

Analyzing passenger loads in the Valle de Aburrá transportation network

Área Metropolitana Del Valle De Aburrá

Project Team No 21

Members: Sandra Ruiz, Christian Velez, Felipe Garcia, Boris Martinez, Sebastian Marin, Carlos Taimal, Jorge Saavedra.

Overview of the industry

Big cities around the world face the problem of optimization of their public transport systems. Public entities have to deal with a complex situation where it converges diverse variables like the growth of the city, number of routes, number of passengers, number of buses, number of stops, schedules, etc. in order to balance demand and supply efficiently. Nowadays, thanks to the capacity to register huge amounts of data about the transport services in real-time, and the increasing computational capacity to process big chunks of data we can use Data Science methodologies that allows us to understand, analyze and visualize these complex systems and therefore improve their planning and management.

Business Impact

The planning of public transportation services is one of the most important areas of city planning. This allows to generate an accurate offer to the passengers without increasing operational costs nor generate traffic jam in the important avenues. With the information on the number of passengers on each route and each point in the city, transportation planning policies can be designed to benefit the community in terms of saving travel time or optimizing transportation routes to expand the coverage of the transportation network.

Specific Problem

The Área Metropolitana del Valle de Aburrá Area seeks to identify high demand patterns in order to implement strategies that improve the competitiveness of public transportation.

The entity wants to measure the passenger demand in the transportation network of Valle de Aburrá Region, particularly, the passenger load in each road network's arc. The expected results must be able to be filtered by date (time of the day, day of the week, or in general, a time interval) and by route, so that way it is easier to observe the demand levels for the public transportation service through space and time parameters.

The methods to reach the result will be georeferenced data analysis, finding through heat maps where the highest passenger density occurs and eventually and with the proper algorithm use the stations' data to find the nearest arc and then associate that demand in order to properly allocate the public transportation resources.

Data Description

The data given by the stakeholder (Area Metropolitana del Valle de Aburra) is stored in CSV and KML files. The CSV files have the historical data captured and transmitted by the GPS's installed on each vehicle. There are

3850 vehicles linked to this system transmitting data each 3 minutes approximately. The dataset for this project has the following fields:

FIELD	TYPE	DESCRIPTION
SECUENCIARECORRIDO	INTEGER	Primary key that identifies the track for a vehicle
RECORRIDOFINALIZADO	INTEGER	Complete / incomplete flag (S/N)
IDVEHICULO	INTEGER	unique identifier for a vehicle
CODIGORUTA	STRING	unique identifier for a track, each of these identifiers are related to a KML file
FECHAREGISTRO	DATE	date and time when the data was recorded
LATITUD	FLOAT	The latitude where passengers board/alight the vehicle*
LONGITUD	FLOAT	The longitude where passengers board/alight the vehicle*
SUBENDELANTERA	INTEGER	Quantity of passengers that board the vehicle through the front door
SUBENTRASERA	INTEGER	Quantity of passengers that board the vehicle through the back door
BAJANDELANTERA	INTEGER	Quantity of passengers that alight the vehicle through the front door
BAJANTRASERA	INTEGER	Quantity of passengers that alight the vehicle through the back door

- The latitude and longitude coordinates are displayed in the WGS84 standard.

We also have the data of the number of passengers who board the bus, either by the entrance or exit door. Each event is georeferenced with the information of the bus GPS and the route they are doing.