Data Analysis of the MSRTC public bus transport system

By Ramya Sharma
Under Prof. Milind Sohoni

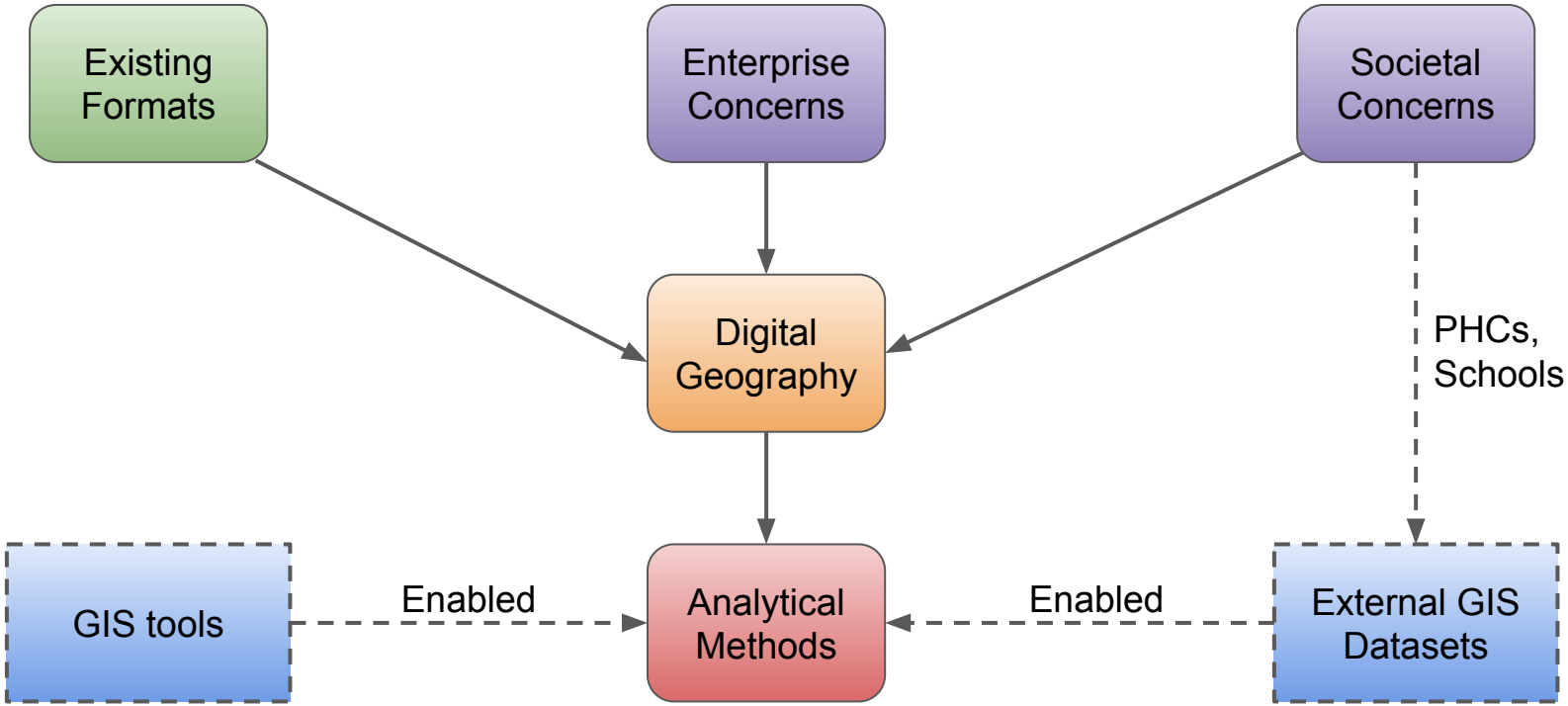
Index

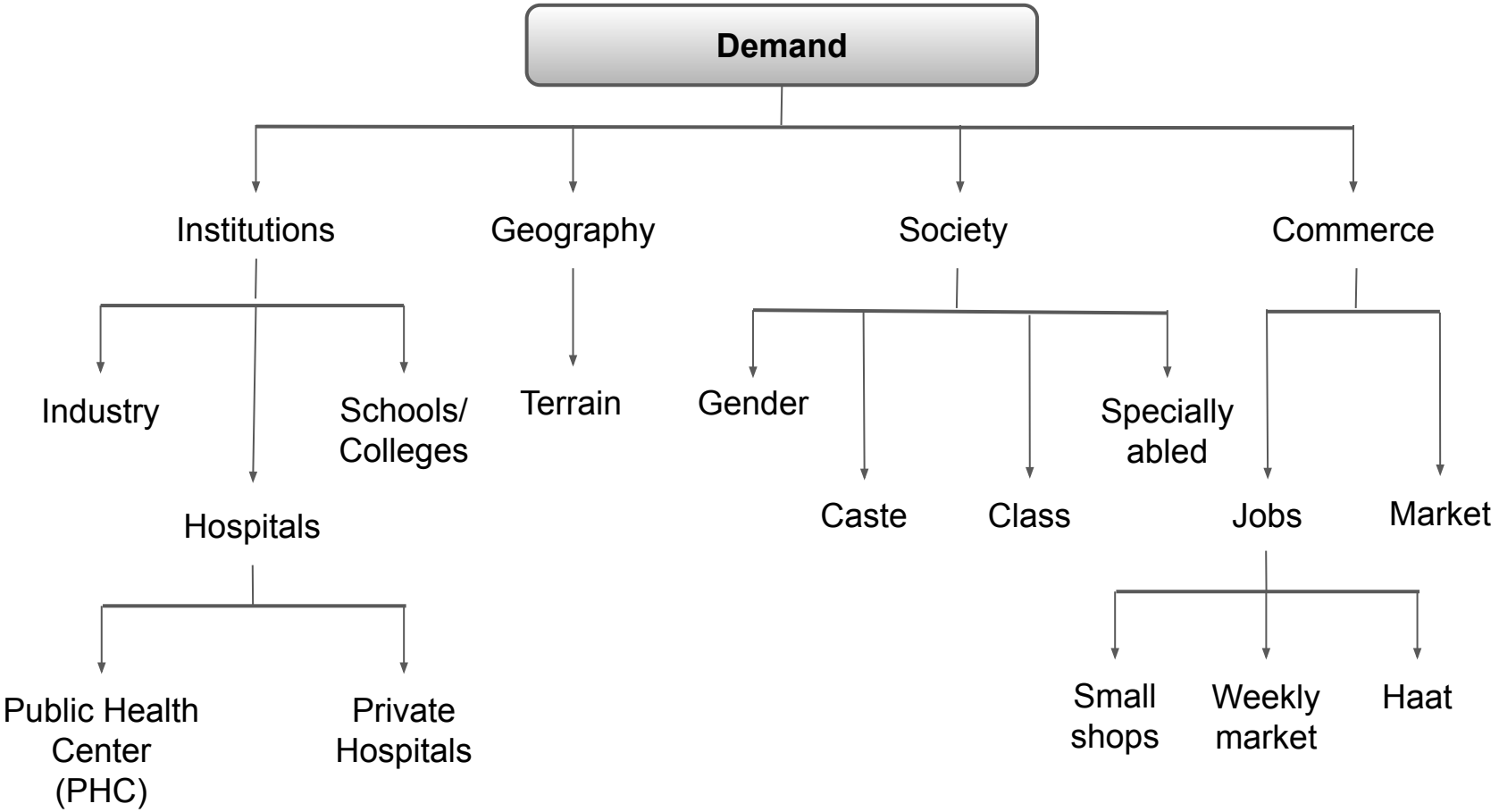
- Objectives of our research
- Earlier work and field work
- Extension of GIS part - successes and failures
- New analytical framework
- New GIS analysis for Covid-19 pandemic
- Conclusion

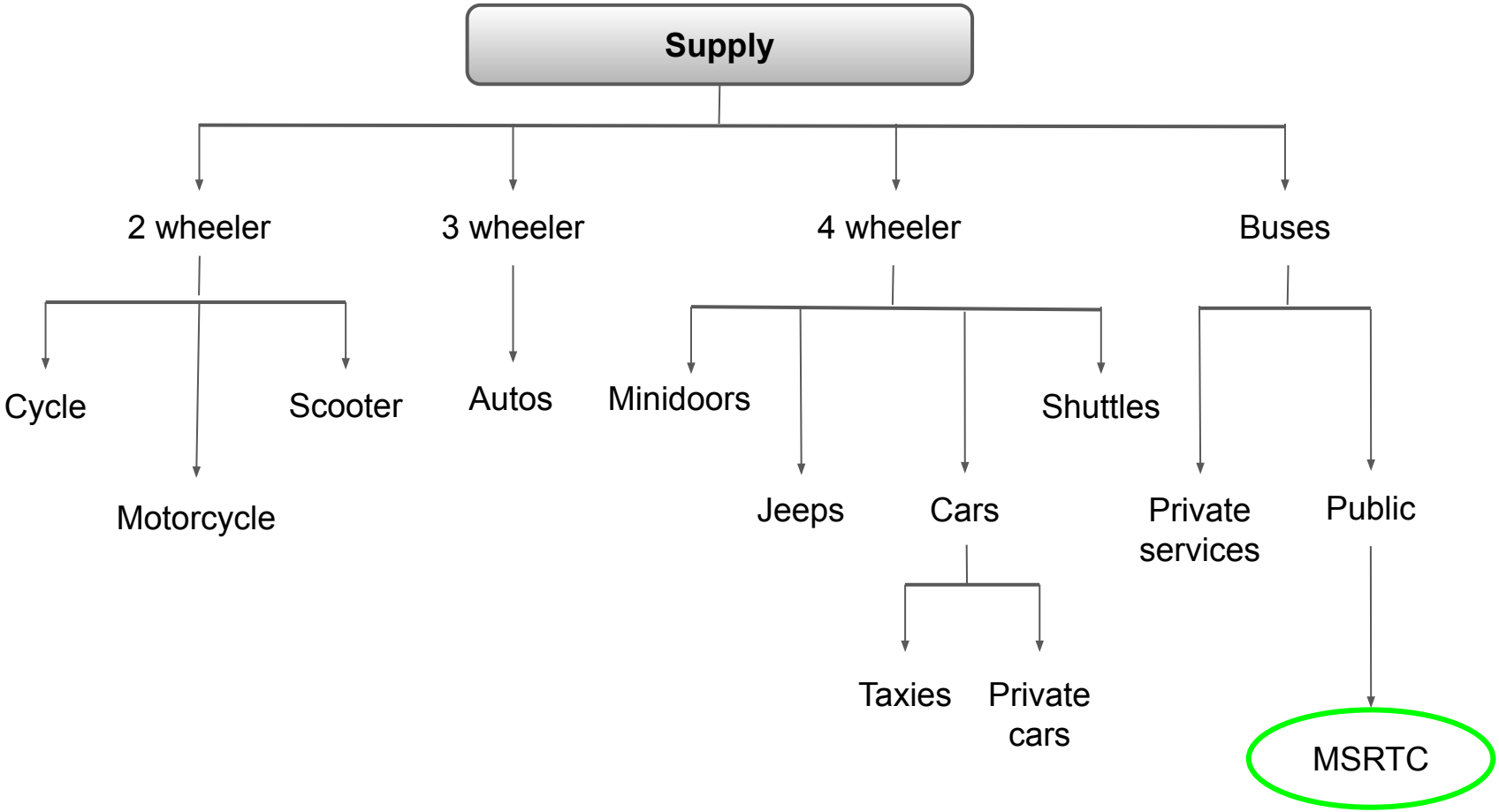
Objectives

- Framework for collecting the data:
  - Form4, Master data, ETIM data and ABC data are not coherent and there is no common denominator between them.
  - Integration of GIS with the route data will lead to data coherency along with locating other GIS features like schools, factories, markets and PHCs.
- Proposing analytical methods to answer certain concerns:
  - The current methodologies might not be enough to observe the route traffic data.
  - Introducing metrics like Ridership and Consistency will help the MSRTC and trimax to observe the traffic and keep the data coherent.

Objectives


Existing Formats
Enterprise Concerns
Societal Concerns
Digital Geography
PHCs, Schools
GIS tools
Enabled
Analytical Methods
Enabled
External GIS Datasets


Demand
Institutions
Geography
Society
Commerce
Industry
Schools/ Colleges
Hospitals
Terrain
Gender
Caste
Class
Specially abled
Jobs
Market
Public Health Center (PHC)
Private Hospitals
Small shops
Weekly market
Haat


Supply
2 wheeler
3 wheeler
4 wheeler
Buses
Cycle
Scooter
Motorcycle
Autos
Minidoors
Jeeps
Cars
Shuttles
Taxies
Private cars
Private services
Public
MSRTC

MSRTC

- Maharashtra State Road Transport Corporation:
  - Fleet strength - 18,449 buses
  - 8.7 million passengers daily
  - One of the main transport facilities for Rural Maharashtra

Earlier Work

- Field work
- Approximate generation of GIS data for Shahapur Taluka

Current work

- Creation of GIS based visual route information system for MSRTC
- Data Analysis of the current MSRTC system:
  - Ridership analysis
  - Consistency of the routes
  - Data inconsistencies
- Creation of a bus timetable for Covid-19
  - Connecting all villages to the PHCs allocated for covid-19 via public transport(MSRTC)

Field Work

- Field Visit Nashik (18-21 Feb 2020) :
  - **Objective:** To gather the technical requirements of the employees of MSRTC which might be useful for the extension of our project.
  - **Observations:**
    - All the work is done manually on excel worksheets.
    - Got the datasets like Form4 and Form 1, which we did not initially have for Sinnar.
    - Learnt how new trips are introduced and existing trips are stopped.
    - Money pits for a bus depot are: Obligatory trips and maintenance work.

Data Used

- We have used the following data provided for Sinnar and Shahapur talukas:
  a. Form4 and Master data: (Source: MSRTC)
     i. These two tables give out the planned schedule information, i.e. the scheduled buses, crew, order of bus stops, ticket pricing scheme, whether manav-vikas bus or not etc.
  b. ETIM data: (Source: TRIMAX)
     i. This data contains the real time information of bus en route, like the ETIM machine number, route number, trip number, ticket issue date and time, from and to stations, number of passengers, type of bus, inspections, tolls etc.
  c. ABC data: (Source: TRIMAX)
     i. This table is summary data of the trips for the entire month and contains information like money earned from the trip, trip departure time, percentage load factor, ABC status of the trip etc.

Objective: To produce GIS framework for the same.

Talukas

- Shahapur taluka (Thane District)
  - 6 towns and 225 villages
  - Major Local train stops: Kasara, Asangaon, Vashind, Atgaon
  - Number of buses: 56
- Sinnar taluka (Nashik District)
  - 6 towns and 129 villages
  - Number of buses: 72

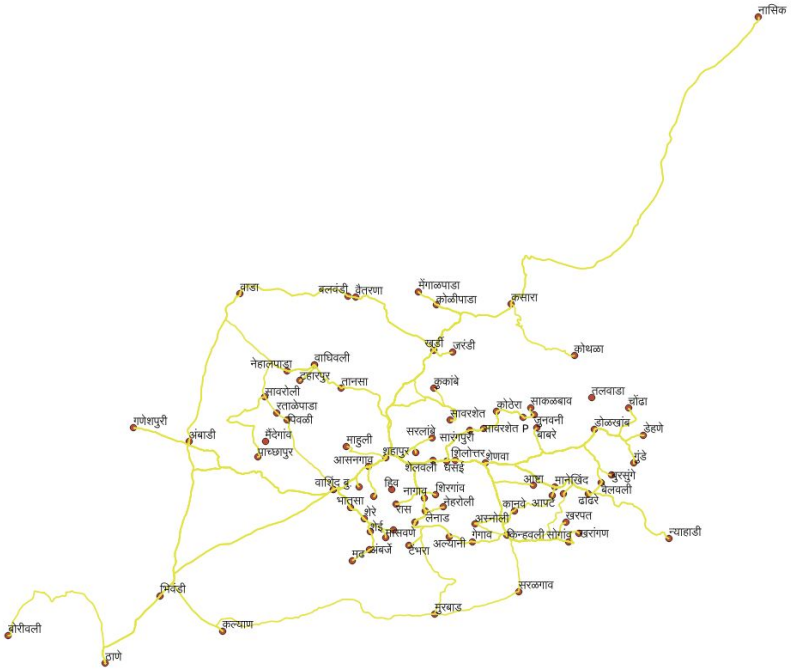
GIS vs GPS

- **GPS** - one of the many ways that are used to pinpoint an exact location on the earth's surface.
  - Mainly used in - Locating positions, Mapping and surveying.
- **GIS -** a computer program that is designed to capture, analyze, interpret and store data that has been transmitted from navigation systems such as GPS and make the information available for use.
  - Used in - Data mapping, Proximity analysis and Location analysis.

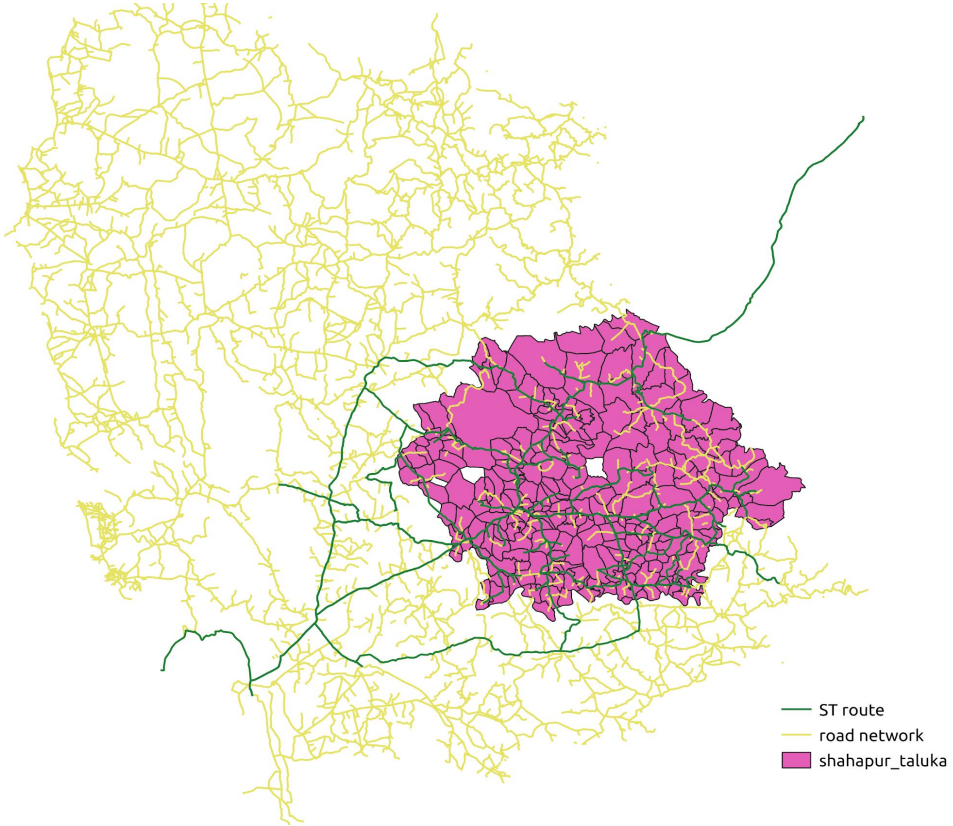
https://grindgis.com/gis/difference-between-gis-and-gps

Earlier work - Terminals

- The terminals were manually searched on Google API and noted down.



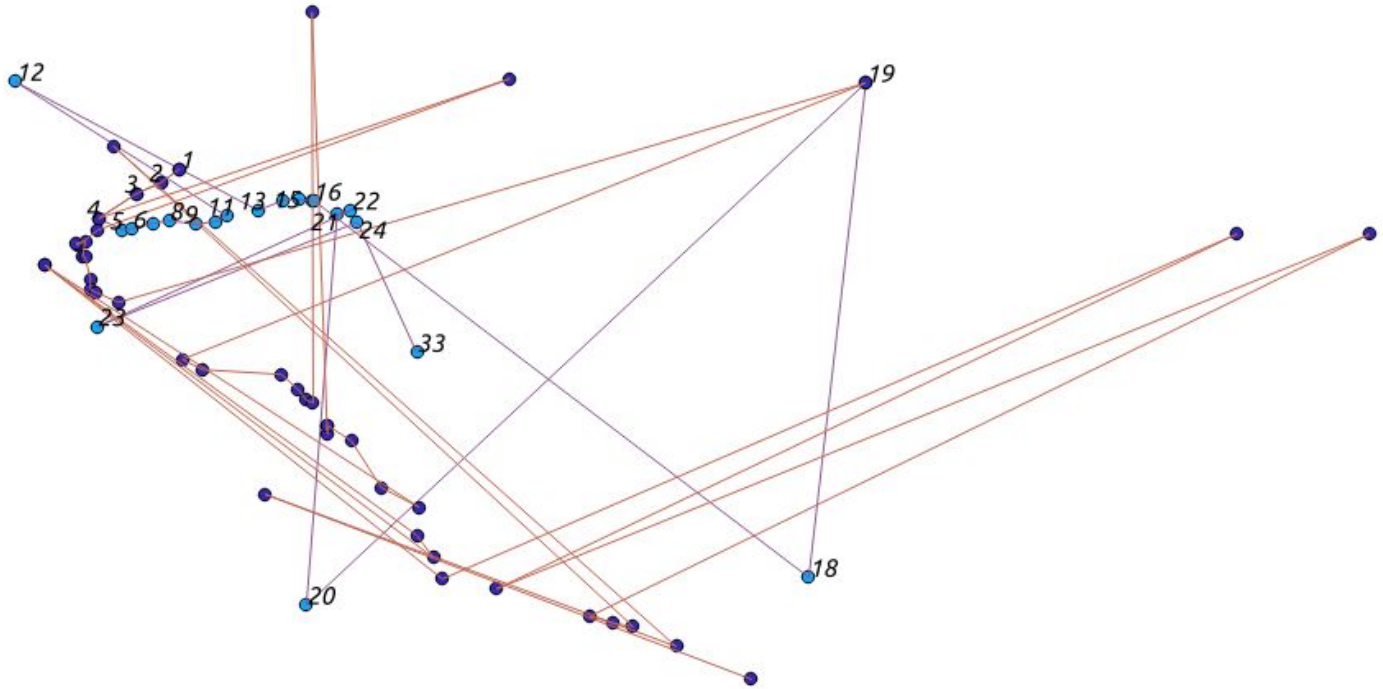
Earlier work - ST bus route in Shahapur



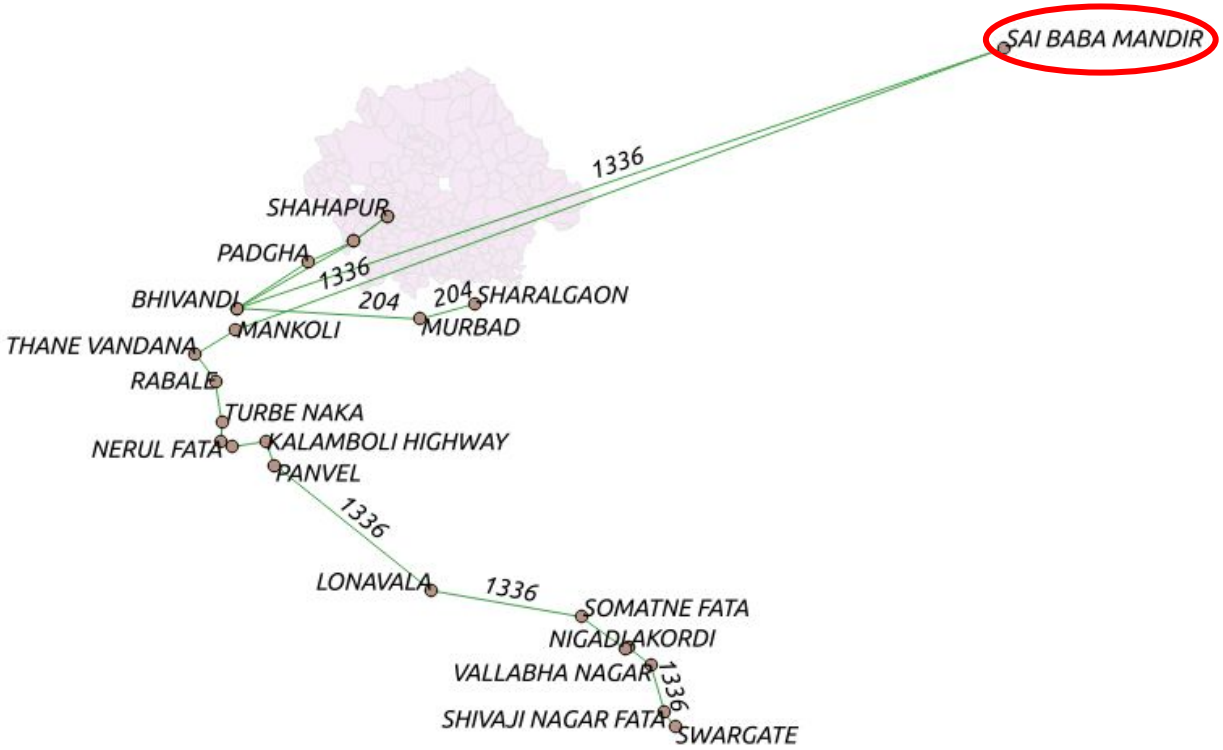
Current work - Bus stands

- When we tried to automate the search for bus stands, following errors were encountered:
  - Multiple bus stations had different code but same name for bus stand.
  - Bus stand names were ambiguous.
  - Even after geolocating the bus station, there was no proof that it would be the correct one.
- The work with bus stand was postponed to when the GIS data would be received from MSRTC.

Current work - Bus stands



Current work - Bus stands



Analysing ETIM data and ABC data

- **Objectives**:
  - Analysing consistency of the routes
  - Analysing ridership on routes
  - Analysing ticket issue patterns on routes
  - Analysing punctuality of trips

Consistency of the routes

- Consistent route:
  - A route is said to be consistent, if the tickets are issued for only those bus stops which are mentioned in the route master file.
- Inconsistent stops:
  - There might be some exceptional bus stops on a route which are not mentioned in the route master file, but tickets are issued on that stop.
- Consistency Checking:
  - Checks the tickets which were issued on the bus stops already stated in the master file.

Consistency of the routes

- Analysis(Shahapur):
  - We have a total of 265 routes for Shahapur for the month of July.
  - Out of these 265 routes, 66 routes are not consistent routes.
  - So, we conclude that 75% routes out of total routes are consistent routes.
  - The max count of the stops where tickets were issued in etim but not in the route master file is 5 stops for a route.

	A	B	D	E	F	G	H	I
1	route_no		no_bus_stops_etim			no_bus_stops_master	bus_stops_etim_not_in_master	no_bus_stops_etim_not_in_master
2	17527		19			20	{'SPDAE', 'CGHAPD', 'KMBAM', 'SLLIFT', 'KUGR'}	5
3	18344		15			18	{'SHEVA', 'MLGST', 'UMARAI', 'STHGN', 'SHST'}	5
4	65117		23			25	{'BVD', 'KBF', 'LALAFT', 'SBMD', 'RABP'}	5
5	3856		17			17	{'SLLIFT', 'CNWDI', 'DAAIFT', 'KUGR'}	4
6	3891		20			32	{'NSKCBS', 'KHRDI', 'PNDRGLT', 'BHGTWDT'}	4
7	13159		24			33	{'RAHD', 'BHGTWDT', 'PWRHOS', 'VINLI'}	4
8	107950		22			34	{'ATGO', 'ZDGA', 'VDLB', 'OZR'}	4
9	203		41			44	{'LONADPS', 'SWR', 'VBLN'}	3
10	3905		11			11	{'KOLIPDBT', 'UMEM', 'AJNP'}	3
11	16555		11			11	{'GLBHNBT', 'VACA', 'AHAMBT'}	3
12	17520		13			19	{'NSKCBS', 'WDVRE', 'GNDE'}	3
13	42722		21			21	{'OAGNN', 'ZDGA', 'VASD'}	3
14	44351		24			30	{'ARVIDD', 'SONNE', 'DHL'}	3
15	102169		27			44	{'KSBSST', 'ATGO', 'RAHD'}	3
16	1341		38			41	{'NGDH', 'NIRA'}	2
17	3606		5			11	{'RBP', 'SBBA'}	2

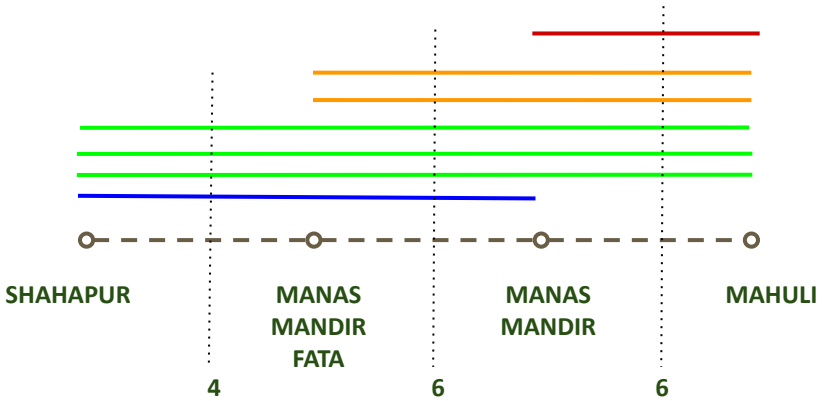
Consistency of the routes

- Analysis(Sinnar):
  - Out of the 228 routes, 19 routes are not consistent routes.
  - So, we conclude that 91% routes out of total routes are consistent routes.
  - The max count of the stops where tickets were issued in etim but not in the route master file is 15 stops for a route.

	A	D	F	G	H	I
1	route_no	no_bus_stops_etim		no_bus_stops_master	bus_stops_etim_not_in_master	no_bus_stops_etim_not_in_master
2	1470	22		85	{'SNGR', 'GHRO', 'DBD', 'NRYGN', 'BHSRI', 'NSKFT', 'NSKO', 'SNNR', 'NAEE', 'NWCBS', 'NSKOLD', 'ALPT', 'SMR', 'BOTA', 'RJGR'}	15
3	1143	22		26	{'MCR', 'GHRO', 'SGFTFT', 'NRYGN', 'NSKFT', 'NAEE', 'SNMNBP', 'NSKOLD', 'NSKRD', 'SMR', 'RJGR'}	11
4	77716	26		31	{'AJINRA', 'MEVN', 'CHOFS', 'KADIBD', 'ZGDEFT', 'NSKMRG'}	6
5	1511	22		22	{'LONIRA', 'KLHARAK'}	2
6	1514	21		22	{'LONIRA', 'KLHARAK'}	2
7	2725	30		30	{'MLANF', 'MSLSN'}	2
8	7116	12		13	{'DNASN', 'CHASSN'}	2
9	37252	11		12	{'BBAMNN', 'TKDSN'}	2
10	87222	29		30	{'NNDT', 'JMHEBRJ'}	2
11	1155	30		36	{'MRJN'}	1
12	2130	30		31	{'KUNW'}	1
13	2197	8		20	{'ITBS'}	1
14	7172	12		12	{'KWDRS'}	1
15	7174	14		14	{'PGLWDTNK'}	1
16	18701	18		18	{'SNNR'}	1
17	81293	31		30	{'SOTDABBE'}	1

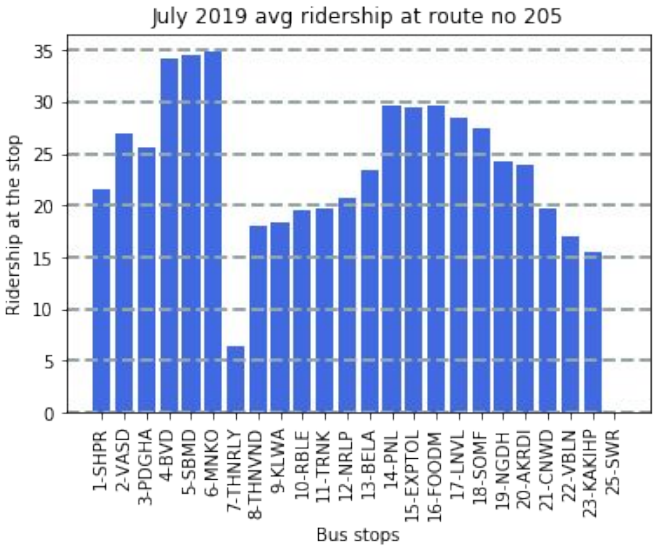
Analysing ridership on routes

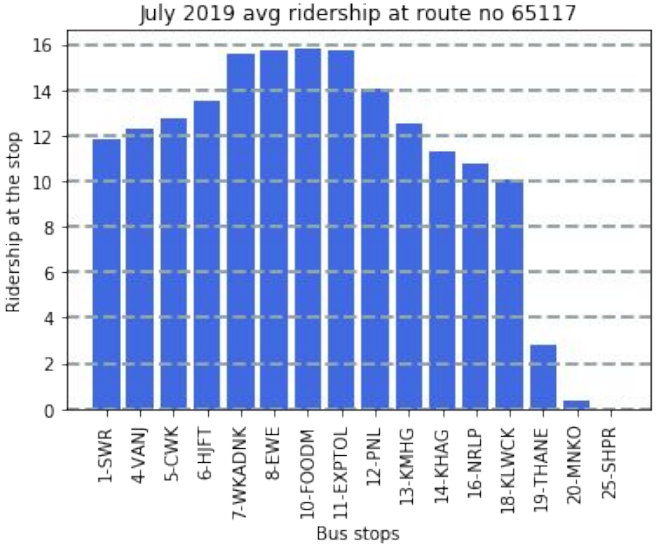
- Ridership:
  - Defines at a time how many passengers are travelling in a bus between consecutive stops.
  - For example, here is the Ridership on 3rd July, 2019 for a trip 00S81428 on route SHAHAPUR-MAHULI:



SHAHAPUR - SWARGATE [Route 205 - 65117]

- Average Ridership for the month of July, 2019:





Ridership analysis

- Weighted avg ridership - The weighted average ridership calculated with distance as weights across all the month.

$$ridership_{weighted-avg} = \Sigma(ridership_{ij} * distance_{ij}) / distance_{ij}$$

- Max ridership - the maximum ridership observed across the entire month.
- Std dev - The standard deviation across the weighted average ridership per day across the entire month.

Ridership analysis

- P1 ridership- Expected number of people (with 82% probability, mean + 1 std dev)according to Gaussian distribution) when compared to the seating capacity of the bus.
- P2 ridership - Expected number of people (with 97.5% probability, mean + 2 std dev) according to Gaussian distribution).
- **Utilization** - Categorization on the basis of what % of bus was occupied:
  - α (alpha) : P1 ridership>=80% capacity
  - β (beta) : 50%<= P1 ridership < 80%
  - γ (gamma) : P1 ridership < 50%

Ridership analysis

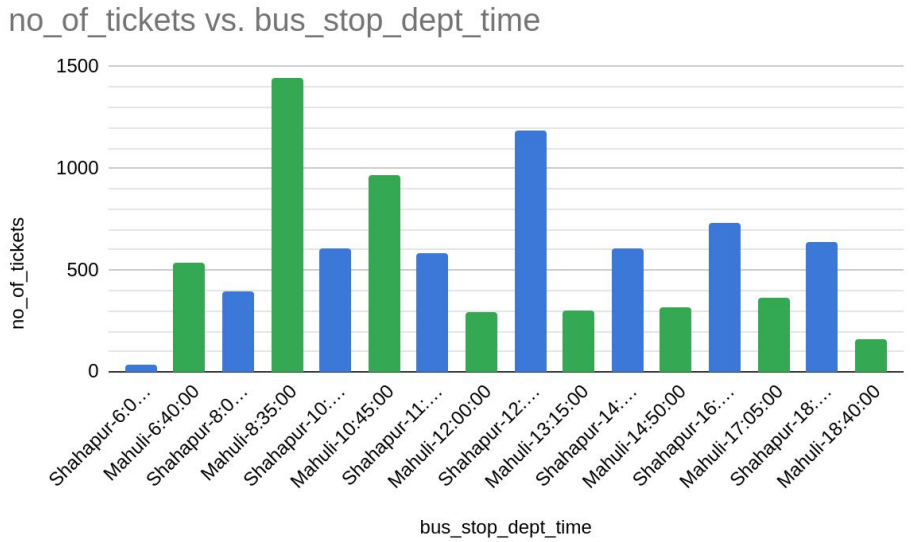
route_no	trip_no	max_ridership	weighted_avg_ridership	std_dev	P1 ridership	P2 ridership	dept_time	from_cd	till_cd	kilometer	abc_status	utilization
205	000M4971	61	18.19	6.60	24.79	31.38	8:50:00	SHAHAPUR	SWARGATE	212.1	B	β
65117	000M4972	49	10.36	6.55	16.92	23.47	17:00:00	SWARGATE	SHAHAPUR	211.1	C	γ
3906	00S81665	4	1.74	1.19	2.93	4.13	6:00:00	SHAHAPUR	MAHULI	9	C	γ
3906	00S81428	22	9.20	4.07	13.27	17.33	8:00:00	SHAHAPUR	MAHULI	9	C	γ
3906	00S81433	71	17.99	13.86	31.85	45.71	10:20:00	SHAHAPUR	MAHULI	9	C	β
3906	00S81435	33	17.89	7.49	25.38	32.86	11:20:00	SHAHAPUR	MAHULI	9	C	β
3906	00S81437	66	35.00	15.56	50.55	66.11	12:40:00	SHAHAPUR	MAHULI	9	B	α
3906	00S81439	51	26.20	11.47	37.67	49.14	14:15:00	SHAHAPUR	MAHULI	9	C	α
3906	00S81720	42	23.05	8.40	31.45	39.85	16:30:00	SHAHAPUR	MAHULI	9	C	β
3906	00S81448	33	19.01	7.82	26.83	34.65	18:05:00	SHAHAPUR	MAHULI	9	C	β
16567	00S81427	35	17.12	9.49	26.61	36.09	6:40:00	MAHULI	SHAHAPUR	9	C	β
16567	00S81429	87	44.30	17.21	61.51	78.72	8:35:00	MAHULI	SHAHAPUR	9	B	α
16567	00S81434	61	30.62	14.57	45.19	59.75	10:45:00	MAHULI	SHAHAPUR	9	B	α
16567	00S81436	29	9.70	7.46	17.16	24.62	12:00:00	MAHULI	SHAHAPUR	9	C	γ
16567	00S81438	20	9.60	5.01	14.61	19.62	13:15:00	MAHULI	SHAHAPUR	9	C	γ
16567	00S81440	32	13.64	6.98	20.62	27.61	14:50:00	MAHULI	SHAHAPUR	9	C	γ
16567	00S81446	67	12.12	17.62	29.74	47.36	17:05:00	MAHULI	SHAHAPUR	9	C	β
16567	00S81721	25	5.79	5.92	11.71	17.63	18:40:00	MAHULI	SHAHAPUR	9	C	γ

Ridership analysis

- We can see that out of 18 trips which ran across the month, 7 were categorized as gamma. Thus, mini buses can be run on these trips with low load.

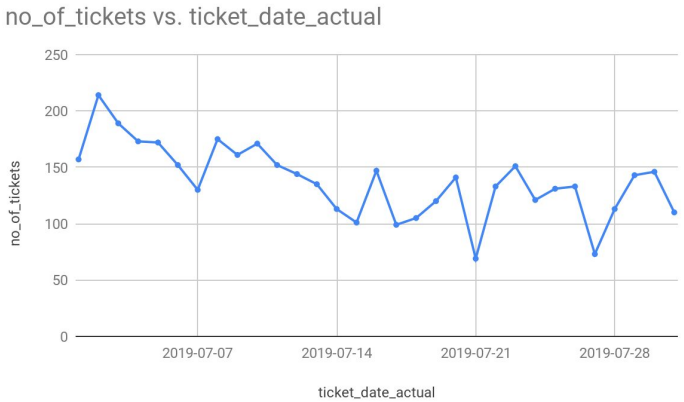
Ticket issue pattern

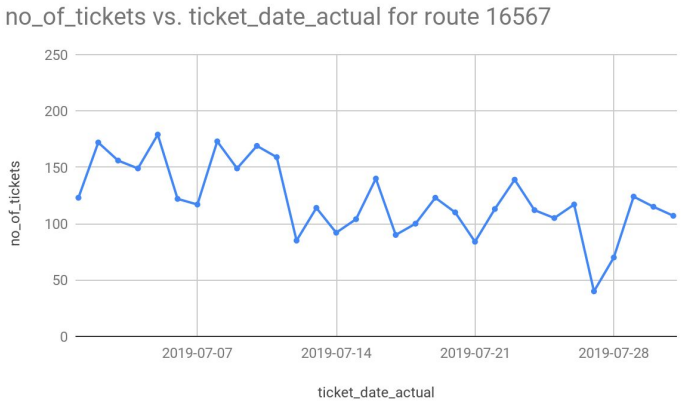
- Ticket issue pattern according to timings:
  - We observed the number of tickets issued per trip for an entire month and have displayed those tickets arranged by the trips' departure time (SHAHAPUR-MAHULI [Route 3906 - 16567]).



Ticket issue pattern

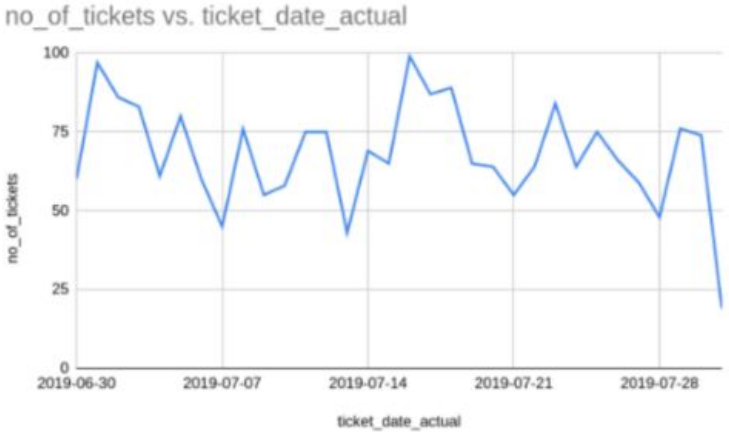
- Number of tickets issued on all the trips per day across the entire month:
  - Less number of tickets issued on weekends, thus we can reduce the number of trips on weekends (SHAHAPUR-MAHULI [Route 3906 - 16567]).

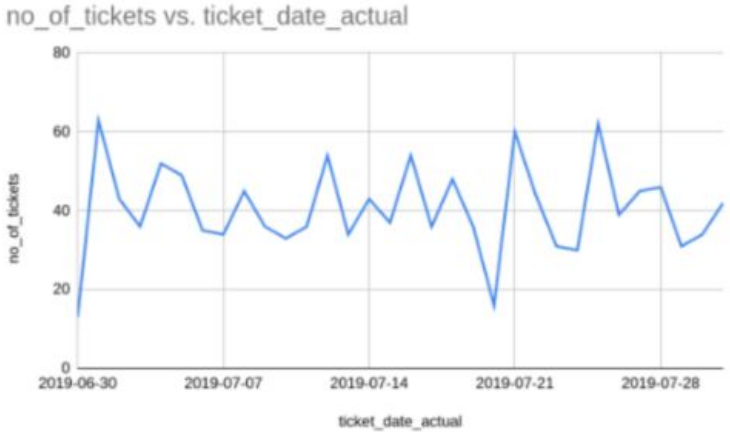




Ticket issue pattern

- Number of tickets issued on all the trips per day across the entire month:
  - Sinnar- Saikheda [route no 37263-37264]





Punctuality of the trips

- The trips can be classified as follows on the basis of the time difference between the first ticket issued and the scheduled time:
  - As Scheduled (0-20 minutes late)
  - Early (upto 1 hour early)
  - Late (20-60 minutes late)
- 68% of the Shahapur trips were as scheduled.

Summary Table for Punctuality: (July,2019)

	trip_status	no_of_etim_july_trip	percentage
1	**as scheduled(10-20mins)**	**2767**	**23.04**
2	**as scheduled(within 10mins)**	**5359**	**44.63**
3	cannot say late/early(9-12hrs)	63	0.52
4	cannot say late/early(more than 12hrs	25	0.2
5	**early**	**830**	**6.91**
6	early by 1-2hrs	118	0.98
7	early by 2-4hrs	58	0.48
8	early by 5-8hrs	125	1.04
9	**late**	**2254**	**18.77**
10	late by 1-2hrs	306	2.54
11	late by 2-4hrs	81	0.67
12	late by 5-8hrs	21	0.17
13	**total**	**12007**	
14			

~68% On time

Early by 20 mins to 1 hr

Late by 20 mins to 1 hr

Findings

- ABC data is not synchronous with ETIM data:
  - Operated number of trips in ETIM may not match the number of trips stated in ABC data.
  - The number of trips operated in ABC varies from what is observed in ETIM (eg 224 trips in ABC data for route 3906, but the total run trips (with extras) is 239).
  - The earnings from extra trips are not added in the ABC data, which leads to high amount if the route is long. (eg. for route 65117, ABC earnings are Rs 190692 , while ETIM earnings with extra trips are Rs 215127, a difference of Rs 24,435. Similarly, for route 205 ABC earnings are Rs 2,97,744 , while ETIM earnings with extra trips are Rs 3,25,220, a difference of Rs 27,476.) This difference remains unaccounted?

Findings

- ABC data is not synchronous with ETIM data
  - 5% of ABC trips are not represented in ETIM and 2.16% of the ETIM trips are not in the ABC data for Sinnar.
- ETIM data is not synchronous with the Master data.
  - 3.28% and 5.26% of the ETIM routes are not represented in the master data for Sinnar and Shahapur respectively.
- Form 4 data is not synchronous with any other data.

Findings

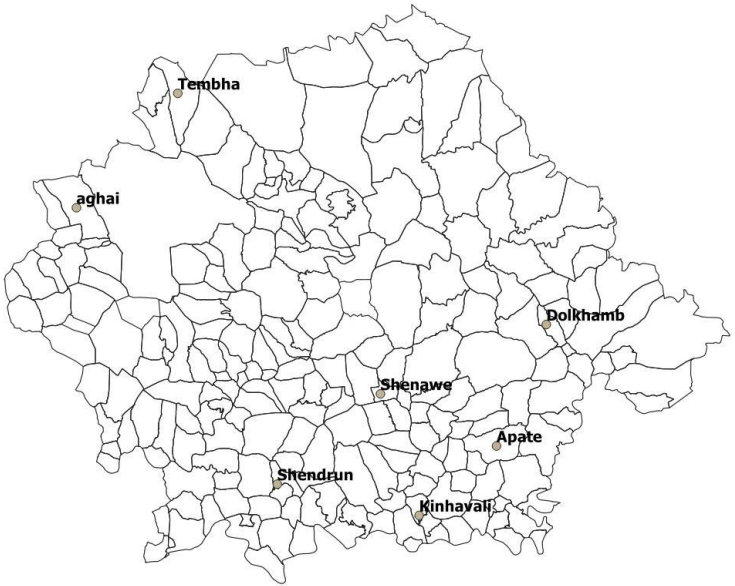
- ETIM data is not synchronous with the Master data:
  - For some routes, tickets are issued on the bus stops which are not present in the master route data. (For example, on route 3906, 887 passenger tickets (674 tickets for 887 passengers) are issued from or till Asangaon fata, which is not a Bus stand in the master data for this route, out of 6064 passenger tickets (4284 tickets for 6064 passengers). This means that 14.63% of the tickets are not consistent for this route and might be discarded.

Findings

- Form 4 data is not synchronous with any other data:
  - The Form 4 data can not be mapped to any other data table as the attributes do not match. The service id is different than given as trip numbers, the timings are different. Form4 data is basically obsolete if we want to use it for observational purposes.

Covid-19 work

- **Objective** - create a timetable for scheduling the buses so that the people can visit the nearest PHC in case of medical issue during COVID-19 breakout.



Covid work - Shahapur

- As we do not have the GIS data for Shahapur, the next best approach, string matching for bus stops, was selected.
- Basic requirements like load balancing were applied so not to overwhelm the PHCs.
- Only active routes, i.e., routes which were mentioned in ETIM, ABC and master data were considered.

Covid work - Shahapur

- Methodology 1 : assigning routes on the basis of villages covered
  - Efficiency : **9.35%**
  - Find out all the villages that need to be covered by each PHC.
  - Select the routes which have *PHC names* (name of the PHC village) in their bus stops. Select the route with the maximum coverage of the list of the villages allocated to each PHC.
  - This method is not efficient, as it covers only **9.35%** of the total bus stands(although 26.94% villages were covered). Hence, it can not be expected that a large number of villages were covered by this method.

Covid work - Shahapur

- Methodology 1 : assigning routes on the basis of villages covered
  - Efficiency : **9.35%**
- Methodology 2: assigning routes on the basis of bus stands covered
  - Efficiency: **36%**
- Methodology 3 : assigning routes on the basis of bus stands covered (using intersecting routes)
  - **1 hop coverage**
  - Problem with this data: majority of the intersecting routes which are selected are selected based on a single PHC (AGHAI).

Covid work - Shahapur

- Methodology 2: assigning routes on the basis of bus stands covered
  - From ETIM, find out all the list of bus stands that need to be covered (i.e. a ticket was issued on them)
  - A route is considered if it has a PHC village as a bus stop. The route matching percentage with the list of all the bus stops is checked. The route with the highest matching is selected on that basis.
  - Efficiency : **36%**

Covid work - Shahapur

- Methodology 3 : assigning routes on the basis of bus stands covered (using intersecting routes)
  - **1 hop coverage**
  - Select one main route for each PHC based on the PHC name (on the basis of maximum matching).
  - A route is considered if it has an intersecting bus stop with some route previously selected PHC route.
  - Problem with this data: majority of the intersecting routes which are selected are selected based on a single PHC (AGHAI).

Covid work - Shahapur

	route_no	from_cd	till_cd	percentage_match	phc_name	bus_stops_included	new_bus_stops	intersecting_bus_stops	intersecting_route
2	3865	ASANGAON	NEHADI	3.921569	AGHAI	TEMBHURLI, KHUTAL, NEHADI, SHIROSHI, PHANGLOSH...	10	TEMBHURLI	106896
3	3877	ASANGAON	GOKULGAON	0.784314	AGHAI	LENAD KHURD, GOKULGAON, BHATPADA	2	BHATPADA	124521
4	3878	ASANGAON	NEHAROLI	0.392157	AGHAI	BHATPADA, NEHAROLI	1	BHATPADA	124521
5	3891	SHAHAPUR	NASIK CBS	0.784314	AGHAI	KHAMBALE, JAWHAR FATA, TAKET FATA, ATGAON RLY....	2	PUNADE	82558
6	13159	NASIK CBS	SHAHAPUR	6.666667	AGHAI	KHAMBALE, KHARDI, JAWHAR FATA, TAKET FATA, ATG...	17	PUNADE	82558
8	16570	DHADRE	SHAHAPUR	0.392157	AGHAI	BHINGURALE, TEMBHURLI, MANJRA, BELAVALI	1	TEMBHURLI, MANJRA, BELAVALI	106896
9	16793	SHAHAPUR	WADA	4.705882	AGHAI	BELWAD, SONALA, PUNADE, AMBIVALI, ATGAON RLY. ...	12	PUNADE	82558

Covid work - Shahapur

- Methodology 3 : assigning routes on the basis of bus stands covered (using intersecting routes)
  - Also, for 90% coverage, >50 buses are required.
  - While trying to create a timetable and schedule buses, this is creating problems. We are not able to implement load balancing and scheduling a bus every 5 minutes might lead to inconsistencies. Even after being thoroughly checked, this solution does not seem correct.
  - The only solution seemed like is using a **GIS based solution for selection of routes** (performed on sinnar) instead of the textual data.

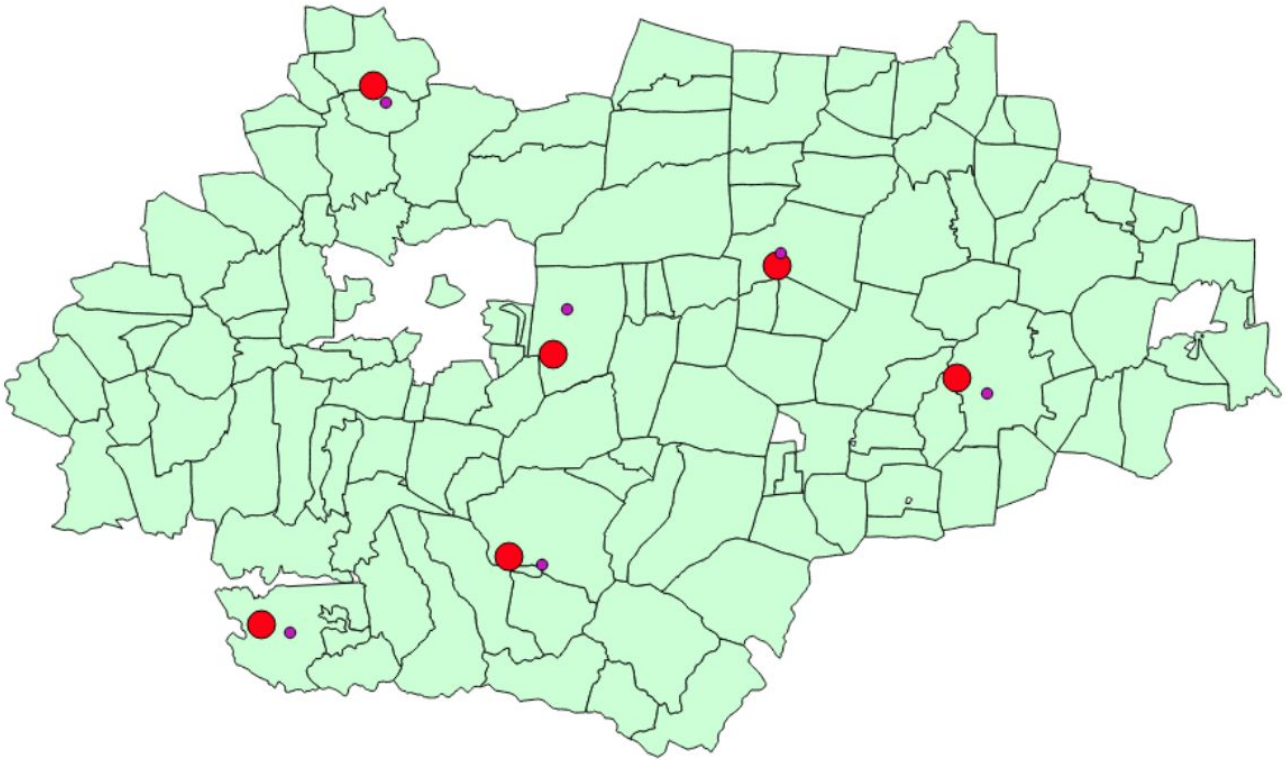
Covid work - Shahapur

- Conclusion:
  - Due to lack of proper locations and GIS data, generating a consistent map or timetable for covid was difficult.
  - Implementing important factors like load balancing and scheduling was generating many errors.
  - Hence, the need for a GIS enabled technology for the bus system is a necessity for the MSRTC system.

Covid work - GIS data for Sinnar

- **1 hop coverage**
  - Take the calculated GIS data (nearest bus stop to the PHC) and the master data.
  - Select the required bus stand % cover for the taluka.
  - Find the routes on which a PHC lies, call these routes the main routes. Assign times to these routes.
  - Define the intersection points.
  - On these intersection points schedule the buses as such that they reach 30 minutes before (for layover) the PHC going bus reaches that stop.
  - Create a time schedule based on these constraints.

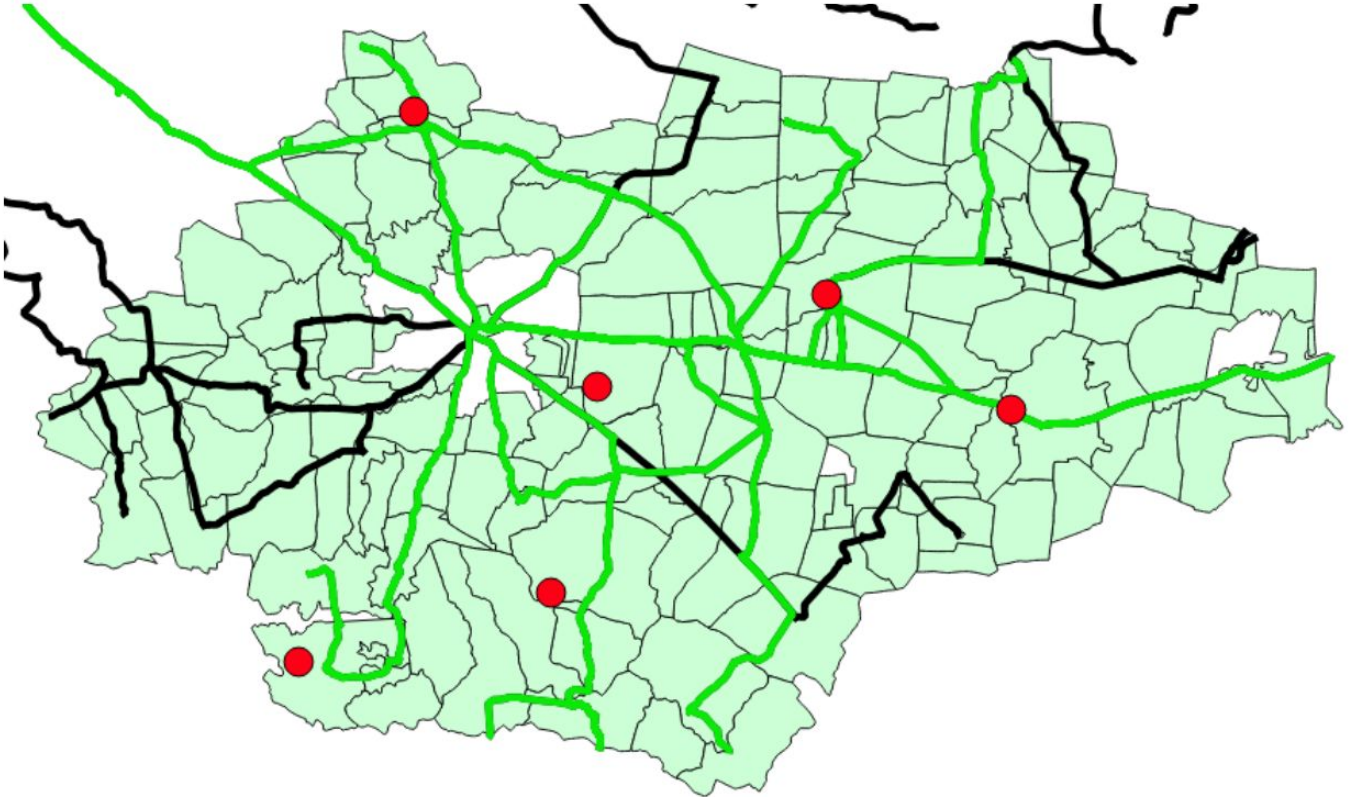
Covid work - GIS data



Covid work - GIS data

PHC Village	Nearest Bus Stop Code
Wavi	VAVIK
Dapur	DULWIA
Thangaon	TTAGSN
Pandhurli	MUSON
Naigaon	JGN
Deopur	DOPUIA

Covid work - GIS data



Conclusion

- We tried to understand the working of Taluka bus-depot and service provisioning in Shahapur Taluka.
- We tried to analyze the various datasets and gaps in them.
- We tried to create a timetable for MSRTC in times of Covid-19 for patients to visit PHCs.

Questions?

Thank you