Planning: How to Complete Ridership Analyses

**Objective/Purpose/ Why:**

Ridership analysis is a key part of almost every transit plan. This SOP details how to process data from transit providers to ensure an accurate and easy-to-understand output of ridership data.

**Subject Matter Expert:**

Tim Simon

Alliance Transportation Group

Office 512.821.2081 | Mobile 541.600.0515  [tsimon@emailatg.com](mailto:tsimon@emailatg.com)

**Process**

1. **PROCESS THE RAW DATA**

Download all data that you’ve received from your transit provider and thoroughly examine it. What information is included in each spreadsheet / document? Your data will need to include the following components:

* **Data from National Transit Database (NTD) -** Make sure you have a copy of whatever the transit agency has reported as their final ridership totals. This ensures you have something with which to compare your results of the raw data analysis. You can also use the final totals for each route to **extrapolate** the data using a smaller sample (this will be explained later).
* **Raw Data –** this data is from a source such as automatic passenger counters (APC) or fare transactions (like [Genfare](https://www.genfare.com/)). This data may be messy or incomplete, but it is important to have it as a snapshot in time to understand ridership patterns. The data should include the following items:
  + Route Numbers (i.e., “1” or “Route 1”)
  + Stop Numbers (i.e., 20011)
  + Stop Names (i.e., Main @ Sonic)
  + Direction (i.e., Inbound or Outbound)
  + Boarding and Alighting Numbers (i.e., BOARD\_ALL, ALIGHT\_ALL)
  + Day of the Week (this could be in a date format, such as 1/10/2020)
  + Time of Day (i.e., 3:30 p.m.)
  + Revenue Data (typically a spreadsheet with financial info that was submitted to the NTD is sufficient, but sometimes it is included in the ridership spreadsheets as well)

\*If you are doing a bus stop analysis, you’ll also want to make sure that the data includes information about the amenities at each bus stop.

Below are some screenshots of APC data:

Table

Description automatically generated

Graphical user interface, application, table

Description automatically generated

And here are some screenshots of data from Genfare:

Table

Description automatically generated Table

Description automatically generated

1. **COMPARE RAW DATA WITH GIS DATA**

In order to complete your analysis, you’ll be cleaning up this data and putting it into ArcPro. Thus, you’ll need to look at whatever map packages you’ve been given and make sure you can do a [**Join**](https://desktop.arcgis.com/en/arcmap/10.3/manage-data/tables/essentials-of-joining-tables.htm)using the GIS data you’ve been given.

To check, just open up the GIS data you received and open the “Bus Stops” layer attribute table. If there is a common identifier in both the raw data and the attribute table, you’ll be able to complete the join (i.e., if the stop numbers are the same in GIS and the ridership data).

If it looks like there are stops on the map that aren’t in the ridership data (or you discover that there are more differences than you realized later on), you’ll need to download the attribute table as an Excel file so you can compare and reconcile any differences. ([Here](file://atg.local/adata/Planning/PLDV-2020.0113%20Las%20Cruces%20RoadRUNNER%20SRTP%20Update/Tasks/1_Existing%20Conditions/Ridership/Boardings%20+%20Alightings%20Analysis/ALL%20ROUTES_with%20stops_working_v9.xlsx) and [here](file://atg.local/adata/Planning/PLDV-2020.0113%20Las%20Cruces%20RoadRUNNER%20SRTP%20Update/Tasks/1_Existing%20Conditions/Ridership/Boardings%20+%20Alightings%20Analysis/ALL%20ROUTES_with%20stops_working_v3.xlsx) are examples if you need to do that.) For the purposes of this SOP, let’s say that it looks like everything matches and you’re good to go.

1. **LOG THE DATA**

In the project’s data log, note that you’ve gone through the data and verified it has what you need. This helps you stay organized throughout the analysis process. Here’s an example of a data log: ["Z:\Planning\PLDV-2020.0113 Las Cruces RoadRUNNER SRTP Update\Data\ATG\_Data Log\_RoadRUNNER SRTP Update.xlsx"](file:///Z:/Planning/PLDV-2020.0113%20Las%20Cruces%20RoadRUNNER%20SRTP%20Update/Data/ATG_Data%20Log_RoadRUNNER%20SRTP%20Update.xlsx)

1. **CLEAN THE DATA**

If the APC data was given to you in multiple spreadsheets (i.e., one workbook for each route), compile it into one big workbook. Do the same thing for your Genfare data or any other data you’ve received. Remember to insert a column that identifies where your data came from when you copy and paste (i.e., the month and year from each individual workbook you’re compiling).

**Graphical user interface, table

Description automatically generated**

Once you’ve compiled that data from each source into one big spreadsheet, you’ll want to turn it into a table. Do this is Excel by clicking Insert > Table.

**Graphical user interface, application

Description automatically generated**

Your data will then look something like this:

Table

Description automatically generated

1. **BOARDING AND ALIGHTING ANALYSIS**

Using the dataset that has the stop numbers and boarding and alighting data, create a **Pivot Table** to get your ridership totals for each route. (If you’ve never used a Pivot Table, it’s very easy and incredibly useful; find a tutorial on YouTube!)

Your Pivot Table should look something like this (mine includes a Slicer tool):

Table

Description automatically generated with low confidence

Take a look at the totals from your Pivot table and compare totals between the datasets you’ve received and between the NTD data. If the numbers are similar to NTD data, you can probably continue to use the APC dataset as is. However, if your APC dataset shows ridership that is *significantly* lower than the NTD totals, you’ll want to extrapolate the data (in the next step) and come back to this list. You may also need to expand other parts of the data even if your ridership numbers look similar, so make sure to read the section about extrapolating data.

**Your** [**final boarding and alighting analysis**](file:///Z:/Planning/PLDV-2020.0113%20Las%20Cruces%20RoadRUNNER%20SRTP%20Update/Tasks/1_Existing%20Conditions/Ridership/Boardings%20+%20Alightings%20Analysis/Final_Boardings_Alightings_20210406.xlsx) **should include the following:**

* Total Boarding for Each Stop
* Total Alighting for Each Stop
* Total Weekday Boarding for Each Stop
* Total Weekday Alighting for Each Stop
* Total Weekend Boarding for Each Stop
* Total Weekend Alighting for Each Stop
* Weekday Daily Boarding Average\* for Each Stop
* Weekday Daily Alighting Average\* for Each Stop
* Weekend Daily Boarding Average\* for Each Stop
* Weekend Daily Alighting Average\* for Each Stop
* Daily Average\* Boardings Overall for Each Stop
* Daily Average\* Alightings Overall for Each Stop
* Weekday Stop Productivity\*\*
* Weekend Stop Productivity\*\*
* Total Stop Productivity\*\*

(\*averages are calculated by total boardings or alightings / the number of days represented in your dataset. Confirm with your supervisor the number of days you’d like to use; do you want to use the number of days represented in the Genfare data you received? Or is there a number of days that you can calculate based on the NTD dataset provided? Perhaps it doesn’t matter because it’s all similar? Also remember that you’ll need to know how many weekdays and weekends occurred to get an accurate number for each. Once you have daily averages, be sure to confirm they match up with the NTD dataset you’ve been given.)

(\*\*productivity is the sum of boardings and alightings at a stop, it shows total activity)

**How to Extrapolate Data**

While the numbers in your dataset may seem way off compared to the NTD you received, they still tell an important story. The numbers in your APC data tell you about a snapshot in time and where riders were getting on and off the bus during that *small* window of time. In other words, if you know the percentage of riders that were getting off on the Main Street stop during this small window of time, you could apply that percentage to a larger sample (based off the NTD data).

***Example: The Bus Stop at Main***

According to the APC dataset sample, there are 100 people total who ride Route 2. Of those 100 people, 10 of them are getting of the bus at the Main Street stop. This means 10% of riders are alighting at the Main Street Stop.

Using this same ratio, let’s apply it to the NTD dataset, which says that 13,783 people ride Route 2 annually. Calculate what 10% of the NTD total would be to find out how many people are getting off at Main Street.

The result would show that 1,378 people are alighting at the Main Street stop.

This same process can be used for other parts of analysis as well. For example, my datasets included days of the week, but only in a separate dataset that didn’t include stop numbers. So, I had to identify the percentage of weekday and weekend ridership out of the totals given in that separate dataset, and then apply it to my larger sample set to get the average weekday and weekend numbers.

You can use this spreadsheet below as a template to extrapolate data. It has all the formulas used for taking a percentage of the smaller dataset and expanding it to a larger one:

[**"Z:\Planning\PLDV-2020.0113 Las Cruces RoadRUNNER SRTP Update\Tasks\1\_Existing Conditions\Ridership\Boardings + Alightings Analysis\Final\_Boardings\_Alightings\_20210406.xlsx"**](file:///Z:/Planning/PLDV-2020.0113%20Las%20Cruces%20RoadRUNNER%20SRTP%20Update/Tasks/1_Existing%20Conditions/Ridership/Boardings%20+%20Alightings%20Analysis/Final_Boardings_Alightings_20210406.xlsx)

Once you’ve got a good dataset, **have someone QC it.** There are lots of formulas and you’ll want to make sure you didn’t miss anything.

1. **JOIN THE BOARDING AND ALIGHTING DATA TO GIS**

Finally, you’ve got the numbers you need and it’s time to visualize that data using GIS! Complete a join (probably based on the Stop Number).

Remember, to complete the join, that you’ll need an individual CSV file for *each* route, and it cannot include formulas. I’d recommend copying the information for each route into a separate workbook (copy values only) and save as a CSV. Then join each CSV to the bus stop layer in GIS. (You’ll complete a join multiple times, once for each route you have.) Having each route as its own layer will give you the ability to view the productivity of each route and see which segments of each route are performing the best.

**A Note on System-Wide Productivity**

If you’d like to look at system-wide productivity as well (i.e., is this segment of Route 1 the most productive *compared to all other parts of the system*, not just compared to the rest of Route 1), I would recommend creating one big CSV that includes information from all routes to join with the bus stops layer.

**Visualizing the Data**

I’d recommend creating the following maps at minimum:

* Boarding and Alighting Pie Chart Maps
* Heat Maps Showing Productivity

You can find examples of this here: ["Z:\Planning\PLDV-2020.0113 Las Cruces RoadRUNNER SRTP Update\Mapping\Task\_1\_Ridership\_Analysis\RidershipAnalysis\_20210413.aprx"](file:///Z:/Planning/PLDV-2020.0113%20Las%20Cruces%20RoadRUNNER%20SRTP%20Update/Mapping/Task_1_Ridership_Analysis/RidershipAnalysis_20210413.aprx)

1. **COMPLETE A PEAK TIME ANALYSIS**

Completing a peak time analysis seemed much simpler in my experience, perhaps just based on the dataset that I had. Essentially, you’ll follow a similar format of compiling the data and then using Pivot Tables to summarize. This analysis helps you understand **when** people are riding the bus most often, and it can help inform recommendations around bus frequency (i.e., maybe you’d like to increase the bus frequency of Route A in the mornings since it is most productive during the morning commute, for example).

Here is an example of our **original data**: ["Z:\Planning\PLDV-2020.0113 Las Cruces RoadRUNNER SRTP Update\Data\2-Ridership\Route Ridership Time of Day Feb 2019\_Feb 2020 GENFARE.zip"](file:///Z:/Planning/PLDV-2020.0113%20Las%20Cruces%20RoadRUNNER%20SRTP%20Update/Data/2-Ridership/Route%20Ridership%20Time%20of%20Day%20Feb%202019_Feb%202020%20GENFARE.zip)

Here is an example of the **combined data**: ["Z:\Planning\PLDV-2020.0113 Las Cruces RoadRUNNER SRTP Update\Tasks\1\_Existing Conditions\Ridership\Peak Time Analysis\ALL ROUTES COMBINED\_raw.xlsx"](file://atg.local/adata/Planning/PLDV-2020.0113%20Las%20Cruces%20RoadRUNNER%20SRTP%20Update/Tasks/1_Existing%20Conditions/Ridership/Peak%20Time%20Analysis/ALL%20ROUTES%20COMBINED_raw.xlsx)

And here is the example of **peak times** **calculated using pivot tables**: ["Z:\Planning\PLDV-2020.0113 Las Cruces RoadRUNNER SRTP Update\Tasks\1\_Existing Conditions\Ridership\Peak Time Analysis\Peak Times and Daily Avg.xlsx"](file:///Z:/Planning/PLDV-2020.0113%20Las%20Cruces%20RoadRUNNER%20SRTP%20Update/Tasks/1_Existing%20Conditions/Ridership/Peak%20Time%20Analysis/Peak%20Times%20and%20Daily%20Avg.xlsx)

Again, just make sure to compare your peak time numbers to the other numbers you’re calculating and make sure they match up.

**Outcome:**

A ridership analysis that includes boarding and alighting information as well as a peak time analysis. The information used in this analysis is typically used to write the Existing Conditions chapter of a transit plan, like this one: ["Z:\Planning\PLDV-2020.0113 Las Cruces RoadRUNNER SRTP Update\Tasks\6\_Final Plan\\_Draft Chapters\Chapter 1\_Existing Conditions\RoadRUNNER-Fixed Route-Chapter 1-20210428-Draft.pdf"](file:///Z:/Planning/PLDV-2020.0113%20Las%20Cruces%20RoadRUNNER%20SRTP%20Update/Tasks/6_Final%20Plan/_Draft%20Chapters/Chapter%201_Existing%20Conditions/RoadRUNNER-Fixed%20Route-Chapter%201-20210428-Draft.pdf)

**Resources:**

Multiple resources are identified throughout the SOP.

**Definitions:**

NTD – National Transit Database

APC – Automatic Passenger Counter