Traffic Jam: Predicting People's Movement into Nairobi

4-5 minutes

Description of the Problem

This challenge asks you to build a model that predicts the number of seats that Mobiticket can expect to sell for each ride, i.e. for a specific route on a specific date and time. There are 14 routes in this dataset. All of the routes end in Nairobi and originate in towns to the North-West of Nairobi towards Lake Victoria.

The towns from which these routes originate are:

1. Awendo
2. Homa Bay
3. Kehancha
4. Kendu Bay
5. Keroka
6. Keumbu
7. Kijauri
8. Kisii

9. Mbita

10. Migori

11. Ndhiwa

12. Nyachenge

13. Oyugis

14. Rodi

15. Rongo

16. Sirare

17. Sori

The routes from these 14 origins to the first stop in the outskirts of Nairobi takes approximately 8 to 9 hours from time of departure. From the first stop in the outskirts of Nairobi into the main bus terminal, where most passengers get off, in Central Business District, takes another 2 to 3 hours depending on traffic.

The three stops that all these routes make in Nairobi (in order) are:

1. Kawangware: the first stop in the outskirts of Nairobi

2. Westlands

3. Afya Centre: the main bus terminal where most passengers disembark

All of these points are mapped [here](#).

Passengers of these bus (or shuttle) rides are affected by Nairobi traffic not only during their ride into the city, but from there they must continue their journey to their final

destination in Nairobi wherever that may be. Traffic can act as a deterrent for those who have the option to avoid buses that arrive in Nairobi during peak traffic hours. On the other hand, traffic may be an indication for people's movement patterns, reflecting business hours, cultural events, political events, and holidays.

This is all for you to explore in the data.

Description of the data

train_revised.csv (zipped) is the dataset of tickets purchased from Mobiticket for the 14 routes from "up country" into Nairobi between 17 October 2017 and 20 April 2018. This dataset includes the variables: ride_id, seat_number, payment_method, payment_receipt, travel_date, travel_time, travel_from, travel_to, car_type, max_capacity.

test_questions.csv is the dataset on which you will apply your model to estimate number of tickets sold by Mobiticket per unique ride. This dataset contains all of the rides offered on the same 14 routes during the two weeks following train.csv, i.e. 21 April 2018 to 9 May 2018. The variables included in this dataset: ride_id, travel_date, travel_time, travel_from, travel_to, car_type, max_capacity.

sample_submission.csv is a table to provide an example of what your submission file should look like. This table has two columns: ride_id, number_of_ticket.

Uber Movement traffic data can be accessed at movement.uber.com. Data is available for Nairobi through

June 2018. (If the data for April-June are not up yet, they will be shortly.) Uber Movement provided historic hourly travel time between any two points in Nairobi. Any tables that are extracted from the Uber Movement platform can be used in your model.

Variables description:

- ride_id: unique ID of a vehicle on a specific route on a specific day and time.

- seat_number: seat assigned to ticket

- payment_method: method used by customer to purchase ticket from Mobiticket (cash or Mpesa)

- payment_receipt: unique id number for ticket purchased from Mobiticket

- travel_date: date of ride departure. (MM/DD/YYYY)

- travel_time: scheduled departure time of ride. Rides generally depart on time. (hh:mm)

- travel_from: town from which ride originated

- travel_to: destination of ride. All rides are to Nairobi.

- car_type: vehicle type (shuttle or bus)

- max_capacity: number of seats on the vehicle

  Have a look at this blog by Johnowhitaker.