Data Science Major Project

Project Topic: - Transport Demand Prediction

Regression

Description: -

Business Context

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 route's end in Nairobi and originate in towns to the northwest of Nairobi towards Lake Victoria.

The towns from which these routes originate are:

- Awendo
- Homa Bay
- Kehancha
- Kendu Bay
- Keroka
- Keumbu
- Kijauri
- Kisii
- Mbita
- Migori
- Ndhiwa
- Nyachenge
- Oyugis
- Rodi
- Rongo

© 2025 LaunchedGlobal. All Rights Reserved.

Made by recycled paper, because we care.

- Sirare
- Sori

The routes from these 14 origins to the first stop in the outskirts of Nairobi take approximately 8 to 9 hours from the time of departure. From the first stop in the outskirts of Nairobi into the main bus terminal, where most passengers get off, in the Central Business District, it 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 on 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 this 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 can avoid buses arriving in Nairobi during peak traffic hours. On the other hand, traffic may be an indication of people's movement patterns, reflecting business hours, cultural events, political events, and holidays.

Data Description:

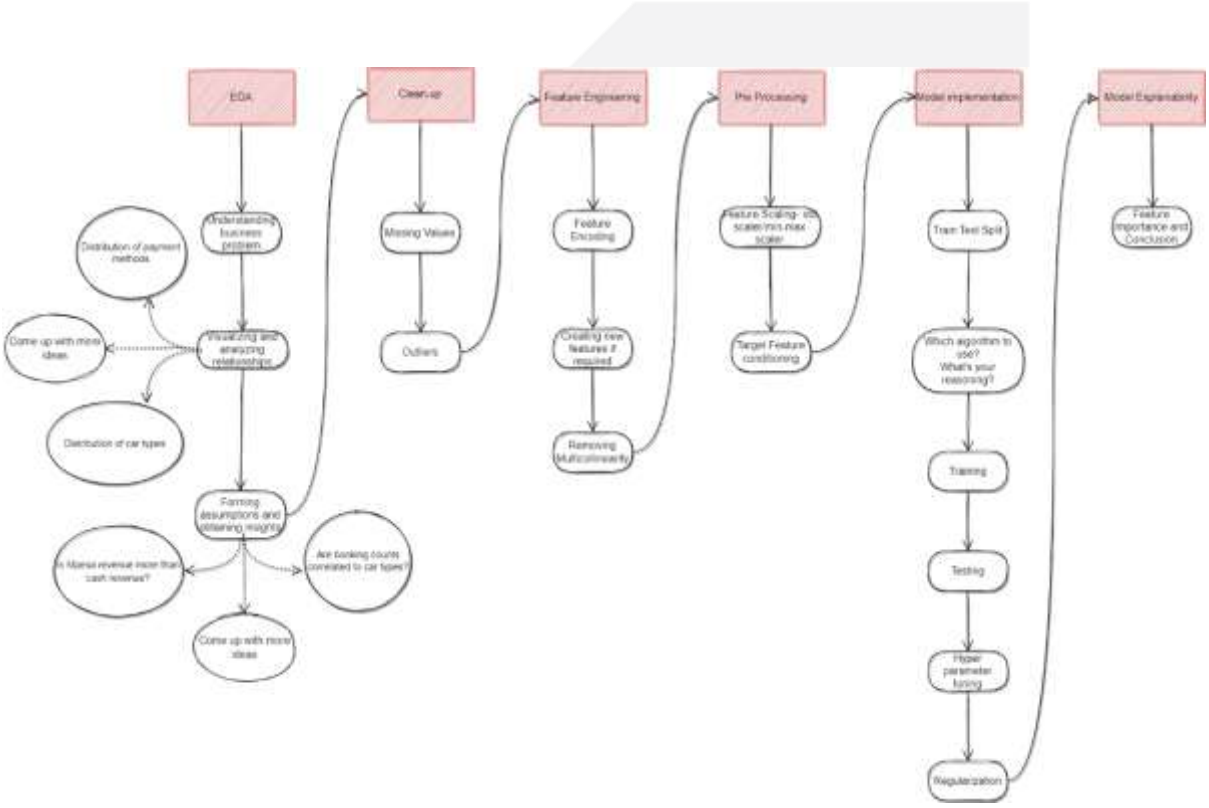
Fields	Description
ride_id	unique ID of a vehicle on a specific route on a specific day and time
seat_number	seat assigned to ticket
payment_method	method used by customer to purchase ticket from Mobiticket (cash or Mpesa)
payment_receipt	unique id number for ticket purchased from Mobiticket
travel_date	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

© 2025 LaunchedGlobal. All Rights Reserved.
Made by recycled paper, because we care.

Main Libraries to be Used:

- Pandas for data manipulation, aggregation
- Matplotlib and Seaborn for visualization and behaviour with respect to the target variable
- NumPy for computationally efficient operations
- Scikit Learn for model training, model optimization, and metrics calculation

Project Architecture:



© 2025 LaunchedGlobal. All Rights Reserved.
Made by recycled paper, because we care.