/F, 9YFA, A1, A2, AMGL, B.PALLY-ATNR, B1, B2, BGLR, CHVL, CPL, CTPL, DVRM, ECIL X ROADS TO GHATKESAR, ECIL-AMALAPURAM, ECIL-GUNTUR, ECIL-RAZOLE, ECIL-TPT, GUNTUR, KADAPA, KNL, KP, M1, MARKAPURAM, MGBS-TIRUPATI, MP-NZB, MYP-GDK, MYP-KRMR, MYP-NZB, MYP-VJA, MYP-WGLL, NDL, NIZAMABAD-ARSP, NIZAMABAD-MBRK, ONGOLE, P1, PALAMANERU, PCPL, PULIVENDULA, THADIPATHRI, VIJAYAWADA, VJA, VJA-ATNR, WARANGAL, WARANGAL-BMRM, WARANGAL-PCM, YGT]

## 4.2. Service\_type

Need to figure out what all these abbreviations stand for.

|  |  |
| --- | --- |
| SERVICE\_TYPE | |
| AC |  |
| AE |  |
| CO |  |
| CS |  |
| CV |  |
| FS |  |
| IB |  |
| IH |  |
| IM |  |
| LA |  |
| LB |  |
| LN |  |
| MA |  |
| ME |  |
| MF |  |
| MO |  |
| MS |  |
| PA |  |
| PC |  |
| PM |  |
| PN |  |
| PO |  |
| PS |  |
| SA |  |
| SN |  |
| SU |  |
| VB |  |
| VC |  |
| VR |  |
| XC |  |
| XF |  |
| XG |  |
| XM |  |
| XS |  |

## 4.3. Depot codes

|  |  |
| --- | --- |
| depot\_code | |
| BGD |  |
| BHL |  |
| BKP |  |
| CCL |  |
| CNT |  |
| DSR |  |
| FMA |  |
| FNR |  |
| HCU |  |
| HN1 |  |
| HN2 |  |
| HPT |  |
| JDM |  |
| KCG |  |
| KGD |  |
| KPL |  |
| MDN |  |
| MHM |  |
| MI1 |  |
| MI2 |  |
| MPM |  |
| MS1 |  |
| MS2 |  |
| RJN |  |
| RN1 |  |
| RN2 |  |
| UPL |  |