AFD – Kampala, Uganda / July 2020

Paratransit and street usage study in preparation for the BRT pilot project in GKMA

Objective 1.2

Final Report

File name	Version	Date	Changes	Study Director	Project Manager
1943_190-rep-ssd-1-final_report_obj1.2.docx	1	04.07.2020	-	J. Allaire	S. Saddier

Contact: Simon Saddier
simon.saddier@transitec.net

Transitec Consulting Engineers
28 rue Sainte-Foy, 75002 Paris, France
+33 1 43 48 36 59
www.transitec.net

Table of contents

Page

List of figures

Page

List of tables

Internal Project Glossary

Term	Acronym	Description
FR	-	Field Researcher
FRM	-	Field Research Manager
GIStoGTFS	-	A module of scripts developed by TfC with the purpose to convert GIS data into a GTFS feed
ODK Collect	ODK app	App used to collect identified raw trips and stages in the identification phase.
OpenStreetMap	OSM	A collaborative open source project enabling the creation of free editable maps.
ProcessedStop	-	RawStop, aligned to the OSM basemap manually and named.
ProcessedTerminal	-	RawTerminal, transformed into a polygon and named.
ProcessedTrip	-	RawTrip, matched to the OSM basemap using RB tool.
RawStop	-	Stop received from field using RO app
RawTerminal	-	Stage received from field using ODK app
RawTrip	-	Trace received from field using the RO app
Route	-	A journey with a specific origin and destination points.
Route Observer App	RO app	App to map out spatial and temporal data developed by TfC
Route Observer Dashboard	ROD	The platform from which mapping can be tracked and new mapping assignments are created
Route Snapper	RS tool	App developed by TfC used to digitize raw trips.
Service	-	A route along with its departure time. Services are what transit agencies offer.
Stage	-	Location where a route begins or ends. Stages are stops and Stage Names are UIDs.
Stop		Location where a route passes along the way and stops.
TerminalTrip	TT	An OD pair as identified by an FR during the identification phase of active routes.
Trip	-	A route mapped once from origin stage to destination stage.

1. Executive Summary

The Greater Kampala Metropolitan Area (GKMA) is faced with the rise of unregulated paratransit services due to a lack of formal public transport provision. The main mode of public transport in Kampala is the taxi, providing an estimated two thirds of passenger trips. Taxis are minibuses operated by individual private operators and managed by a diversity of associations, none of which are officially recognized by public authorities in Kampala. Boda-Boda motorcycle service provide the other third of passenger trips. A separate report deals directly with Boda-Boda services.

This report is part of the Paratransit and street usage study in preparation for the BRT pilot project in GKMA. It documents the data collection campaign and its survey protocols, faced challenges, data outcomes, data processing and analysis. The campaign consisted of Onboard Surveys, Frequency Survey's and Section Counts.

110 taxi stages were observed across the GKMA. An estimated **587 forward or return routes** operate out of these stages. The frequency of the majority of those was collected from the field. Video-based classified section counts were implemented to create a rapid demand assessment.

The data deliverables out of the data collection campaign consist of **GIS files** describing the transport network (routes, stops, stages, etc.); **excel files** with route frequencies, and a **GTFS feed** that enables trip planning and advanced analysis. The map deliverables consist of **Transit Maps** covering intracity and intercity services.

The report further documents capacity building workshops held to ensure that mapping knowledge is transmitted to local actors.

2. Introduction

The Greater Kampala Metropolitan Area (GKMA) is faced with the rise of unregulated paratransit services due to a lack of formal public transport provision. The main mode of public transport in Kampala is the taxi. Taxis are minibuses operated by individual private operators and managed by a diversity of associations, none of which are officially recognized by public authorities in Kampala.

The unmet demand from the taxi industry along with increasing levels of congestion led to the rise of unconventional modes of transport such as moto-taxis known as boda-bodas. This can be traced back to the collapse in the early 1990s of the bus transport services, namely Uganda Transport Company (UTC) and People's Transport Company (PTC), leading to the rise of informal service providers. Hence, there is a need for introduction of new public transport services in the GKMA.

This project aims to fulfill one of the objectives of the first stage of assessment in a three-stage approach developed by the Government of Uganda (GoU) to prepare for the implementation of a Bus-Rapid Transit (BRT) system. This first stage aims to conduct a preliminary assessment of the current state of the paratransit sector and other street usage. Through the mapping of the public transport network, the consultant will be responsible for the provision of informed operational data necessary to understand the paratransit services in the GKMA area.

For the purpose of our assignment, a review of existing relevant data was crucial. The consultant has conducted a systematic canvassing and collection of relevant PDF files, maps and GIS shapefiles with aggregated geographically referenced data. These include:

Table 1 Datasets surveyed as part of secondary data review

Source	Year	Description
ITDP UN-Habitat Mobility Mapping study	2016	PT Network, Stages and Stops datasets
Multi-Modal Urban Transport Master Plan for GKMA	2018	consulted for the PT Network, Stages, Stops and Modal Share figures
KCCA Taxi Origin Destination Survey Data	2019	
KCCA Taxi Parks	2014	providing the consultant with PT Stages
Atlas of Urban Expansion, Uganda Bureau of Statistics (UBOS), Database of Global Administrative Areas (GADM)	-	Different sources for Administrative Boundaries and Urban extent Border shapefiles

3. Objectives & Scope of Work

This report includes the following information:

- **Data outcomes** including Stages, Trips, Stops datasets (find the detailed deliverable files listed below) and the Schematic Map;
- **Faced challenges** during field research and post-field;
- **Field research strategy** used and a concise weekly time plan for FR performance;
- **Protocols** of Onboard Survey and Frequency Survey **protocols**;
- **Glossary** Detailed **glossary** of the terms used;
- **Methodology** of processing and analysis conducted with the field data.

The **spatial scope** of the Data Collection campaign is the GKMA. The GKMA is defined as all within 20 km radius of the City Center. A detailed definition of the urban extent to be surveyed was presented in the Inception report and can be found in the Appendix B: Definition of Boundaries of the GKMA.

The **temporal conventions** used throughout this study are:

- The morning peak period (MPP) is the time interval from 6:00 am - 9:00 am.
- The evening peak period (EPP) is the time interval from 4:00 pm - 7:00 pm.
- The off-peak period consists of three intervals: 5:00 am – 6:00 am; 9:00 am – 4:00 pm and 7:00 pm – 11:00 pm.

OffP_1	5:00 AM	6:00 AM
MPP	6:00 AM	9:00 AM
OffP_2	9:00 AM	4:00 PM
EPP	4:00 PM	7:00 PM
OffP_3	7:00 PM	11:00 PM

Figure 1 Time intervals for data collection

The Objectives of the field research and the different Surveys are to produce the following deliverables:

- GIS files with the raw instances from the field.
- GIS file with the processed instances from the field.
- GIS file with the unique routes comprising the network
- GIS file with the processed stops from the field
- Excel files with raw metrics (GIS attribute data) collected from the field.
- Excel file with the raw collected frequency data by route
- Excel file with the processed interpolated frequency data by route
- GTFS feed of the network
- Excel files with the processed estimates of occupancy collected from the field.

These Deliverables fall under Objective 1.2. of the project, which consist in collecting data and mapping taxi routes in GKMA, with the view to further the knowledge basis available to Ugandan authorities on the characteristics of the existing public transport supply. This data will be provided in OSM and GTFS format to inform upcoming evolutions in the paratransit regulatory regime.

4. Data Collection Methodology

4.1 Field Research Strategy

The field research was divided into cycles, each consisting of two weeks:

- **Morning-Shift Week**: A week where all shifts are in the morning, covering the MPP. The shift is 6 am – 12 am.
- **Evening-Shift Week**: A week where all shifts are in the evening, covering the EPP. The shift is 2 pm – 8 pm.

Assignment: An assignment consists of either a frequency count survey or the mapping of a trip assigned to the FR through a mobile app installed on GPS-enabled smartphones.

A local mapping firm "Map Uganda" was contracted by TfC to provide Field Researchers (FR) and a Field Research Manager (FRM) as well as a Coordinator to facilitate field research activities and to maintain a working office in Kampala. An independent consultant (Lydia Letaru) from Makarere University was also hired for data processing and validation.

The **roll-out of the field research capacity** was progressive starting with 5 FRs in the first week, followed by 10 FRs the second week and full capacity of 20 FRs was reached for the following weeks. With each new introduction of FRs or a new survey, a training session was held with TfC members, the FRM and the FRs in question in order to familiarize the FR team with the tools necessary and the appropriate survey protocol.

An **automated assignment strategy** was adopted: through the mobile Route Observer (RO) app developed by TfC, the FR was able to pick an assignment (with detailed instructions for an Onboard Survey or a Frequency Survey) available near their location. Each trip generates a single assignment in a drop-down menu in the RO app per time period (MPP, EPP, Offpeak) per for survey type (onboard and frequency).

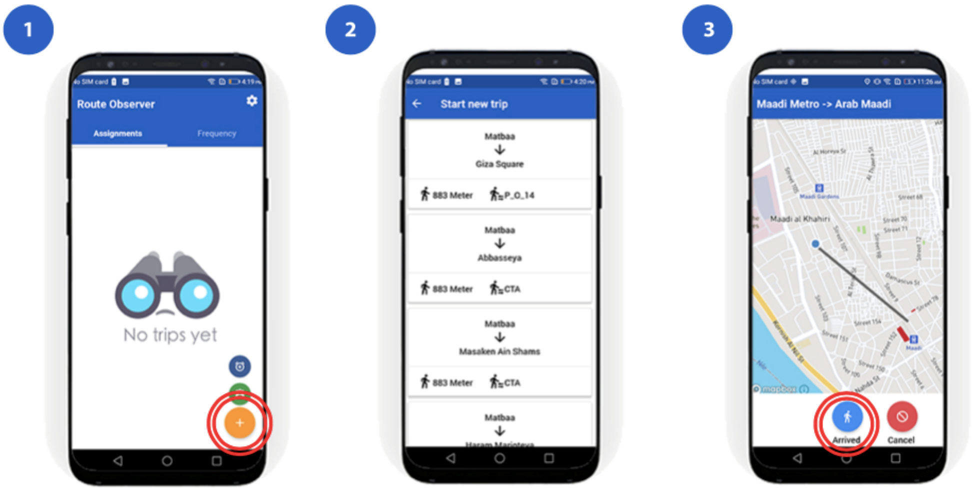
The RO Dashboard had the option to prioritize routes for both Onboard and Frequency assignments. This option was used in the final cycle of Data Collection in order to disable the routes which were neither attempted for the Onboard nor Frequency Surveys, so as to prioritize the routes with incomplete datapoints necessary for analysis purposes.

In addition, manual management of targets was adopted with the help of the FRM: in the second cycle C2 of Data Collection, the FRM communicated to the FRs to complete at least 4 Frequency counts a day, in order to boost Frequency Survey performance and incoming data.

4.2 Survey Types

- **Onboard Survey:**

The Onboard Survey assignments are available on the mobile Route Observer (RO) app and the FR picks an assignment from the available ones near his location. Here are the first few steps the FR is asked to follow:



Click on the "+" button on the main menu and choose the Onboard survey button

Pick an onboard assignment from the RO app from a drop-down menu, listing available assignments nearby and click "Start"

Go to the origin stage indicated and press "Arrived" once you identify the trip

Figure 2 The first steps in choosing an onboard assignment

The FR is then asked to mark when s/he boards the vehicle, when the vehicle boards and to enter the number of passengers at departure. At each stop, the FR enters the number of passengers who boarded and alighted. When the destination is reached, the FR indicates it on the RO app in order to end the assignment. Any additional notes by the FR may be entered at the end.

In the case where the route is not found or not completed, the FR also indicates so on the RO app and proceeds to pursue another available assignment. The FRs were in constant contact with the FRM who advised them on how to proceed in specific cases when help was needed.

- **Frequency Survey:**

Similar to the Onboard Survey, the Frequency Survey is available on the mobile RO app through a dropdown menu based on the location of the FR. For each trip and each time period (MPP, EPP, Offpeak), a Frequency Survey with a duration of an hour was generated.

The FR heads towards the stage indicated in the assignment and for the duration of one hour, marks the arrival and departures of different vehicles for the same route observed. The FR also marks when a vehicle starts boarding and the occupancy of departed vehicles.

- **Section Counts:**

Field researchers were deployed at ten critical points to conduct corridor counts along Jinja road, Entebbe road and Kampala road. The corridor counts are used to estimate modal share as well as daily passenger flow, when combined with route frequency data, along the Pilot BRT corridor. The number of manual 1-hour corridor counts per time period. For each selected location, at least one hour was counted for each time period.

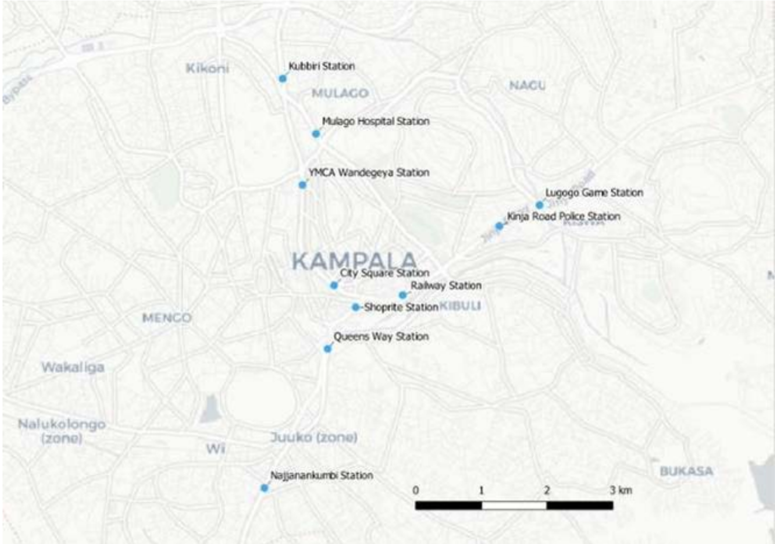


Figure 3 Section Counts locations

4.3 Weekly FR Performance

The weekly figures for both the Onboard and Frequency Surveys are detailed below:

Week (Project)	13	14	15	16	17	18	19	20	21	22
Week (Year)	3	4	5	6	7	8	9	10	11	12
Date	Jan	Jan	Jan	Feb	Feb	Feb	Feb	Mar	Mar	Mar
	12	19	26	2	9	16	23	1	8	15
TIME PERIOD	Whichever	MPH/EPH	MPH/EPH	MPH	EPH	MPH	EPH	MPH	EPH	
identificaiton (ID) or frequency? Mixed with onboard		**ID**			**Frequency**			**Frequency**		**TOTAL**
weekly Average number of active FRs	**5.75**	**10**	**18.6**	**18.6**	**19.8**	**18.8**	**19.6**	**17.6**	**18**	
Onbaord instances per FR, per day	4.04	4.86	4.17	3.51	2.44	1.01	0.6	1.65	2.51	
Instances Mapped per week	93	243	388	326	242	95	59	145	226	1817
Valid Instances Mapped (per	74	216	323	261	182	77	43	128	188	**1492**
# MPP onboard instances	24	51	31	119	0	39	0	68	0	308
# Offpeak onboard instances	66	145	238	201	97	55	30	77	154	1063
# EPP onboard instances	3	47	119	6	145	1	29	0	72	419
Frequency counts per FR per day				2.2	3.74	5.03	5.27	3.74	3.69	
Total Frequency counts made (per week)				205	370	473	516	329	332	**2225**
# man-days used per week	23	50	93	93	99	94	98	88	90	**728**

Figure 4 Weekly FR performance

The total Onboard instances mapped of 1817 exceed the planned total of 1725 onboard trips. The total of **valid** onboard trips constitutes about 86% of the target total. The total of frequency counts also comfortably exceeds the planned 1500 instances by 725 additional instances.

Redundancy:

Out of the 587 routes, 149 routes (approximately, **25% of the total**) were mapped for a total of at least **4 instances** while 55% of the routes were mapped at least twice (a total of 323 routes).

TRANSITEC Transport for Cairo مواصلة للقاهرة

5. Data Outcomes

5.1 Summary

Field research activities commenced on the 13th of January 2020 and were completed after 9 weeks on the 13th of March. A team of 20 FRs spent over 3200 hours in the field collecting data all over the GKMA. Data collected can be summarized per survey as such:

Table 2 Summary of Data Collected

Survey	Outcome	Data Points
Identification	- 110 Stages - 444 TerminalTrips	- Stage boundaries - Stage capacity - Where vehicles are parked (inside or outside a stage cluster) - Size of route's queue at the time of observation
Onboard	- 1817 RawTrips - 1492 ProcessedTrips - 587 Unique ProcessedTrips (routes)	- Itinerary of Trips - Travel time duration - Boarding and Alighting figures - Location of stops - Fare
Frequency	- 1685 observation hours - 379 routes observed	- Number of vehicles at the beginning of the observation period - Occupancy for vehicles that were already loading before observation started - Occupancy of each vehicle on departure - Loading time for each vehicle - Arrival & departure times for each vehicle
Section Counts	- 80 Hours for 9 observation points	- Classified counts of bodas, taxis and buses passing through during observation

5.2 Stages & Terminal Trips

Stages are defined as any area where transit trips start and/or end. The three main characteristics of stages are its boundaries, common name, and capacity.

5.2.1 Preliminary Identification

A preliminary stages identification was conducted by the TfC Team using both the data collected during the Pilot mapping and secondary sources including:

a. ITDP Mobility Mapping study data;
b. OpenStreetMap (OSM) database of bus terminals;
c. Multi-Modal Urban Transport Master Plan for Greater Kampala Metropolitan Area data;
d. KCCA Taxi Parks 2014 GIS layer.

From the trips data of the ITDP & Master plan databases we identified origin and destination points. These points were overlaid with stages polygons from other sources such as OSM and KCCA to see which stages are in common. We also double checked the existence of stages in specific areas using interactive panorama imagery provided through Google's Street View and Mapillary services.

A total of 25 main stages were identified mainly from the Master Plan data, which included a lot of transit stages. This list can be found in Appendix A: List of Stages.

5.2.2 Field Identification

The identification process was performed iteratively throughout the duration of the project. Field Researchers were equipped with an open-source application (ODK Collect) that allowed them to capture coordinates and fill in a form for each route queue they found at that terminal. The form included the following data:

- Location:GPS Coordinates
- Origin
- Mid-Stop
- Destination
- Vehicle type
- Queue size (range)
- Queue Photo

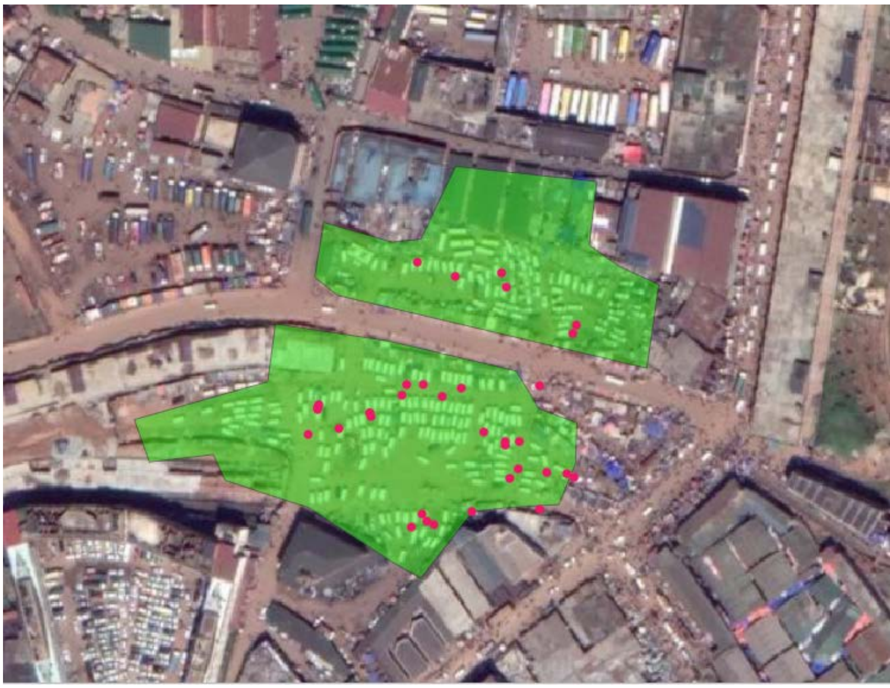


Figure 5 TerminalTrip points (red) overlaid with a digitized "Kisenyi" stage (green) and satellite imagery

The data output of this exercise is a list of Route Identification points, named in the database as "**TerminalTrips**". These are later used for digitizing the stage on GIS as well as create Onboard and Frequency assignments for further field research surveys.

5.2.3 TerminalTrips Standardization

Table 3 Duplicate TerminalTrip points after standardization (Green) & a discarded point that has a transit stage as a destination (Orange)

Raw Origin	Raw Destination	Processed Origin	Processed Destination
City Square	Kira	City Square	Kira Town
City Square (behind centenary Bank)	Namugongo Stage	City Square	Namugongo
City Square (behind centenary Bank)	Namugongo	City Square	Namugongo
Kirombe	Luzira	Kirombe Taxi Stage	Luzira Stage
Old taxi park	Kiruddu hospital	Old Taxi Park	

In order to get a proper list of assignments, there is a need to standardize the raw input from the "TerminalTrips". Field Researchers give different names to route origins and destinations. This means that there might be a duplication of routes. Also, some of these routes might be ending up at transit stages, which are not actual stages but major stops along a given route.

Origin & destination names are standardized based on the list of digitized stages; thus, a stage needs to be digitized first so that its associated "TerminalTrips" could be standardized and converted into assignments. Routes that end up at transit stages are discarded.

5.2.4 Assignment creation

The standardized list of TerminalTrips is periodically uploaded to the Route Observer Dashboard (ROD), which generates assignments accessible in the Route Observer (RO) app. Redundancy is defined for each route and for each time interval. These time intervals are pre-defined as per the field research strategy. So, for example, given a taxi route going from Kirombe Stage to Luzira, and three time intervals (6 am to 9 pm; 9 am to 4 pm; and 4 pm to 7 pm) and a redundancy of 2 for each, these inputs will generate a total of 6 assignments.

Therefore, the sampling strategy is controlled through the ROD by controlling time intervals and redundancy figures for each route.

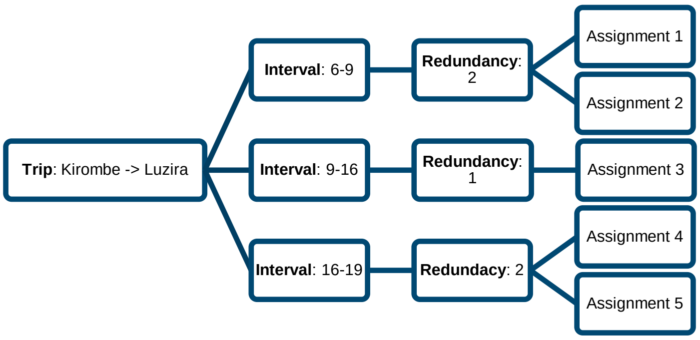


Figure 6 Assignment generation example

5.3 Trips

"Trips" as a dataset encompasses different information that include geography, time, demand and other information (fares etc.). Data is integrated from Onboard and Frequency Surveys to fully describe Trips.

5.3.1 Geography

A Trip in terms of geography is a line (itinerary) starting from an origin stage and ending up at another stage, this information is captured during the Onboard Survey. When a field researcher picks an Onboard assignment and starts recording on the RO app, the app records both the GPS trace of the trip as well as travel time.

Another data point captured is boarding and alighting figures at specific points along the trip. Whenever passengers board or alight the vehicle, the field researcher records the location of this stop and the number of people boarding and alighting. This data point is called “Raw Stops” and each stop is timestamp-ed so that the boarding and alighting figures can be co-related with location and time to measure demand.

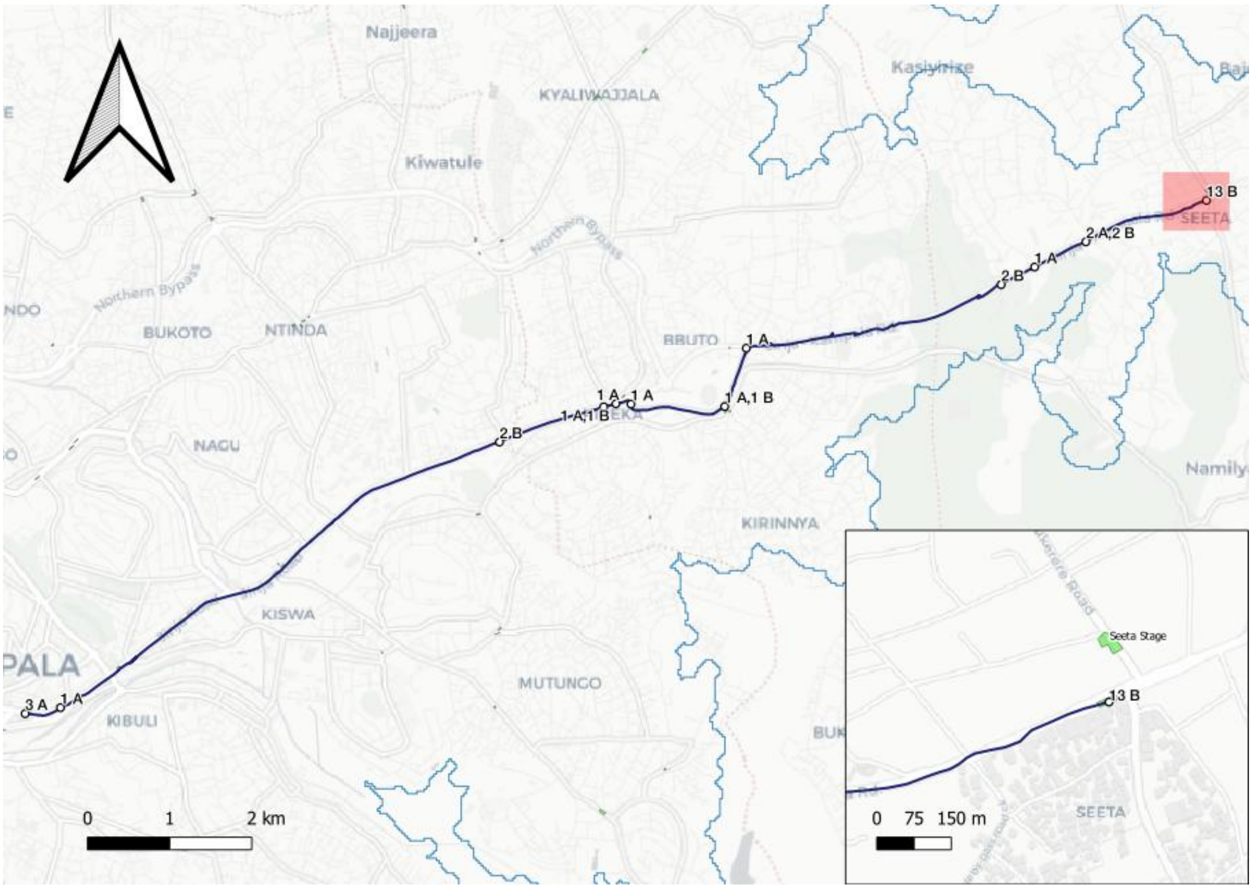


Figure 7 Raw trace of a trip from Seeta to downtown, along with raw stops. Each stop has total number of people alighting (A) and boarding (B)

5.3.2 Time

Temporal aspect of a Trip includes information such as:

- Travel time, captured during Onboard Surveys for different periods (travel time is combined with spatial attributes to estimate commercial speed)
- Frequency, captured during Frequency Surveys for different periods

Travel time has been aggregated to two-hour periods starting from 6:00 AM to 8:00 PM, these values are projected onto segments where each segment has a point of origin and a point of destination.

Figure 8 Commercial speed across segments going from Namanve to Seeta on Kampala-Jina Rd. Values are for 9:00 AM - 11:00 AM period

Frequency data is also aggregated for the same time intervals for each route to come up with headway estimates. Combining headway and travel time data allows building virtual schedules for these paratransit routes.

5.3.3 Processing of Raw Trips

Each "Raw Trip" resulting from an Onboard Survey received is validated and processed individually. Route Snapper, a software developed in-house by TfC, allows for batch processing of incoming raw traces and migrating the processed validated to the "Processed Trips" layer in the database.

Validation determines which traces qualify for processing and which ones need to be mapped again and thus, re-assigned to field researchers. The validation statuses are outlined in the table below.

Table 4 Trips Validation Statuses

Code	Status Name	Status Description
VLD	Trip Valid	The trace matched the origin and destination assigned.
ODM	Origin Destination Mismatch	Trips with endpoints not matching the origin and destination of the assignment. Some traces can overshoot the assigned destination. This is usually the case when drivers find demand for a stop beyond the planned destination point.
TNV	Trip Not Valid	Trips where the GPS trace is inaccurate. The GPS trace can cut off for a few minutes making it hard to trace the path the field researcher took, rendering the trace invalid.
INC	Incomplete	Trips with a partial fraction of the itinerary registered. This is typically due to either vehicles breaking down in the middle of the trip or drivers ending the trip early to avoid congestion.
NTA	Not a trip	This is the trace of a trip that the FR mapped but was not initially assigned.

The only traces considered for processing are the ones assigned a “VLD” or “ODM” status. The others are discarded.

Aligning raw traces received from the field to the OSM road network is done systematically alongside the Data Collection. During this process, repeatedly identified road changes from raw traces which are not visible on the OSM network will be added, after confirming them with satellite images or infrastructure changes information received from the Government of Uganda.

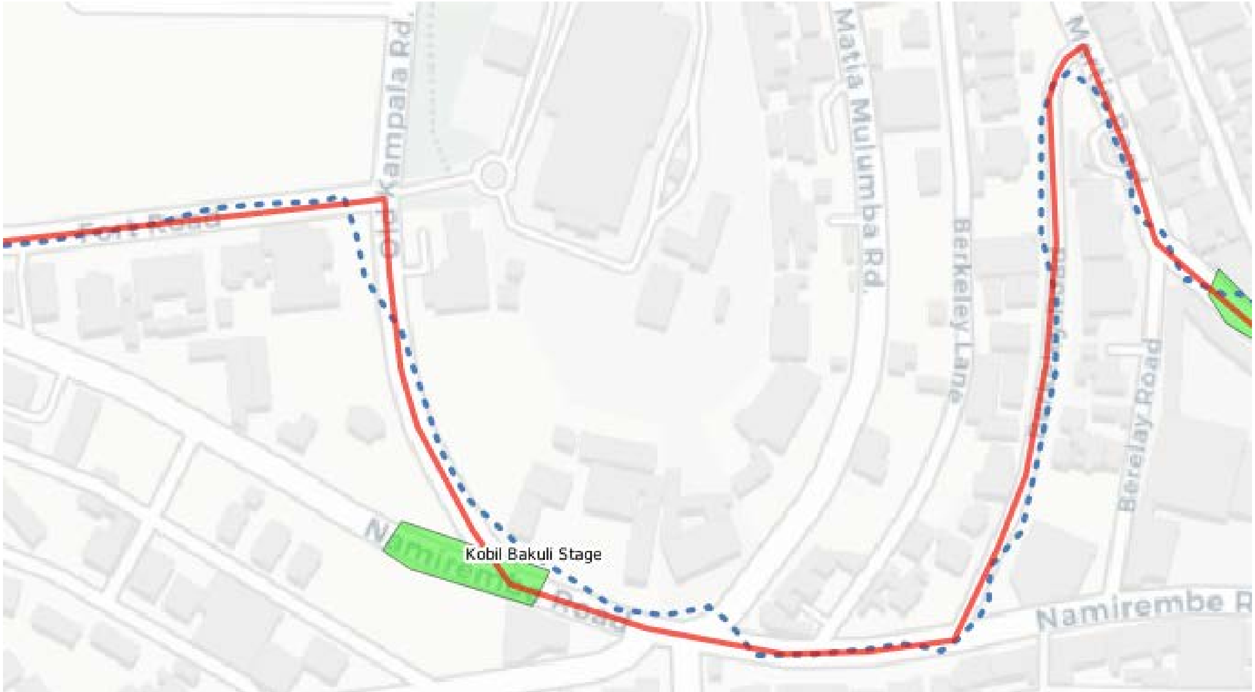


Figure 9 Raw Trace (Dotted Blue) overlaid with its processed trace aligned with OSM (Red)

5.3.4 Generating Route IDs

Each Route ID consists of three components: the agency ID, a geographic component indicator and a serial number:

- **Agency ID:**

The agency ID is an indicator for trips’ vehicle types and service operator.

Table 5 Agency ID codes

Code	Name	Description	Capacity
TX14	Minibus Taxi	Informal – transport minibus taxi	14
P50	Pioneer Bus	Informal – Pioneer bus agency	50
A30	Awakula Ennume	Informal – Awakula Ennume agency	30

- **Geo-component:**

Table 6 Trips Geo-codes

Trip	Geocode
The OD pair lies within the same district	KA for Kampala WA for Wakiso MU for Mukono
Inter-district trip	IC

A code is assigned to each trip based on which districts the Origin Destination (OD) pair of the trip lies in. There are three districts in the studied GKMA: Kampala, Wakiso and Mukono.

- **Serial number generation:**

We generate the serial number using the following script. After creating a matrix with all possible stage pairs, we go over the following steps:

1. Fix a pair of stages.
2. Group trips whose OD pair is the chosen stage pair by agency_id and direction. There can be more than one trip with the same OD pair, agency_id and direction; they may have distinct itineraries. In this step, we determine a list of *heaviest trips* in this trip group.
   a. If we have **only two trips in the same agency-direction group,** then compute their geographic intersection. If this intersection is 70% or higher, pick either trip to add to the list. Otherwise, they are both added to the list as separate trips.
   b. If we have **more than two trips in the same agency-direction group,** then:
      i. Pick the trip which intersects all the others most. We will call this trip the *heaviest trip*. To pick the trip: for each instance, we calculate the intersection percentages with other trips and get an average intersection percentage. We pick the trip with the highest average intersection percentage to be the heaviest trip. Add this chosen trip to the list.
      ii. Compute the intersection percentages of all trips in this agency-direction group with the designated heaviest trip.
         1. If min intersection >= 70%, skip the following step.
         2. Otherwise, that is, if min intersection <70%, get the average percentage and split the trips into two groups according to whether their intersection numbers exceed 70% or not. Repeat steps (a) and (b) for each group. Note that if a group consists of a unique trip or two trips, we pick whichever to be the heaviest trip in step (a). This process ought to terminate of course.

At this point, we get a list of heaviest trips among trips with the same OD pair, same direction and agency_id.

3. For each trip, get its intersection with the trips in the list of heaviest trips. Pick the heaviest trip which intersects it most and group it with it. We then get trip groups with the same OD pair, same direction and agency_id. Do the previous process with the trip group with the same OD pair, agency and opposite direction to form trip groups.
4. Pair trip groups (groups with the same trip id) with different directions together to create routes:

Calculate the intersection of the heaviest trips from each trip group in origin direction (O) with the heaviest trips from the return direction (R) and match O trip groups with R trip groups. For example:

Table 7 Merging Trips into Routes

	Trip C (direction R)	Trip D (direction R)
Trip A (direction O)	35%	70%
Trip B (direction O)	60%	25%

Note: if the number of trip groups in one direction is not equal to the number of trip groups in the opposite direction, we form the maximum number of pairs we can using the matching process mentioned above in the table.

5. The paired trip groups forming a route are given the same serial number. The serial number itself is not unique, it is a counter starting from 001 for each geographic component. The trip groups which are not paired are also given a distinct serial number.
6. The trip_id is the same as the route_id with an additional component for direction: an appended O or R.

5.3.5 Unique Trips & Routes

After each mapped instance of an active trip is processed, a unique processed trip is chosen for the creation of the GTFS feed. This is done by calculating for each instance from the trip groups with identical trip id, its intersection with the other instances and we attach to it an average intersection number accordingly. The highest average intersection number corresponds to the instance which will be chosen for the unique trip layer.

	Instance A	Instance B	Instance C	Instance D	Avg Intersection
Instance A		78%	82%	92%	84%
Instance B	78%		76%	80%	78%
Instance C	82%	76%		84%	81%
Instance D	92%	80%	84%		85%

Table 8 Determination of the UniqueTrip layer

5.4 Stops

Creating a virtual network of stops represented as points is a necessity for modeling any transportation network. This dataset contributes to GTFS creation and, as aforementioned, segmentation of itineraries to project travel time. Stops digitization is divided into:

- **Placement,** in which a clustering of Raw Stops occurs to give indication on the location of actual stops,
- **Naming:** the most frequent name occurring in a cluster is chosen. If this name is not available, we resort to the nearest landmark and finally, the team's local knowledge.
- **Additional Description:** this is appended in the case that the stop name itself does not suffice to describe it. In the case where a stop is duplicated to accommodate to different road directions, the terms (Eastbound, Westbound, Northbound, Southbound) are appended to the stop name to indicate the direction of trips that pass by it.

A lot of stops have been integrated from ITDP's 2016 data which was helpful in determining naming as well as placement for areas that overlapped this study's study area.

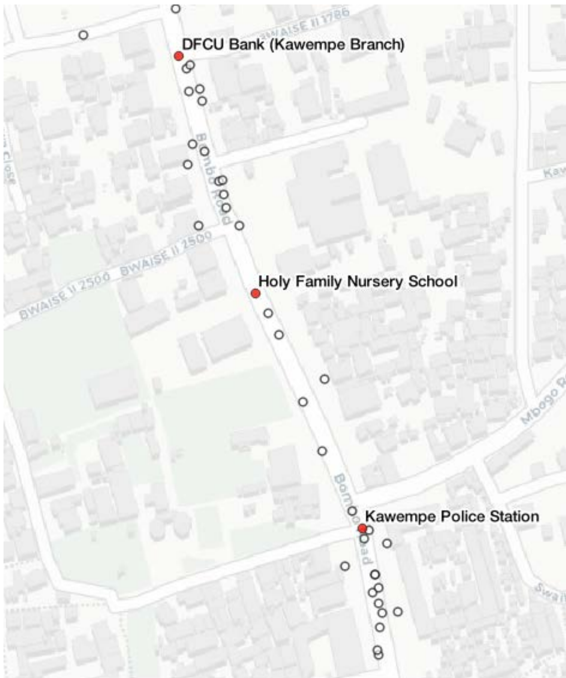


Figure 10 Digitized stops (Red) based on Raw Stops clusters (White)

5.5 GTFS Feed

General Transit Feed Specification (GTFS) Feeds were created from the GIS layers and gathered temporal information mentioned above. GTFS is a data format to describe transportation networks in terms on location and time.

- Trips & stops are based on a complete GIS database & finalized data collection.
- Headway and Schedules estimations are based on collected Frequency Surveys.
- Travel time between stop pairs are based on big data traffic sources, calibrated with GPS data from the field.

The generated feed is ready for implementation. It can be used by local trip-planning app developers, students, urban planners and researchers to plan journeys, compute accessibility or assess different scenarios. Below is an example of how the GTFS feed can be used to produce isochrone maps, illustrating travel time to different parts of the city from a central location using public transport.

Figure 11 Travel-Time reach polygons from city square, Kampala at 9:00 AM using public transport

5.6 Transit Map

5.6.1 Base Map

The Kampala Bus Map followed a hybrid approach between a geographic and a schematic one, having more aspects of a geographic map. The base map was designed to follow the geographic arrangement of the city, while abstracting and simplifying it. This design choice was made based on three aspects:

1. This is a bus map; thus, a geographic approach is more valuable than a schematic one

A geographic map does not distort distances, whereas the schematic maps are usually for readers mainly interested in connectivity and not geography, making it more suitable for rail-based transport such as the subway or Monorail.

2. The nature and dynamics of the streets and accordingly the trips

This is where the « abstraction » element in schematic map design comes in handy and helps relieve the map from too much complexity arising from Kampala's dense network

3. Geographic Transit Maps are more easily readable and relatable, especially for users who are not used to transit maps

5.6.2 Parishes

The administrative boundaries of the parishes are used as a base for the map. These parishes were grouped into categories, each category is drawn as one parish polygon. Ex.: Kololo I + Kololo II + Kololo III are shown as one polygon with the name "Kololo".

5.6.3 Streets

The streets shown on the map are divided into four categories based on the street hierarchy, each category has a different opacity level based on the street's importance. Regional Roads are darker than secondary roads for example. Only the most important streets (according to the scale of the map), as well as the streets on which the trips are placed, are shown on the map.

5.6.4 Routes

To be able to design a comprehensive and readable map, routes need to be aggregated to be readable on the map. This was achieved through two techniques:

1. **Routes filtration Process:**

All routes of the map went through a 5-step filtration process, enabling us to arrive from approx. 600 routes to approx. 139 routes.

Filtration Process	Description	Number of Routes
0. Without Filtration		~ 600 routes
1. Parent vs Child	Through a script, all routes (children) that lie within a longer route (parent) having a 70% identical path and operate on the same direction were removed	235 routes
2. Less than 3 km	All routes less than 3 km were removed	223 routes
3. Route categorization of routes with same OD	All routes having the same Origin-Destination (OD) are grouped together. Each group is seen as one route, however having several diversions	223 routes in 133 categories (some categories contain only one route, some several routes)
4. Categories' filtration	All duplicate routes (having the same OD) that pass by the same stages or within a 500m buffer from the same stages are removed	149 routes in 133 OD categories
5. Similar routes filtration with different OD	All similar routes (having different OD but within 500m buffer) that pass by the same stages or within a 500m buffer from the same stages are removed	139 routes in 104 OD categories (intercity: 60 routes in 40 categories / intracity: 79 routes in 64 categories)

2. **Dividing the map into two maps:**

- **Intracity map:** showing routes within inner Kampala
- **Intercity map:** showing routes that originate or arrive at Inner Kampala, but head to or come from the greater Kampala region.

5.6.5 Stages

The stages shown on the map are either origins or destinations of the filtered routes. This enabled us to filter the stages from 120 stages to 75 stages.

Stages located in Central Kampala area and have proximity to one another, are grouped together. This helped in avoiding route cluttering in this area, as many the routes originate or terminate in one of those stages.

- **Intracity map:** Based on the scale of the map, all central Kampala stages are combined
- **Intercity map:** Based on the scale of the map, only Old Taxi Park and City Square are combined

5.6.6 Style and Colors

To differentiate the routes, two techniques are used, line style and line color. The different line styles represent the two different route agencies (taxi and bus).

As for the colors, each different OD category is being represented with a different color. To try and minimize the usage of similar shades of colors, a route grouping technique is implemented. All routes under which ALL the following criteria apply, are being grouped together in terms of line quantities (several route lines with the same trunk are being combined into one line) and color:

- Origin of the routes are the same
- Routes are geographically allocated in the same direction (North, west, east...)
- Routes share a common trunk (where several lines can be combined into one)
- OD categories that already contain several routes are excluded

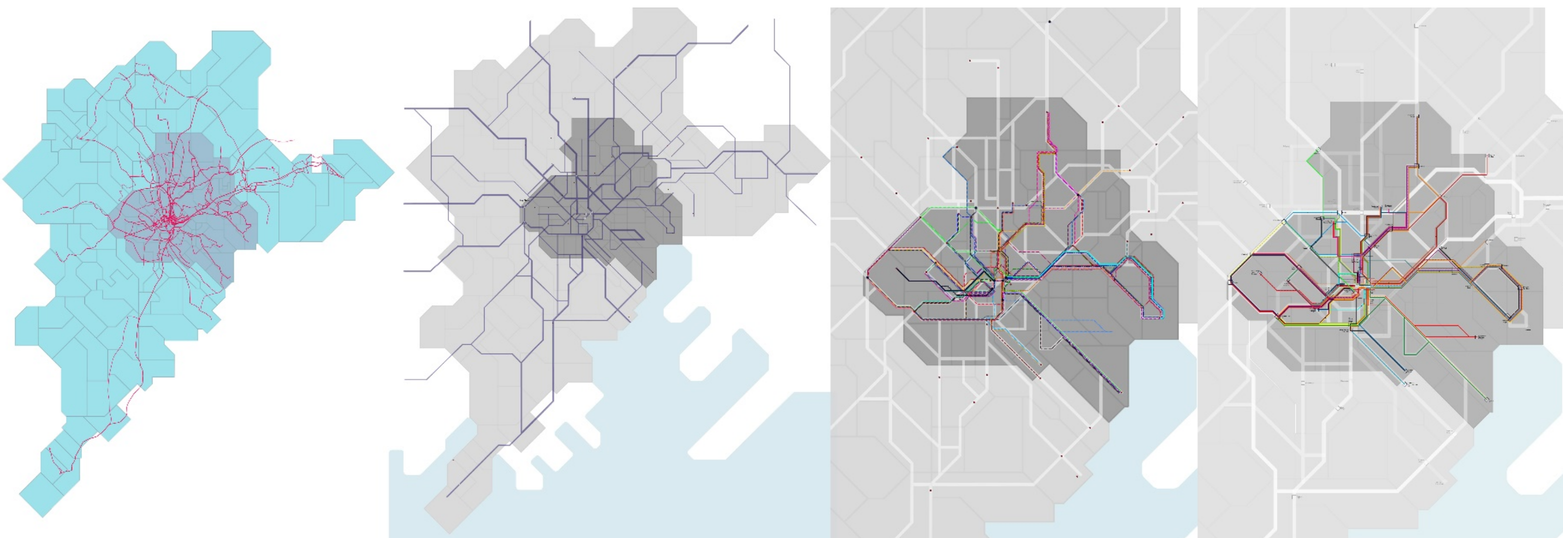


Figure 12 Kampala Transit Map Progress

5.7 Dissemination

5.7.1 DigitalTransport4Africa

DigitalTransport4Africa (DT4A) is part of the Digital Africa ecosystem, a collaborative initiative designed for African Entrepreneurs. DT4A has a rich resources section with a Knowledge center, Open tools, documentation and Open data.

Processed data is to be published on the platform under a Creative Common (BY) license which means users may share, copy, change and build on the dataset as long as proper attribution is given to the data creators and owners.

5.7.2 Uploading Data to OpenStreetMap

Sharing transit data on the OSM platform will help ensure data sustainability as well as empowering the local OSM community. The processed GIS layers of both routes & stops are to be extracted, transformed and loaded to OSM.

OSM handles public transport data in the form of "nodes" for stops and "relations" for routes. However, the OSM community are also weary of uploading data in bulk due to multiple reasons, one of which is data duplication. To avoid duplication, an extract from OSM within the study area will be downloaded locally and compared on GIS with the processed GIS layers for proximity and similarity in itineraries. Only unmatched data from the processed layers will go through to the transformation process. Transformation and loading of data are done based on the process and guidelines outlined by OSM and using JOSM software.

6. Faced Challenges

Some challenges were faced during the field research activities which, while disrupting the consultant's workflow, also provided useful insights into the workings of the public transportation sector in GKMA.

Primary input was the feedback received from the field researchers whether through the RO App "notes" section, or direct communication with the FRM. Also, feedback from the data processing made some patterns visible on the network scale.

6.1 Field Research Challenges

When an FR finishes an assignment on the mobile Route Observer (RO) app, additional notes may be entered. Below are some of the most occurring notes categorized based on the main cause of challenge faced.

6.1.1 Traffic

- The taxi driver chose to end the trip without reaching the destination due to a small number of passengers or in order to avoid traffic jam

As show in Figure 13, A trip was intended to go from "Nansana Yesu Amala" Stage to Makarere University. Due to traffic congestion the driver chose to end the trip at "Nakulabye Market" which is another stage about 2 km away (orange dotted line) from the intended destination.

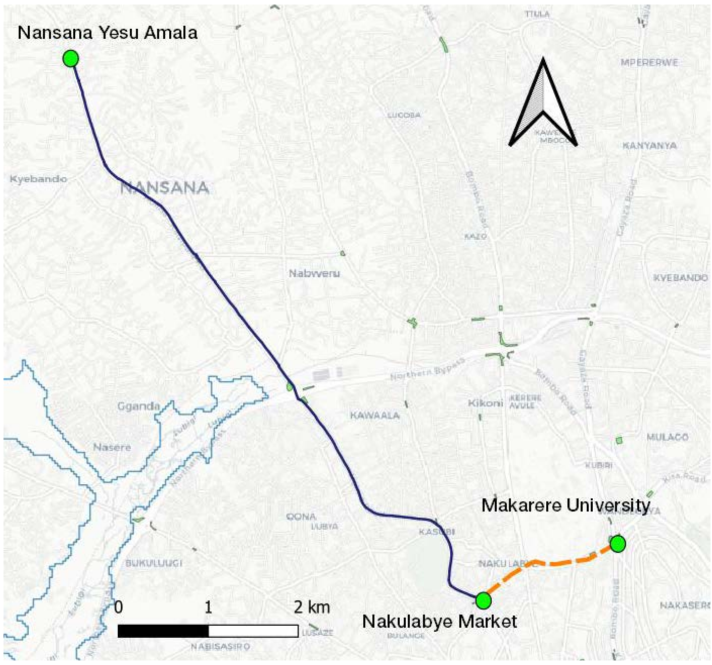


Figure 13 Incomplete trip due to traffic congestion

- Traffic jam causing driver to opt for a shortcut.
- The driver does not enter a taxi park but rather stops at a roadside opposite the park.

6.1.2 Drivers

- Driver overshoots the assigned destination.
- This is usually due to passenger demand and it's an ad-hoc decision taken by the driver.

- The driver loads passengers and proceeds to another stage after the final destination of the designated trip.
- Driver uses the destination stage as a "Transit" stage for another partial trip, responding usually to increased demand on the partial trip's road segment.
- Taxi driver going somewhere changes course midway to go to a completely different stage instead and exchanges passengers with another taxi.
- Reasons for this behavior ranged between personal reasons (driver was on the phone with someone), change in demand, or traffic jam.
- Taxi driver stopped by traffic officers due to lack of papers.

6.1.3 Vehicle and Road infrastructure

- Taxi breaking down in the middle of the road.
- More than one occasion of a tire exploding or engines overheating, indicating a lack of routine maintenance and rather a reactive approach taken by taxi owners.
- A destination stage was moved due to construction work.
- Weather conditions such as rain delaying trips and limiting passenger demand.
- The origin stage being a transit stage at the given time period of the assignment.

Figure 14 Namungoona Stage moved from area outlined in red to queue underlined in blue, due to construction works

6.2 Inconsistent Routes

Some services in the network were inconsistent in multiple ways, observations can be summarized as follows:

- **Irregular destinations:** Trips may sometimes end at one stage, sometimes end at another nearby stage and sometimes the driver would choose to completely change course and go to an irrelevant stage.
- **Rapid Change in supply:** At well-organized stages such as Old Taxi Park, services are somewhat regular in terms of frequency and supply. This is not the case for other stages where supply is completely left for drivers and stage supervisors to organize. Frequency data showed important differences between different days in the same time periods at these stages.
- **Itinerary variation:** Discussed in more detail in "7.1.1" – trips do not systematically follow a similar itinerary every time

6.3 Stages definition

In Kampala, there are main stages that are designated by KCCA and have regular services originating from them. There are, however, what is commonly called "Transit" stages: stages where taxis stop to load passengers during ongoing trips. Defining which is an actual "main" stage and which is a "transit" stage was challenging for the following reasons:

- Some stages have regular services and act as main hubs whilst not necessarily being designated by KCCA
- Some drivers choose to start loading passengers at transit stages, making that stage as the origin rather than a mid-stop. This behavior without a clear pattern makes it harder to define whether this stage is an actual stage or not
- A definition of a route is that it must start and end at a stage. Some observed routes would end up at transit stages before or after their intended destination stage due to change in demand or congestion
- Following up on the last point, when this route behavior was a pattern, this meant that there might be "stages" that act only as points of destinations. For example, a trip coming from Kitoro Taxi Park (Entebbe) to Usafi Taxi park will almost always end at "Shoprite" downtown. But Shoprite itself is not a main stage.

7. Data Analysis

7.1 Route Characteristics

7.1.1 Trips Variation Index

Variation of trips itinerary was measured based on a similarity index relying basically on the length of a common trip. Comparing multiple trips together means we must compare similarity of each trip to all other trips within the same group (Saddier 2017[1]). Grouping was based on origin, destination and agency. The trip variation index (TVI) is calculated as:

$$TVI_r = \frac{\sum_i^{t_r}\sum_j^{t_r} P_{i,j}}{{t_r}^2 - t_r} \quad \forall j \neq i$$ where

$P_{i,j}$= proportion of trip *i* that is shared with trip *j* (length of trip *j* shared with *i* divided by total length of trip *i*),

r = routes, and

t_r= number of trips for a given route.

A TVI of 1 means trips are completely identical to each other, while 0 means trips do not share any common itinerary segments even though they have the same origin and same destination. Out of 1492 processed trips, 720 had a TVI between 0.9 and 1 while 6 trips had a TVI between 0 and 0.1.

TVI mean for all processed trips is 0.831 showing an overall similarity in itineraries for routes. Comparing this to the unique list of trips (without taking redundancy into account) gives a similar result 0.818.

This high similarity could be because Kampala's main corridors are well defined, and most routes originate from around Old Taxi park and City square, thus trips spend most of their itineraries on these main roads with occasional shortcuts.

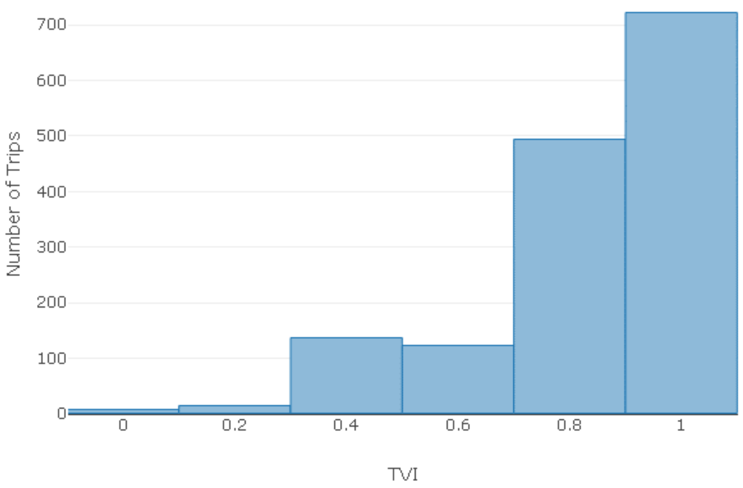


Figure 15 TVI value range distribution for processed trips

An example is shown (Figure 16 Speed profile for Two Trips with low TVI) where two trips (KA021 & KA024) from a group with a TVI of 0.46 are compared. Both trips were taken between 5 and 6 pm, but with around 20 minutes in total travel time difference, mainly due to the change in itinerary. Number of recorded stops was similar, 5 & 7 respectively, but demand may still be a factor in choice of itinerary rather than travel time only.

[1] Fickle or Flexible? : Assessing Paratransit Reliability with Smartphones in Accra, Ghana. Simon Saddier, Zachary Patterson, Alex Johnson, and Natalie Wiseman

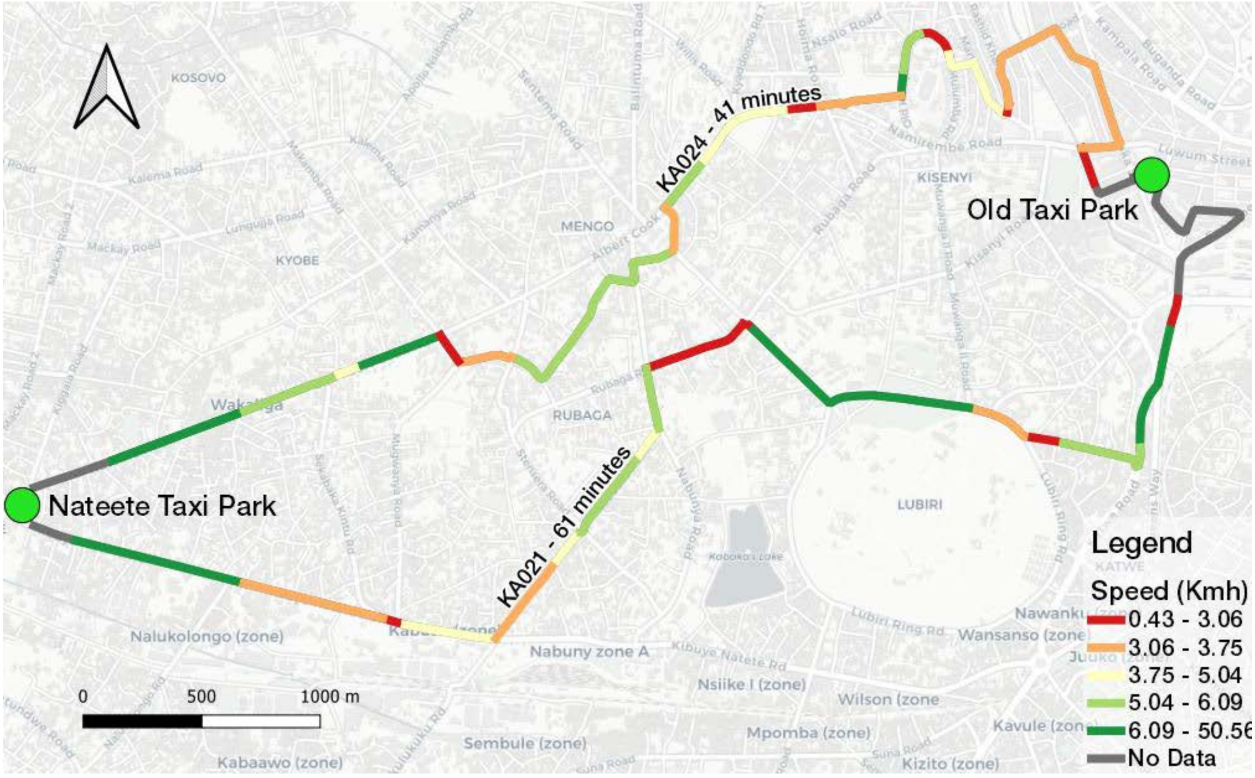


Figure 16 Speed profile for Two Trips with low TVI

7.1.2 Route Length

The estimated length of the route in km is used in further calculations in this section in the rotation analysis. A simple classification of routes based on length using the 3-quantiles is as follows:

Table 9 - Route length classification

Route Length	Class
0 km - 7.4 km	Short
7.4 km – 13.5 km	Medium
>13.5 km	Long

The average route length is 11.9 km with the shortest route of 1.2 km being “Bwaise – Kawaala Junction” and the longest route reaches 38 km being “Nateete Taxi Park – Kitoro Taxi Park”.

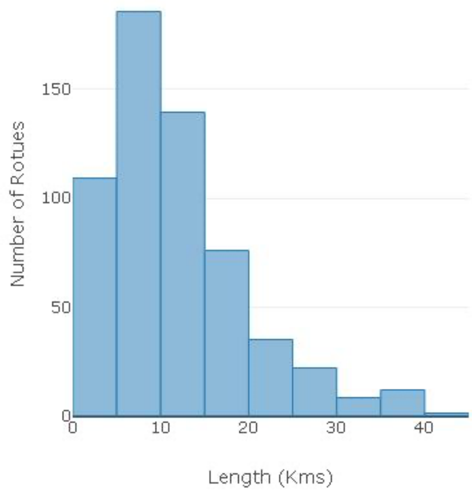


Figure 17 Routes length classification

7.2 Frequency and Headway Analysis

Grouping and Stratification:

A preliminary step of grouping the routes based on Origin, Destination and Agency yielded a total of 364 routes. Following this grouping, multiple strata were created for the routes for which there is collected frequency count in at least one peak and another count in the off-peak (these constitute our preliminary sample, which is 82% of all the routes), for the purpose of the headway analysis:

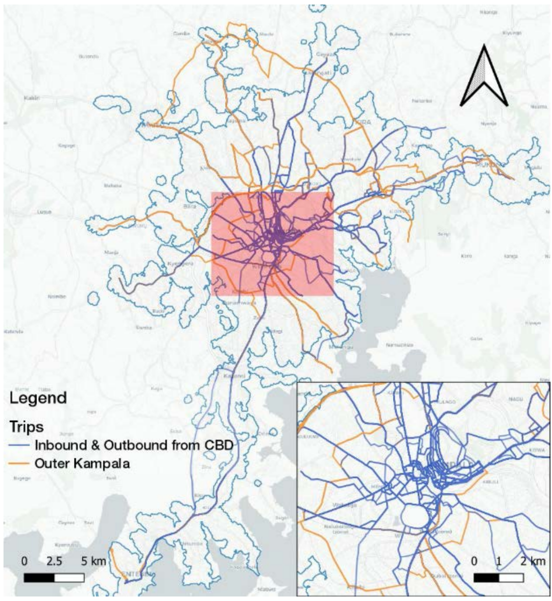


Figure 18 Spatial Levels of trips within GKMA

Table 10 - Spatial stratification of the routes

Spatial Level	Description	Number of Sample Routes
Inbound	Route reaching the CBD.	49
Outbound	Route originating from the CBD.	38
Outer	Neither Origin nor Destination is in the CBD.	211
Total		**298**

For each of the sample routes, an average daily headway is obtained from all the collected frequency instances, which leads to the second level of stratification:

Table 11 - Temporal stratification of the routes

Estimated Average Headway	Frequency	Number of Sample Routes
0 - 5 mins	High	84
5 mins –10 mins	Medium	114
>10 mins	Low	100
Total		**298**

For each of our sample routes, an average headway per peak and in the off-peak is also obtained separately. Thus, for the 9 created strata, the following estimates are based on these averages:

Table 12 - Estimated headway figures per strata

Frequency	Spatial Level	Average MPP headway (mins)	Average EPP headway (mins)	Average OPP headway (mins)	Number of Sample Routes
High	Inbound	4.3	3.5	3.6	17
	Outbound	4.4	4.8	4.0	3
	Outer	3.7	3.8	3.3	64
Low	Inbound	12.1	11.2	13.8	11
	Outbound	22.0	15.7	14.9	25
	Outer	15.9	12.4	14.3	64
Medium	Inbound	5.7	9.7	6.7	21
	Outbound	11.5	5.6	8.0	10
	Outer	7.8	7.1	7.6	83

Based uniquely on the spatial level, the estimated average headway per time period is detailed below:

Table 13 - Estimated headway figures based on spatial level

	Average MPP headway (mins)	Average EPP headway (mins)	Average Offpeak headway (mins)
Inbound	7.4	8.1	8.0
Outbound	12.6	8.7	9.0
Outer	9.1	7.8	8.4

The final step of this analysis is to assign these estimates for all of the routes based on the strata they fit into.

7.3 Passenger Flow Analysis

To understand daily ridership activity for different routes, we perform a daily passenger flow analysis using both our frequency and onboard collected data from surveys.

Approach:

Input Data	Mode of Acquisition
Number of departures per peak period per route	From the frequency and headway analysis above.
Average number of boarding passengers per peak period per route	From the recorded numbers of passengers boarding at the origin stage and at each stop along the route (with the assumption that no passengers board at the destination stage). These figures are obtained from the onboard data.
Seat turnover	Average number of boarding passengers per capacity.
Output Data	Average ridership per PP per route. Daily ridership per route.

From our frequency analysis, we estimated the number of departures per route for each direction during the MPP, EPP and off-peak. This analysis is performed in the following steps:

1. For each route, we separate all onboard recorded instances into three categories according to the mentioned time intervals. We proceed with the routes which have at least one onboard instance per time period.
2. Subsequently, for a given time interval, for each route, we calculate the total number of boarding passengers (which board at any stop along the trip, including the origin stage) and average over the instances of the same route of the same interval. We obtain the average boarding number in the given interval. Divide this average by vehicle capacity to get the seat turnover.
3. For a fixed time interval, we get daily ridership by multiplying the seat turnover from step 2 by the number of departures estimated in this interval along with capacity.

For instance, here is the calculation for the off-peak period.

$$Daily\ ridership_{offpeak} = Seat\ turnover_{offpeak} \times number\ of\ departures\ within\ offpeak\ period \times capacity,$$

4. We sum the previous figure for each time interval to get the daily ridership number per directed route.

Note: The number of boarding passengers including the initial number of passengers onboard at the origin stage is equal to the number of alighting passengers including the passengers getting off at the destination stage. It is thus adequate to only take into consideration the boarding passengers' number for a passenger flow analysis.

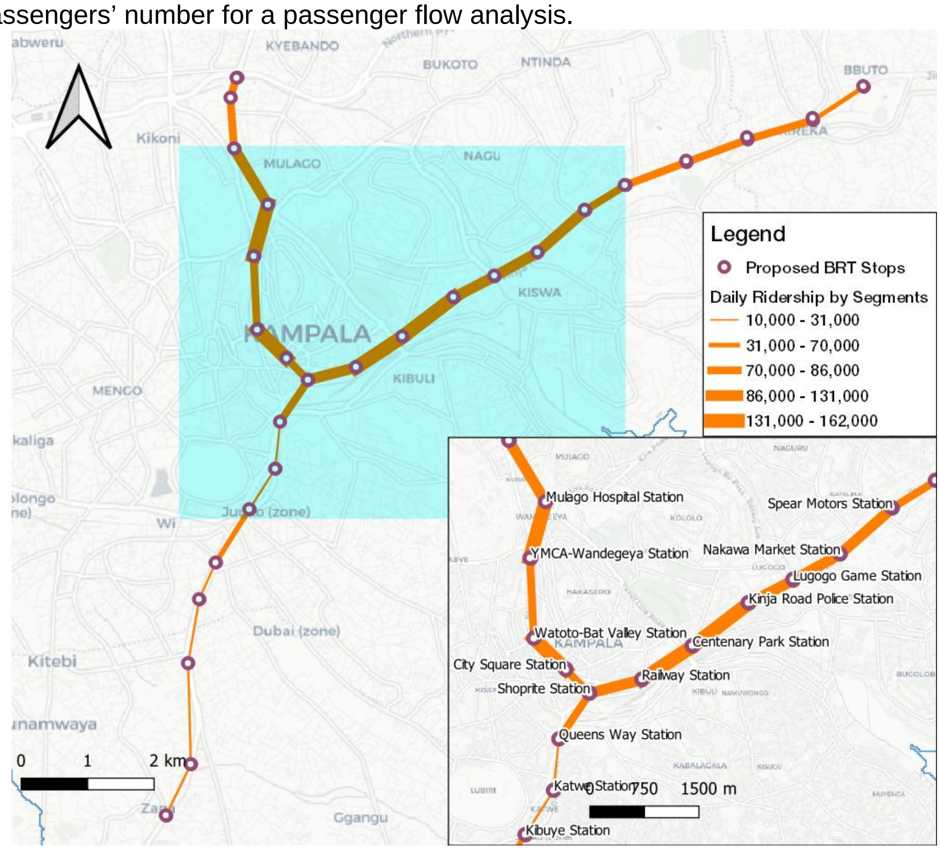


Figure 19 Daily Ridership across planned pilot BRT corridors

7.3.1 Validation from video feeds

Automated section counts were conducted by using high quality video feeds supplied by KCCA's Directorate of Engineering. These video feeds covered 6 geographical areas (Bwaise, Fairway, Kabira II, KatiKatti, Kira Rd, Naguru) broken down as follows:

Table 14 Scope of video feeds

Area	Total Number of Videos	Total Length of Footage (hh:mm:ss)	In morning peak?	In evening peak?
Kira Rd	64	23:38:57	Yes	Yes
Naguru	90	19:44:31	N/A*	N/A*
KatiKatti	27	6:16:09	N/A*	N/A*
Fairway	23	4:59:43	No	No
Bwaise	15	3:04:41	Yes	No
Kabira II	10	2:10:38	No	Yes

A machine learning model was developed to count bodas, buses and taxis as separate classes. These numbers would serve different purposes of this study, one of them is to validate the numbers of departures estimated from the frequency figures using the methodology aforementioned.

Logic of validation can be summarized as:

- Identify "Gates" (exact location on the road where vehicles were counted)
- Identify which routes pass through each gate
- Compare the total number of daily departures estimated for these particular routes against the number of taxis counted at the same gate during the same time period

Kira Road video feeds were chosen as a validator since it was the most covered area with over 20 hours of video feeds in different days (both weekdays).

Figure 20 Gates (A & B) placement and Vehicle counts from Kira Road video feed

Table 15 Kira Road Validation figures

Gate	Estimated Daily Departures	Day 1 Video Taxis Count	Day 2 Video Taxis Count	Error Percentage
A - Kira Road (Ntinda Direction)	1867	2016	2429	15%
B- Kira Road (City Square Direction)	2816	2906	3006	5%

Scope and Results:

The purpose of this analysis is to estimate the ridership activity per route per time period. Note that ridership here indicates the passengers who have boarded a given trip at any particular stop, including the Origin stage. The passenger flow analysis is performed first on a sample of the routes[2]. These routes serve to get average numbers of boarding passengers per time period. The latter are used for the routes with missing onboard data in a given time period.

Table 16 - Average number of boarding passengers based on the sample routes

	Average number of boarding passengers in MPP	Average number of boarding passengers in EPP	Average number of boarding passengers in Offpeak
mean	18.5	19.4	18.7

The following figure shows the routes which exhibit ridership higher than 6000 passengers daily, among our sample with a breakdown of ridership per peak period:

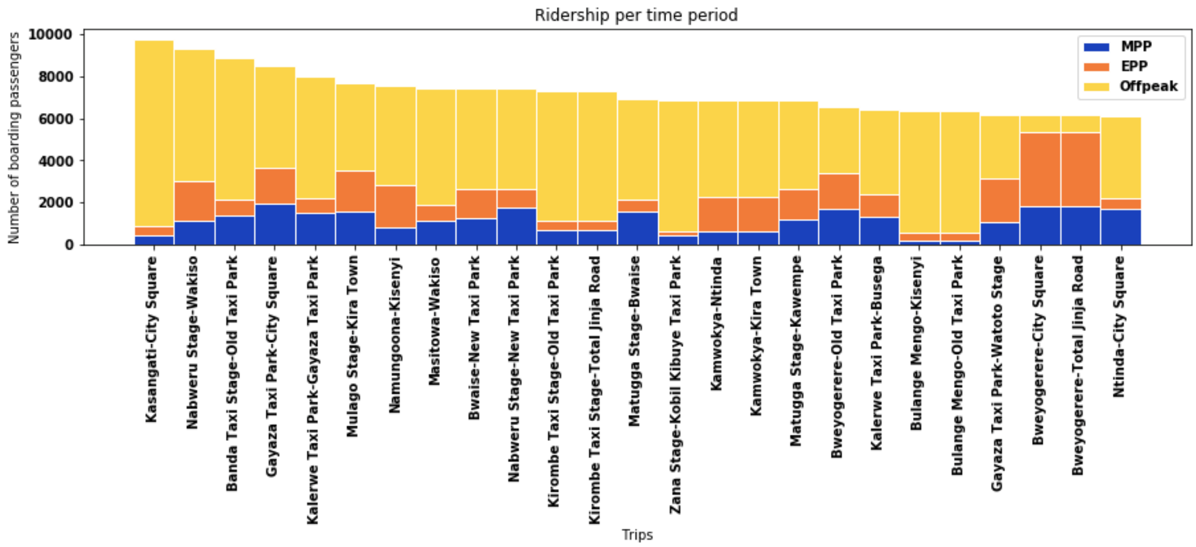


Figure 21 Ridership per peak period for routes with highest daily ridership activity.

For the total of the routes, the average ridership figure per MPP is 594 while the EPP mean is 623. The daily ridership is 2959 passengers on average ranging from 525 to 9751 passengers. Thus, the average ridership figures in the peaks consists of about 41% of the average daily ridership.

Table 17 Summary of indicators used in the passenger flow analysis

	Average number of boarding passengers in MPP	Average number of boarding passengers in EPP	Average number of boarding passengers in Offpeak	Number of departures in MPP	Number of departures in EPP	Number of departures in Offpeak	Ridership MPP	Ridership EPP	Daily Ridership

[2] Out of a total of 364 trips, 144 trips have boarding data in all 3 time periods (MPP, Offpeak, EPP). These are the trips on which the first stage of the passenger flow analysis was performed, namely, about 40% of the total.

mean	18.7	19.0	18.9	32.0	33.0	93.2	594.5	623.7	2959
min	2	4	7.5	3.8	3.9	14.2	45.0	69.2	525
max	55	100	42	112.2	118.2	363.6	3877	3533	9751

8. Workshops

8.1 OSM Mapathon

Partnering up with the World Resource Institute (WRI), an OpenStreetMap workshop (Mapathon) was conducted on the 25th of January 2020 at Makarere University, College of Engineering Design Art and Technology.

The aim of the workshop was to build capacity for students and fresh graduates on working with Open Data and OpenStreetMap and the impact of such platforms and concepts on society, giving incentive for participants to become contributors and active members of their relevant local mapping communities.

There was a filtration process to select attendees to make sure participation was diverse and inclusive as much as possible, 21 students were selected. The day's agenda included:

- **Presentation**: Introduction to DigitalTransport4Africa data sharing platform
- **Presentation**: Introduction to Open Data & OSM
- **Workshop**: iD Editor Walkthrough (a web software that allows users to edit OSM data and map different areas around the globe remotely)
- **Workshop**: Introduction to HOT Tasking Manager (a web tool to divide mapping tasks among volunteers to assist humanitarian projects, making sure data is valid and there is no duplication of effort)
- **Workshop**: Mapping Venerable communities (students created maps for villages in northern Uganda vulnerable to floods to assist with relief efforts)

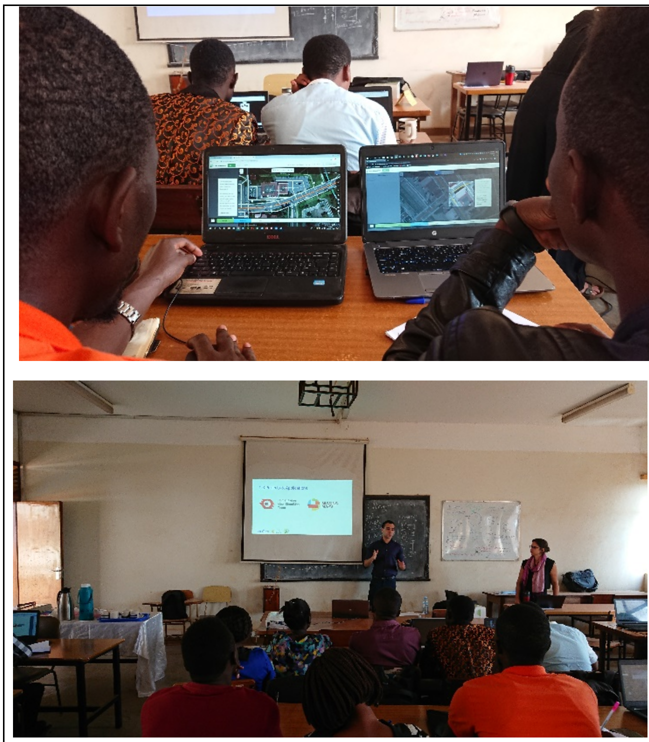


Figure 22 OSM Workshop

8.2 GTFS Workshop – Beyond Mapping

"Beyond Mapping" was a workshop aimed for Government of Uganda, namely KCCA & Ministry of Works and Transport (MoWT) mainly to raise awareness on the use of GTFS data in planning and the importance of building and maintaining a comprehensive Transportation Information System (TIS).

The workshop was made possible by another partnership WRI and was hosted both by KCCA & MoWT. It featured presentations from different use cases of GTFS and TIS in government entities around Africa, namely Ethiopia, Ghana and Egypt.

The workshop resulted in action items such as taking measures to ensure data sustainability within government, communicating with local mapping organizations and universities to increase government capacity and drafting an action plan to implement GTFS and TIS in the government's work stream.

Figure 23 Beyond Mapping GTFS Workshop

9. Tools

9.1 Data Collection

9.1.1 Route Observer - Developed by TfC

This is the main app used in Onboard Surveys. It allows surveyors to map out the spatial and temporal data of routes. The app allows them to capture the coordinates and names of stops, enter numeric data such as alighting and boarding numbers at each stop, trip fares and initial number of boarding passengers. Survey questions are also available on the conditions of the vehicle such as AC availability and wheelchair accessibility. An updated version of the app automates the assignment process itself giving the surveyor the chance to pick an assignment of his choosing based on proximity and input parameters from the Dashboard.

9.2 Data Processing

9.2.1 Route Beautify – Developed by TfC

This app helps with the digitization of raw trips, captured on Route Observer. The raw trips are aligned and matched to the OpenStreetMap (OSM) road network to generate the processed GIS route data. This app provides a practical alternative to the time-consuming manual digitization of raw trips, consisting of vertices matching on GIS.

9.2.2 ODK Collect – Open Source Toolbox

This is an open-source Android app which replaces paper forms in data collection processes. It supports entry constraints, multiple choice, numeric answers and geolocations.

It is the main tool used in the identification phase, through which static field researchers register their GPS coordinates and enter the origin, destination, midpoint and vehicle type.

9.2.3 GIS2GTFS

GIStoGTFS is a series of R scripts functioning together as a whole to convert GIS data into a GTFS feed. It is a module composed of multiple functions, where the main functions are responsible for generating the feed files.

10. References

- Transitec / ODA / Suez / TfC. Study of the Paratransit Transport System and Street Usage in the Greater Kampala Metropolitan Area. (Inception Report). AFD.
- Kumar, A., 2011. Understanding the emerging role of motorcycles in African cities - A political economy perspective (SSATP Discussion paper No. 13). SSATP.

11. Appendix

11.1 Appendix A: List of Stages

Name
Bwaise
Kalerwe (Gayaza)
Makarere University Main Gate
Wandegeya
Usafi Taxi Park
Kampala Old Taxi Park
New City Park
Kisenyi
Ndejje
Nateete Taxi Park
Nakulabye Market
Gayaza Taxi Park
Banda Taxi Stage
Namayiba Taxi Park
Kira Town
Kawempe
Kitoro Taxi Park
Kyaliwajjala
Kagoma
Matugga
Kajjansi
Mukono Taxi Park
Kumuwenda
Wakiso
Ntinda

Figure 24 List of main stages compiled

11.2 Appendix B: Definition of the Urban Extent

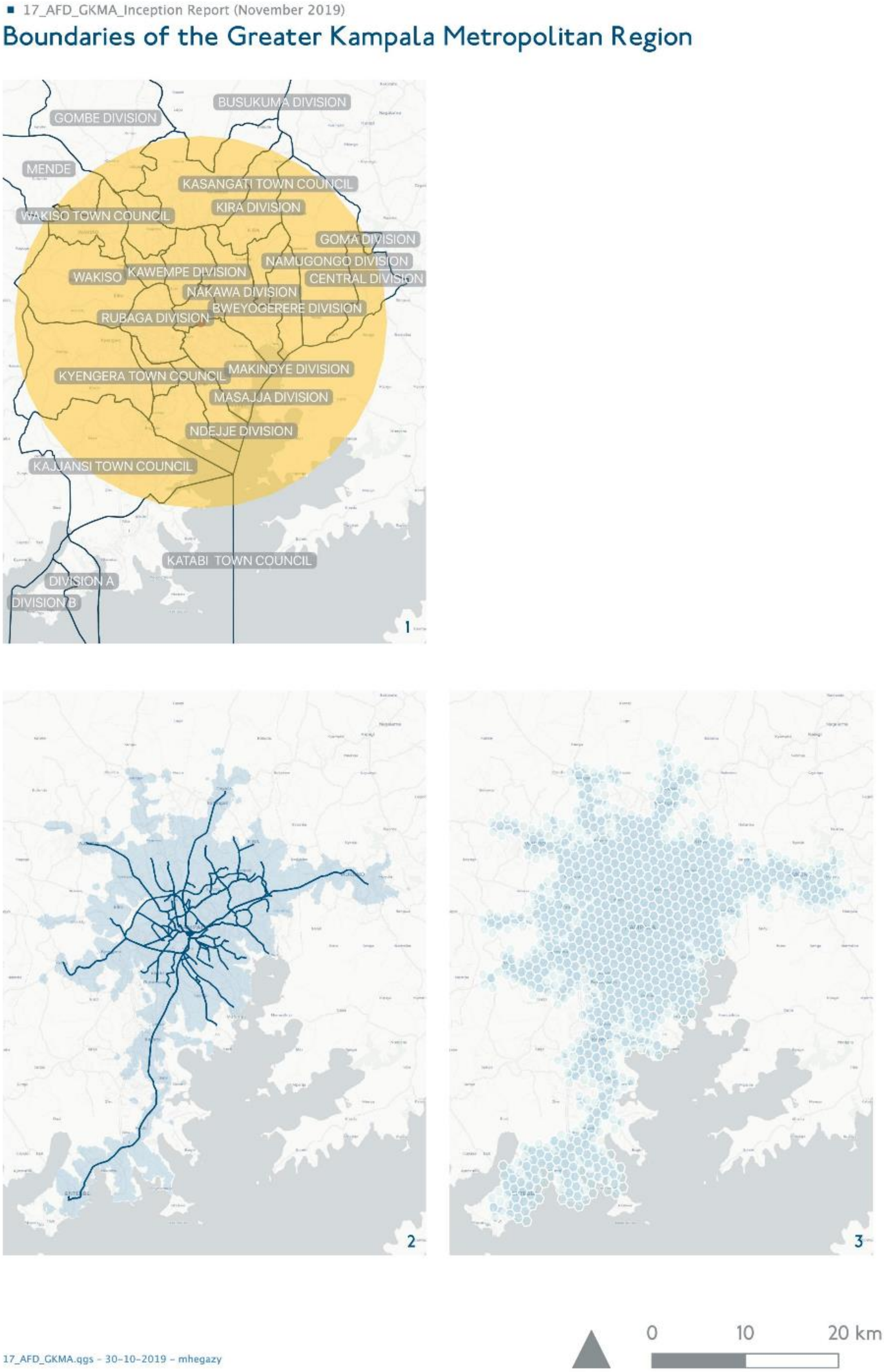


Figure 25 Definition of Boundaries of the GKMA