Acknowledgment

It gives me great pleasure to present this paper, which is a requirement for the Computer Science and Engineering course CS-798 and is titled "**Analyzing and Standardizing MSRTC bus depot data**."

My deepest thanks go out to my mentor **Prof. Milind A. Sohoni** for mentoring me and assisting me in better understanding the subject. All the meetings and discussions were highly helpful in gaining critical insights into the work. A special thanks to Prof. Jitendra Shah for his valuable suggestions and helpful discussions throughout the MTP work.

Also, a special thanks to **Mr. Rahul Gokhale**, **Mr. Asim,** and **Mr. Shubhranshu Maurya** for their help throughout my project work.

Avijit Kumar Dash

IIT Bombay

November 2022

Abstract

In this report we look at rural public transport as a development service and ways of representing it, analyzing it, and doing its social accounting. In this work, we tried to improve transportation efficiency and ways of representing it. We have focused on Shahapur and Sinnar taluka in our study. We look at the many data formats that have been received and try to figure out how they are related. Maharashtra State Road Transportation Corporation (MSRTC) is the primary provider of rural transportation services. We will go through the most important data formats and how they're related. To depict the key data set, Form-4, we employed graph theory as a tool. This is referred to as digital geography. We have tried to find out how data is related to each other as well as if there is any mismatch among them. Besides this, we tried to formulate and examine if it is possible to express our data in GTFS format. Finally, some suggestions for the extra problems identified in the field are provided in this report.

Stage-I Summary

Shahapur Taluka

Shahapur is one of the seven talukas of the Thane district. It operates around 65 buses on 270 routes for approximately 80 villages in Taluka.

Fig: Google Street Map 3D image of Shahapur Depot

ST Route Networks:

Fig: Road Network Thane Division

Bus Stops:

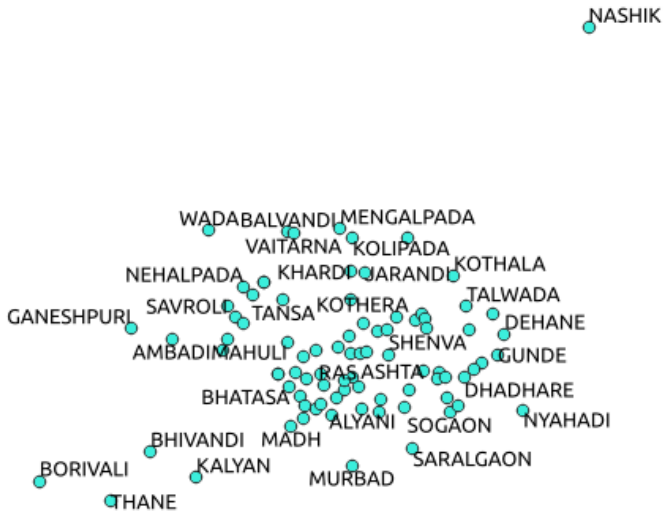


Fig: Stop Location

Shahapur Taluka Polygon:

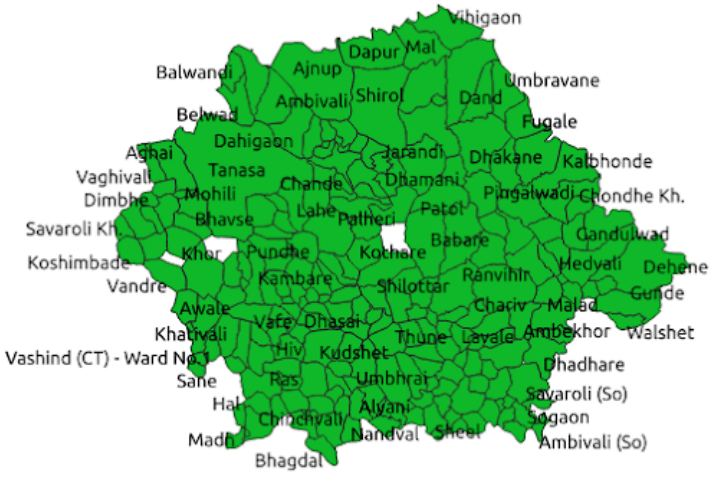


Fig: Shahapur Taluka boundary (census polygon data)

Route network for Shahapur taluka:

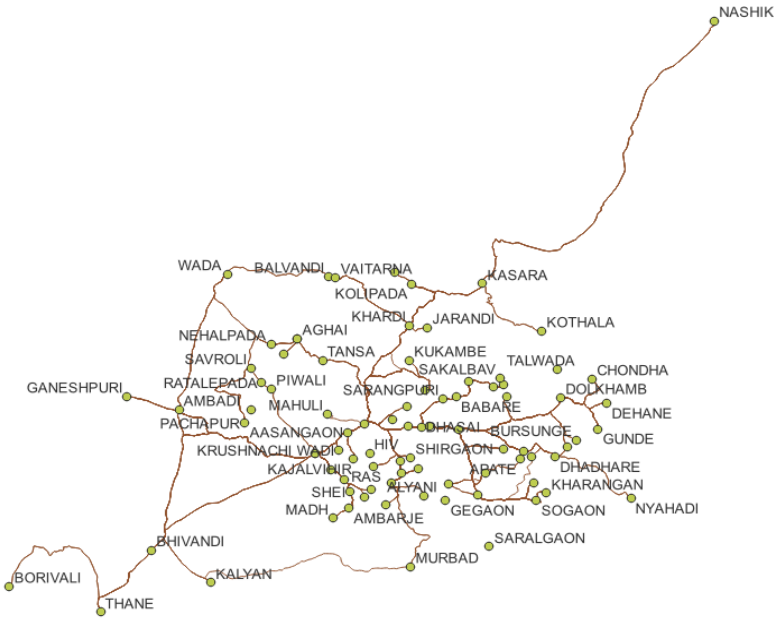


Fig: Generated Complete Shapefile of the route segments

We employed a network analysis processing toolkit to develop the road segment routes, and we mainly produced the digital geography of the route network using the Shortest path (point-to-point) analysis approach.

Trips passing through any particular segments

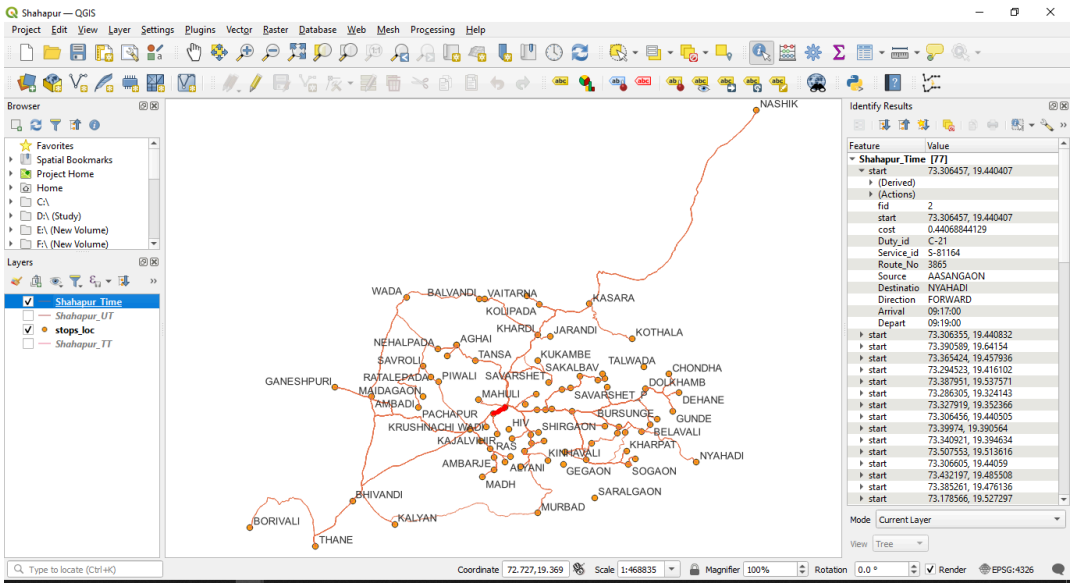


Fig: Trips passing through the selected route segment

The lines in the figure with orange colors represent the Shahapur Route Network. The red line indicates a specific route segment about which we want to find out which bus will come and when. All details linked to the selected route segment are provided in the right panel under "Identify Features." Simply click on the line displayed on the map to select a route segment.

Introduction

The current project endeavors to replicate and improve upon the work of Mr. Sudanshu Kulkarni and Anshul Goel's public transportation work in the Shahapur taluka of Thane district and Sinnar Taluka in Nashik District, Maharashtra. The project concentrates on public transportation and considers it as a developmental service. The methods of reproducing, analyzing, and improving the existing tools and techniques employed in a bus depot so that core-level administrators can make better and improved judgments in public transportation as a development service. In that case, numerous analyses have been done on various datasets such as form-4, ticketing (ETIM) data, route (Master) data, stop location data, geographical information, and related data.

Scope of the Project:

In this project, we are restricted to bus transportation in Shahapur Taluka and Sinnar Taluka and related data formats used in bus terminals. This project covered Shahapur taluka and Sinnar Taluka and surrounding talukas if and only if a bus service is available to the other taluka from Shahapur and Sinnar.

Chapter Organization

******** Chapter 1 summarizes the stage-i project. Chapter 2 is the introduction to the project. How datasets can be standardized describe in chapter 3. Shahapur data analysis represent in chapter 4. The digital geography of Shahapur firefly narrates in chapter 5. Chapter 6 is all about Sinnar taluka. Depict of Sinnar Digital geography and a description of all the datasets are in chapter 7. Chapter 8 is all about Sinnar taluka data analysis. What are the issues with datasets are described in chapter 9. The conclusion is in chapter 10 and chapter 11 is about future works.**

Research Question(s)

1. How can data structures be used to represent a rural public transportation system?
2. What are the applications of the proposed GIS-based framework (Digital geography)?
3. How is it beneficial in transportation?
4. Dataset needed to develop the framework: –

a. Shahapur census data and thane district road network data.
b. Shahapur taluka Form 4, Master data, ETIM data, and Terminals location data.
c. Sinnar taluka census data and thane district road network data.
d. Sinnar taluka Form 4, Master data, ETIM data, and Terminals location data.

Objective

1. To create Digital geography i.e., proposed GIS-based framework for Sinnar Taluka, Nashik, Maharashtra, India
2. To map Form 4 data in a GIS.
3. To enable social benefits accounting of services offered by Shahapur and Sinnar Taluka Bus depot.

Methodology

1. Basic analysis of the Development of Digital Geography of Shahapur Taluka, Maharashtra.

Standardizing Dataset

Standardizing Dataset

New approaches to data representation, visualization, storage, and querying have been developed in response to the requirement to handle data on the Web more easily. The concepts behind open data, linked data, and the semantic web have considerably increased the transmission of information and data. The idea of "open data" refers to the idea that as data created by governments, their institutions, and other public bodies are, by definition, public, it should be made available in an open, raw, and machine-readable format. Then, using applications that maximize the value of the data, it is feasible to utilize, reuse, publish, and distribute this data [1]. GTFS makes it possible to create a wide range of practical applications using and integrating public datasets. To represent our data in a standard way, we can use the GTFS [2] format. However, the GTFS format of the data obligates data to be represented in strictly formed CSV files.

GTFS (General Transit Feed Specification)
GTFS is a format for standardizing datasets for road networks. It started in 2005, and Google introduced it. GTFS was founded in 2005 and has grown to be a crucial component of the transit data environment, making it worthwhile to more people than only Google and transit agencies. At first, it was called *Google Transit Feed Specification* Because of Google. A few years later, in 2009, they presented it as *General Transit Feed Specification,* the standard format for analyzing road data. Since the adoption of GTFS has grown, companies like TransLoc have harnessed that data to create cutting-edge mobile applications that enable transit agencies to provide better service to their customers.
GTFS feeds require the following fields to be included [3]:

- **agency.txt:** The agency that provides the data in this feed contains information about one or more transit agencies.
- **routes.txt:** It contains information about the routes of the transit agency and these routes available to riders within a single service.
- **trips.txt:** A sequence of two or more stops that occurs at a particular time.
- **stops_times.txt:** The specific times that a vehicle arrives and departs from a stop location.
- **stops.txt:** The individual locations where vehicles pick up and drop off passengers and information about all the stops.
- **calendar.txt:** A schedule of when the service is available.

It also has some optional tables; those are

- calendar_dates.txt

- fare_attributes.txt
- fare_rules.txt
- shapes.txt
- frequencies.txt
- transfers.txt
- feed_info.txt

Data format for GTFS:

agency
agency_name
agency_url
agency_timezone

routes	
Primary Key	route_id
Other attributes	route_short_name
	route_long_name
	route_type

trips	
Primary Key	trip_id
Foreign key	route_id
	service_id
Optional attributes	block_id

stop_times

Primary Key	stop_id
Foreign key	trip_id
Other attributes	arrival_time
	departure_time
	stop_sequence

stops	
Primary Key	stop_id
Other attributes	stop_name
	stop_lon
	stop_lat

calendar	
Primary Key	service_id
Other attributes	Monday
	Tuesday
	Wednesday
	Thursday
	Friday
	Saturday
	Sunday
	start_date
	end_date

Available dataset and their attributes for standardizing dataset according to GTFS:

agency	Information
agency_name	MSRTC
agency_url	https://msrtc.maharashtra.gov.in/
agency_timezone	(GMT+5:30)

routes		Table that contains information
Primary Key	route_id	Route No is available in the master set.
Other attributes	route_short_name	This information is available in the master dataset as well.
	route_long_name	Similarly, this information is also available in the master dataset.
	route_type	Route type is available in the master set.

Trips		Table that contains information
Primary Key	trip_id	ETIM has this information
Foreign key	route_id	ETIM and master dataset have this information
	service_id	The service_id weekend service is listed
Optional attributes	block_id	Not available.

stop_times		Table that contains information
Primary Key	stop_id	We can provide every stop with an id.
Foreign key	trip_id	We can get information from ETIM
Other attributes	arrival_time	From-IV has this information
	departure_time	From-IV has this information
	stop_sequence	This sequence we can achieve from master data

stops		**Table that contains information**
Primary Key	stop_id	We can provide every stop with an id.
Other attributes	stop_name	This information is available on stop location data
	stop_lon	We can get this information from stop location data
	stop_lat	This information is also available to stop location data

calendar		
Primary Key	service_id	The service_id weekend service is listed
Other attributes	Monday	Instead of a day, the date is known to the ETIM dataset.
	Tuesday	
	Wednesday	
	Thursday	
	Friday	
	Saturday	
	Sunday	
	start_date	This information can be extracted from ETIM.
	end_date	This information is also available on ETIM.

Though we can extract much information using form-4, ETIM, and Master data, there are also some issues, so making our datasets in GTFS format is challenging. Like some route numbers are available in the master dataset that is not included in the ETIM also vice versa. The ticketing date is available, but the day is not mentioned, so we have converted every date to the day format. The service number is in the form-iv, but it is tough to map it to ETIM data.

stop_times dataset required fields are stop_id (primary key), trip_id (foreign key), arrival_time, departure_time, and stop_sequence, and we have all the available fields, but what makes it challenging to standardize is these fields we can get from three different datasets, and it is hard to map with one another. So it is like an impossible task to do.

GTFS Tables Mapping with our data

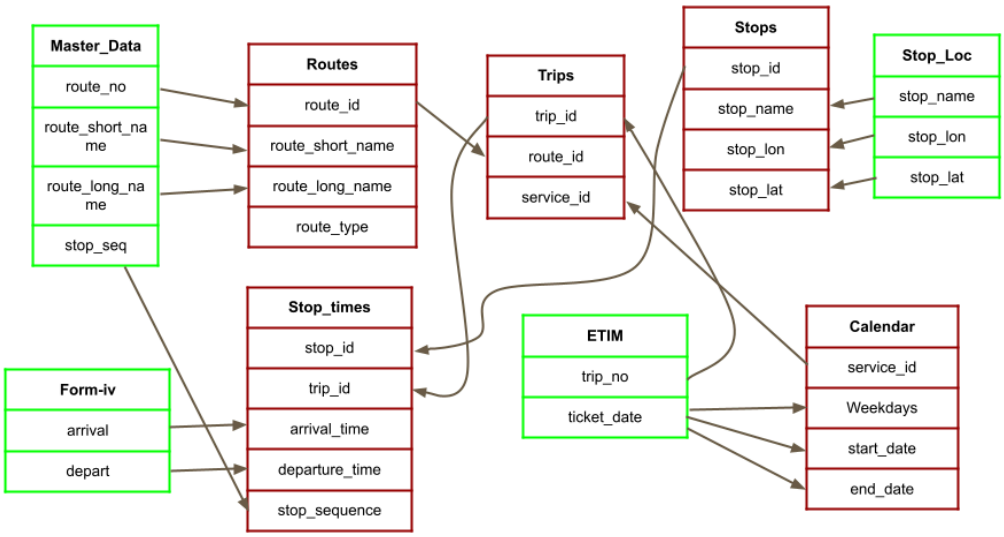
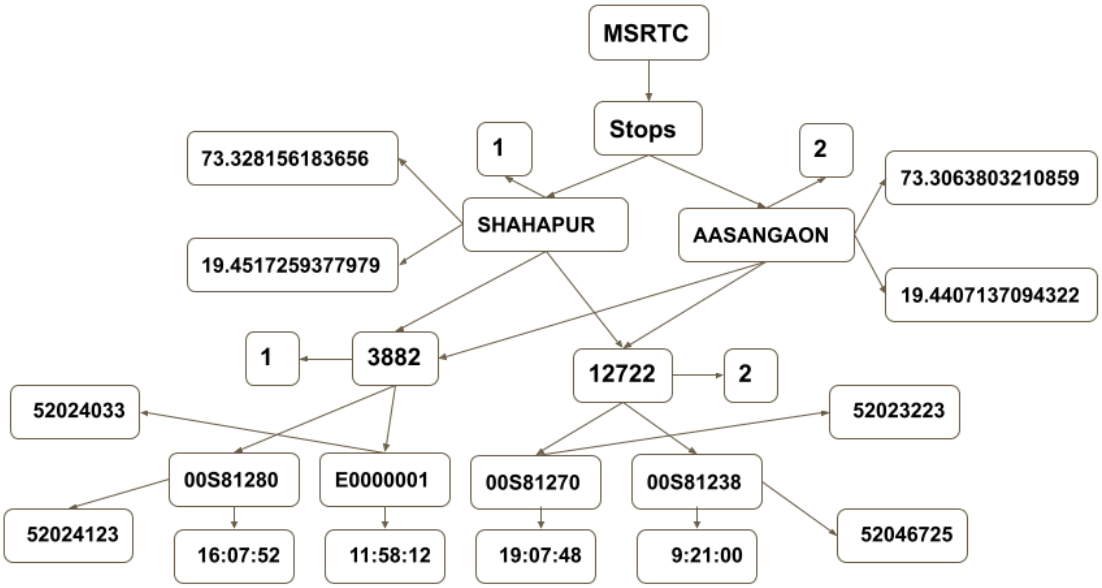


Fig: GTFS Tables Mapping with our data

service_id: Identifies a set of dates when service is available for one or more routes. This one is a little bit tricky to extract from our data. We need to group trip no to get a particular day and from the day we can get which day it is.



Shahapur Data Analysis

Traffic Analysis

A road traffic survey counts the number of cars or people that travel along geographically separated sites, routes, highways, trails, or crossroads during a set period of time using either manual or automatic observation. Traffic studies have a variety of applications, including helping to address national, international, and local traffic issues. Particularly, the information gathered is crucial in directing the choice-making process for transportation planning.
Data plays an important role in traffic analysis. Data can accurately depict the local traffic condition. There is numerous information that traffic analysis gathers, in addition to the number of vehicles utilizing a road or the travel time. In the case of MSRTC traffic analysis, it can be done using ETIM data and can generate numerous information like bus condition, profit generation, bus schedule analysis, etc. This analysis can also be done to make them profitable based on important locations like tourist places, and economic places if available nearby.

Traffic analysis according to route no:

According to Shahapur ETIM data, 274 routes are there. Every month many passengers travel those routes. The highest number of passenger-bought tickets for a route was 24070, and the

route number was 12708, but there are two such routes where only one ticket was sold, which are 123395 and 48947.

	route_no	traffic_count
3	12699	4604
4	12702	1582
5	12704	4147
6	12706	1069
7	12708	24070
8	12712	1802
9	12713	3370
0	12714	3046
1	127154	55
2	12717	1335
3	12719	172
4	12720	2097
5	12721	2486
6	12722	3454

Fig: Route-wise traffic flow-1

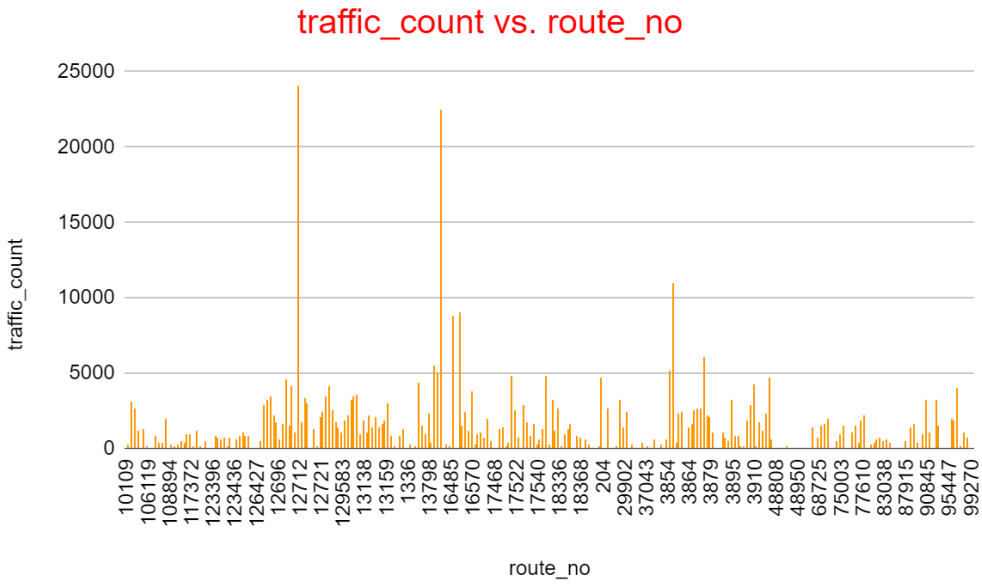


Fig: Route-wise traffic flow visualization

Issues with traffic analysis according to route no:

There are few such routes where less than ten tickets were sold for a month. This is a more significant issue because if a few number tickets were sold on those routes, those are loss-making

routes. If it was approximately ten par days, then it was supposed to be near about 300 tickets per month, then those routes may not be faced a considerable loss.

route_no	traffic_count
108863	7
119363	5
123395	1
126402	7
17525	9
3862	2
3883	3
48945	2
48947	1
48955	2
72365	2
85210	8
87915	5
99270	9

Fig: Routes that are abnormal

Traffic analysis according to source-destination:

There mentioned 4429 source-to-destination roads in ETIM data. From all those Shahapur (SHPR) to Kinnavali (KINV), 10395 tickets were sold for a month, the most sold for a source destination. So these are the most profitable roads according to ETIM data. This is the only road where more than ten thousand tickets were sold. However, other than these, there are a few roads where more than three thousand tickets were sold.

A	B	C
from_stage_cod	till_stage_code	count
SHPR	TLVLIL	10395
RSGY	HVAPB	8405
CNWD	BVD	8279
ADB	KNTIR	7021
DSIVL	NSHD	6679
SHST	STTRST	5347
GRWDA	MSVLS	4102
WLE	PHL	4025
SHIHI	VLFTL	3968
KLEAST	CGHAPD	3592
STTRST	SPGNST	3516
SONALAWT	TMBHA	3472
BHTSIST	VASD	3325
VEALI	TTHR	3268
NDI	SHEVA	3234
PDGHA	SHPR	3188
CNHAO	CGHAPD	3117

Fig: Ticket sold according to the source to destination

Apart from these, there are some source-to-destination roads where less than ten tickets were sold for a month. These roads may be intermediate, so they sold tickets this way.

A	B	C
from_stage_	till_stage_co	count
ABAKAR	MUSDI	1
ABAKAR	KUGR	2
ABAKAR	MSVLS	2
ABAKAR	KRNC	2
ABAKAR	GRWDA	5
ABAKAR	SHPR	8
AC	THNVND	3
ADB	VJWR	1
ADB	SHPR	1
ADB	KAIVBT	1
ADB	KSNEST	1
ADB	ANGN	1
ADB	SVRL	1
ADB	KUDS	1
ADB	GNPR	2
ADB	KMLAFTB	6
ADB	KRIVBT	8

Fig: Road where less than ten tickets were sold.

Reasonableness of routes

Consistent Route:

If tickets are only issued for the bus stops listed in the master data, those routes are said to be consistent.

Inconsistent Route:

Tickets may be given at some exceptional bus stops on a route even though they are not listed in the master data; those are known as inconsistent routes. Some routes mentioned in the ETIM are not listed in the master data. There are a few consistent routes in Shahapur taluka.

route_no		inconsistence
1510		Not in master data
2724		Not in master data
3057		Not in master data
1878		Not in master data
375		Not in master data
1401		Not in master data
2129		Not in master data
7646		Not in master data
1880		Not in master data

Fig: Inconsistent routes

Cause for inconsistency:

Stop not present in master data

The bus driver might stop the bus at such a stop where the tickets were issued, but there may not be an entry in the master data (i.e., not an official stop for that route).

Route not present in master data

The route may not be included in the master data. However, the bus stop might be on the official route because only the ETIM machine could produce a ticket to that stop.

Solution for the inconsistent route:

Route added to master data:

If numerous tickets are sold on a route not mentioned in the master data, it can be permanently added to master data.

Ignore a route:

A stop can be ignored if less number of tickets are being sold so that it cannot be an anomaly.

Shahapur Area Digital Geography

Centroid

The arithmetic mean position of all the points on the surface of a solid or planar figure is known as the centroid, also known as the geometric center or center of the figure. Centroid can be helpful for measuring the nearest stop location from the village center.

To create the village centroid by using the *centroid* tools that are already there in QGIS.

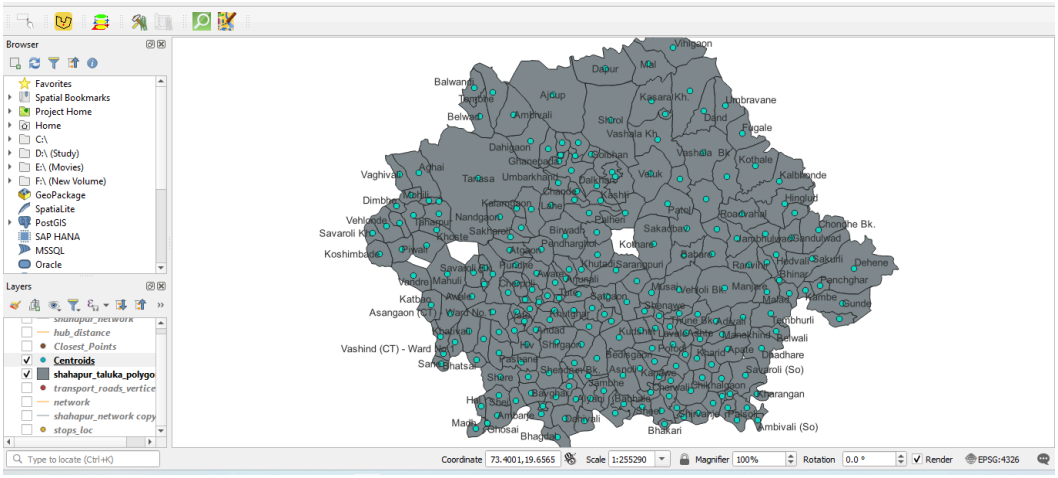


Fig: Village centroid

Stop Location

Stop location is the actual bus stop location that is operating stops within Shahapur.

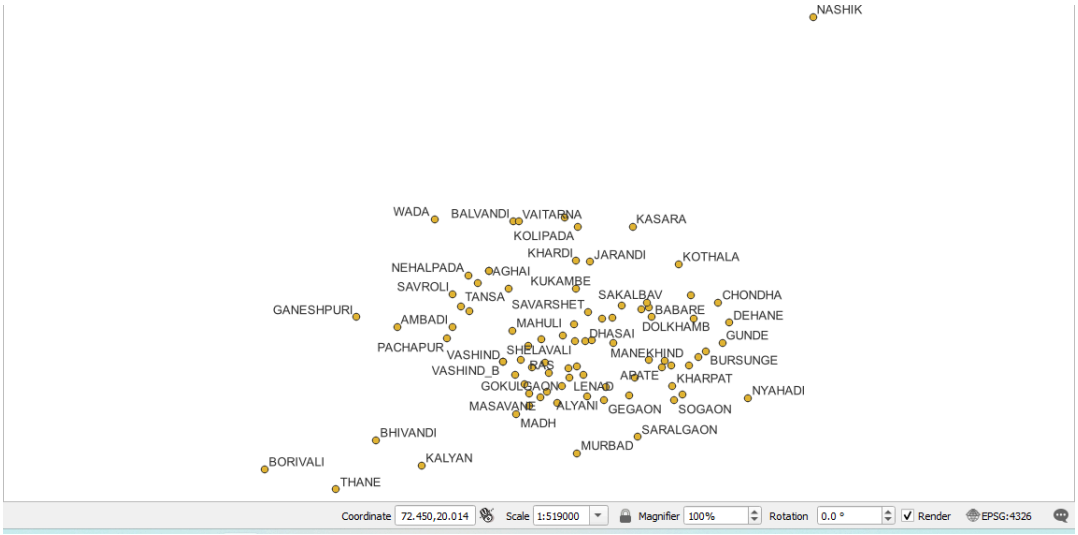


Fig: Stop location

Hub Distance

Hub points are actually the stop locations. We can measure the distance between the village centroid and the stop location. By analyzing hub distance we can see which one is the nearest hub for that particular village.
Here sky blue color circles are village centroids and orange color circles are stop locations or hubs.

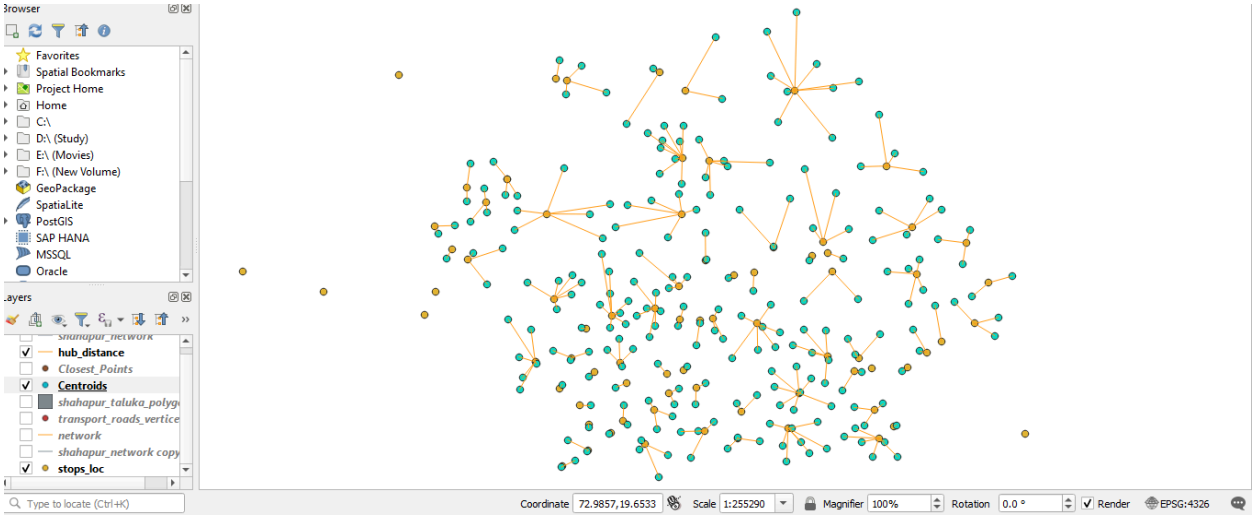


Fig: Hub distance from centroid

Below the image is the sample output for hub distance.

hub_distance — Features Total: 223, Filtered: 223, Selected: 0

	id	census_201	district_c	district_n	taluka_cod	taluka_nam	ward_no	area_name	area_type	HubDist	hh_tot_tot	hh_tot_g	hh_tot_l
1	93966	552767	517	Thane	4167	Shahapur	0	Patol	Rural	4855.00759562078	100	30.6	53.2
2	38921	552913	517	Thane	4167	Shahapur	0	Kasara Bk. (CT)	Urban	672.4418163103...	100	72.6	23.3
3	94387	552820	517	Thane	4167	Shahapur	0	Shilottar	Rural	984.347045492274	100	34.9	59.7
4	93968	552791	517	Thane	4167	Shahapur	0	Hedvali	Rural	1394.205822297...	100	79.3	13.2
5	76403	552909	517	Thane	4167	Shahapur	0	Dahivali	Rural	2273.345578566...	100	8	84.2
6	36671	552713	517	Thane	4167	Shahapur	0	Koshimbade	Rural	960.7771649910...	100	40.5	59
7	60240	552710	517	Thane	4167	Shahapur	0	Vedvahal	Rural	2229.279421307...	100	6.6	89.8
8	76406	552756	517	Thane	4167	Shahapur	0	Vashala Bk	Rural	5535.086800794...	100	77.6	17.4
9	62759	552690	517	Thane	4167	Shahapur	0	Aghai	Rural	2011.843545299...	100	64	31.3
10	62761	552794	517	Thane	4167	Shahapur	0	Ranvihir	Rural	2552.293143255...	100	58.5	28.9
11	62762	552879	517	Thane	4167	Shahapur	0	Khutghar	Rural	2708.589480860...	100	81.5	16.1
12	94799	552754	517	Thane	4167	Shahapur	0	Umbravane	Rural	5766.277649103...	100	15	55
13	66221	552895	517	Thane	4167	Shahapur	0	Hal	Rural	2070.212070228...	100	80.7	12
14	36694	552796	517	Thane	4167	Shahapur	0	Babare	Rural	1470.338625884...	100	46	49.2
15	62765	552904	517	Thane	4167	Shahapur	0	Vithobachegaon	Rural	1228.530830504...	100	79.4	10.5
16	36695	552883	517	Thane	4167	Shahapur	0	Nadgaon (Lonad)	Rural	1287.676188267...	100	42.2	55.4
17	80685	552725	517	Thane	4167	Shahapur	0	Pundhe	Rural	4124.806724675...	100	65.4	31.3
18	36815	552704	517	Thane	4167	Shahapur	0	Sakharoli	Rural	5491.140910150...	100	59	40.4
19	63030	552711	517	Thane	4167	Shahapur	0	Vehlonde	Rural	1795.428872124...	100	16.3	69.2
20	80992	552908	517	Thane	4167	Shahapur	0	Kalgaon	Rural	2785.064192475...	100	35.8	62.2

Show All Features

Fig: Hub distance attributes

Edge

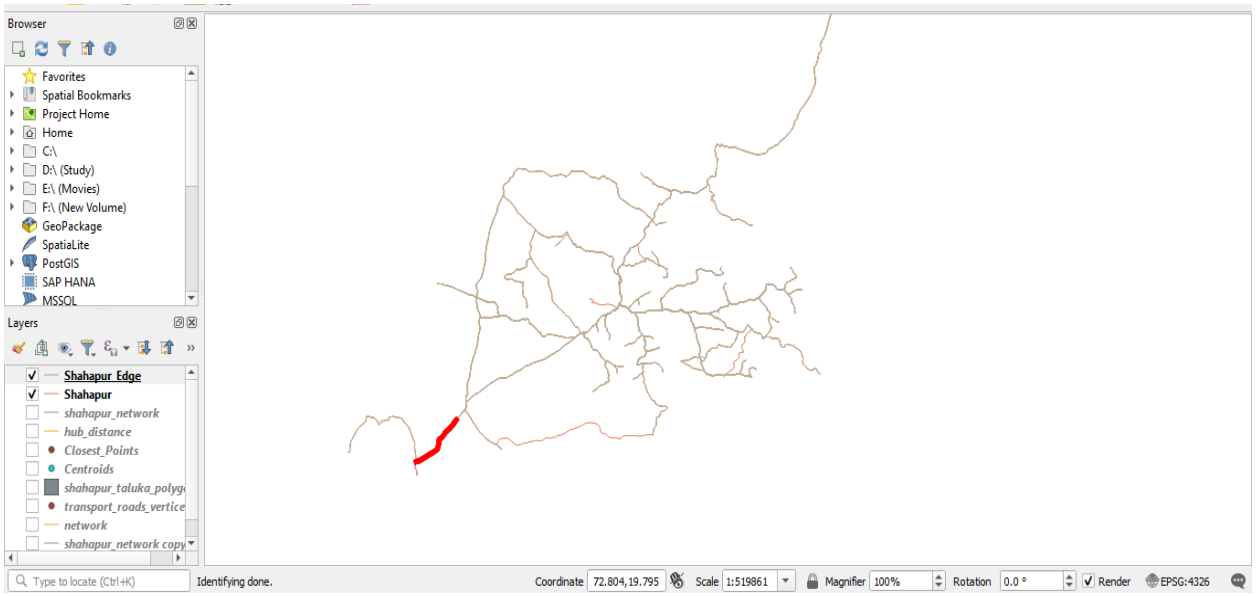
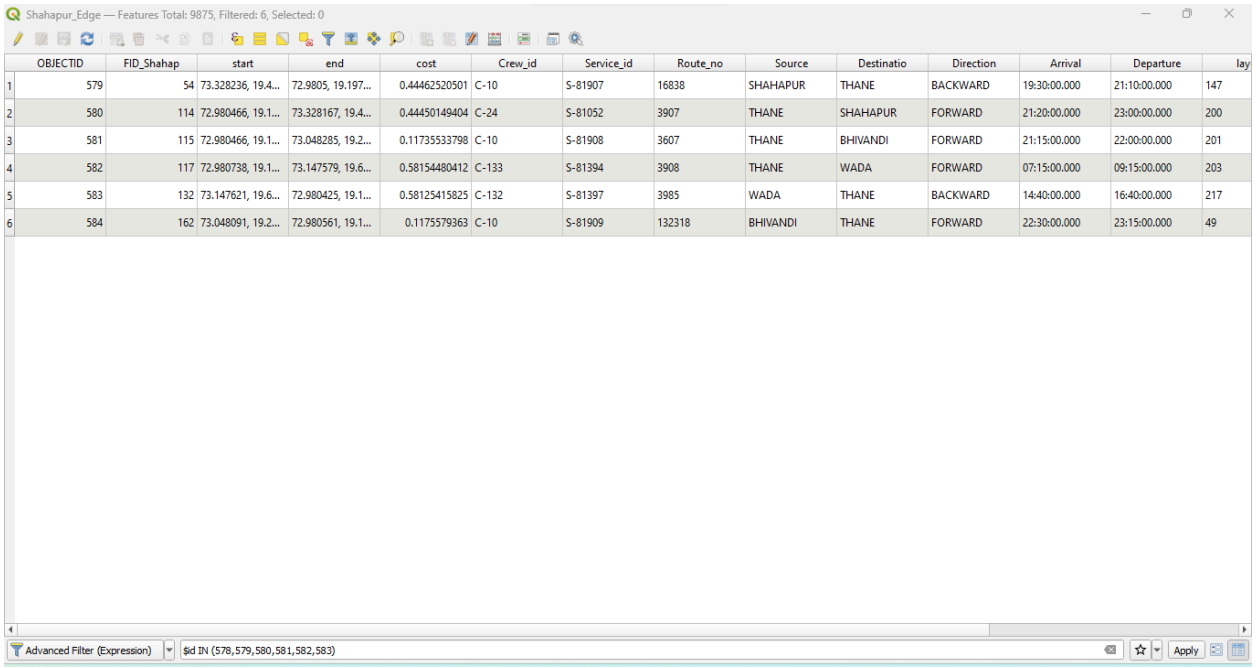


Fig: One particular selected edge



	OBJECTID	FID_Shahap	start	end	cost	Crew_id	Service_id	Route_no	Source	Destinatio	Direction	Arrival	Departure	lay
1	579	54	73.328236, 19.4...	72.9805, 19.197...	0.44462520501	C-10	S-81907	16838	SHAHAPUR	THANE	BACKWARD	19:30:00.000	21:10:00.000	147
2	580	114	72.980466, 19.1...	73.328167, 19.4...	0.44450149404	C-24	S-81052	3907	THANE	SHAHAPUR	FORWARD	21:20:00.000	23:00:00.000	200
3	581	115	72.980466, 19.1...	73.048285, 19.2...	0.11735533798	C-10	S-81908	3607	THANE	BHIVANDI	FORWARD	21:15:00.000	22:00:00.000	201
4	582	117	72.980738, 19.1...	73.147579, 19.6...	0.58154480412	C-133	S-81394	3908	THANE	WADA	FORWARD	07:15:00.000	09:15:00.000	203
5	583	132	73.147621, 19.6...	72.980425, 19.1...	0.58125415825	C-132	S-81397	3985	WADA	THANE	BACKWARD	14:40:00.000	16:40:00.000	217
6	584	162	73.048091, 19.2...	72.980561, 19.1...	0.1175579363	C-10	S-81909	132318	BHIVANDI	THANE	FORWARD	22:30:00.000	23:15:00.000	49

Fig: Selected edge attributes

Sinnar Taluka Bus Depot

About

Sinnar is one of the 15 talukas of Nashik district (census 2011). It operates around 80 buses as of February 19, 2020, on 228 routes in July 2019 on 685 trips in January 2020.

Fig: Google Street Map 3D image of Sinnar Depot

Organization of Bus-Depot

A depot manager oversees each bus station. The depot manager is in charge of the entire bus depot and its departments.

i) ***Traffic:*** The cashier, clerical section, traffic control office, announcement section, ticket vending section, rest houses, and waiting area are all sections of the traffic department. The traffic department is in charge of crew shift scheduling, ensuring that the depot adheres to the division office's timetable (Form 4). Several operational aspects, as well as report generation, have been restored.

ii) ***Workshop:*** The workshop department is in charge of maintaining the buses that are currently in service. They are also responsible for assisting any bus in the event of a catastrophe, tyre blowout, or other similar situations.

Digital Work and Dataset available

Tools and Technology used for the project

Application: QGIS 3.22 Biatowieza

Tools: pgadmin4

DBMS: PostgreSQL

OS: Windows 11

System Configuration

a. QGIS and PostGIS must be connected by a DB connection.
b. Python 3 is already installed in the system.
c. All the shapefiles are imported in the form of PostGIS tables.

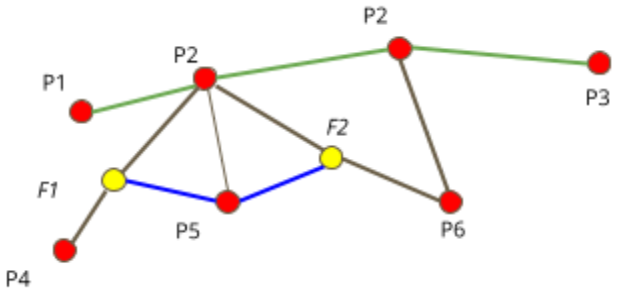
Problem with existing system

a. Lack of GIS data at the taluka level.
b. Lack of convergence of transportation, GIS, and demographics data.
c. Absence of a unified database for demand estimation and service provisioning.
d. Insufficient data to carry out location-based analysis
e. Integration of GIS with the route data will lead to data coherency along with locating other GIS features like schools, factories, markets, and PHCs.
f. Mismatch of data.

Digital Geography

- Model used by GIS, which helps us provide analysis and model geospatial data.
- Mathematically, it is an undirected, planar graph G=(V, E) where V is a set of vertices and E is a set of edges.
    - Dots represent the locations.
    - Lines represent that route exists between the two points.
    - Based on the properties/attributes, these can be assigned different colors for better representation.

- Digital Geography in simple words is a GIS-based graphical interface along with the demography of the region.
- On a high level, digital geography has the following key components:
    - Vertex: These are the points* which represent a particular location in the underlying graphical representation.
    - Edge: It is a poly-line* connecting two points, signifying that there exists a path between them.




```
{v} = {P1, P2, P3, P4, P5, P6, F1, F2}
{E} = {E1, E2, E3, E4, E5, E6, E7, E8, E9}
```

Elements of Digital Geography

Mapping Locations: Vertex

- Each vertex is a Point geometry.
- Each vertex has a latitude and longitude.
- A vertex may be an important location like Bus Stop, school, PHC, Sub-PHC, tourist location, marketplace, etc.

Mapping Paths: Edges

- Each edge is a Polyline Geometry.
- Each edge e is an ordered set of vertices (vi, vj) such that vi and vj $\in$ Vertex set.
- For our case, an edge is a list of points when connected forms the desired edge.
- An edge subsequently can have further additional properties (attributes) like the type of road, length of the edge, traffic on the route segment, the profitability of the route segment, etc.

Datasets are available to us

We have several datasets for analyzing and standardizing. Previously, we created digital geography using stop location, form-iv, and master dataset for Shahapur. However, we have used digital geography in a similar manner to check whether it is applicable to Sinnar or not. Now for further analysis, we have used the ETIM dataset. ETIM dataset can be used to analyze traffic flow, profit, etc.

Sr.	Dataset Name	Brief Description
1	Form-iv	Form 4 is the Operational data calculated by using the previous year's travel data, which is used by MSRTC for scheduling crew members and trips.
2	Stop location	The stop location data describes the stops and their geometry information.
3	ETIM	ETIM data is the real-time ticketing information generated while issuing tickets by the conductor and stored in ETIM machines.
4	Master data	The Master data describes the route information.

Available shapefiles:

1. Sinnar taluka village Polygon.
2. Road network polygon.

Dataset Used

Form-4

Sr.	Attribute Name	Values	Description
1	arrival	Time	Arrival-time of a bus at a given time at a particular terminal.
2	duty_id	Character Varying	A crew member's duty_id is a kind of id that provides so that the crew does not change.
3	d_eng	Place Names	Destination terminal name in English.

4	d_mar	Place Names	Destination terminal name in Marathi.
5	depart	Time	Departure time of a bus at a given time at a particular terminal.
6	distance	Kilometer	Distance between source and destination terminal.
7	road_seg	Name	This is the same route segmentation.
8	service_id	Character Varying	A service id is a particular service between terminals at a given time.
9	s_eng	Place Names	Source terminal name in English.
10	s_mar	Place Names	Source terminal name in Marathi.
11	schedule_id	Numeric	A schedule_id is an id that is given so that the bus does not change.

Stops

Sr.	Attribute Name	Values	Description
1	name	Place Names	Stop name
2	name_mar	Place Names	Stop name in Marathi.
3	taluka	Place Names	Taluka name of a stop.
4	village	Place Names	Village name of a stop
4	wkt	Geom	Latitude-Longitude for a stop

ETIM

Sr.	Attribute Name	Values	Description
1	etim_no	Character Varying	ETIM no from which the ticket has been sold.
2	from_stage_code	Place Names	Source terminal name (Short Form)
3	route_no	Numeric	Route no for every terminal to terminal
4	ticket_id	Numeric	Auto-generated id no for tickets.
5	ticket_number	Numeric	Individual ticket no for every person
6	till_stage_code	Place Names	Destination terminal name (Short Form)
7	total_amt	Numeric	Ticket selling amount for a day
8	ticket_date	Date	Ticket selling date
9	ticket_date_actual	Date	Ticket selling date (formatted)
10	ticket_time	Time	Ticket selling time
11	trip_no	Character Varying	Trip no for every route no

Master Data

Sr.	Attribute Name	Values	Description
1	from_stop_cd	Place Names	Source terminal name (Short Form)
2	Route_kms	Kilometer	Distance of that route
3	route_no	Numeric	Route no for every terminal to terminal
4	till_stop_cd	Place Names	Destination terminal name (Short Form)
5	via_stop_cd	Place Names	Intermediate terminal between source and destination.

List of Tables and their attributes

Sr.	Table Name	Unique Key Attribute	Other Attribute	Added Attribute	Table name of that additional attributes
1	Form-IV	**(schedule, duty id, and service id)**	source, destination, arrival time, departure time, distance	Geometry point. Route No.	Those attributes taken from - Stops Location, and Master File table
2	Stop Location	**Wkt point (lat-long)**	Name, name_mar, village, taluka, taluka_mar		
3	ETIM	**(ticket_id, trip_no, route_no)**	etim_no, ticket_number, from_stage_code, till_stage_cod, total_amt, ticket_date, ticket_date_actual, ticket_time		
4	Master	**route_no**	Route_kms, from_stop_cd,till_stop_cd, via_stop_cd		

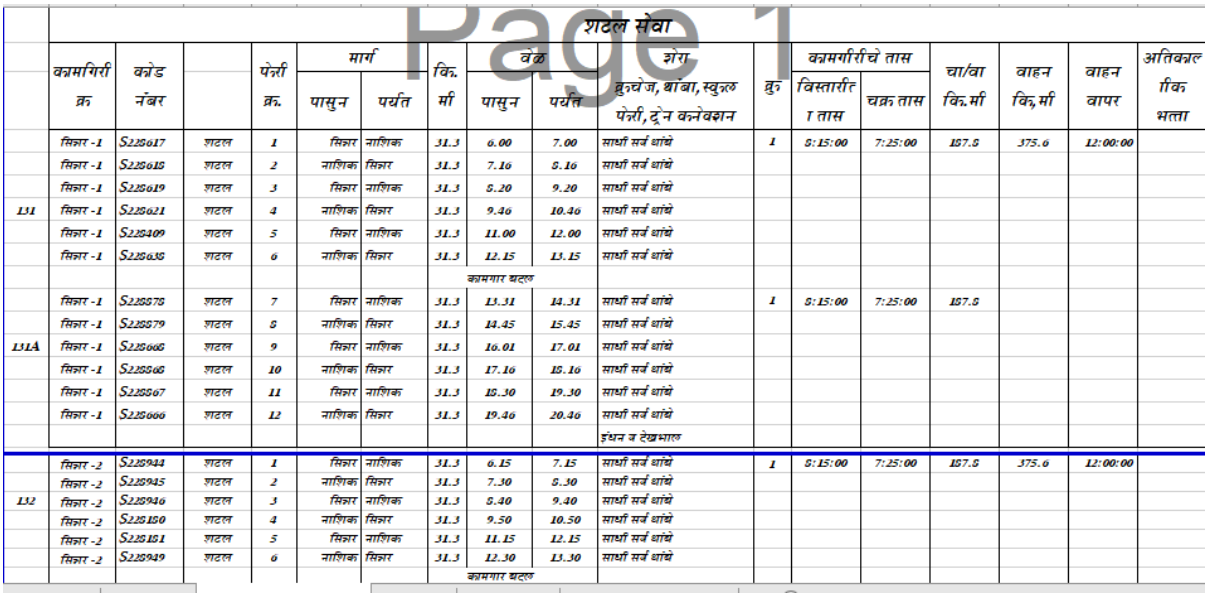
Form-IV:

Form 4 is a bus timetable created and maintained at the Division level with the help of the Depot manager of that specific taluka. It is one of the most important datasets for our

research. It has several attributes: schedule id, Duty id, service id, and places are the essential attributes.

In the dataset, every row contains information about the schedule id, duty_id of crew members, service_id of a bus, Source termini, destination termini, the arrival time of the bus, departure time of the bus, distance from source to destination, and the route segment name.

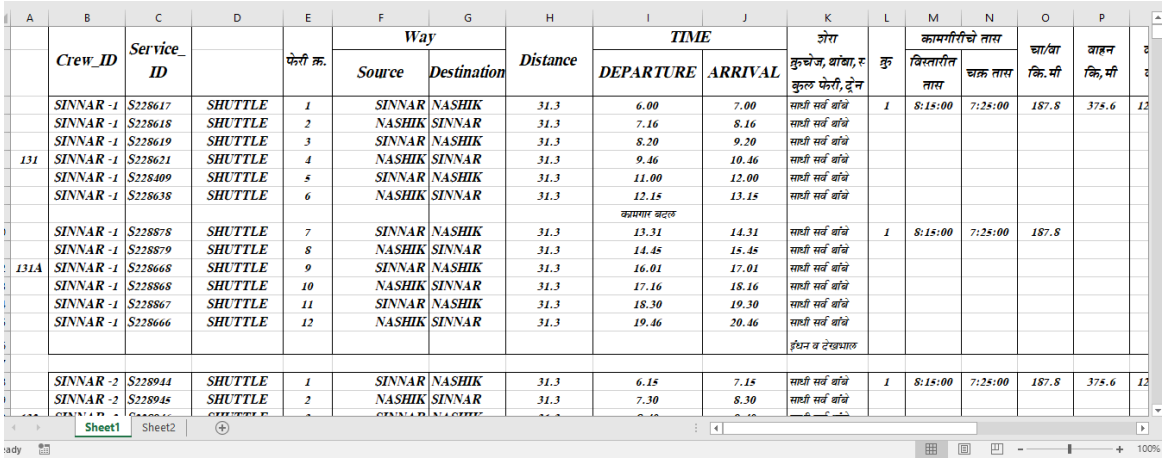
The form-iv we received was in Marathi so before using that dataset we translated it into English.



शटल सेवा

	कामगिरी क्र	व्होड नंबर		फेरी क्र.	मार्ग पासुन	मार्ग पर्यंत	कि. मी	वेळ पासुन	वेळ पर्यंत	शेरा हुप्पेज, थांबा, स्कूल फेरी, ट्रेन कनेक्शन	कु	कामगीरीचे तास विस्तारीत तास	कामगीरीचे तास चक्र तास	चा/वा कि.मी	वाहन कि. मी	वाहन वापर	अतिरिक्त किं भत्ता
	सिन्नर -1	S228617	शटल	1	सिन्नर	नाशिक	31.3	6.00	7.00	साथी सर्व थांबे	1	8:15:00	7:25:00	187.8	375.6	12:00:00	
	सिन्नर -1	S228618	शटल	2	नाशिक	सिन्नर	31.3	7.16	8.16	साथी सर्व थांबे							
	सिन्नर -1	S228619	शटल	3	सिन्नर	नाशिक	31.3	8.20	9.20	साथी सर्व थांबे							
131	सिन्नर -1	S228621	शटल	4	नाशिक	सिन्नर	31.3	9.46	10.46	साथी सर्व थांबे							
	सिन्नर -1	S228409	शटल	5	सिन्नर	नाशिक	31.3	11.00	12.00	साथी सर्व थांबे							
	सिन्नर -1	S228638	शटल	6	नाशिक	सिन्नर	31.3	12.15	13.15	साथी सर्व थांबे							
								कामगार बदल									
	सिन्नर -1	S228878	शटल	7	सिन्नर	नाशिक	31.3	13.31	14.31	साथी सर्व थांबे	1	8:15:00	7:25:00	187.8			
	सिन्नर -1	S228879	शटल	8	नाशिक	सिन्नर	31.3	14.45	15.45	साथी सर्व थांबे							
131A	सिन्नर -1	S228668	शटल	9	सिन्नर	नाशिक	31.3	16.01	17.01	साथी सर्व थांबे							
	सिन्नर -1	S228868	शटल	10	नाशिक	सिन्नर	31.3	17.16	18.16	साथी सर्व थांबे							
	सिन्नर -1	S228867	शटल	11	सिन्नर	नाशिक	31.3	18.30	19.30	साथी सर्व थांबे							
	सिन्नर -1	S228666	शटल	12	नाशिक	सिन्नर	31.3	19.46	20.46	साथी सर्व थांबे							
										इंधन व देखभाल							
	सिन्नर -2	S228944	शटल	1	सिन्नर	नाशिक	31.3	6.15	7.15	साथी सर्व थांबे	1	8:15:00	7:25:00	187.8	375.6	12:00:00	
	सिन्नर -2	S228945	शटल	2	नाशिक	सिन्नर	31.3	7.30	8.30	साथी सर्व थांबे							
132	सिन्नर -2	S228946	शटल	3	सिन्नर	नाशिक	31.3	8.40	9.40	साथी सर्व थांबे							
	सिन्नर -2	S228950	शटल	4	नाशिक	सिन्नर	31.3	9.50	10.50	साथी सर्व थांबे							
	सिन्नर -2	S228951	शटल	5	सिन्नर	नाशिक	31.3	11.15	12.15	साथी सर्व थांबे							
	सिन्नर -2	S228949	शटल	6	नाशिक	सिन्नर	31.3	12.30	13.30	साथी सर्व थांबे							
								कामगार बदल									

Fig: Form-4 original



	Crew_ID	Service_ ID		फेरी क्र.	Way Source	Way Destination	Distance	TIME DEPARTURE	TIME ARRIVAL	शेरा हुप्पेज, थांबा, स्कूल फेरी, ट्रेन	कु	कामगीरीचे तास विस्तारीत तास	कामगीरीचे तास चक्र तास	चा/वा कि. मी	वाहन कि. मी
	SINNAR -1	S228617	SHUTTLE	1	SINNAR	NASHIK	31.3	6.00	7.00	साथी सर्व थांबे	1	8:15:00	7:25:00	187.8	375.6
	SINNAR -1	S228618	SHUTTLE	2	NASHIK	SINNAR	31.3	7.16	8.16	साथी सर्व थांबे					
	SINNAR -1	S228619	SHUTTLE	3	SINNAR	NASHIK	31.3	8.20	9.20	साथी सर्व थांबे					
131	SINNAR -1	S228621	SHUTTLE	4	NASHIK	SINNAR	31.3	9.46	10.46	साथी सर्व थांबे					
	SINNAR -1	S228409	SHUTTLE	5	SINNAR	NASHIK	31.3	11.00	12.00	साथी सर्व थांबे					
	SINNAR -1	S228638	SHUTTLE	6	NASHIK	SINNAR	31.3	12.15	13.15	साथी सर्व थांबे					
								कामगार बदल							
	SINNAR -1	S228878	SHUTTLE	7	SINNAR	NASHIK	31.3	13.31	14.31	साथी सर्व थांबे	1	8:15:00	7:25:00	187.8	
	SINNAR -1	S228879	SHUTTLE	8	NASHIK	SINNAR	31.3	14.45	15.45	साथी सर्व थांबे					
131A	SINNAR -1	S228668	SHUTTLE	9	SINNAR	NASHIK	31.3	16.01	17.01	साथी सर्व थांबे					
	SINNAR -1	S228868	SHUTTLE	10	NASHIK	SINNAR	31.3	17.16	18.16	साथी सर्व थांबे					
	SINNAR -1	S228867	SHUTTLE	11	SINNAR	NASHIK	31.3	18.30	19.30	साथी सर्व थांबे					
	SINNAR -1	S228666	SHUTTLE	12	NASHIK	SINNAR	31.3	19.46	20.46	साथी सर्व थांबे					
										इंधन व देखभाल					
	SINNAR -2	S228944	SHUTTLE	1	SINNAR	NASHIK	31.3	6.15	7.15	साथी सर्व थांबे	1	8:15:00	7:25:00	187.8	375.6
	SINNAR -2	S228945	SHUTTLE	2	NASHIK	SINNAR	31.3	7.30	8.30	साथी सर्व थांबे					

Fig: Original form-4 after translating it to English

	A	B	C	D	E	F	G
1	Crew_ID	Service_ID	Source	Destination	Distance	DEPARTURE	ARRIVAL
2	SINNAR -1	S228617	SINNAR	NASHIK	31.3	6.00	7.00
3	SINNAR -1	S228618	NASHIK	SINNAR	31.3	7.16	8.16
4	SINNAR -1	S228619	SINNAR	NASHIK	31.3	8.20	9.20
5	SINNAR -1	S228621	NASHIK	SINNAR	31.3	9.46	10.46
6	SINNAR -1	S228409	SINNAR	NASHIK	31.3	11.00	12.00
7	SINNAR -1	S228638	NASHIK	SINNAR	31.3	12.15	13.15
8	SINNAR -1	S228878	SINNAR	NASHIK	31.3	13.31	14.31
9	SINNAR -1	S228879	NASHIK	SINNAR	31.3	14.45	15.45
10	SINNAR -1	S228668	SINNAR	NASHIK	31.3	16.01	17.01
11	SINNAR -1	S228868	NASHIK	SINNAR	31.3	17.16	18.16
12	SINNAR -1	S228867	SINNAR	NASHIK	31.3	18.30	19.30
13	SINNAR -1	S228666	NASHIK	SINNAR	31.3	19.46	20.46
14	SINNAR -2	S228944	SINNAR	NASHIK	31.3	6.15	7.15
15	SINNAR -2	S228945	NASHIK	SINNAR	31.3	7.30	8.30
16	SINNAR -2	S228946	SINNAR	NASHIK	31.3	8.40	9.40
17	SINNAR -2	S228180	NASHIK	SINNAR	31.3	9.50	10.50
18	SINNAR -2	S228181	SINNAR	NASHIK	31.3	11.15	12.15
19	SINNAR -2	S228949	NASHIK	SINNAR	31.3	12.30	13.30
20	SINNAR -2	S228019	SINNAR	NASHIK	31.3	13.45	14.45
21	SINNAR -2	S228951	NASHIK	SINNAR	31.3	15.00	16.00
22	SINNAR -2	S228067	SINNAR	NASHIK	31.3	16.15	17.15

Sheet1 Form-4

Fig: Extracted necessary information for original form-4

It is still not the completed form-4 we need for our work. We need to add some other attributes to this dataset to make it a complete form-4.

1	id	Source	Destination	Crew_ID	Service_ID	Route_No	Direction	Departure	Arrival	Distance
2	1	AADWADI	SINNAR	SINNAR -21	S228150	9705	FORWARD	11:10:00	12:10:00	28.5
3	2	AADWADI	SINNAR	SINNAR -51	S228580	9705	FORWARD	10:50:00	11:50:00	28.5
4	3	AADWADI	SINNAR	SINNAR -60	S228198	9705	FORWARD	6:15:00	7:15:00	28.5
5	5	AADWADI	SINNAR	SINNAR -64	S228681	9705	FORWARD	17:15:00	18:15:00	29.4
6	4	AADWADI	THANGAON	SINNAR -64	S228679	127106	FORWARD	16:50:00	16:55:00	2.3
7	6	AGASKHIND	SINNAR	SINNAR -50	S228741	91411	FORWARD	16:15:00	17:00:00	22.7
8	7	AHMEDNAGAR	NASHIK	SINNAR -73	M-10076	1511	FORWARD	8:15:00	12:30:00	171.7
9	8	AHMEDNAGAR	NASHIK	SINNAR -74	M-10242	1511	FORWARD	8:30:00	12:45:00	171.7
10	9	AHMEDNAGAR	NASHIK	SINNAR -75	M-10086	1511	FORWARD	5:30:00	9:45:00	171.7
11	10	AHMEDNAGAR	NASHIK	SINNAR -76	M-10188	1511	FORWARD	13:00:00	17:15:00	171.7
12	11	AKOLE	SINNAR	SINNAR -11	S224858	9718	FORWARD	15:45:00	17:15:00	42.7
13	12	AKOLE	SINNAR	SINNAR -11	S224870	9718	FORWARD	19:45:00	21:15:00	42.7
14	13	AKOLE	SINNAR	SINNAR -12	S224849	9718	FORWARD	7:30:00	9:00:00	42.7
15	14	AKOLE	SINNAR	SINNAR -12	S228270	9718	FORWARD	11:15:00	12:45:00	42.7
16	15	AKOLE	SINNAR	SINNAR -14	S224837	9718	FORWARD	8:45:00	10:15:00	42.7
17	16	AKOLE	SINNAR	SINNAR -14	S224863	9718	FORWARD	12:15:00	13:45:00	42.7
18	17	AKOLE	SINNAR	SINNAR -14	S224871	9718	FORWARD	15:45:00	17:15:00	42.7
19	18	AKOLE	SINNAR	SINNAR -14	S224801	9718	FORWARD	19:10:00	20:40:00	42.7
20	19	AKOLE	SINNAR	SINNAR -21	S224843	9718	FORWARD	15:15:00	16:45:00	42
21	20	AKOLE	SINNAR	SINNAR -25	S224843	9718	FORWARD	10:35:00	12:05:00	42
22	21	AKOLE	SINNAR	SINNAR -28	S228320	9718	FORWARD	10:15:00	12:15:00	59.7
23	22	AKOLE	SINNAR	SINNAR -36	S228338	9718	FORWARD	18:10:00	19:40:00	42.7

Form-4

Fig: Complete form-4 with necessary attributes

Bus stop location:

This dataset contains information about the bus stops and their geometry that is necessary for graphically visualizing the data. This dataset includes information on the geometry point of termini, the name of those termini, the village name where the termini are situated, and the taluka name.

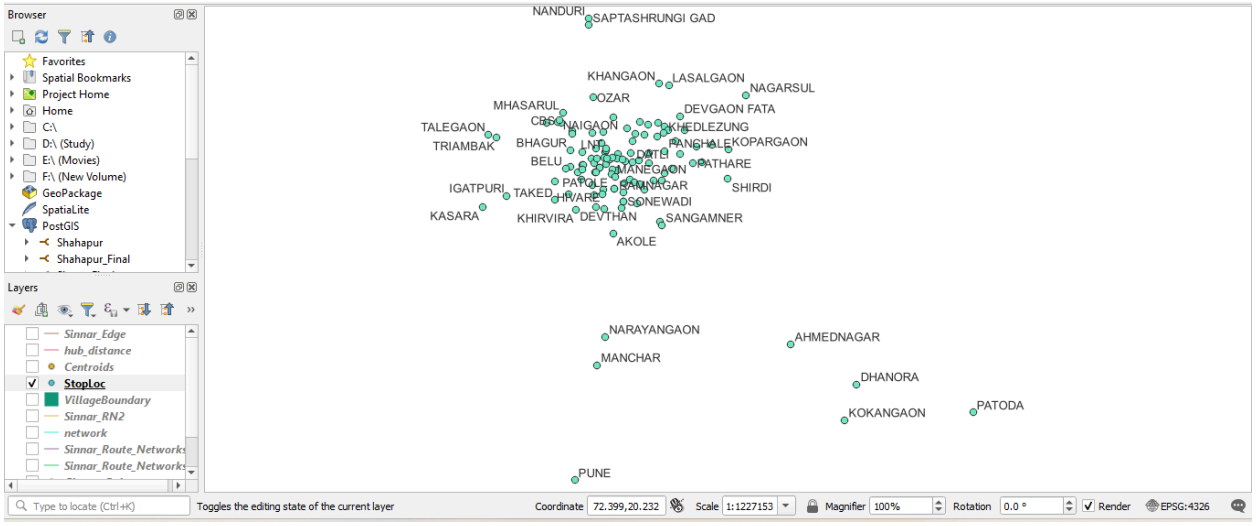


Fig: Bus stop location

	A	B
	Stop_Name	WKT
2	AADWADI	POINT (73.92358 19.7378421)
3	AGASKHIND	POINT (73.8342839 19.8117985)
4	AHMEDNAGAR	POINT (74.7363011 19.0860023)
5	AKOLE	POINT (74.0068377 19.5412291)
6	AUNDHWADI	POINT (73.8617345 19.7888826)
7	BELU	POINT (73.8123396 19.8065792)
8	BHAGUR	POINT (73.8331873 19.8816958)
9	BHATWADI	POINT (73.9771589 19.8251823)
10	BHENDALI	POINT (74.0640005 19.9700673)
11	BHOKANI	POINT (74.1559256 19.8268692)
12	CHAPADGAON	POINT (74.0512826 19.7067804)
13	CHINCHOLI FATA	POINT (73.9374109 19.8880536)
14	DAHIWADI	POINT (74.2632323 19.9174023)
15	DAPUR	POINT (74.0526955 19.7445753)
16	DARDE FATA	POINT (74.3696045 19.8325935)
17	DATTANAGAR	POINT (74.0757038 19.7594235)
18	DATLI	POINT (74.0861615 19.8322763)
19	DEVTHAN	POINT (74.0425301 19.6432322)
20	DEVGAON FATA	POINT (74.2817504 20.0185788)
21	DEVPUR FATA	POINT (74.1530693 19.8672397)
22	DAGADWADI	POINT (74.026284930758 19.8652026401392)
23	DHAMORI FATA	POINT (74.2989688 19.9619382)

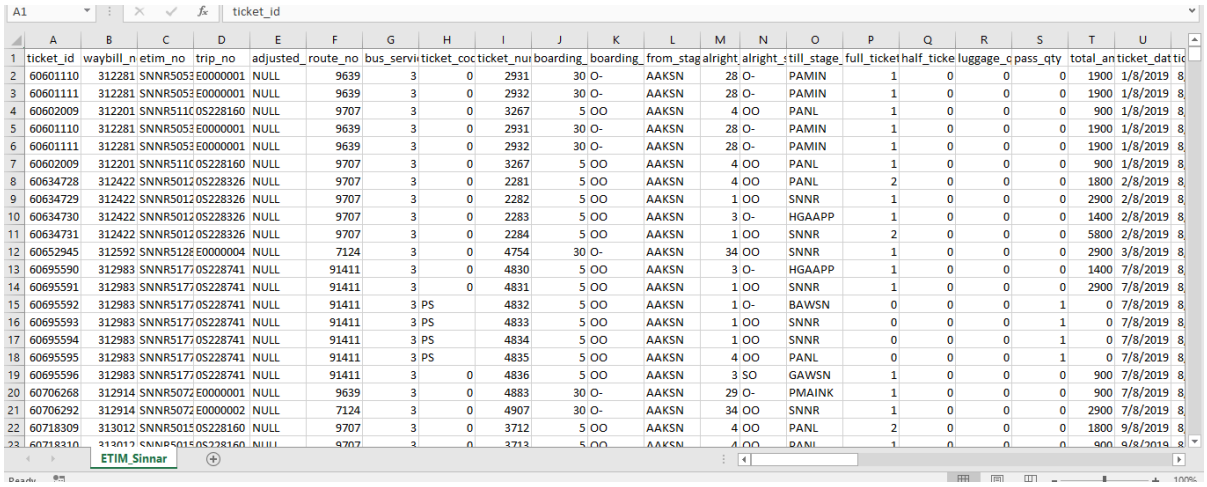
Stop Loc

Fig: Attribute table for Stop Location dataset

ETIM:

Every conductor for any bus service must obtain an ETIM from the cashier, as well as cash change. Many charging connections may be found in the cashier's cabin, where unused ETIMs can be charged. TRIMAX is responsible for the data handling of these ETIM. ETIM data is real-time ticketing information created by the conductor when issuing tickets and saved in ETIM

equipment administered by TRIMAX. The information has been compiled (ticket id, waybill number, ETIM number, trip number, route number, ticket code, ticket number, stage code, till stage code, total amt, ticket date actual, and ticket time).

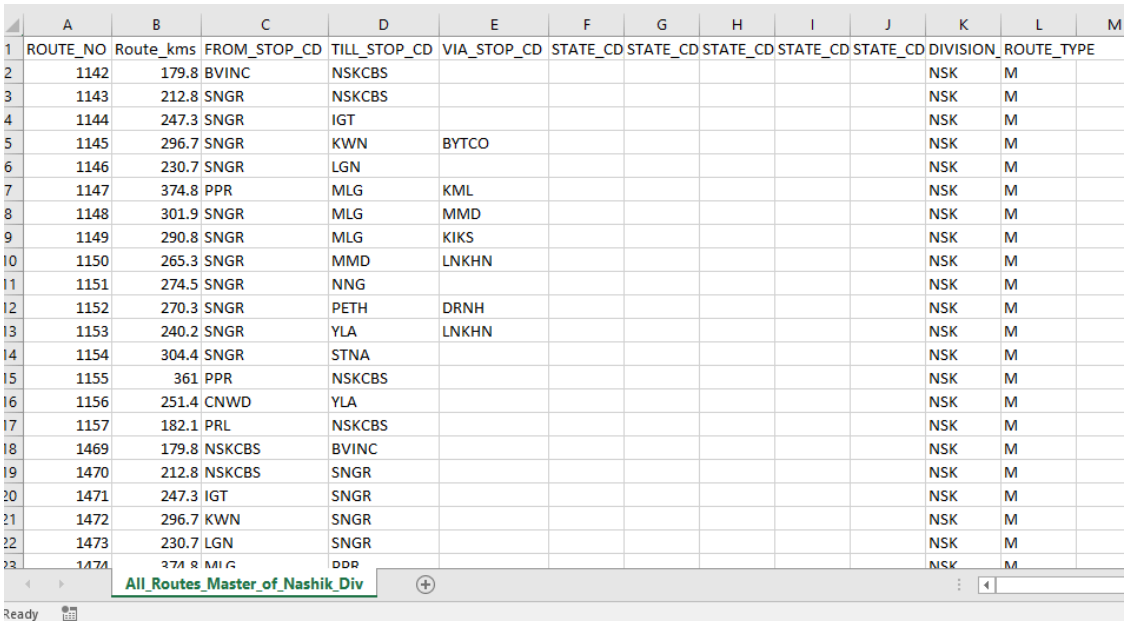


ticket_id	waybill_n	etim_no	trip_no	adjusted_	route_no	bus_servi	ticket_coc	ticket_nur	boarding_	boarding_	from_stag	alright	alright_s	till_stage_	full_ticket	half_ticke	luggage_c	pass_qty	total_an	ticket_dat	tic
60601110	312281	SNNR5053	E0000001	NULL	9639	3	0	2931	30	O-	AAKSN	28	O-	PAMIN	1	0	0	0	1900	1/8/2019	8
60601111	312281	SNNR5053	E0000001	NULL	9639	3	0	2932	30	O-	AAKSN	28	O-	PAMIN	1	0	0	0	1900	1/8/2019	8
60602009	312201	SNNR5110	S228160	NULL	9707	3	0	3267	5	OO	AAKSN	4	OO	PANL	1	0	0	0	900	1/8/2019	8
60601110	312281	SNNR5053	E0000001	NULL	9639	3	0	2931	30	O-	AAKSN	28	O-	PAMIN	1	0	0	0	1900	1/8/2019	8
60601111	312281	SNNR5053	E0000001	NULL	9639	3	0	2932	30	O-	AAKSN	28	O-	PAMIN	1	0	0	0	1900	1/8/2019	8
60602009	312201	SNNR5110	S228160	NULL	9707	3	0	3267	5	OO	AAKSN	4	OO	PANL	1	0	0	0	900	1/8/2019	8
60634728	312422	SNNR5012	S228326	NULL	9707	3	0	2281	5	OO	AAKSN	4	OO	PANL	2	0	0	0	1800	2/8/2019	8
60634729	312422	SNNR5012	S228326	NULL	9707	3	0	2282	5	OO	AAKSN	1	OO	SNNR	1	0	0	0	2900	2/8/2019	8
60634730	312422	SNNR5012	S228326	NULL	9707	3	0	2283	5	OO	AAKSN	3	O-	HGAAPP	1	0	0	0	1400	2/8/2019	8
60634731	312422	SNNR5012	S228326	NULL	9707	3	0	2284	5	OO	AAKSN	1	OO	SNNR	2	0	0	0	5800	2/8/2019	8
60652945	312592	SNNR5128	E0000004	NULL	7124	3	0	4754	30	O-	AAKSN	34	OO	SNNR	1	0	0	0	2900	3/8/2019	8
60695590	312983	SNNR5177	S228741	NULL	91411	3	0	4830	5	OO	AAKSN	3	O-	HGAAPP	1	0	0	0	1400	7/8/2019	8
60695591	312983	SNNR5177	S228741	NULL	91411	3	0	4831	5	OO	AAKSN	1	OO	SNNR	1	0	0	0	2900	7/8/2019	8
60695592	312983	SNNR5177	S228741	NULL	91411	3	PS	4832	5	OO	AAKSN	1	O-	BAWSN	0	0	0	1	0	7/8/2019	8
60695593	312983	SNNR5177	S228741	NULL	91411	3	PS	4833	5	OO	AAKSN	1	OO	SNNR	0	0	0	1	0	7/8/2019	8
60695594	312983	SNNR5177	S228741	NULL	91411	3	PS	4834	5	OO	AAKSN	1	OO	SNNR	0	0	0	1	0	7/8/2019	8
60695595	312983	SNNR5177	S228741	NULL	91411	3	PS	4835	5	OO	AAKSN	4	OO	PANL	0	0	0	1	0	7/8/2019	8
60695596	312983	SNNR5177	S228741	NULL	91411	3	0	4836	5	OO	AAKSN	3	SO	GAWSN	1	0	0	0	900	7/8/2019	8
60706268	312914	SNNR5072	E0000001	NULL	9639	3	0	4883	30	O-	AAKSN	29	O-	PMAINK	1	0	0	0	900	7/8/2019	8
60706292	312914	SNNR5072	E0000002	NULL	7124	3	0	4907	30	O-	AAKSN	34	OO	SNNR	1	0	0	0	2900	7/8/2019	8
60718309	313012	SNNR5015	S228160	NULL	9707	3	0	3712	5	OO	AAKSN	4	OO	PANL	2	0	0	0	1800	9/8/2019	8

Fig: Attribute table for ETIM dataset

Master Data:

The Master data describes the route information. The route information comprises the order of bus stops on the route and the distance between bus stops.



ROUTE_NO	Route_kms	FROM_STOP_CD	TILL_STOP_CD	VIA_STOP_CD	STATE_CD	STATE_CD	STATE_CD	STATE_CD	STATE_CD	DIVISION_	ROUTE_TYPE
1142	179.8	BVINC	NSKCBS							NSK	M
1143	212.8	SNGR	NSKCBS							NSK	M
1144	247.3	SNGR	IGT							NSK	M
1145	296.7	SNGR	KWN	BYTCO						NSK	M
1146	230.7	SNGR	LGN							NSK	M
1147	374.8	PPR	MLG	KML						NSK	M
1148	301.9	SNGR	MLG	MMD						NSK	M
1149	290.8	SNGR	MLG	KIKS						NSK	M
1150	265.3	SNGR	MMD	LNKHN						NSK	M
1151	274.5	SNGR	NNG							NSK	M
1152	270.3	SNGR	PETH	DRNH						NSK	M
1153	240.2	SNGR	YLA	LNKHN						NSK	M
1154	304.4	SNGR	STNA							NSK	M
1155	361	PPR	NSKCBS							NSK	M
1156	251.4	CNWD	YLA							NSK	M
1157	182.1	PRL	NSKCBS							NSK	M
1469	179.8	NSKCBS	BVINC							NSK	M
1470	212.8	NSKCBS	SNGR							NSK	M
1471	247.3	IGT	SNGR							NSK	M
1472	296.7	KWN	SNGR							NSK	M
1473	230.7	LGN	SNGR							NSK	M
1474	374.8	MLG	PPR							NSK	M

Fig: Attribute table for Master dataset

Sinnar taluka village Polygon

Polygon file of Villages in Sinnar Taluka with the Census data. This is as follows:

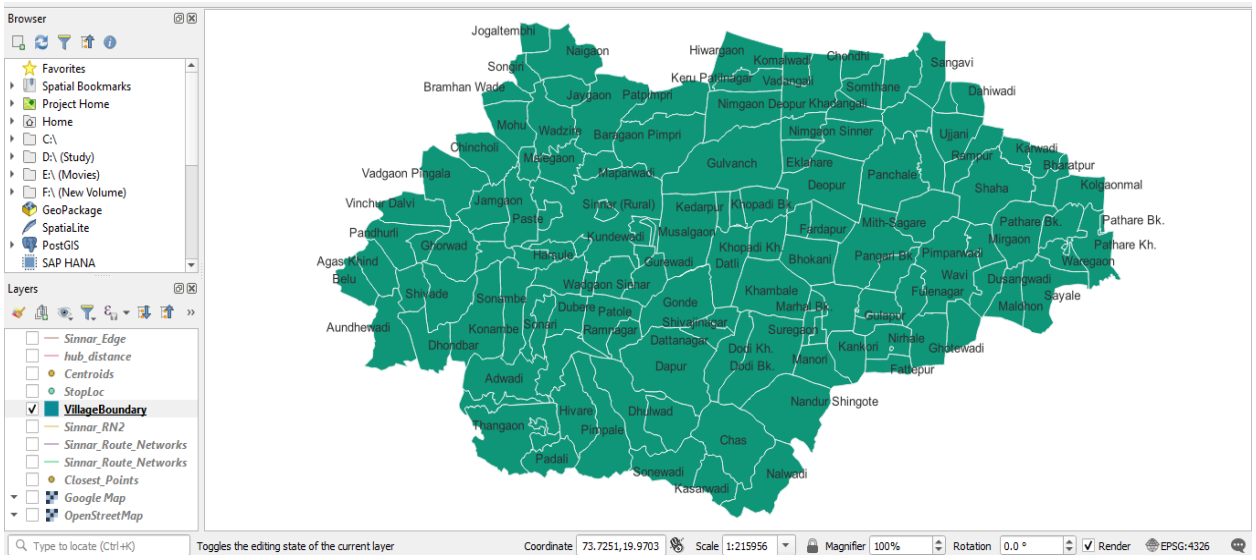
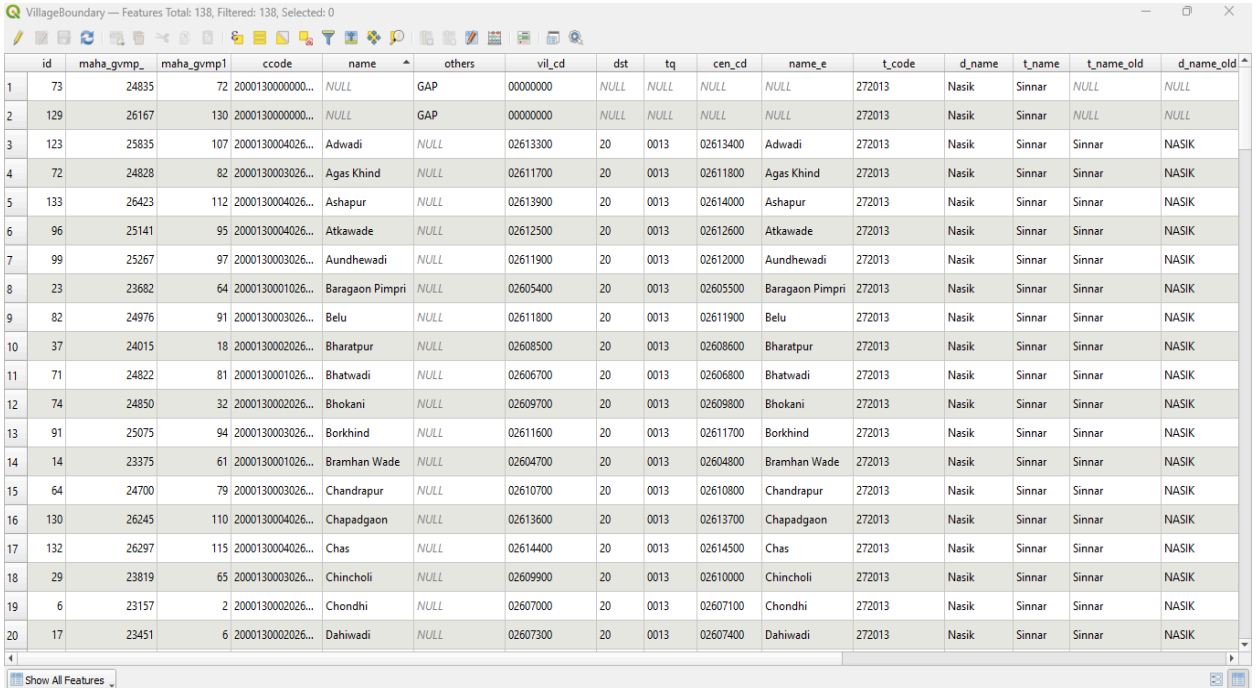


Fig: Sinnar Taluka boundary (census polygon data)



	id	maha_gvmp_	maha_gvmp1	ccode	name	others	vil_cd	dst	tq	cen_cd	name_e	t_code	d_name	t_name	t_name_old	d_name_old
1	73	24835	72	2000130000000...	NULL	GAP	00000000	NULL	NULL	NULL	NULL	272013	Nasik	Sinnar	NULL	NULL
2	129	26167	130	2000130000000...	NULL	GAP	00000000	NULL	NULL	NULL	NULL	272013	Nasik	Sinnar	NULL	NULL
3	123	25835	107	2000130004026...	Adwadi	NULL	02613300	20	0013	02613400	Adwadi	272013	Nasik	Sinnar	Sinnar	NASIK
4	72	24828	82	2000130003026...	Agas Khind	NULL	02611700	20	0013	02611800	Agas Khind	272013	Nasik	Sinnar	Sinnar	NASIK
5	133	26423	112	2000130004026...	Ashapur	NULL	02613900	20	0013	02614000	Ashapur	272013	Nasik	Sinnar	Sinnar	NASIK
6	96	25141	95	2000130004026...	Atkawade	NULL	02612500	20	0013	02612600	Atkawade	272013	Nasik	Sinnar	Sinnar	NASIK
7	99	25267	97	2000130003026...	Aundhewadi	NULL	02611900	20	0013	02612000	Aundhewadi	272013	Nasik	Sinnar	Sinnar	NASIK
8	23	23682	64	2000130001026...	Baragaon Pimpri	NULL	02605400	20	0013	02605500	Baragaon Pimpri	272013	Nasik	Sinnar	Sinnar	NASIK
9	82	24976	91	2000130003026...	Belu	NULL	02611800	20	0013	02611900	Belu	272013	Nasik	Sinnar	Sinnar	NASIK
10	37	24015	18	2000130002026...	Bharatpur	NULL	02608500	20	0013	02608600	Bharatpur	272013	Nasik	Sinnar	Sinnar	NASIK
11	71	24822	81	2000130001026...	Bhatwadi	NULL	02606700	20	0013	02606800	Bhatwadi	272013	Nasik	Sinnar	Sinnar	NASIK
12	74	24850	32	2000130002026...	Bhokani	NULL	02609700	20	0013	02609800	Bhokani	272013	Nasik	Sinnar	Sinnar	NASIK
13	91	25075	94	2000130003026...	Borkhind	NULL	02611600	20	0013	02611700	Borkhind	272013	Nasik	Sinnar	Sinnar	NASIK
14	14	23375	61	2000130001026...	Bramhan Wade	NULL	02604700	20	0013	02604800	Bramhan Wade	272013	Nasik	Sinnar	Sinnar	NASIK
15	64	24700	79	2000130003026...	Chandrapur	NULL	02610700	20	0013	02610800	Chandrapur	272013	Nasik	Sinnar	Sinnar	NASIK
16	130	26245	110	2000130004026...	Chapadgaon	NULL	02613600	20	0013	02613700	Chapadgaon	272013	Nasik	Sinnar	Sinnar	NASIK
17	132	26297	115	2000130004026...	Chas	NULL	02614400	20	0013	02614500	Chas	272013	Nasik	Sinnar	Sinnar	NASIK
18	29	23819	65	2000130003026...	Chincholi	NULL	02609900	20	0013	02610000	Chincholi	272013	Nasik	Sinnar	Sinnar	NASIK
19	6	23157	2	2000130002026...	Chondhi	NULL	02607000	20	0013	02607100	Chondhi	272013	Nasik	Sinnar	Sinnar	NASIK
20	17	23451	6	2000130002026...	Dahiwadi	NULL	02607300	20	0013	02607400	Dahiwadi	272013	Nasik	Sinnar	Sinnar	NASIK

Fig: Attribute table of Sinnar taluka polygon census file.

Road network of Undivided Thane source: OSM

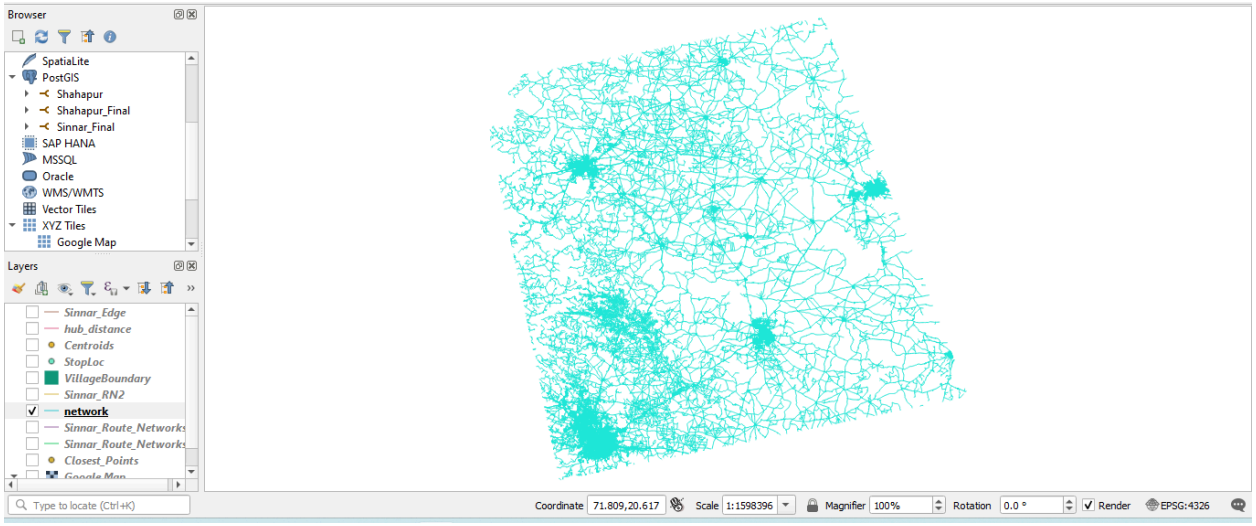
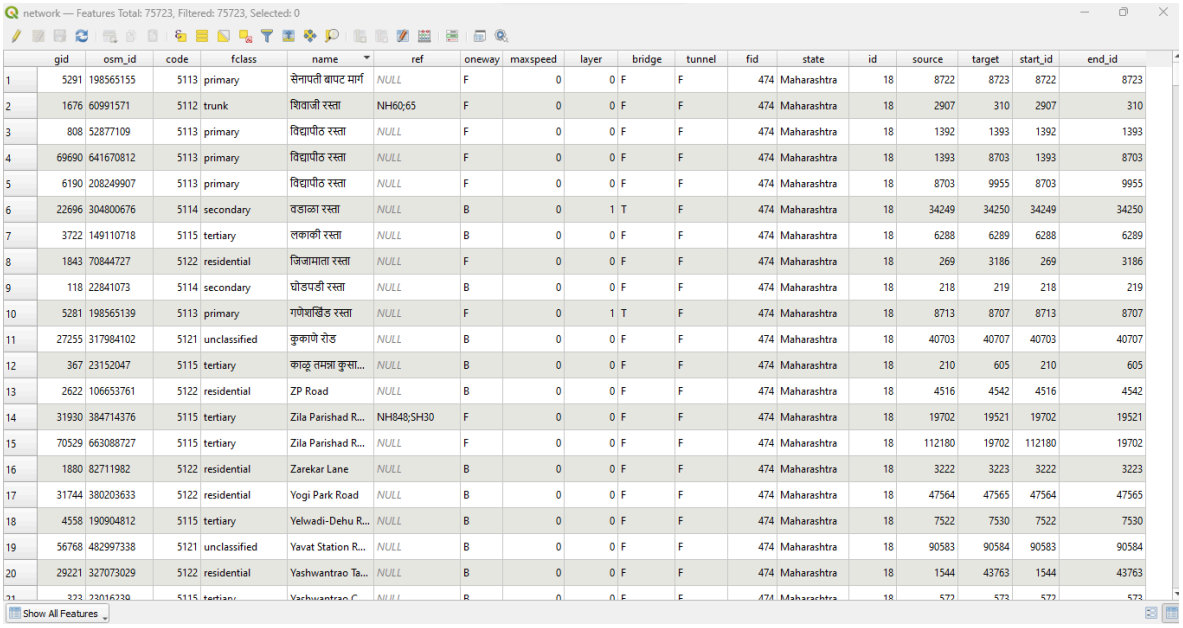


Fig: Road Network



	gid	osm_id	code	fclass	name	ref	oneway	maxspeed	layer	bridge	tunnel	fid	state	id	source	target	start_id	end_id
1	5291	198565155	5113	primary	सेनापती बापट मार्ग	NULL	F	0	0	F	F	474	Maharashtra	18	8722	8723	8722	8723
2	1676	60991571	5112	trunk	शिवाजी रस्ता	NH60;65	F	0	0	F	F	474	Maharashtra	18	2907	310	2907	310
3	808	52877109	5113	primary	विद्यापीठ रस्ता	NULL	F	0	0	F	F	474	Maharashtra	18	1392	1393	1392	1393
4	69690	641670812	5113	primary	विद्यापीठ रस्ता	NULL	F	0	0	F	F	474	Maharashtra	18	1393	8703	1393	8703
5	6190	208249907	5113	primary	विद्यापीठ रस्ता	NULL	F	0	0	F	F	474	Maharashtra	18	8703	9955	8703	9955
6	22696	304800676	5114	secondary	वडाळा रस्ता	NULL	B	0	1	T	F	474	Maharashtra	18	34249	34250	34249	34250
7	3722	149110718	5115	tertiary	लकाकी रस्ता	NULL	B	0	0	F	F	474	Maharashtra	18	6288	6289	6288	6289
8	1843	70844727	5122	residential	जिजामाता रस्ता	NULL	F	0	0	F	F	474	Maharashtra	18	269	3186	269	3186
9	118	22841073	5114	secondary	घोडपडी रस्ता	NULL	B	0	0	F	F	474	Maharashtra	18	218	219	218	219
10	5281	198565139	5113	primary	गणेशखिंड रस्ता	NULL	F	0	1	T	F	474	Maharashtra	18	8713	8707	8713	8707
11	27255	317984102	5121	unclassified	कुकाणे रोड	NULL	B	0	0	F	F	474	Maharashtra	18	40703	40707	40703	40707
12	367	23152047	5115	tertiary	काळू तमन्ना कुसा...	NULL	B	0	0	F	F	474	Maharashtra	18	210	605	210	605
13	2622	106653761	5122	residential	ZP Road	NULL	B	0	0	F	F	474	Maharashtra	18	4516	4542	4516	4542
14	31930	384714376	5115	tertiary	Zila Parishad R...	NH848;SH30	F	0	0	F	F	474	Maharashtra	18	19702	19521	19702	19521
15	70529	663088727	5115	tertiary	Zila Parishad R...	NULL	F	0	0	F	F	474	Maharashtra	18	112180	19702	112180	19702
16	1880	82711982	5122	residential	Zarekar Lane	NULL	B	0	0	F	F	474	Maharashtra	18	3222	3223	3222	3223
17	31744	380203633	5122	residential	Yogi Park Road	NULL	B	0	0	F	F	474	Maharashtra	18	47564	47565	47564	47565
18	4558	190904812	5115	tertiary	Yelwadi-Dehu R...	NULL	B	0	0	F	F	474	Maharashtra	18	7522	7530	7522	7530
19	56768	482997338	5121	unclassified	Yavat Station R...	NULL	B	0	0	F	F	474	Maharashtra	18	90583	90584	90583	90584
20	29221	327073029	5122	residential	Yashwantrao Ta...	NULL	B	0	0	F	F	474	Maharashtra	18	1544	43763	1544	43763
21	333	23016239	5115	tertiary	Yashwantrao C...	NULL	B	0	0	F	F	474	Maharashtra	18	572	573	572	573

Figure: Road network attribute table

These are all the datasets available to us; This data can be used to illustrate Sinnar route networks as well as in other analysis.

Centroid

The village centroid is the center point of any village. We have used centroid tools to find out every village's centroid.

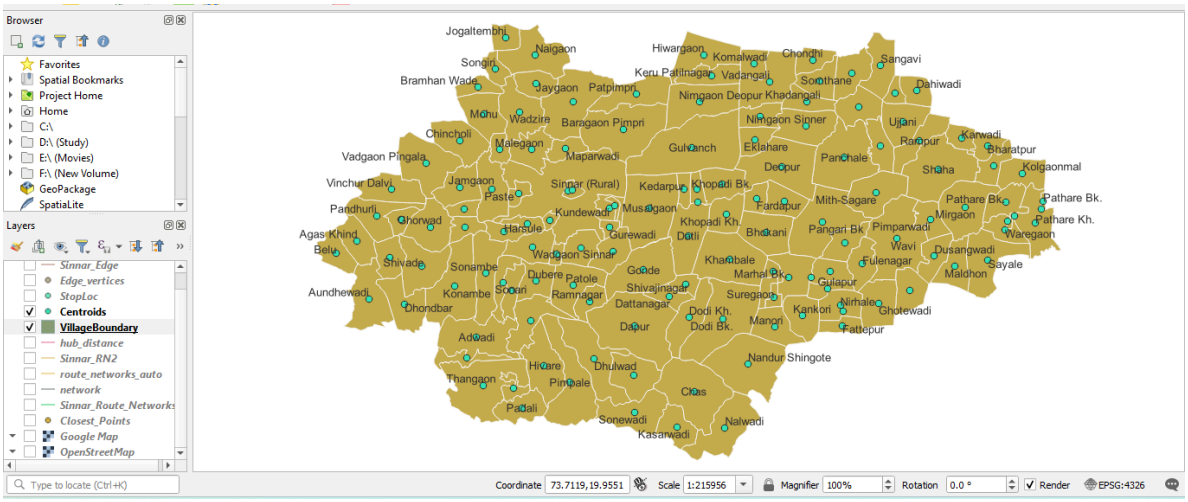


Fig: The sky blue color represents every village centroid

Hub Distance

Hub distance is the distance from the village centroid to the nearest bus stop. We have used hub distance tools to find out hub distance.

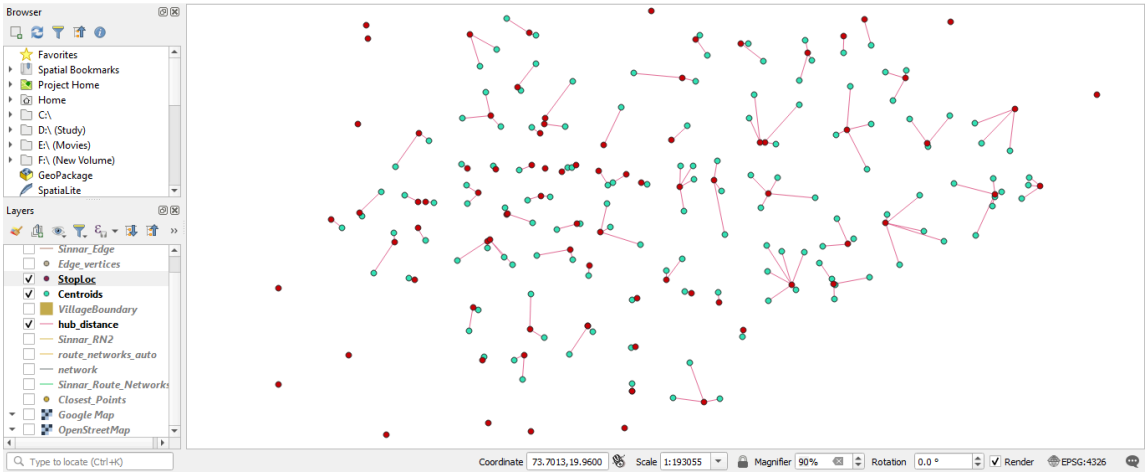


Fig: The line in between two vertices showing the hub distance between the village centroid and bus stop

Sinnar Route Network

To create a route network, we have used QGIS. In QGIS, there is a toolbox called “Network Analysis”. We can use this tool to obtain the shortest path of a route. We can use the shortest path (point to layer) to get the most straightforward way.

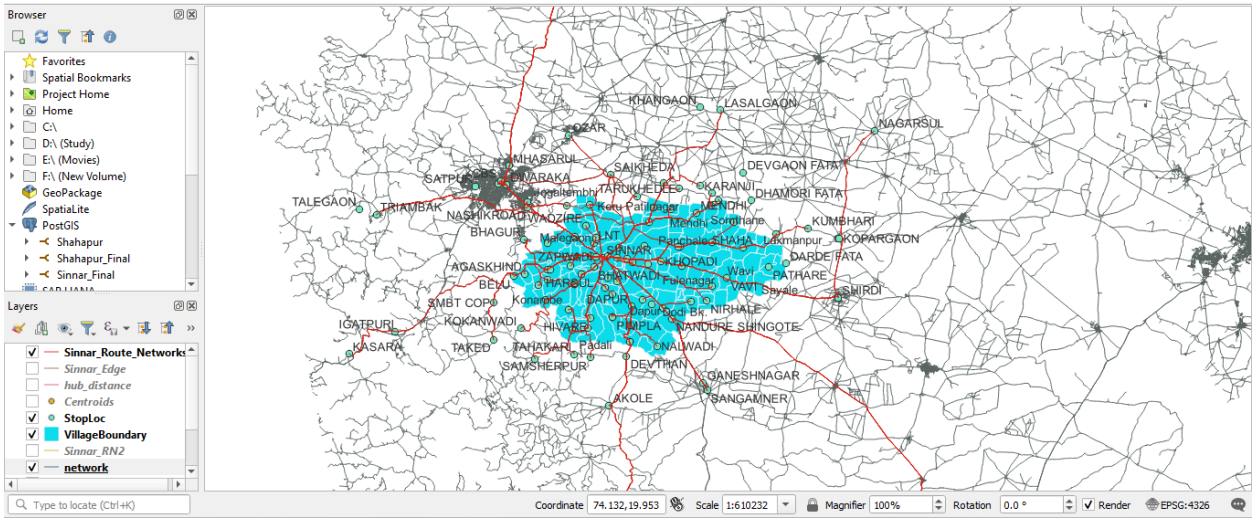


Fig: the combination of all three census data

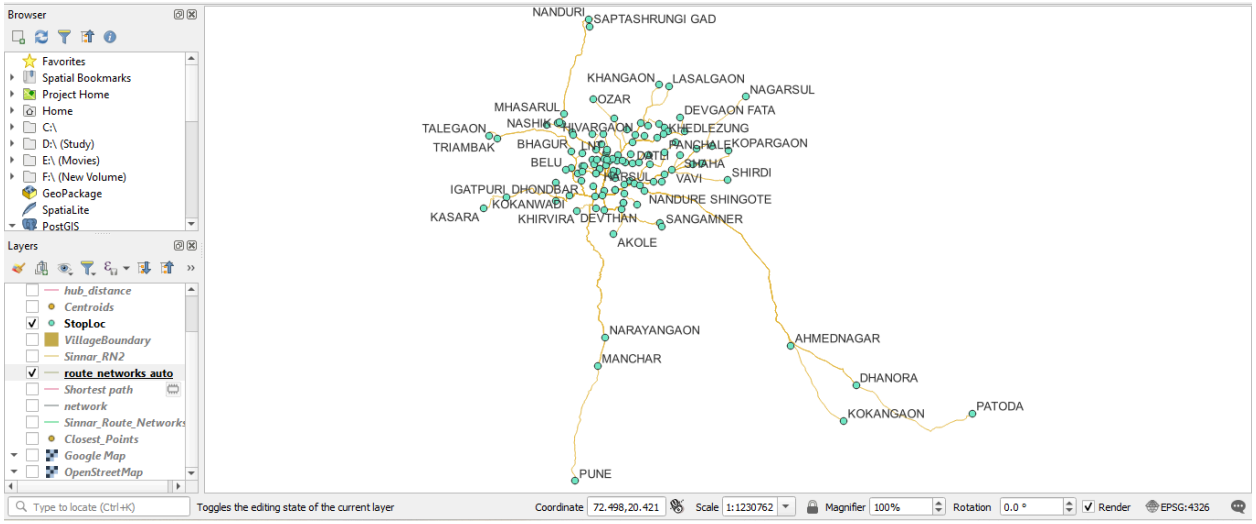


Fig: Generated road network using “Network Analysis tool”

The generated road network was created using the “Network Analysis Tool” shortest path. It helps to build road networks quickly; it needs to select the start and end points for making the network. After that, it will start processing the road network. But after creating the network, we found that the network is not completed, it has some missing roads already there in the form-4,

and some different routes were also created that are not mentioned in the form-4. So we decided to do it another way, described below.

Sinnar Route networks creation:

We recreated the Shahapur taluka road network using form-4, Road network census data, and Shahapur village census data using QGIS.

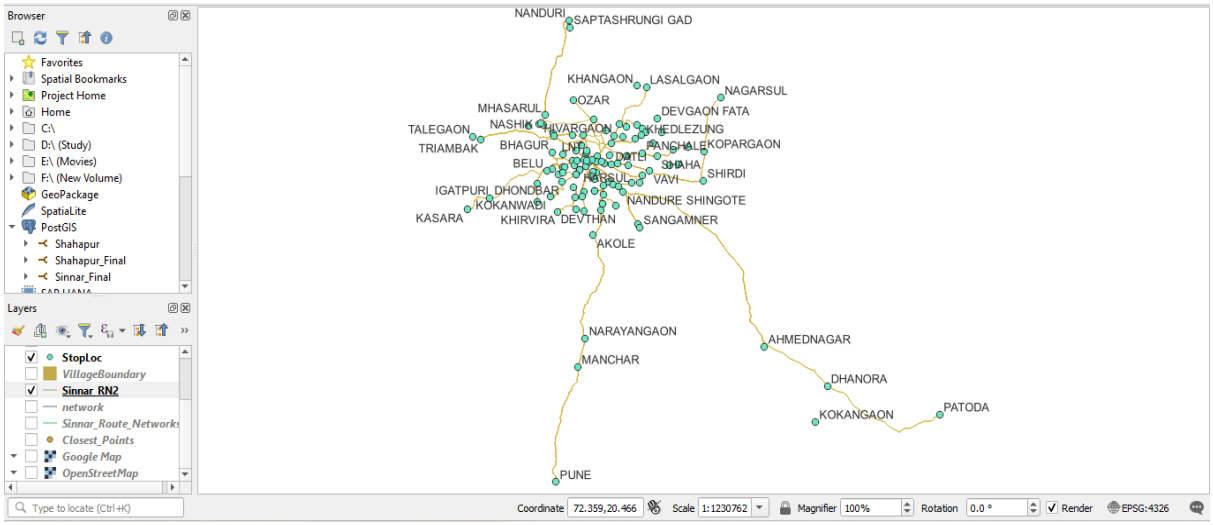


Figure: Generated Complete route networks of the route segments.

A network analysis processing toolkit was utilized to develop the road segment, and the digital geography of the route network was mainly produced using the Shortest path (point-to-point) analysis approach. Instead of using a single procedure, we adopted a batch process to speed up the process.

We can filter terminal locations by Shahapur taluka only and create those routes that lie only within Sinnar taluka.

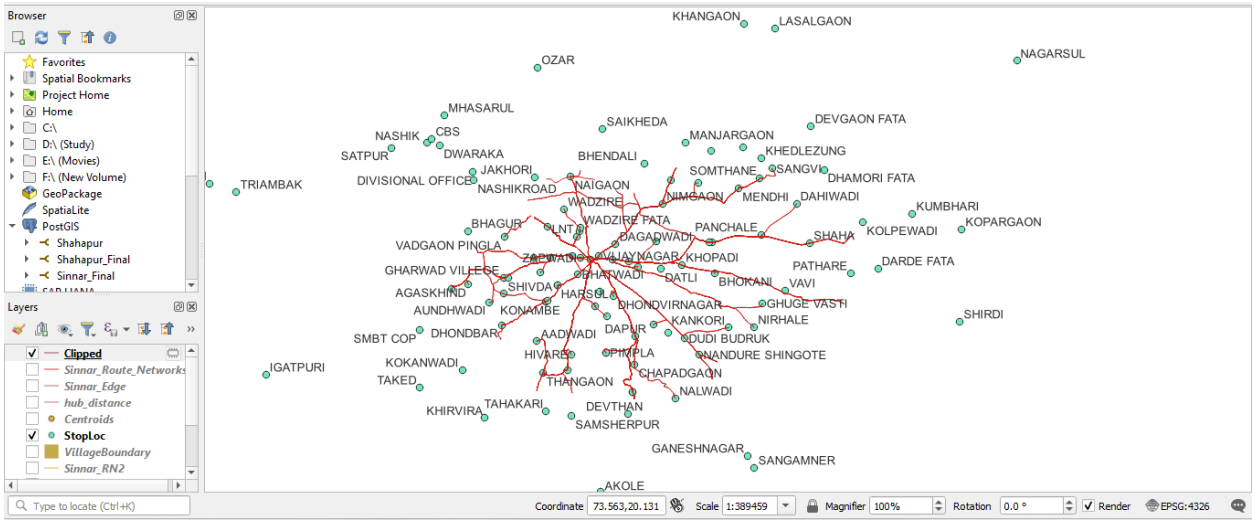


Figure: Route network for those who fall under Shahapur taluka

Edge Creation

After completing the route network, we can divide the route into the edge where any route overlaps. To create the edge, we need to use three tools.

1. Dissolve,
2. Multipart to single part
3. Line intersection

Dissolve

One of the most popular tools for generalizing data in GIS systems is the dissolve tool. The fundamental objective is to burn adjacent borders based on a shared characteristic. However, it can also be applied to developing multipart features for lines and points.

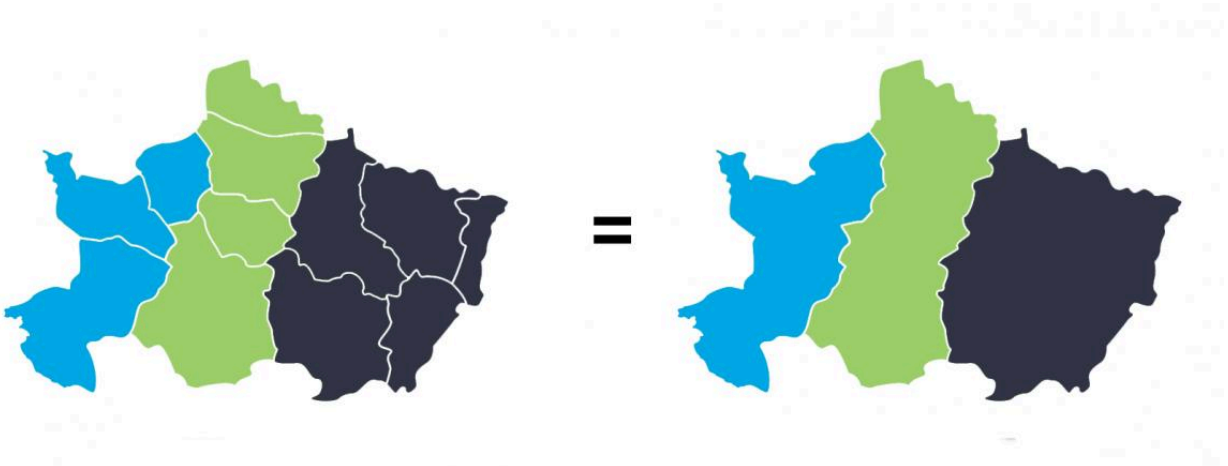


Fig: Dissolve (Left: Original, Right: After a dissolve, Image Source: Google)

Multipart to single part

Separates multipart input information to produce a new layer with single-part features. The output feature class will treat each component of a multipart input feature as a separate single-part feature. Existing single-part features won't be impacted.

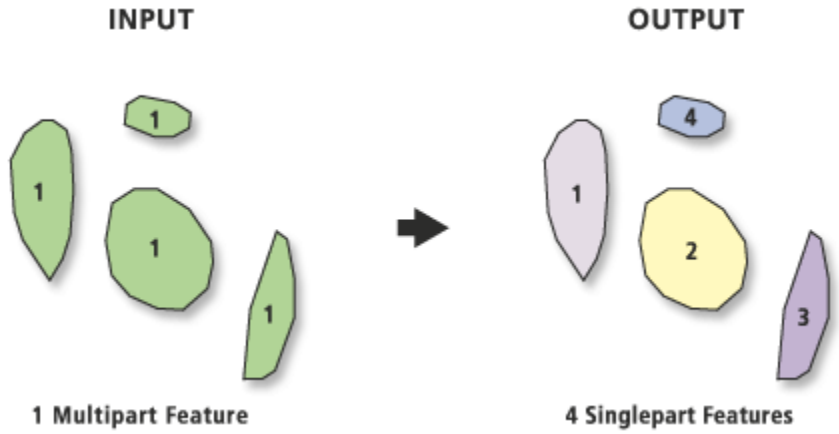


Fig: Multipart to single part (Image source: Google)

Usage:

- A line or polygon layer might be the input layer.
- All input layer names must be distinct.
- The output layer will keep the characteristics of the input features.
- An output layer will treat each component of a multipart input feature as a separate single-part feature. Existing single-part features won't be impacted.
- The output feature class will retain the characteristics of the input feature attributes. The output feature class will gain a new property called ORIG FID, filled with the input feature IDs.
- Use the Dissolve tool to rebuild multipart features from single-part features using a standard field value, such as ORIG FID.

Line Intersection:

The tool for line intersection split line characteristics are seen when lines intersect. The split operation uses the feature class's default attribute values to produce a new feature while updating the form of the old feature.

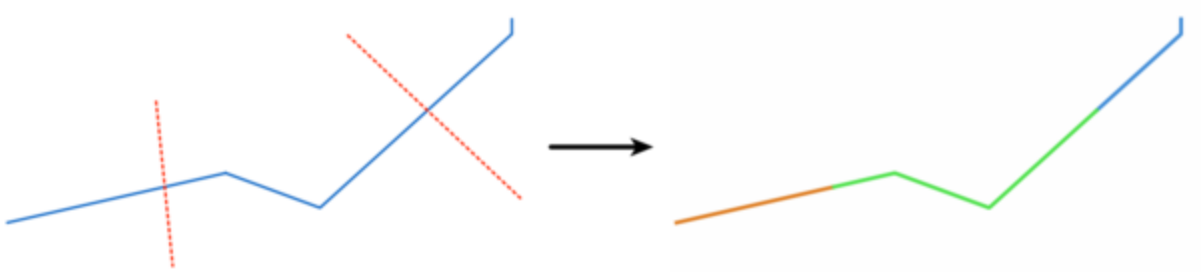


Fig: Line Intersection (Image Source: Google)

Sinnar Route Edge

First, we need to dissolve all the route segments to get a single polyline route network. After that, multipart to single-part tools will help us to create multipart input information to produce a new layer with single-part features. Finally, line intersection helps to build vertices for all those starts and ends of an edge.

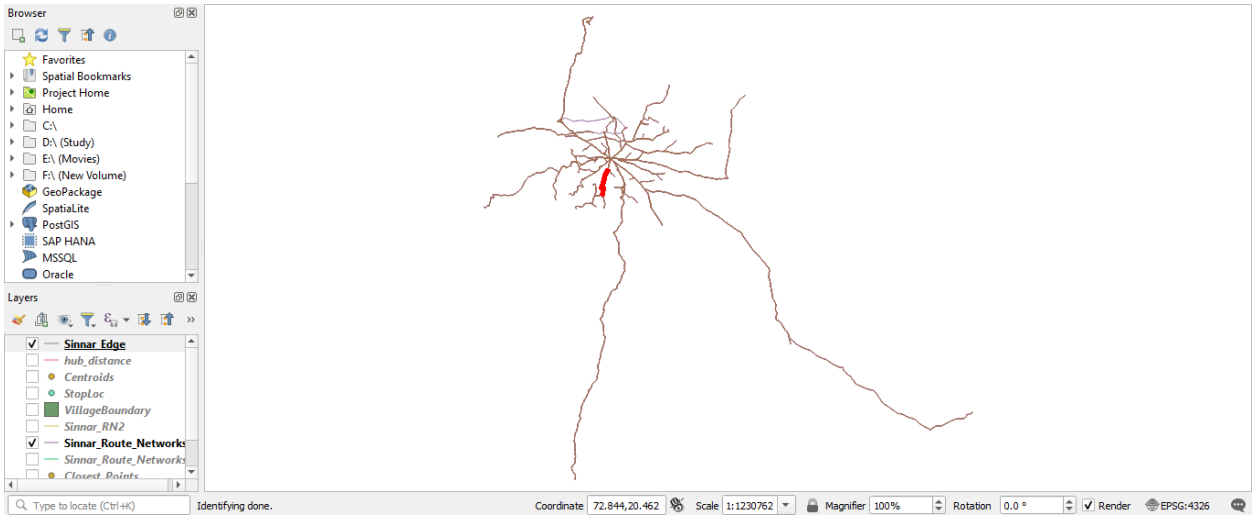
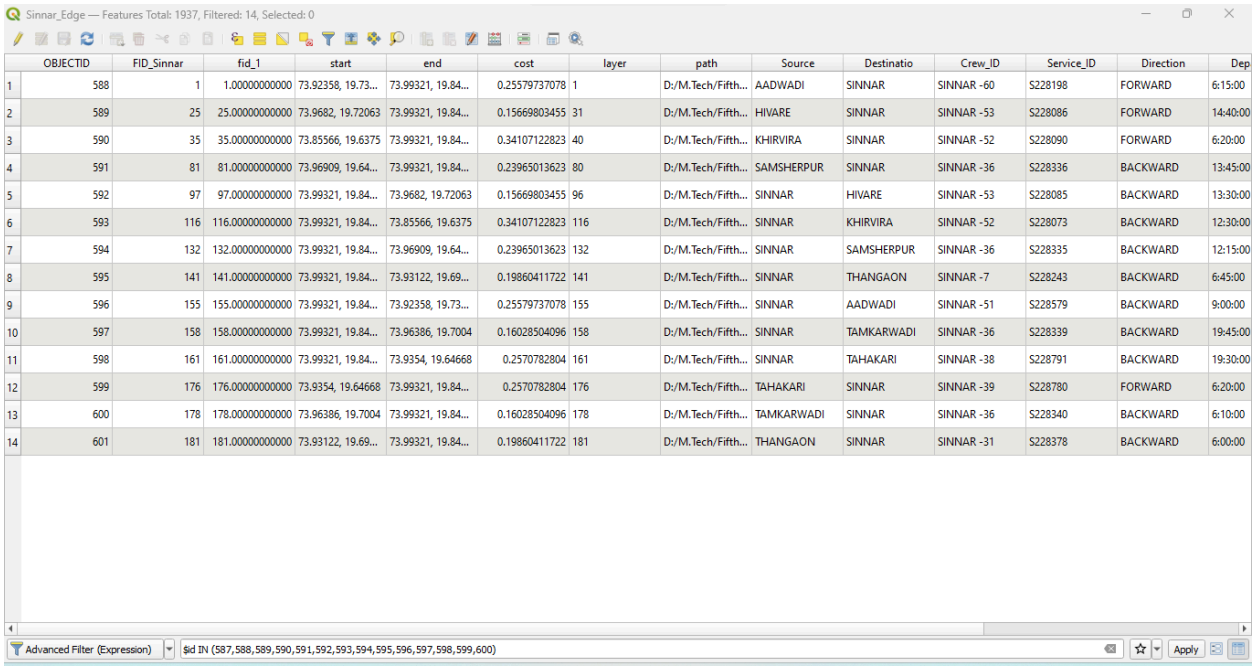


Fig: Edge of Sinnar route networks



	OBJECTID	FID_Sinnar	fid_1	start	end	cost	layer	path	Source	Destinatio	Crew_ID	Service_ID	Direction	Dep
1	588	1	1.00000000000	73.92358, 19.73...	73.99321, 19.84...	0.25579737078	1	D:/M.Tech/Fifth...	AADWADI	SINNAR	SINNAR -60	S228198	FORWARD	6:15:00
2	589	25	25.00000000000	73.9682, 19.72063	73.99321, 19.84...	0.15669803455	31	D:/M.Tech/Fifth...	HIVARE	SINNAR	SINNAR -53	S228086	FORWARD	14:40:00
3	590	35	35.00000000000	73.85566, 19.6375	73.99321, 19.84...	0.34107122823	40	D:/M.Tech/Fifth...	KHIRVIRA	SINNAR	SINNAR -52	S228090	FORWARD	6:20:00
4	591	81	81.00000000000	73.96909, 19.64...	73.99321, 19.84...	0.23965013623	80	D:/M.Tech/Fifth...	SAMSHERPUR	SINNAR	SINNAR -36	S228336	BACKWARD	13:45:00
5	592	97	97.00000000000	73.99321, 19.84...	73.9682, 19.72063	0.15669803455	96	D:/M.Tech/Fifth...	SINNAR	HIVARE	SINNAR -53	S228085	BACKWARD	13:30:00
6	593	116	116.00000000000	73.99321, 19.84...	73.85566, 19.6375	0.34107122823	116	D:/M.Tech/Fifth...	SINNAR	KHIRVIRA	SINNAR -52	S228073	BACKWARD	12:30:00
7	594	132	132.00000000000	73.99321, 19.84...	73.96909, 19.64...	0.23965013623	132	D:/M.Tech/Fifth...	SINNAR	SAMSHERPUR	SINNAR -36	S228335	BACKWARD	12:15:00
8	595	141	141.00000000000	73.99321, 19.84...	73.93122, 19.69...	0.19860411722	141	D:/M.Tech/Fifth...	SINNAR	THANGAON	SINNAR -7	S228243	BACKWARD	6:45:00
9	596	155	155.00000000000	73.99321, 19.84...	73.92358, 19.73...	0.25579737078	155	D:/M.Tech/Fifth...	SINNAR	AADWADI	SINNAR -51	S228579	BACKWARD	9:00:00
10	597	158	158.00000000000	73.99321, 19.84...	73.96386, 19.7004	0.16028504096	158	D:/M.Tech/Fifth...	SINNAR	TAMKARWADI	SINNAR -36	S228339	BACKWARD	19:45:00
11	598	161	161.00000000000	73.99321, 19.84...	73.9354, 19.64668	0.2570782804	161	D:/M.Tech/Fifth...	SINNAR	TAHAKARI	SINNAR -38	S228791	BACKWARD	19:30:00
12	599	176	176.00000000000	73.9354, 19.64668	73.99321, 19.84...	0.2570782804	176	D:/M.Tech/Fifth...	TAHAKARI	SINNAR	SINNAR -39	S228780	FORWARD	6:20:00
13	600	178	178.00000000000	73.96386, 19.7004	73.99321, 19.84...	0.16028504096	178	D:/M.Tech/Fifth...	TAMKARWADI	SINNAR	SINNAR -36	S228340	BACKWARD	6:10:00
14	601	181	181.00000000000	73.93122, 19.69...	73.99321, 19.84...	0.19860411722	181	D:/M.Tech/Fifth...	THANGAON	SINNAR	SINNAR -31	S228378	BACKWARD	6:00:00

Fig: Sinnar route networks edge attributes

Sinnar Village Data Analysis

Traffic analysis according to route no:

Data from Sinnar ETIM indicates that there are 228 routes. Many people use such routes each month. There are two such routes 7112 and 9678 where more than fifty thousand tickets were sold, but there is also one such route 114139 where only one ticket was sold.

	A	B
1	route_no	traffic_count
2	1143	6836
3	1155	5482
4	1383	1226
5	1470	5418
6	1482	5574
7	1511	9810
8	1514	9452
9	2128	4017
10	2129	130
11	2130	10966
12	2131	24
13	2197	22
14	2723	4118
15	2724	127
16	2725	12127
17	4891	37
18	7100	18626

Fig: Snapshot of route-wise traffic flow

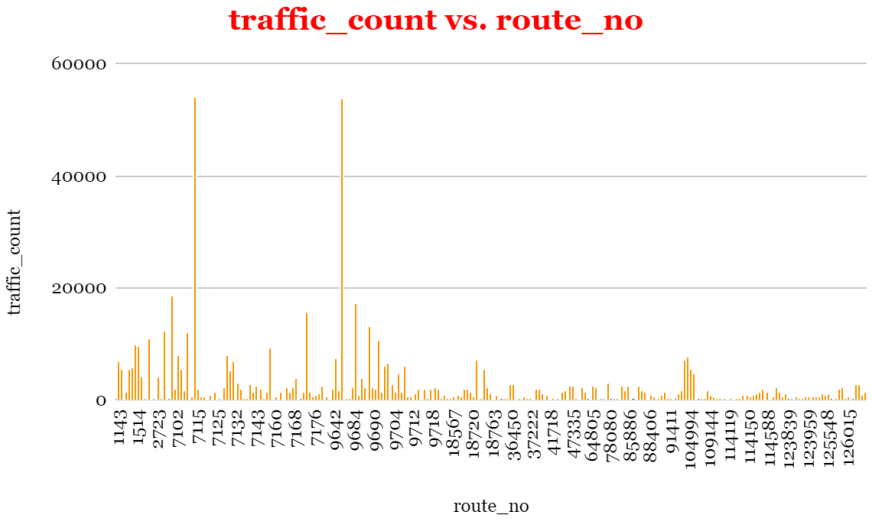


Fig: Route-wise traffic flow visualization

Issues with traffic analysis according to route no:

There are hardly any such routes where for a month, fewer than 10 tickets were sold. This is a bigger problem since such routes will lose money if too few tickets are sold on them. Also, this could not be a problem If those are intermediate routes.

	A	B
1	route_no	traffic_count
118	18791	4
153	80819	7
154	80821	4
189	114139	1
207	123840	2
225	126953	6
230		
231		

Fig: Routes where less than 10 tickets were sold

Traffic analysis according to source-destination:

In the ETIM data, 6204 source-to-destination roads were indicated. The most tickets ever sold for a source-destination were 28944 from all those roads from Sinnar (SNNR) to Nashik Old (NSKOLD) in a single month. According to ETIM statistics, these are the highways that are the most lucrative. In contrast to these, there are a few additional highways where more than 10,000 tickets were sold.

	A	B	C
1	from_stage_	till_stage_co	count
2	SNNR	NSKOLD	28944
3	NSKOLD	SNNR	27576
4	NSKO	SNNR	15627
5	SNNR	NSKO	14745
6	SNNR	TTAGSN	7833
7	SNNR	MLGSNK	6390
8	TTAGSN	SNNR	5763
9	SNNR	SNMBE	5538
10	SNNR	DBREWD	4437
11	SNMBE	SNNR	4387
12	SNNR	AKLE	3863
13	SMR	NSKCBS	3739
14	SNNR	PATOLE	3575
15	SNNR	SMR	3558
16	DBRE	SNNR	3405
17	AKLE	SNNR	3377
18	SMR	SNNR	3099

Fig: Ticket sold according to the source to destination

Inconsistent Route in Sinnar taluka:

Tickets may be given at some exceptional bus stops on a route even though they are not listed in the master data; those are known as inconsistent routes. Some routes mentioned in the ETIM are not listed in the master data. There are a few consistent routes in Sinnar taluka.

	A	B	C
1	route_no		inconsistence
4	75092		Not in master data
8	50842		Not in master data
0	42135		Not in master data
1	37625		Not in master data
07	52495		Not in master data
14	75091		Not in master data
38	7576		Not in master data
58	4891		Not in master data
63	41718		Not in master data
81	42136		Not in master data
84	75085		Not in master data
87	1383		Not in master data
70			

Fig: Sinnar taluka inconsistent routes

Issues with Data and Necessary Data

Data Issues

- **ABC data is not synchronous with ETIM data**. Since the number of trips operated in ABC varies from what is observed in ETIM.
- **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.
- **Form 4 data is not synchronous with any other data.**
- **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.
- **Form 4 data is obsolete** if we want to use it for observational purposes.

For better analysis

- Updated Form 4
- Exact termini location lat-long
- Updated Master data
- Consistent ETIM data
- GTFS format data

Moreover the data needed to be in GTFS format.

Conclusion

MSRTC is one of the essential services for rural Maharashtra's socio-economic development. The purpose of this report was to establish Sinnar taluka's Digital Geography in Nashik, Maharashtra. The process for developing digital geometry has been designed in such a way that the majority of it may be replicated by anybody who has little knowledge of it. Through a GIS-based graphical interface, this digital geography will assist us in better understanding the operation of the Shahapur Taluka bus depot. This Digital Geography is an attempt to digitalize the operation of Maharashtra Public Bus Transport and assess the process of the Bus Transportation System, which may aid us in making various decisions for the depot. Besides Sinnar taluka digital geography we tried various analyses on Shahapur taluka as well as the Sinnar taluka dataset that are available to us.

It is quite possible to scale up digital geography, and problems involving optimization, maximum coverage, and other graph theory problems that might directly influence decision-making can be worked out. The protocol has to be strengthened and the technique created earlier needs to be thoroughly tested in the field with people from diverse regions and demographic groups.

Future Work

1. Collection of updated MSRTC data to carry out a better analysis.
2. **Development of a web-based application:** If MSRTC develops a web-based solution, the traffic department's workload will be considerably reduced because it is a web-based platform that can be accessed from any computer system anywhere.
3. **Vehicle tracking system:** MSRTC is undertaking a Vehicle Tracking System at the District level. It can be studied and expertise can be expanded on the subject matter.

References

1. T. Berners-Lee, N. Shadbolt, “There’s gold to be mined from all our data,” The Times, 2012.
2. Elena Mishevska, Bojan Najdenov, Milos Jovanovik, Dimitar Trajanov, “Open Public Transport Data in Macedonia”, The 11th International Conference for Informatics and Information Technology, April 2014.
3. https://en.wikipedia.org/wiki/GTFS
4. https://transloc.com/blog/what-is-gtfs-why-does-it-matter-public-transit/
5. Sudhansu Kulkarni, “Analysis of public transport as a development service,” 2018.
6. Anshul Kumar, “GIS Framework for Taluka Bus Transportation Analysis and Provisioning,” 2020.
7. Ramya Sharma, “Data Analysis of the MSRTC public bus transport system,” 2022.
8. Sunny Johari: Evaluating State Bus Transportation, 2020