NYC TLC Taxi Fare Estimator Project

Executive summary report prepared by Automatidata

OVERVIEW

The New York City Taxi & Limousine Commission (TLC) has requested that Automatidata develop a model to estimate taxi fares before the ride. They have provided data that they collected from the past year from their taxi and limousine licensees.

PROJECT STATUS

- Using data from nearly 22,000 rides, mean distances and mean trip durations were estimated for different combinations of pick-up and drop-off zones.
- A multiple linear regression model was developed to predict the fare based on trip distance and mean duration. The model achieved a high accuracy and can account for 89.5% of the fare amount, based on these two variables.

NEXT STEPS

The NYC TLC can now implement the developed model in order to estimate taxi and limousine fares before the ride.

It is important to note that rides to JFK airport have been excluded from the model. In those cases, the standard $52 fare should be applied.

KEY INSIGHTS

- Using the trip distance (which a driver can estimate with a GPS) and mean duration of trips based on pick-up and drop-off zones, the final fare can be estimated to a high degree of accuracy:
  - The average difference between predicted fare and true fare is only $0.04.
  - Half of the predicted fares fall within $1.00 of the true fare.
  - 95% of predicted fares fall within $5.00 of the true fare.
- Rush hour does not correlate well with the fare amount and was not considered in the model.